ISSN 1342-2804

Research Reports on Mathematical and Computing Sciences

Worst-Case Violation of Sampled Convex Programs for Optimization with Uncertainty

Takafumi Kanamori and Akiko Takeda

February 2006, B–425

Department of Mathematical and Computing Sciences
Tokyo Institute of Technology

SERIES B: Operations Research

B-425 Worst-Case Violation of Sampled Convex Programs for Optimization with Uncertainty

Takafumi Kanamori† and Akiko Takeda‡
February 2006

Abstract. Sampled convex programs are studied to deal with convex optimization problems including uncertainty. A deterministic approach called robust optimization is commonly applied to solve these problems. On the other hand, sampled convex programs are randomized approach based on constraint sampling. Calafiore and Campi have proposed sufficient number of samples such that only small portion of original constraints are violated at randomized solution. Our main concern is not only the probability of violation, but also the degree of violation, that is, the worst-case violation. We derive an upper bound of the worst-case violation for sampled convex programs under general uncertainty set, and provide the relation between the probability of violation and the worst-case violation. As well as the probability of violation, the degree of violation is also assured to be depressed to small value with sufficiently large number of random samples.

Key words.

Uncertainty, Sampled Convex Programs, Worst-Case Violation, Violation Probability, Uniform Bound.

† Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2-12-1-W8-28 Oh-Okayama, Meguro-ku, Tokyo 152-8552 Japan. *kanamori@is.titech.ac.jp*

‡ Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2-12-1-W8-29 Oh-Okayama, Meguro-ku, Tokyo 152-8552 Japan. *takeda@is.titech.ac.jp*

1 Introduction

Uncertain programs have been developed to deal with optimization problems including inexact data, that is, uncertainty. In practical situations, say, decision-making environments, usually we do not know exact forms of objective functions or constraint functions, since optimization problems are constructed based on observed data for which contamination is inevitable. Even some information in the future may be required to formalize exact cost functions or constraints. Uncertain convex program (UCP) is formalized as a convex program under uncertain constraints. Without loss of generality, objective function is assumed to be linear without uncertainty parameters, and only constraints include uncertainty parameters, $\boldsymbol{u} \in \mathcal{U}$. When a realized value of uncertainty parameter is $\boldsymbol{u} \in \mathcal{U}$, the constraint for optimization problem is given as the form of $f(\boldsymbol{x}, \boldsymbol{u}) \leq 0$, where $\boldsymbol{x}$ is optimization variable and f is convex in $\boldsymbol{x}$. We do not know which uncertainty parameter will be realized, and thus, we need to deal with whole constraints, $f(\boldsymbol{x}, \boldsymbol{u}) \leq 0$ ($\boldsymbol{u} \in \mathcal{U}$) according to some optimization criteria.

Robust convex program (RCP) is one of the common approaches for UCP, see for instance [1, 2, 10, 11, 12, 13]. In RCP approach, one looks for a solution that is feasible under all possible realizations of uncertainty parameters. This means that the worst-case constraints are taken into account. Feasible region for RCP is described as $\{\boldsymbol{x} : f(\boldsymbol{x}, \boldsymbol{u}) \leq 0, \ \forall \boldsymbol{u} \in \mathcal{U}\}$. In some applications, the worst-case scenario needs to be considered, if the violation of constraints causes significant detriments. From the viewpoint of computation, though RCP is still convex optimization problem, in general it is numerically hard to solve the problem, because RCP involves infinitely many constraints. If constraint function f and uncertainty set $\mathcal{U}$ satisfy some nice properties, RCP can be reduced to standard convex problems that are tractable. For example, robust linear programs may result in second order cone programs [15], and robust second order cone programs may result in semidefinite programs. For details, see [19].

Sampled convex program (SCP) is a practical alternative to RCP for problems with uncertainty. The purpose of SCP is to find an approximated solution for RCP which satisfies almost all uncertainty constraints, while tractability is retained. SCP uses a probability distribution P on uncertainty set $\mathcal{U}$. The set of constraints for SCP is defined by random samples. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N$ be N independent and identically distributed random samples over $\mathcal{U}$, extracted according to P. The feasible region for SCP is defined as $\{\boldsymbol{x} : f(\boldsymbol{x}, \boldsymbol{u}_i) \leq 0, \, i = 1, \ldots, N\}$, which depends on the realization of random samples. SCP is a convex optimization problem with finite number of constraints, and thus, it is tractable for wide range of UCP. This is an advantage of SCP over RCP, while resulting solutions do not necessarily satisfy all constraints. Random sampling has been well-established technique in practical situations. The significant issue is to estimate the number of random samples, N, to guarantee that resulting solution violates only a small portion of constraints. Calafiore and Campi [4, 6] defined *violation probability* for randomized solutions, and proposed some practical lower bounds for the number of random samples to achieve small violation probability. If the number of random samples, N, is chosen properly according to the criterion given by Calafiore and Campi, obtained solution satisfies almost all constraints with high probability. Although the fact that violation probability is equal to zero does not necessarily denote that all constraints are satisfied, solutions with small violation probability are enough acceptable in many practical situations.

Chance constrained program (CCP) is one of the classical approaches for optimization problems with uncertainty [8], and CCP also uses probability distributions over uncertainty set $\mathcal{U}$, as well as SCP. Typically, constraint set for CCP is nonconvex, and thus, it is difficult to solve the problems exactly. Relation between CCP and SCP is also discussed (see [5, 7]).

Previous works of Calafiore and Campi have mainly focused on how much percentage of uncertain constraints are violated at an optimal solution of SCP. In this paper, we investigate statistical properties for function values of uncertain constraints at the optimal solutions of SCP. It is important to assess how large the solution violates each constraint function. In many applications, degree of violation at given solutions, measured by the value of each constraint function, has significant role to assess the validity of the approximated solution. This is the main reason that we focus on the values of constraint functions for UCP.

Degree of violation at a given solution $\boldsymbol{x}$ for SCP is governed by the worst-case violation defined as $\max_{\boldsymbol{v}\in\mathcal{U}} f(\boldsymbol{x}, \boldsymbol{v})$. Thus, our main concern is to provide an upper bound of the worst-case violation at a solution of SCP. To assess the worst-case violation, we need to evaluate the tail probability of $f(\boldsymbol{x}, \boldsymbol{u})$, that is, the probability such that $f(\boldsymbol{x}, \boldsymbol{u})$ takes values around $\max_{\boldsymbol{v}\in\mathcal{U}} f(\boldsymbol{x}, \boldsymbol{v})$. A uniform lower bound of the tail probability is derived under some conditions, and it is applied to evaluate the value of the worst-case violation stochastically. Moreover, we study the relation between the violation probability defined by Calafiore and Campi, and the worst-case violation. As well as violation probability, the worst-case violation is assured to take small value with high probability under a large enough number of random samples. By using uniform lower bound of the tail probability, we can estimate the number of random samples to achieve an approximated solution with high accuracy before solving SCP. This is a priori evaluation. Min-max problems are a common application of UCP, and our results can be applied to assess optimal values of min-max problems.

A criterion for a-posteriori assessments is also proposed. After computing an optimal solution of SCP, one can make an a-posteriori assessment with low computation cost for the value of the worst-case violation at the solution. The assessment is very useful to evaluate the worst-case violation with high accuracy. In general, uniform lower bound of the tail probability is not sufficiently tight in a priori evaluation, and as the result, the number of samples required to achieve high accuracy becomes quite large. Even if it is intractable to solve SCP with such a large amount of constraint functions given by random samples, a-posteriori assessment is still tractable, since we do not need to solve any optimization problems for a-posteriori assessment, but what we need is to compute the values of constraint functions.

The paper is organized as follows. In Section 2, first, we introduce the results of Calafiore and Campi, and derive an upper bound of the worst-case violation. A posteriori assessment for the worst-case violation is also proposed. To derive an upper bound of the worst-case violation, we need to evaluate an uniform lower bound for the tail probability of $f(\boldsymbol{x}, \boldsymbol{u})$. Under some conditions, the uniform lower bound is calculated in Section 3. We also provide a practical way of estimating the uniform lower bound based on constraint function f and uncertainty set $\mathcal{U}$. In Section 4, we derive a tighter uniform lower bound of tail probability for particular uncertainty set. In Section 5, we show some numerical simulations, and Section 6 is devoted to concluding remarks.

2 Worst-case Violation for General Robust Optimization

General robust convex program is described as

$$(RCP) \quad \left| \begin{array}{l} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad f(\boldsymbol{x}, \boldsymbol{u}) \le 0 \quad \forall \boldsymbol{u} \in \mathcal{U} \subset \mathbb{R}^d, \end{array} \right.$$

where $f(\boldsymbol{x}, \boldsymbol{u})$ is convex function in $\boldsymbol{x}$, and $\mathcal{X}$ is a convex subset in $\mathbb{R}^m$. Calafiore and Campi [4, 6] have proposed to approximately solve problem (RCP). In this approach, an approximation problem

$$(SCP_N) \quad \left| \begin{array}{l} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad f(\boldsymbol{x}, \boldsymbol{u}_i) \le 0 \quad i = 1, \dots, N, \end{array} \right.$$

is solved instead of original problem (RCP), where $\boldsymbol{u}_i, \ i = 1, \dots, N$, are N independent identically distributed samples on $\mathcal{U}$. The above problem is called sampled convex program (SCP). Fundamental question concerning with randomized optimization problems is how many samples are needed to achieve accurate solutions. The feasible region of (SCP_N) is regarded as an approximation for that of original robust optimization problem (RCP) which consists of infinitely many constraints. Note that feasible region of (SCP_N) depends on random variables, and thus, statistical discussions are useful to deal with convergence properties of optimal solutions.

Generally, optimal solutions of (SCP_N) are not necessarily feasible for (RCP). The ratio of violated constraints at $\boldsymbol{x} \in \mathcal{X}$ is measured by violation probability [4] defined below.

Definition 2.1. Let P be a probability measure on $\mathcal{U}$. Violation probability at $\boldsymbol{x} \in \mathcal{X}$ is defined by

$$V(\boldsymbol{x}) = P\{\boldsymbol{u} \in \mathcal{U} : f(\boldsymbol{x}, \boldsymbol{u}) > 0\},$$

where $f(\boldsymbol{x}, \cdot) : \mathcal{U} \to \mathbb{R}$ is assumed to be measurable function for each $\boldsymbol{x}$.

For example, if P is the uniform distribution on $\mathcal{U}$, violation probability is calculated from the volume of the set, $\{\boldsymbol{u} \in \mathcal{U} : f(\boldsymbol{x}, \boldsymbol{u}) > 0\}$. Even if an optimal solution of (SCP_N) is not feasible for (RCP), the solution with small violation probability would be practically acceptable as an approximation of optimal solution for (RCP).

Calafiore and Campi [6] have proposed that a solution of (SCP_N) fails to satisfy only a small portion of the original constraints, when sufficient number of samples N are taken.

Theorem 2.2 (Calafiore and Campi [6]). *For a level parameter, $\epsilon \in (0,1)$, and a confidence parameter, $\eta \in (0,1)$, let $N(\epsilon,\eta)$ be*

$$N(\epsilon,\eta) = \frac{2}{\epsilon}\log\frac{1}{\eta} + 2m + \frac{2m}{\epsilon}\log\frac{2}{\epsilon},$$

where m is the dimension of $\mathcal{X}$. Let $\hat{\boldsymbol{x}}_N$ be an optimal solution of (SCP_N). Then, for any positive integer $N > N(\epsilon,\eta)$, the inequality

$$P^N\{V(\hat{\boldsymbol{x}}_N) > \epsilon\} \le \eta,$$

holds, where $\mathrm{P}^N\{\cdots\}$ denotes the probability over N independent random samples, $\boldsymbol{u}_i \in \mathcal{U}, i = 1,\ldots,N$.

In Theorem 2.2, sufficient number of samples $N(\epsilon,\eta)$ is determined by level parameter, ϵ, confidence parameter, η, and the dimension of $\mathcal{X}$. Note that these parameters are specified before solving the optimization problems.

In Theorem 2.2, the value of $f(\hat{\boldsymbol{x}}_N,\boldsymbol{u})$ is not taken into account, though the probability of violation $f(\hat{\boldsymbol{x}}_N,\boldsymbol{u}) > 0$ is considered. The aim of this section is to derive a bound of constraint function $f(\boldsymbol{x},\boldsymbol{u})$ at an optimal solution of (SCP_N). Throughout this paper, we call $\max_{\boldsymbol{u}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{u})$ *the worst-case violation* of $\boldsymbol{x}$, when $\max_{\boldsymbol{u}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{u}) > 0$. At an optimal solution $\hat{\boldsymbol{x}}_N$ of (SCP_N), $\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N,\boldsymbol{u})$ is generally larger than zero, but it would converge to non-positive value as N goes to infinity. Now, we show how to decide the number of samples, N, so that the worst-case violation at $\hat{\boldsymbol{x}}_N$ is within a given tolerance level. For uncertainty set $\mathcal{U}$ and constraints $f(\boldsymbol{x},\boldsymbol{u})$, the following conditions are assumed.

Assumption 2.3. (a) Let $\mathcal{U}$ be a compact set in $\mathbb{R}^d$. Suppose that there exists d-dimensional hypersphere S in $\mathcal{U}$ which satisfies the condition,

$$\mathrm{conv}(\{\boldsymbol{u}\} \cup S) \subset \mathcal{U}, \quad \text{for any } \boldsymbol{u} \in \mathcal{U},$$

where $\mathrm{conv}(A)$ denotes the convex hull of set A.

(b) Let the function $f : \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ be convex in $\boldsymbol{x}$, and Lipschitz continuous on $\mathcal{U}$, that is, there exists a constant L such that

$$|f(\boldsymbol{x},\boldsymbol{u}) - f(\boldsymbol{x},\boldsymbol{v})| \le L\|\boldsymbol{u}-\boldsymbol{v}\|,$$

holds for $\forall \boldsymbol{x} \in \mathcal{X}$ and $\forall \boldsymbol{u},\boldsymbol{v} \in \mathcal{U}$, where $\|\cdot\|$ denotes the Euclidean norm.

A compact convex set in $\mathbb{R}^d$ including an open subset satisfies the condition (a), and an example of function with property (b) is shown in Section 3.2.

To estimate the worst-case violation of $f(\boldsymbol{x}, \boldsymbol{u})$, the tail probability of $f(\boldsymbol{x}, \boldsymbol{u})$ has important role. Let us define $p(\delta, \boldsymbol{x})$ as

$$p(\delta, \boldsymbol{x}) = P\left\{ \boldsymbol{u} \in \mathcal{U} : \max_{\boldsymbol{v} \in \mathcal{U}} f(\boldsymbol{x}, \boldsymbol{v}) - \delta < f(\boldsymbol{x}, \boldsymbol{u}) \right\}.$$

Under conditions (a) and (b), the tail probability, $p(\delta, \boldsymbol{x})$, has uniform lower bound.

Lemma 2.4. *Suppose conditions (a) and (b). Under the uniform distribution over $\mathcal{U}$, there exist a positive constant B and an increasing function $q(\delta)$ such that the inequality,*

$$0 \le q(\delta) \le p(\delta, \boldsymbol{x}), \quad 0 \le \delta \le B, \tag{1}$$

holds for any $\boldsymbol{x} \in \mathcal{X}$.

The function $q(\delta)$ which satisfies (1) is called uniform lower bound of tail probability for $0 \le \delta \le B$. The proof of Lemma 2.4 is shown in Section 3. In the proof, concrete expressions for $q(\delta)$ and B are indicated: $q(\delta)$ and B depend on the volume of $\mathcal{U}$, the radius of hypersphere $S \subset \mathcal{U}$, the diameter of $\mathcal{U}$, and a Lipschitz constant L.

The uniform distribution in Lemma 2.4 can be replaced by other probability distributions with some regularity conditions, such that the probability distribution has probability density function which is bounded below by some positive constant. For the sake of simplicity, we assume the uniform distribution over $\mathcal{U}$.

An uniform bound of $f(\hat{\boldsymbol{x}}_N, \boldsymbol{u})$ is given by the following theorem.

Theorem 2.5. *Let $q(\delta)$ be a uniform lower bound of tail probability for $0 \le \delta \le B$, and let $\epsilon \in (0, q(B))$, $\eta \in (0,1)$ and $N \ge N(\epsilon, \eta)$. An optimal solution of (SCP_N), $\hat{\boldsymbol{x}}_N$, satisfies the following inequalities simultaneously with probability of at least $1 - \eta$,*

$$\begin{aligned} &(i) \quad V(\hat{\boldsymbol{x}}_N) \le \epsilon, \\ &(ii) \quad \max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) \le q^{-1}(\epsilon), \end{aligned}$$

that is,

$$P^N\left\{ V(\hat{\boldsymbol{x}}_N) \le \epsilon, \ \max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) \le q^{-1}(\epsilon) \right\} \ge 1 - \eta$$

holds.

Proof. From Theorem 2.2, the inequality

$$P^N\{V(\hat{\boldsymbol{x}}_N) \le \epsilon\} \ge 1 - \eta,$$

holds, if $N \geq N(\epsilon, \eta)$. For $\hat{\boldsymbol{x}}_N$ such as $V(\hat{\boldsymbol{x}}_N) \leq \epsilon$, we derive an upper bound of the worst-case violation of $\hat{\boldsymbol{x}}_N$. Note that the assumption of $\epsilon < q(B)$ leads to $\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) \leq B$. Indeed, if $\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) > B$ holds, we have

$$\epsilon < q(B) \leq P\left\{\boldsymbol{u} \in \mathcal{U} : \max_{\boldsymbol{v}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{v}) - B < f(\hat{\boldsymbol{x}}_N, \boldsymbol{u})\right\} \leq P\left\{\boldsymbol{u} \in \mathcal{U} : 0 < f(\hat{\boldsymbol{x}}_N, \boldsymbol{u})\right\} = V(\hat{\boldsymbol{x}}_N) \leq \epsilon,$$

and this is contradiction. When $\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) < 0$ holds, it is clear that the inequality (ii) holds. Therefore, it is sufficient to consider the case of $\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) \geq 0$. Applying Lemma 2.4, we have

$$q\left(\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u})\right) \;\leq\; p\left(\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}), \hat{\boldsymbol{x}}_N\right) \;=\; P\left\{\boldsymbol{u} \in \mathcal{U} : 0 < f(\hat{\boldsymbol{x}}_N, \boldsymbol{u})\right\} \;=\; V(\hat{\boldsymbol{x}}_N) \;\leq\; \epsilon.$$

Therefore, we obtain a bound $\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) \leq q^{-1}(\epsilon)$ under the condition of $V(\hat{x}_N) \leq \epsilon$. ■

When the parameter $\epsilon > 0$ is fixed, there exists a sequence $\eta_N \to 0$ as $N \to \infty$, and with probability of at least $1 - \eta_N$, (i), and (ii) holds. Hence, in probability, $V(\hat{\boldsymbol{x}}_N)$ converges to zero, and $\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u})$ converges to a non-positive value, when the number of sampled constraints goes to infinity.

Note that the bound $q^{-1}(\epsilon)$ of $f(\hat{\boldsymbol{x}}_N, \boldsymbol{u})$ does not depend on $\hat{\boldsymbol{x}}_N$, and thus, before solving (SCP_N), Theorem 2.5 guarantees that the worst-case violation at an optimal solution of (SCP_N) is less than $q^{-1}(\epsilon)$ with high probability if $N(\geq N(\epsilon, \eta))$ samples are drawn. Therefore, we can set a parameter ϵ and η, which determine sample number N, so that the worst-case violation is assured to be small with high probability.

Next we study a-posteriori assessment. Once an optimal solution $\hat{\boldsymbol{x}}_N$ of (SCP_N) is computed, one can make a-posteriori assessment of the worst-case violation by Monte-Carlo methods. Suppose that new samples, $\mathcal{U}^{(M)} = \{\tilde{\boldsymbol{u}}_1, \ldots, \tilde{\boldsymbol{u}}_M\}$, are independently and identically distributed from the distribution P. Following theorem gives a probabilistic bound of $\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u})$.

Theorem 2.6. *Let $\hat{\boldsymbol{x}}_N$ be an optimal solution of (SCP_N). Let $M \geq \frac{\ln \eta}{\ln(1-q(\delta))}$ for fixed $\delta \in (0, B]$ and $\eta \in (0, 1)$, where $q(\delta)$ denotes uniform lower bound of tail probability for $0 \leq \delta \leq B$. Then, the inequality*

$$\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) < \max_{\tilde{\boldsymbol{u}}\in\mathcal{U}^{(M)}} f(\hat{\boldsymbol{x}}_N, \tilde{\boldsymbol{u}}) + \delta$$

holds with probability of at least $1 - \eta$.

Proof. With probability at most $1 - q(\delta)$, the inequality,

$$\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \delta \geq f(\hat{\boldsymbol{x}}_N, \tilde{\boldsymbol{u}})$$

holds when $\tilde{\boldsymbol{u}}$ is drawn from the distribution P, since we have $1 - p(\delta, \hat{\boldsymbol{x}}_N) \leq 1 - q(\delta)$ for $\delta \in (0, B]$. Thus, following inequalities holds,

$$\begin{aligned} & P^M\{\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \delta < \max_{\tilde{\boldsymbol{u}}\in\widetilde{\mathcal{U}}^{(M)}} f(\hat{\boldsymbol{x}}_N, \tilde{\boldsymbol{u}})\} \\ = \ & 1 - P\{\max_{u\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \delta \geq \max_{\tilde{\boldsymbol{u}}\in\widetilde{\mathcal{U}}^{(M)}} f(\hat{\boldsymbol{x}}_N, \tilde{\boldsymbol{u}})\} \\ = \ & 1 - \prod_{i=1}^{M} P\{\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \delta \geq f(\hat{\boldsymbol{x}}_N, \tilde{\boldsymbol{u}}_i)\} \\ \geq \ & 1 - (1 - q(\delta))^M \ \geq \ 1 - \eta. \end{aligned}$$

■

Note that a-posteriori assessment does not require computation for optimization, and thus, even if the number of random samples, M, is large, computation of a-posteriori assessment is still tractable. Also, in a-posteriori assessment, we can utilize $\bar{q}(\delta, \hat{\boldsymbol{x}}_N)$, defined for the solution $\hat{\boldsymbol{x}}_N$, satisfying $0 \leq \bar{q}(\delta, \hat{\boldsymbol{x}}_N) \leq p(\delta, \hat{\boldsymbol{x}}_N)$ instead of uniform lower bound $q(\delta)$, since a-posteriori assessment is carried out after the solution $\hat{\boldsymbol{x}}_N$ is obtained. The use of $\bar{q}(\delta, \hat{\boldsymbol{x}}_N)$ would decrease the number of samples, M, in comparison to that of $q(\delta)$. In numerical simulations in Section 5, we show how to construct $\bar{q}(\delta, \hat{\boldsymbol{x}}_N)$.

3 Uniform Upper Bound for Worst-case Violation

3.1 Evaluation of Uniform Lower Bound of Tail Probability

We evaluate a uniform lower bound of the tail probability, when P is the uniform distribution over $\mathcal{U}$. Under the assumptions, (a) and (b), Lemma 2.4 assures that there exist an increasing function $q_0(\delta)$ and a positive constant B_0 such that inequality,

$$^\forall\delta \in [0, B_0], \quad 0 \leq q_0(\delta) \leq p(\delta, \boldsymbol{x}),$$

holds for all $\boldsymbol{x} \in \mathcal{X}$. To show concrete form of q_0, we need the volume formula for the intersection of two d-dimensional hyperspheres. The volume formula is provided in Lemma 3.1.

Lemma 3.1. *Let d-dimensional hypersphere of radius s centered at the origin be S_1 and that of radius r centered at an boundary point of S_1 be S_2. Then, for $r \in [0, 2s]$, the volume of the intersection of two hyperspheres is given as*

$$D_d(r, s) = V_{d-1}(1) \left\{ s^d \int_0^{\cos^{-1}(1 - \frac{r^2}{2s^2})} (\sin x)^d dx + r^d \int_{\cos^{-1}(-\frac{r}{2s})}^{\pi} (\sin x)^d dx \right\}, \tag{2}$$

where $V_d(r)$ denotes the volume of d-dimensional hypersphere with radius r.

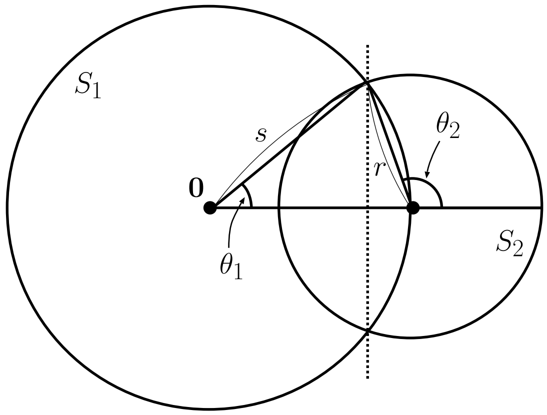

Figure 1: The set $S_1 \cap S_2$ is divided into two parts, one is the right part of the dotted line, and the other is left part of the dotted line. Volume of each part is calculated by integration for a part of hypersphere. Definitions of angles, θ_1 and θ_2, are illustrated.

Proof. We have

$$\begin{aligned}\mathrm{Vol}(S_1 \cap S_2) &= \int_0^{\theta_1} V_{d-1}(s \sin x) s \sin x dx + \int_{\theta_2}^{\pi} V_{d-1}\left(r \sin x\right) r \sin x dx, \\ &= V_{d-1}(1) \left\{ s^d \int_0^{\theta_1} (\sin x)^d dx + r^d \int_{\theta_2}^{\pi} (\sin x)^d dx \right\},\end{aligned}$$

where θ_1 and θ_2 is defined by

$$s \sin \theta_1 = r \sin \theta_2, \quad s \cos \theta_1 = s + r \cos \theta_2.$$

The geometrical interpretation of these equations is illustrated in Figure 1. This volume formula holds for $0 \le r \le 2s$. Angles, θ_1 and θ_2, are given by

$$\cos \theta_1 = 1 - \frac{r^2}{2s^2}, \quad \cos \theta_2 = -\frac{r}{2s},$$

and thus, we reach the conclusion. ■

Next, we show the proof of Lemma 2.4.

Proof of Lemma 2.4. For a fixed $\boldsymbol{x} \in \mathcal{X}$, let $\bar{\boldsymbol{u}} \in \mathcal{U}$ be an optimal solution of

$$\left| \; \max_{\boldsymbol{u}} f(\boldsymbol{x}, \boldsymbol{u}), \quad \text{s.t. } \boldsymbol{u} \in \mathcal{U}. \right.$$

Note that the compactness of $\mathcal{U}$ and continuity of f on $\mathcal{U}$ assure the existence of $\bar{\boldsymbol{u}}$. By Lipschitz continuity, $|f(\boldsymbol{x}, \bar{\boldsymbol{u}}) - f(\boldsymbol{x}, \boldsymbol{u})| < \delta$ holds, when $\|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \frac{\delta}{L}$. Thus, we have inequality,

$$\begin{aligned}P\left\{\boldsymbol{u} \in \mathcal{U} : \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \frac{\delta}{L}\right\} &\le P\{\boldsymbol{u} \in \mathcal{U} : |f(\boldsymbol{x}, \bar{\boldsymbol{u}}) - f(\boldsymbol{x}, \boldsymbol{u})| < \delta\} \\ &= P\{\boldsymbol{u} \in \mathcal{U} : f(\boldsymbol{x}, \bar{\boldsymbol{u}}) - f(\boldsymbol{x}, \boldsymbol{u}) < \delta\} \\ &= p(\delta, \boldsymbol{x}).\end{aligned}$$

Thus, lower bound of the probability such as $\|\bar{\boldsymbol{u}}-\boldsymbol{u}\|<\delta/L$ also provides that of $p(\delta,\boldsymbol{x})$. Since the probability distribution over $\mathcal{U}$ is the uniform distribution, the probability of $\|\bar{\boldsymbol{u}}-\boldsymbol{u}\|<\delta/L$ is given by

$$P\left\{\boldsymbol{u}\in\mathcal{U}:\|\bar{\boldsymbol{u}}-\boldsymbol{u}\|<\frac{\delta}{L}\right\}=\frac{\mathrm{Vol}(\{\boldsymbol{u}\in\mathcal{U}:\|\bar{\boldsymbol{u}}-\boldsymbol{u}\|<\delta/L\})}{\mathrm{Vol}(\mathcal{U})},$$

where $\mathrm{Vol}(\cdot)$ denotes volume under the Lebesgue measure on $\mathbb{R}^d$. We derive a lower bound of $\mathrm{Vol}(\{\boldsymbol{u}\in\mathcal{U}:\|\bar{\boldsymbol{u}}-\boldsymbol{u}\|<\delta/L\})$. The radius of S in assumption (a) is denoted by r, and let the diameter of $\mathcal{U}$ defined by $\sup_{\boldsymbol{u},\boldsymbol{v}\in\mathcal{U}}\|\boldsymbol{u}-\boldsymbol{v}\|$ be R. Let $S(\boldsymbol{y},w)$ be hypersphere of radius w at center $\boldsymbol{y}\in\mathbb{R}^d$,

$$S(\boldsymbol{y},w)=\left\{\boldsymbol{x}\in\mathbb{R}^d\ :\ \|\boldsymbol{x}-\boldsymbol{y}\|\le w\right\}.$$

Let us define $A_d(\theta)$ for $\theta\in[0,\pi]$ as surface area of a part of d-dimensional unit hypersphere defined by

$$\left\{\boldsymbol{x}=(x_1,\dots,x_d)\in\mathbb{R}^d\ :\ \|\boldsymbol{x}\|=1,\ \cos\theta\le x_1\right\}.$$

In other words, $A_d(\theta)$ is solid angle in $\mathbb{R}^d$. In Figure 2 (a), $A_d(\theta)$ for $d=2$ is illustrated. From assumption on $\mathcal{U}$, there exist a hypersphere, $S(\boldsymbol{c},r)\subset\mathcal{U}$.

When $\bar{\boldsymbol{u}}\in S(\boldsymbol{c},r)$, we have

$$\mathrm{Vol}(\{\boldsymbol{u}\in\mathcal{U}:\|\bar{\boldsymbol{u}}-\boldsymbol{u}\|<\delta/L\})\ge\mathrm{Vol}(\{\boldsymbol{u}\in S(\boldsymbol{c},r):\|\bar{\boldsymbol{u}}-\boldsymbol{u}\|<\delta/L\}).\tag{3}$$

The right-hand side is the volume of the intersection of two hyperspheres such that each radius is respectively equal to r and δ/L. The volume of these intersection decreases as the distance between centers ($\boldsymbol{c}$ and $\bar{\boldsymbol{u}}$) of these hyperspheres increases, and therefore, the minimum volume is attained when $\bar{\boldsymbol{u}}$ exists in the boundary of $S(\boldsymbol{c},r)$. According to Lemma 3.1, the right-hand side of (3) is lower bounded by $D_d(\delta/L,r)$ for $0\le\delta/L\le 2r$. Thus, for $\bar{\boldsymbol{u}}\in S(\boldsymbol{c},r)$, a lower bound,

$$\frac{1}{\mathrm{Vol}(\mathcal{U})}D_d\left(\frac{\delta}{L},r\right)\le P\left\{\boldsymbol{u}\in\mathcal{U}:\|\bar{\boldsymbol{u}}-\boldsymbol{u}\|<\frac{\delta}{L}\right\},\quad 0\le\frac{\delta}{L}\le 2r,$$

is derived.

When $\bar{\boldsymbol{u}}\notin S(\boldsymbol{c},r)$, there exists a hypersphere $S(\boldsymbol{d}_\alpha,\alpha r)\subset S(\boldsymbol{c},r)$ for $0<\alpha<1$, where $\boldsymbol{d}_\alpha$ is defined as

$$\boldsymbol{d}_\alpha=\bar{\boldsymbol{u}}+\frac{\boldsymbol{c}-\bar{\boldsymbol{u}}}{\|\boldsymbol{c}-\bar{\boldsymbol{u}}\|}\left(\|\boldsymbol{c}-\bar{\boldsymbol{u}}\|+(1-\alpha)r\right).$$

Let subsets K_1 and K_2 be

$$\begin{aligned}K_1&=\mathrm{conv}(\{\bar{\boldsymbol{u}}\}\cup S(\boldsymbol{d}_\alpha,\alpha r))\cup S(\boldsymbol{c},r),\\K_2&=\left\{\bar{\boldsymbol{u}}+\beta(\boldsymbol{v}-\bar{\boldsymbol{u}})\in\mathbb{R}^d:\boldsymbol{v}\in S(\boldsymbol{d}_\alpha,\alpha r),\ \beta\ge 0\right\},\end{aligned}$$

respectively, where K_2 is a cone with vertex at $\bar{\boldsymbol{u}}$. The function $\ell(\alpha)$ is defined as

$$\ell(\alpha)=\sup\{\ell:S(\bar{\boldsymbol{u}},\ell)\cap K_2\subset K_1\}.$$

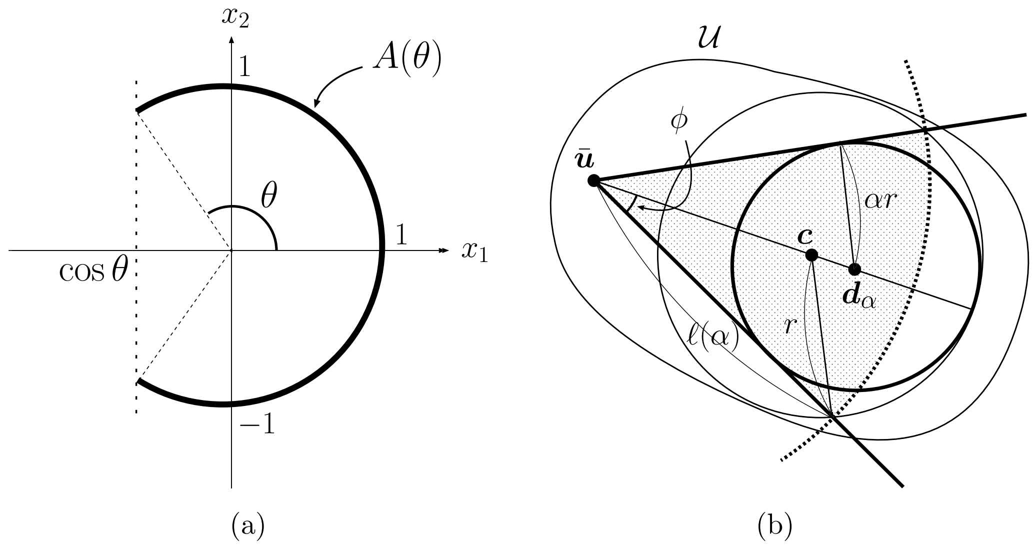

Figure 2: (a) Surface area, $A_d(\theta)$, of a part of unit hypersphere, $\{\boldsymbol{x} = (x_1, \cdots, x_d) \in \mathbb{R}^d \ : \ \|\boldsymbol{x}\| = 1, \ \cos\theta \le x_1\}$. In the figure, surface area for $d = 2$, which denotes the length of the thick arch, is drawn. (b) Definitions of $S(\boldsymbol{c}, r)$, $S(\boldsymbol{d}_\alpha, \alpha r)$, ϕ, and $\ell(\alpha)$ are illustrated. The shaded set denotes $S(\bar{\boldsymbol{u}}, \ell(\alpha)) \cap K_2$.

The definition of $\ell(\alpha)$ is illustrated in Figure 2 (b). Note that for $0 \le \ell \le \ell(\alpha)$, the inclusion relation, $S(\bar{\boldsymbol{u}}, \ell) \cap K_2 \subset K_1 \subset \mathcal{U}$, holds. Thus, for $0 \le \delta/L \le \ell(\alpha)$, the volume of $\{\boldsymbol{u} \in \mathcal{U} \ : \ \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}$ is lower bounded by that of $\{\boldsymbol{u} \in S(\bar{\boldsymbol{u}}, \ell) \cap K_2 \ : \ \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}$ which is equal to $V_d(\delta/L)$ times $A_d(\phi)/A_d(\pi)$, where $V_d(a)$ denotes the volume of d-dimensional hypersphere with radius a, and ϕ is defined as $\sin\phi = \alpha r/\|\bar{\boldsymbol{u}} - \boldsymbol{d}_\alpha\|$. Since $\|\bar{\boldsymbol{u}} - \boldsymbol{d}_\alpha\| + \alpha r \le R$ holds, ϕ is lower bounded by $\sin^{-1}\frac{\alpha r}{R - \alpha r}$. As the result, a lower bound of the volume of $\{\boldsymbol{u} \in \mathcal{U} \ : \ \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}$ for $0 \le \delta/L \le \ell(\alpha)$ is given by

$$\mathrm{Vol}(\{\boldsymbol{u} \in \mathcal{U} : \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}) \ \ge \ \frac{A_d(\phi)}{A_d(\pi)} V_d\left(\frac{\delta}{L}\right) \ \ge \ \frac{A_d(\sin^{-1}\frac{\alpha r}{R - \alpha r})}{A_d(\pi)} V_d\left(\frac{\delta}{L}\right).$$

Therefore, we obtain

$$\frac{1}{\mathrm{Vol}(\mathcal{U})} \frac{A_d(\sin^{-1}\frac{\alpha r}{R - \alpha r})}{A_d(\pi)} V_d\left(\frac{\delta}{L}\right) \ \le \ P\left\{\boldsymbol{u} \in \mathcal{U} : \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \frac{\delta}{L}\right\} \tag{4}$$

for $0 \le \delta/L \le \ell(\alpha)$. Applying the cosine formula, we have

$$r^2 = \|\bar{\boldsymbol{u}} - \boldsymbol{c}\|^2 + \ell(\alpha)^2 - 2\|\bar{\boldsymbol{u}} - \boldsymbol{c}\| \, \ell(\alpha) \cos\phi.$$

From the definition of $\ell(\alpha)$, larger solution of the quadratic equation is equal to $\ell(\alpha)$. Substituting

$$\sin\phi = \frac{\alpha r}{\|\bar{\boldsymbol{u}} - \boldsymbol{d}_\alpha\|} = \frac{\alpha r}{\|\bar{\boldsymbol{u}} - \boldsymbol{c}\| + (1 - \alpha) r}$$

to the solution, we obtain

$$\ell(\alpha) \ = \ \|\bar{\boldsymbol{u}} - \boldsymbol{c}\| \sqrt{1 - \left(\frac{\alpha r}{\|\bar{\boldsymbol{u}} - \boldsymbol{c}\| + (1 - \alpha) r}\right)^2} + \sqrt{r^2 - \alpha^2 r^2 \left(\frac{\|\bar{\boldsymbol{u}} - \boldsymbol{c}\|}{\|\bar{\boldsymbol{u}} - \boldsymbol{c}\| + (1 - \alpha) r}\right)^2}.$$

For $\|\bar{\boldsymbol{u}} - \boldsymbol{c}\| \geq r$, the first term of $\ell(\alpha)$ is increasing function of $\|\bar{\boldsymbol{u}} - \boldsymbol{c}\|$. Thus, we have the inequality,

$$
\begin{aligned}
\ell(\alpha) &\geq r\sqrt{1-\left(\frac{\alpha r}{r+(1-\alpha)r}\right)^2} + \sqrt{r^2-\alpha^2 r^2\left(\frac{\|\bar{\boldsymbol{u}}-\boldsymbol{c}\|}{\|\bar{\boldsymbol{u}}-\boldsymbol{c}\|+(1-\alpha)r}\right)^2} \\
&\geq r\sqrt{1-\left(\frac{\alpha r}{r+(1-\alpha)r}\right)^2} + \sqrt{r^2-\alpha^2 r^2} \\
&= r\left(\sqrt{1-\alpha^2}+\frac{2\sqrt{1-\alpha}}{2-\alpha}\right).
\end{aligned}
$$

Therefore, inequality (4) holds for $0 \leq \frac{\delta}{L} \leq r\left(\sqrt{1-\alpha^2}+\frac{2\sqrt{1-\alpha}}{2-\alpha}\right)$, regardless of $\bar{\boldsymbol{u}} \in \mathcal{U}\backslash S(\boldsymbol{c}, r)$. It is easy to see $0 \leq \sqrt{1-\alpha^2}+\frac{2\sqrt{1-\alpha}}{2-\alpha} \leq 2$, for $0 \leq \alpha \leq 1$.

In summary, for any $\alpha \in (0,1)$, inequality

$$
\begin{aligned}
&q_0(\delta) := \frac{1}{\mathrm{Vol}(\mathcal{U})}\min\left\{\frac{A_d(\sin^{-1}\frac{\alpha r}{R-\alpha r})}{A_d(\pi)}V_d\left(\frac{\delta}{L}\right),\ D_d\left(\frac{\delta}{L}, r\right)\right\} \ \leq\ p(\delta, \boldsymbol{x}), \\
&0 \leq \delta \leq B_0 := rL\left(\sqrt{1-\alpha^2}+\frac{2\sqrt{1-\alpha}}{2-\alpha}\right).
\end{aligned} \tag{5}
$$

is satisfied. It is clear that the lower bound is strictly increasing function with respect to δ. ■

Though there would be more tight uniform lower bound for tail probability than $q_0(\delta)$, calculation of volume would be more complicated for such lower bound.

The preferable α in (5), denoted by α^*, is given as follows. Typically, inequality

$$
\frac{A_d(\sin^{-1}\frac{\alpha r}{R-\alpha r})}{A_d(\pi)}V_d\left(\frac{\delta}{L}\right) \ \leq\ D_d\left(\frac{\delta}{L}, r\right)
$$

will hold. In Theorem 2.5, the value of ϵ is chosen from $(0, q_0(B_0))$, where $B_0 = rL\left(\sqrt{1-\alpha^2}+\frac{2\sqrt{1-\alpha}}{2-\alpha}\right)$, and then, $q_0(B_0)$ should take large value for practical use. Thus, α^* is determined by

$$
\begin{aligned}
\alpha^* &= \operatorname*{arg\,max}_{0\leq\alpha\leq 1} A_d\left(\sin^{-1}\frac{\alpha r}{R-\alpha r}\right)V_d\left(r\left(\sqrt{1-\alpha^2}+\frac{2\sqrt{1-\alpha}}{2-\alpha}\right)\right) \\
&= \operatorname*{arg\,max}_{0\leq\alpha\leq 1}\left(\sin^{-1}\frac{\alpha r}{R-\alpha r}\right)^{d-1}\left(\sqrt{1-\alpha^2}+\frac{2\sqrt{1-\alpha}}{2-\alpha}\right)^{d}.
\end{aligned} \tag{6}
$$

When R and r are given, numerical computation of α^* is easily performed.

3.2 Estimation of Uniform Bound

In practical problems, we need to estimate the function $q_0(\delta)$ and a constant B_0 in (5) to guarantee the accuracy of resulting randomized solution. We propose a simple way of estimating $q_0(\delta)$ for given uncertainty set $\mathcal{U}$ and constraint function f. As shown in the proof of Lemma 2.4, the lower bound $q_0(\delta)$ depends on volume of $\mathcal{U}$, radius of hypersphere $S \subset \mathcal{U}$, diameter of $\mathcal{U}$, and Lipschitz constant L. Here, for specific $\mathcal{U}$ and $f(\boldsymbol{x}, \boldsymbol{u})$, we show how to estimate those parameters.

Lipschitz constant L: We consider a quadratic constraint function $f(\boldsymbol{x},\boldsymbol{u})$ in $\boldsymbol{x}$ which is linearly perturbed in terms of $\boldsymbol{u}$, that is, $f(\boldsymbol{x},\boldsymbol{u}) := \boldsymbol{x}^\top \boldsymbol{Q}(\boldsymbol{u})\boldsymbol{x} + \boldsymbol{q}(\boldsymbol{u})^\top \boldsymbol{x} + \gamma(\boldsymbol{u})$, where $\boldsymbol{Q}(\boldsymbol{u}) := \boldsymbol{Q}_0 + \sum_{j=1}^d u_j \boldsymbol{Q}_j$, $\boldsymbol{q}(\boldsymbol{u}) := \boldsymbol{q}_0 + \sum_{j=1}^d u_j \boldsymbol{q}_j$, and $\gamma(\boldsymbol{u}) := \gamma_0 + \sum_{j=1}^d u_j \gamma_j$ for $\boldsymbol{u} \in \mathcal{U}$. This function is also rewritten as $f(\boldsymbol{x},\boldsymbol{u}) = \boldsymbol{d}(\boldsymbol{x})^\top \boldsymbol{u} + (\boldsymbol{x}^\top \boldsymbol{Q}_0 \boldsymbol{x} + \boldsymbol{q}_0^\top \boldsymbol{x} + \gamma_0)$, where $\boldsymbol{d}(\boldsymbol{x}) := (\boldsymbol{x}^\top \boldsymbol{Q}_1 \boldsymbol{x} + \boldsymbol{q}_1^\top \boldsymbol{x} + \gamma_1, \ldots, \boldsymbol{x}^\top \boldsymbol{Q}_d \boldsymbol{x} + \boldsymbol{q}_d^\top \boldsymbol{x} + \gamma_d)^\top$. Then we have

$$|f(\boldsymbol{x},\boldsymbol{u}) - f(\boldsymbol{x},\boldsymbol{v})| = |\boldsymbol{d}(\boldsymbol{x})^\top (\boldsymbol{u}-\boldsymbol{v})| \leq \max_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{d}(\boldsymbol{x})\| \times \|\boldsymbol{u}-\boldsymbol{v}\|.$$

It is possible to estimate roughly an upper bound for $\max_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{d}(\boldsymbol{x})\|$ as

$$L := \sqrt{\sum_{j=1}^{d} (\sigma_{max}(\boldsymbol{Q}_j) r_x^2 + \|\boldsymbol{q}_j\| r_x + \gamma_j)^2}$$

with the maximum eigenvalue $\sigma_{max}(\boldsymbol{Q})$ of $\boldsymbol{Q}$ and and the diameter r_x of $\mathcal{X}$ such as $\max_{\boldsymbol{x}_1,\boldsymbol{x}_2\in\mathcal{X}} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \leq r_x$.

Diameter R of $\mathcal{U}$: If the diameter R of $\mathcal{U}$ is not easily available, we can compromise a hypercube including $\mathcal{U}$ and use the length of diagonal line for hypercube as an upper bound of R. For $\mathcal{U}$ described by convex quadratic functions, it is solvable to find a hypercube including $\mathcal{U}$. Indeed, such a hypercube is obtained by solving $2d$ quadratic programs

$$\left| \; \max \ \boldsymbol{e}_i^\top \boldsymbol{u} \quad \text{s.t. } \boldsymbol{u} \in \mathcal{U} \right. , \tag{7}$$

and the ones with the object function $-\boldsymbol{e}_i^\top \boldsymbol{u}$ for $i = 1, \ldots, d$.

Radius r of inscribed hypersphere: We show how to calculate the radius of inscribed hypersphere for $\mathcal{U}$. At first, we assume that uncertainty set $\mathcal{U}$ is described as

$$\mathcal{U} = \left\{ \boldsymbol{u} \in \mathbb{R}^d \ : \ \boldsymbol{a}_i^T \boldsymbol{u} \leq b_i, \ i = 1, \ldots, p \right\}.$$

Let $\mathcal{B} = \{ r\boldsymbol{v} + \boldsymbol{c} \in \mathbb{R}^d \ : \ \|\boldsymbol{v}\| \leq 1 \}$ be a hypersphere of radius r. The uncertainty set $\mathcal{U}$ contains $\mathcal{B}$ if and only if

$$\boldsymbol{a}_i^T (r\boldsymbol{v} + \boldsymbol{c}) \leq b_i, \ i = 1, \ldots, p$$

holds for all $\|\boldsymbol{v}\| \leq 1$. Note that equality,

$$\sup_{\boldsymbol{v}:\|\boldsymbol{v}\|\leq 1} \boldsymbol{a}_i^T (r\boldsymbol{v} + \boldsymbol{c}) = r\|\boldsymbol{a}_i\| + \boldsymbol{a}_i^T \boldsymbol{c},$$

holds. In consequence, the maximum radius of inscribed hypersphere is given by the optimal value of the linear program,

$$\left| \begin{array}{ll} \max\limits_{r,\boldsymbol{c}} \ r & \\ \quad \text{s.t.} \ \ r\,\|\boldsymbol{a}_i\| + \boldsymbol{a}_i^T \boldsymbol{c} \leq b_i, & i = 1, \ldots, p, \\ \qquad\qquad -r \leq 0. & \end{array} \right.$$

We can apply same technique for quadratic constraints. Assume that $\mathcal{U}$ is given as

$$\mathcal{U} = \left\{ \boldsymbol{u} \in \mathbb{R}^d \ : \ \begin{array}{l} \boldsymbol{a}_i^T \boldsymbol{u} \le b_i, \ i = 1, \ldots, p, \\ \boldsymbol{u}^T A_j \boldsymbol{u} + 2\boldsymbol{d}_j^T \boldsymbol{u} + e_j \le 0, \ j = 1, \ldots, q \end{array} \right\},$$

where A_j denotes positive definite matrix, and we would like to find the maximum radius, r, such as $\mathcal{B} \subset \mathcal{U}$. We first work out the condition under which

$$\boldsymbol{u}^T A_j \boldsymbol{u} + 2\boldsymbol{d}_j^T \boldsymbol{u} + e_j \le 0,$$

holds for all $\boldsymbol{u} \in \mathcal{B}$. This occurs if and only if

$$\sup_{\boldsymbol{v}:\, \|\boldsymbol{v}\| \le 1} (r\boldsymbol{v} + \boldsymbol{c})^T A_j (r\boldsymbol{v} + \boldsymbol{c}) + 2\boldsymbol{d}_j^T (r\boldsymbol{v} + \boldsymbol{c}) + e_j \le 0,$$

and this is equivalent to the condition that there exists $\lambda_j \ge 0$ such as

$$\begin{pmatrix} -\lambda_j - e_j - \boldsymbol{d}_j^T A_j^{-1} \boldsymbol{d}_j & \boldsymbol{0}^T & (\boldsymbol{c} + A_j^{-1}\boldsymbol{d}_j)^T \\ \boldsymbol{0} & \lambda_j I & rI \\ \boldsymbol{c} + A_j^{-1}\boldsymbol{d}_j & rI & A_j^{-1} \end{pmatrix} \succeq 0. \tag{8}$$

The detail of the derivation refers to [3]. Therefore, for uncertainty set described by quadratic functions and linear functions, finding the maximum radius of inscribed hypersphere is solvable by semi-definite programs.

If the set $\mathcal{U}$ is convex but is not the one described above, we can utilize the following technique: at first, from $\mathcal{U}$ we pick up several points arbitrarily (say, $C := \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N\}$) and then, find a minimum volume ellipsoid covering C by solving a convex program. Then a smaller ellipsoid shrunk by a factor of the dimension d of $\mathcal{U}$ about its center is guaranteed to lie inside the convex hull of C, that is, inside the assumed $\mathcal{U}$. See also [3] for the details.

Volume $\mathrm{Vol}(\mathcal{U})$ **of** $\mathcal{U}$**:** We suppose that $\mathcal{U}$ is included in hypercube $\mathcal{C} := \prod_{i=1}^d [\underline{c}_i, \overline{c}_i] \subset \mathbb{R}^d$, which is available by the technique proposed for diameter R of $\mathcal{U}$. Monte Carlo method is a simple way of estimating $\mathrm{Vol}(\mathcal{U})$. For some positive number n, let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ be random samples identically and independently distributed according to the uniform distribution over hypercube $\mathcal{C}$. The volume of $\mathcal{U}$ is estimated by $\widehat{w} \times \mathrm{Vol}(\mathcal{C})$, where $\widehat{w}$ is defined by

$$\widehat{w} = \frac{1}{n} \sum_{i=1}^n I(\boldsymbol{v}_i \in \mathcal{U}) \tag{9}$$

where $I(\cdot)$ is the indicator function. Note that calculation of $\mathrm{Vol}(\mathcal{C})$ is easy. Clearly, $\widehat{w} \times \mathrm{Vol}(\mathcal{C})$ is an unbiased estimator of $\mathrm{Vol}(\mathcal{U})$, that is, mean value of $\widehat{w} \times \mathrm{Vol}(\mathcal{C})$ with respect to random samples, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, is equal to $\mathrm{Vol}(\mathcal{U})$. Computational cost of the estimator depends on how difficult it is to determine if $\boldsymbol{v}_i \in \mathcal{U}$ or not. When the dimension of $\mathcal{U}$, denoted by d, is not so high, the computation of $\widehat{w} \times \mathrm{Vol}(\mathcal{C})$ will be easily performed. The deviation of $\widehat{w}$ from $\mathrm{Vol}(\mathcal{U})/\mathrm{Vol}(\mathcal{C})$ is evaluated by Hoeffding's inequality [9]: for any $\tau > 0$,

$$P_U\left\{\widehat{w} \le \frac{\mathrm{Vol}(\mathcal{U})}{\mathrm{Vol}(\mathcal{C})} - \tau\right\} \;\le\; e^{-2\tau^2 n},$$

holds, where $P_U\{\cdots\}$ denotes the probability distribution over random samples $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$, each of which is generated from the uniform distribution over $\mathcal{C}$.

As mentioned above, the radius r of hypersphere $S \subset \mathcal{U}$, an upper bound of diameter of $\mathcal{U}$ denoted by R, and Lipschitz constant L of function $f(\boldsymbol{x}, \boldsymbol{u})$ on $\mathcal{U}$ are available. Moreover, the volume of $\mathcal{U}$ is estimated by $\widehat{w} \times \mathrm{Vol}(\mathcal{C})$ via random sampling over hypercube including $\mathcal{U}$. As the result, we have an estimator of the uniform lower bound of tail probability, $q_0(\delta)$. The following two theorems are analogs of Theorems 2.5 and 2.6, and moreover, probabilistic deviation of the estimator is taken into account. The function $h(\delta)$ and the constant B_0 are defined as

$$h(\delta) = \min\left\{\frac{A_d(\sin^{-1}\frac{\alpha^* r}{R-\alpha^* r})}{A_d(\pi)} V_d\left(\frac{\delta}{L}\right),\; D_d\left(\frac{\delta}{L}, r\right)\right\},$$

$$B_0 = rL\left(\sqrt{1-\alpha^{*2}} + \frac{2\sqrt{1-\alpha^*}}{2-\alpha^*}\right),$$

respectively, where α^* is the value given by (6). Note that $h(\delta)/\mathrm{Vol}(\mathcal{U})$ is also a lower bound of $p(\delta, \boldsymbol{x})$, even if the diameter of $\mathcal{U}$ is replaced by its upper bound, and the radius of inscribed hypersphere S is replaced by its lower bound.

Theorem 3.2. *Let $n \ge \frac{1}{2\tau^2}\log\frac{1}{\beta}$ for arbitrary $\tau > 0$ and $\beta \in (0,1)$, and let $\widehat{w} = \frac{1}{n}\sum_{i=1}^n I(\boldsymbol{v}_i \in \mathcal{U})$ be an unbiased estimator of $\frac{\mathrm{Vol}(\mathcal{U})}{\mathrm{Vol}(\mathcal{C})}$ calculated from n random samples which are uniformly distributed over hypercube $\mathcal{C}$. Also, let $N \ge N(\epsilon, \eta)$ for $\epsilon \in \left(0,\, \frac{h(B_0)}{(\widehat{w}+\tau)\mathrm{Vol}(\mathcal{C})}\right)$ and $\eta \in (0,1)$. Then, the following inequalities hold simultaneously with probability at least $(1-\eta)\times(1-\beta)$,*

$$(i)\quad V(\hat{\boldsymbol{x}}_N) \;\le\; \epsilon,$$
$$(ii)\quad \max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) \;\le\; h^{-1}((\widehat{w}+\tau)\mathrm{Vol}(\mathcal{C})\epsilon).$$

Proof. From the condition of n, inequality,

$$\mathrm{Vol}(\mathcal{U}) < (\widehat{w}+\tau)\mathrm{Vol}(\mathcal{C})$$

holds in probability of at least $1-\beta$, and then, we have

$$0 \;<\; \epsilon \;<\; \frac{h(B_0)}{(\widehat{w}+\tau)\mathrm{Vol}(\mathcal{C})} \;<\; \frac{h(B_0)}{\mathrm{Vol}(\mathcal{U})}$$

For such ϵ, inequalities, $V(\hat{\boldsymbol{x}}_N) \le \epsilon$ and $\max_{\boldsymbol{u}\in\mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) \le h^{-1}(\mathrm{Vol}(\mathcal{U})\epsilon)$, are simultaneously satisfied in probability of at least $1-\eta$, as shown in the proof of Theorem 2.5. When $\mathrm{Vol}(\mathcal{U}) \le (\widehat{w}+\tau)\mathrm{Vol}(\mathcal{C})$ is valid, we have

$$h^{-1}(\mathrm{Vol}(\mathcal{U})\epsilon) \;\le\; h^{-1}((\widehat{w}+\tau)\mathrm{Vol}(\mathcal{C})\epsilon).$$

In consequence, (i) and (ii) are valid in probability of at least $(1-\eta) \times (1-\beta)$. ■

We can also derive a-posteriori probability for assessments of optimal solution based on the estimator of $q_0(\delta)$. Given an optimal solution $\hat{\boldsymbol{x}}_N$ of (SCP_N), we apply Monte-Carlo methods to assess the worst-case violation. Suppose that samples $\mathcal{U}^{(M)} = \{\tilde{\boldsymbol{u}}_1, \dots, \tilde{\boldsymbol{u}}_M\}$ are identically and independently distributed over $\mathcal{U}$. The theorem below provides a probabilistic bound of $\max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u})$.

Theorem 3.3. *Let $\hat{\boldsymbol{x}}_N$ be an optimal solution of (SCP_N). Suppose that n samples are used to estimate* $\mathrm{Vol}(\mathcal{U})$*, where n satisfies inequality $n \geq \frac{1}{2\tau^2} \log \frac{1}{\beta}$ for fixed $\tau > 0$ and $\beta \in (0,1)$. Let $M \geq \frac{\ln \eta}{\ln(1-\hat{q}_0(\delta))}$, where $\widehat{q}_0(\delta) = \frac{h(\delta)}{(\widehat{w}+\tau)\mathrm{Vol}(\mathcal{C})}$ for fixed $\delta \in (0, B_0]$ and $\eta \in (0,1)$. Then, the inequality,*

$$\max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) < \max_{\tilde{\boldsymbol{u}} \in \tilde{\mathcal{U}}} f(\hat{\boldsymbol{x}}_N, \tilde{\boldsymbol{u}}) + \delta$$

holds with probability at least of $(1-\beta) \times (1-\eta)$.

Proof of the theorem is omitted, since it is almost same as that of Theorem 2.6.

4 Worst-case Violation for Particular Robust Optimization

4.1 Min-max optimization problems

Min-max optimization problem is a common application of robust convex program. Let $f(\boldsymbol{x}, \boldsymbol{u})$ be an objective function to be minimized in $\boldsymbol{x}$, where uncertainty is represented by parameter $\boldsymbol{u}$ varying among uncertainty set $\mathcal{U} \subset \mathbb{R}^d$. The constraints on $\boldsymbol{x}$ are specified as $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^m$. We would like to minimize the objective function under the most disadvantage condition. Thus, the optimization problem is formalized as min-max problem such as

$$\left| \ \min_{\boldsymbol{x}} \max_{\boldsymbol{u}} f(\boldsymbol{x}, \boldsymbol{u}), \quad \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{u} \in \mathcal{U}, \right.$$

which is also written as

$$(P) \quad \left| \begin{array}{l} \displaystyle\min_{\boldsymbol{x} \in \mathcal{X}, t \in R} \ t \\ \qquad \text{s.t.} \ \ f(\boldsymbol{x}, \boldsymbol{u}) - t \leq 0 \quad \forall \boldsymbol{u} \in \mathcal{U}. \end{array} \right.$$

The min-max problem (P) is approximated by random sampling such as

$$(P_N) \quad \left| \begin{array}{l} \displaystyle\min_{\boldsymbol{x} \in \mathcal{X}, t \in R} \ t \\ \qquad \text{s.t.} \ \ f(\boldsymbol{x}, \boldsymbol{u}_i) - t \leq 0, \quad i = 1, \dots, N. \end{array} \right.$$

Set of random samples, $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_N\}$, is denoted by $\mathcal{U}^{(N)}$, and let $V(\hat{\boldsymbol{x}}_N, \hat{t}_N)$ be the violation probability of a feasible solution $(\hat{\boldsymbol{x}}_N, \hat{t}_N)$ for (P_N).

Note that for min-max problems, the values of objective function are directly connected to those of constraint functions. Thus, it is possible to assess the optimal value of (P) by applying results in Section 2. A uniform lower bound of tail probability for (P) can be calculated from $\mathcal{U}$ and f as shown in Lemma 2.4.

Theorem 4.1. *Let $q(\delta)$ be a uniform lower bound of tail probability for $0 \le \delta \le B$. Let $\epsilon \in (0, q(B))$, $\eta \in (0,1)$, $N \ge N(\epsilon,\eta)$ and $(\hat{\boldsymbol{x}}_N, \hat{t}_N)$ be an optimal solution of (P_N). Then, following inequalities hold simultaneously with probability of at least $1-\eta$,*

$$\begin{aligned} &(i) \quad V(\hat{\boldsymbol{x}}_N, \hat{t}_N) \le \epsilon \\ &(ii) \quad 0 \le \mathrm{opt}(P) - \mathrm{opt}(P_N) \le q^{-1}(\epsilon), \end{aligned}$$

where $\mathrm{opt}(\cdot)$ *denotes the optimal value of optimization problem.*

Proof. Firstly, we have $0 \le \mathrm{opt}(P) - \mathrm{opt}(P_N)$, since inequality,

$$\max_{\hat{\boldsymbol{u}} \in \mathcal{U}^{(N)}} f(\boldsymbol{x}, \hat{\boldsymbol{u}}) \ \le\ \max_{\boldsymbol{u} \in \mathcal{U}} f(\boldsymbol{x}, \boldsymbol{u}),$$

holds for any $\boldsymbol{x} \in \mathcal{X}$ from the incursion relation $\mathcal{U}^{(N)} \subset \mathcal{U}$. Next, we derive an upper bound of $\mathrm{opt}(P) - \mathrm{opt}(P_N)$ in probability. Suppose that $V(\hat{\boldsymbol{x}}_N, \hat{t}_N) \le \epsilon$, then, we have

$$p\left(\max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \mathrm{opt}(P_N), \hat{\boldsymbol{x}}_N\right) \le \epsilon$$

because $\hat{t}_N = \mathrm{opt}(P_N)$ holds. The above inequality and assumption of $\epsilon < q(B)$ lead to $\max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \mathrm{opt}(P_N) \le B$. Thus, applying Lemma 2.4, we have

$$q\left(\max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \mathrm{opt}(P_N)\right) \ \le\ p\left(\max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \mathrm{opt}(P_N), \hat{x}_N\right) \ \le\ \epsilon.$$

In consequence, we obtain

$$\mathrm{opt}(P) - \mathrm{opt}(P_N) \ \le\ \max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \mathrm{opt}(P_N) \ \le\ q^{-1}(\epsilon),$$

that is valid whenever $V(\hat{\boldsymbol{x}}_N, \hat{t}_N) \le \epsilon$ holds. That is, with probability of at least $1-\eta$, inequalities,

$$V(\hat{\boldsymbol{x}}_N, \hat{t}_N) \le \epsilon \quad \text{and,} \quad 0 \ \le\ \mathrm{opt}(P) - \mathrm{opt}(P_N) \ \le\ q^{-1}(\epsilon)$$

are satisfied for $N > N(\epsilon, \eta)$. ■

Next, we study a-posteriori assessment of optimal value for (P_N) as well as the worst-case violation of (SCP_N). Once an optimal solution $(\hat{\boldsymbol{x}}_N, \hat{t}_N)$ of (P_N) is computed, one can make a-posteriori assessment of optimal values by Monte-Carlo methods. Suppose that new samples $\mathcal{U}^{(M)} = \{\tilde{\boldsymbol{u}}_1, \ldots, \tilde{\boldsymbol{u}}_M\}$ are generated independently and identically. Following theorem gives a probabilistic bound of $\mathrm{opt}(P) - \mathrm{opt}(P_N)$ as a-posteriori assessment. Note that a-posteriori assessment does not require the computation for optimization and even if the number of samples in $\mathcal{U}^{(M)}$ is large, the computation is still tractable.

Theorem 4.2. *Let $q(\delta)$ be a uniform lower bound of tail probability for $0 \le \delta \le B$, and let $\hat{\boldsymbol{x}}_N$ be an optimal solution of (P_N). For $M \ge \frac{\ln \eta}{\ln(1-q(\delta))}$ induced from $\delta \in (0, B]$ and $\eta \in (0,1)$, the inequality*

$$0 \le \mathrm{opt}(P) - \mathrm{opt}(P_N) < \max_{\boldsymbol{v} \in \mathcal{U}^{(N)} \cup \mathcal{U}^{(M)}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{v}) - \max_{\boldsymbol{u} \in \mathcal{U}^{(N)}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) + \delta$$

holds with probability at least of $1 - \eta$.

Proof. With probability at least $1 - \eta$, the inequality

$$\max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \delta < \max_{\tilde{\boldsymbol{u}} \in \mathcal{U}^{(M)}} f(\hat{\boldsymbol{x}}_N, \tilde{\boldsymbol{u}})$$

holds as shown in Theorem 2.6. Hence with probability at least $1 - \eta$, we see

$$\begin{aligned} \max_{\tilde{\boldsymbol{u}} \in \mathcal{U}^{(N)} \cup \mathcal{U}^{(M)}} f(\hat{\boldsymbol{x}}_N, \tilde{\boldsymbol{u}}) - \max_{\hat{\boldsymbol{u}} \in \mathcal{U}^{(N)}} f(\hat{\boldsymbol{x}}_N, \hat{\boldsymbol{u}}) &\ge \max_{\tilde{\boldsymbol{u}} \in \mathcal{U}^{(M)}} f(\hat{\boldsymbol{x}}_N, \tilde{\boldsymbol{u}}) - \max_{\hat{\boldsymbol{u}} \in \mathcal{U}^{(N)}} f(\hat{\boldsymbol{x}}_N, \hat{\boldsymbol{u}}) \\ &> \max_{\boldsymbol{u} \in \mathcal{U}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) - \delta - \max_{\hat{\boldsymbol{u}} \in \mathcal{U}^{(N)}} f(\hat{\boldsymbol{x}}_N, \hat{\boldsymbol{u}}) \\ &\ge \min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{u} \in \mathcal{U}} f(\boldsymbol{x}, \boldsymbol{u}) - \delta - \max_{\hat{\boldsymbol{u}} \in \mathcal{U}^{(N)}} f(\hat{\boldsymbol{x}}_N, \hat{\boldsymbol{u}}) \\ &= \mathrm{opt}(P) - \mathrm{opt}(P_N) - \delta, \end{aligned}$$

and $0 \le \mathrm{opt}(P) - \mathrm{opt}(P_N)$ is clear. This is the inequality to be proved. ■

Theorem 4.1 and Theorem 4.2 are easily modified for estimated uniform lower bound including $\widehat{w}$ in (9).

4.2 Common robust optimization problems

In robust optimization, one of the big issues is to find appropriate uncertainty set, though this is an application dependent question. Now we introduce some common uncertainty sets $\mathcal{U} \subset \mathbb{R}^d$ and function $f(\boldsymbol{x}, \boldsymbol{u})$ for robust optimization problems.

- The ellipsoidal uncertainty set $\mathcal{U} = \{\boldsymbol{u} : \|\boldsymbol{u}\| \le 1, \boldsymbol{u} \in \mathbb{R}^d\}$ is proposed
  - in [2] for an uncertain linear program with a linear uncertain constraint $f(\boldsymbol{x}, \boldsymbol{u}) := \boldsymbol{q}(\boldsymbol{u})^\top \boldsymbol{x}$, where $\boldsymbol{q}(\boldsymbol{u}) = \boldsymbol{q}_0 + \sum_{j=1}^{d} u_j \boldsymbol{q}_j$, and
  - in [1] for an uncertain quadratically constrained convex quadratic programs (QCQP) with a quadratic uncertain constraint $f(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{x}^\top \boldsymbol{A}(\boldsymbol{u})^\top \boldsymbol{A}(\boldsymbol{u}) \boldsymbol{x} + \boldsymbol{q}(\boldsymbol{u}) \boldsymbol{x} + \gamma(\boldsymbol{u})$, where $(\boldsymbol{A}(\boldsymbol{u}), \boldsymbol{q}(\boldsymbol{u}), \gamma(\boldsymbol{u})) = (\boldsymbol{A}_0, \boldsymbol{q}_0, \gamma_0) + \sum_{j=1}^{d} u_j (\boldsymbol{A}_j, \boldsymbol{q}_j, \gamma_j)$.
- The norm-constrained uncertainty set,

$$\mathcal{U}_k := \left\{ \boldsymbol{u} = \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix} : \|\boldsymbol{u}\| \le 1, \boldsymbol{v}_1 \ge \boldsymbol{0}, \boldsymbol{v}_1 \in \mathbb{R}^k, \boldsymbol{v}_2 \in \mathbb{R}^{d-k} \right\},$$

  is proposed in [12] for a QCQP problem

– with a quadratic uncertain constraint $f(\boldsymbol{x},\boldsymbol{u}) := \boldsymbol{x}^\top \boldsymbol{Q}(\boldsymbol{v}_1)\boldsymbol{x}+\boldsymbol{q}(\boldsymbol{v}_1)^\top \boldsymbol{x}+\gamma(\boldsymbol{v}_1)$, where $\boldsymbol{u}=\boldsymbol{v}_1$ and $(\boldsymbol{Q}(\boldsymbol{u}),\boldsymbol{q}(\boldsymbol{u}),\gamma(\boldsymbol{u})) = (\boldsymbol{Q}_0,\boldsymbol{q}_0,\gamma_0)+\sum_{j=1}^d u_j(\boldsymbol{Q}_j,\boldsymbol{q}_j,\gamma_j)$, and also,
– with a quadratic uncertain constraint $f(\boldsymbol{x},\boldsymbol{u}) := \boldsymbol{x}^\top \boldsymbol{Q}(\boldsymbol{v}_1)\boldsymbol{x}+\boldsymbol{q}(\boldsymbol{v}_2)^\top \boldsymbol{x}+\gamma(\boldsymbol{v}_2)$, where $\boldsymbol{Q}(\boldsymbol{u}) = \boldsymbol{Q}_0+\sum_{j=1}^k u_j\boldsymbol{Q}_j$, $(\boldsymbol{q}(\boldsymbol{u}),\gamma(\boldsymbol{u})) = (\boldsymbol{q}_0,\gamma_0)+\sum_{j=d-k+1}^d u_j(\boldsymbol{q}_j,\gamma_j)$.

The robust optimization problem (RCP) with $f(\boldsymbol{x},\boldsymbol{u})$ and $\mathcal{U}$ described above is tractable whenever $\mathcal{X}$ is convex set, and thus, it is not necessary to obtain an approximated solution via the sampled problem (SCP_N). However, the evaluation of a uniform lower bound of tail probability for the above uncertainty sets is useful to solve other kinds of non-tractable problems, for example, the minimax relative regret problem (robust deviation problem): $\min_{\boldsymbol{x}\in\mathcal{X}}\max_{\boldsymbol{u}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{u}) := g(\boldsymbol{x},\boldsymbol{u})-\min_{\boldsymbol{y}\in\mathcal{X}} g(\boldsymbol{y},\boldsymbol{u})$, where $g(\boldsymbol{x},\boldsymbol{u})$ is a convex quadratic function in $\boldsymbol{x}$. Kouvelis and Yu [14] proposed a scenario based approach to represent the input data uncertainty $\boldsymbol{u}$ in the minimax relative regret problem.

The proof of Lemma 2.4 provides expressions for B_0 and $q_0(\delta)$, where B_0 and q_0 depend on volume of $\mathcal{U}$, radius of hypersphere $S\subset\mathcal{U}$, diameter of $\mathcal{U}$, and Lipschitz constant L. However, for common robust optimization problems presented above, it is possible to provide tighter estimations for uniform lower bound of the tail probability.

Lemma 4.3. *Under the uniform distribution on ellipsoidal uncertainty set* $\mathcal{U} = \{\boldsymbol{u}\in\mathbb{R}^d : \|\boldsymbol{u}\|\le 1\}$*, the inequality*

$$q_1(\delta) := \frac{1}{V_d(1)} D_d\left(\frac{\delta}{L},1\right) \;\le\; p(\delta,\boldsymbol{x}), \quad 0\le\delta\le B_1 := 2L$$

holds for any $\boldsymbol{x}\in\mathcal{X}$*, where* $V_d(r)$ *denotes the volume of* d*-dimensional hypersphere with radius* r*, and* $D_d(r,s)$ *is defined as (2).*

Proof. For a given $\boldsymbol{x}\in\mathcal{X}$, let $\bar{\boldsymbol{u}}$ be an optimal solution in $\arg\max_{\boldsymbol{u}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{u})$. Since $\|\boldsymbol{u}-\bar{\boldsymbol{u}}\| < \frac{\delta}{L}$ implies $|f(\boldsymbol{x},\bar{\boldsymbol{u}})-f(\boldsymbol{x},\boldsymbol{u})| \le L\|\boldsymbol{u}-\bar{\boldsymbol{u}}\| < \delta$, we have

$$\begin{aligned} P\left\{\boldsymbol{u}\in\mathcal{U} : \|\boldsymbol{u}-\bar{\boldsymbol{u}}\| \le \frac{\delta}{L}\right\} &= P\left\{\boldsymbol{u}\in\mathcal{U} : \|\boldsymbol{u}-\bar{\boldsymbol{u}}\| < \frac{\delta}{L}\right\} \\ &\le P\left\{\boldsymbol{u}\in\mathcal{U} : \max_{\boldsymbol{v}} f(\boldsymbol{x},\boldsymbol{v})-\delta < f(\boldsymbol{x},\boldsymbol{u})\right\}. \end{aligned}$$

The probability on the left-hand side is described as $\mathrm{Vol}(\{\boldsymbol{u}\in\mathcal{U} : \|\boldsymbol{u}-\bar{\boldsymbol{u}}\| \le \frac{\delta}{L}\})/\mathrm{Vol}(\mathcal{U})$. The denominator, $\mathrm{Vol}(\mathcal{U})$, is equal to $V_d(1)$. Next, we evaluate a lower bound of $\mathrm{Vol}(\{\boldsymbol{u}\in\mathcal{U} : \|\bar{\boldsymbol{u}}-\boldsymbol{u}\| \le \frac{\delta}{L}\})$. This is the volume of the intersection of two hyperspheres such that each radius is respectively equal to 1 and $\frac{\delta}{L}$, and the distance between two centers is equal to $\|\bar{\boldsymbol{u}}\|$. For any two hyperspheres, the volume of the intersection decreases as the distance between centers of two hyperspheres increases. Applying this fact, we find that $\mathrm{Vol}(\{\boldsymbol{u}\in\mathcal{U} : \|\bar{\boldsymbol{u}}-\boldsymbol{u}\| \le \frac{\delta}{L}\})$ takes minimum value when $\|\bar{\boldsymbol{u}}\|=1$, because of $\bar{\boldsymbol{u}}\in\mathcal{U}$. When $\|\bar{\boldsymbol{u}}\|=1$, we have $\mathrm{Vol}(\{\boldsymbol{u}\in\mathcal{U} : \|\bar{\boldsymbol{u}}-\boldsymbol{u}\| \le \frac{\delta}{L}\}) = D_d\left(\frac{\delta}{L},1\right)$ according to Lemma 3.1, and thus, reach the conclusion. ■

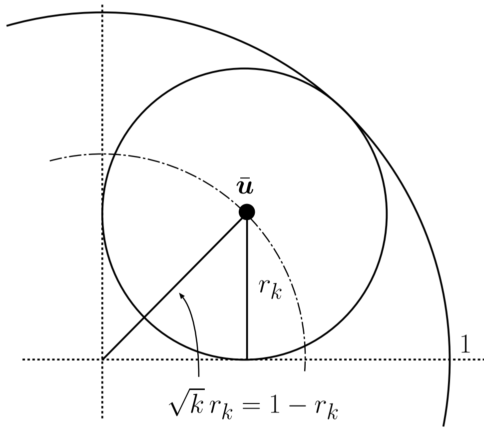

Figure 3: If $\|\bar{\boldsymbol{u}}\| \leq 1 - r_k$, the hypersphere of radius r_k centered at $\bar{\boldsymbol{u}}$ is included in unit hypersphere.

The value of ϵ should be included in the interval $(0, q_1(B_1))$, when we apply Theorem 2.5 or Theorem 4.1 to obtain upper bound of the worst-case violation. For the function q_1, however, $q_1(B_1) = \frac{1}{V_d(1)} D_d(2,1) = 1$ holds, and thus, we can take any value of $(0,1)$ for probability ϵ. Since function $q_1(\delta)$ is continuous and increasing with respect to δ, it is easy to calculate $q_1^{-1}(\epsilon)$ that is a uniform bound of the worst-case violation.

We extend Lemma 4.3 for the norm-constrained uncertainty set $\mathcal{U}_k$.

Lemma 4.4. *Under the uniform distribution on norm-constrained uncertainty set*

$$\mathcal{U}_k := \left\{ \boldsymbol{u} = \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix} : \|\boldsymbol{u}\| \leq 1, \boldsymbol{v}_1 \geq \boldsymbol{0}, \boldsymbol{v}_1 \in \mathbb{R}^k, \boldsymbol{v}_2 \in \mathbb{R}^{d-k} \right\},$$

the inequality

$$q_1(\delta) \leq p(\delta, \boldsymbol{x}), \quad 0 \leq \delta \leq \frac{L}{\sqrt{k}+1}$$

holds for any $\boldsymbol{x} \in \mathcal{X}$, where the function q_1 is defined in Lemma 4.3.

Proof. Let r_k be $\frac{1}{\sqrt{k}+1}$, and s be $\frac{\delta}{L}$, for the sake of simplicity. For any $\bar{\boldsymbol{u}} \in \mathcal{U}_k$, we estimate an uniform lower bound of the volume of $\{\boldsymbol{u} \in \mathcal{U}_k : \ \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| \leq s\}$ which is described as $\mathcal{W} \cap V_k$, where $\mathcal{W} := \{\boldsymbol{u} : \ \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| \leq s, \|\boldsymbol{u}\| \leq 1\}$ and $V_k := \{\boldsymbol{u} \in \mathbb{R}^d : u_i \geq 0, \quad i = 1, \ldots, k\}$. Note that the set $\mathcal{W}$ is the one considered in Lemma 4.3, while the range $[0, r_k L]$ of δ is smaller than before. The proof is decomposed into two parts. The case of $\|\bar{\boldsymbol{u}}\| \leq 1 - r_k$ is firstly studied, and secondly, the case of $1 - r_k < \|\bar{\boldsymbol{u}}\| \leq 1$ is considered. Geometrical meaning of these conditions is illustrated in Figure 3.

First, we prove $\frac{\mathrm{Vol}(\mathcal{W} \cap V_k)}{\mathrm{Vol}(\mathcal{U}_k)} \geq q_1(\delta)$ under the condition of $\|\bar{\boldsymbol{u}}\| \leq 1 - r_k$. Let $\boldsymbol{u}$ be an element of $\mathcal{W} \cap \{\bar{\boldsymbol{u}} + \boldsymbol{x} : x_1, \ldots, x_k \geq 0\}$, then, $u_i \geq \bar{u}_i$ holds for $i = 1, \ldots, k$. Since $\bar{\boldsymbol{u}}$ lies in $\mathcal{U}_k$, we have inequalities $\bar{u}_i \geq 0, i = 1, \ldots, k$. Thus, inequalities $u_i \geq \bar{u}_i \geq 0, i = 1, \ldots, k$ are valid. This denotes $\boldsymbol{u} \in \mathcal{W} \cap V_k$. As the consequence, we obtain $\mathrm{Vol}(\mathcal{W} \cap V_k) \geq \mathrm{Vol}(\mathcal{W} \cap \{\bar{\boldsymbol{u}}+\boldsymbol{x} : x_1, \dots, x_k \geq 0\})$. Since $\{\boldsymbol{u} : \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| \leq s\} \subset \{\boldsymbol{u} : \|\boldsymbol{u}\| \leq 1\}$ holds under the conditions $\|\bar{\boldsymbol{u}}\| \leq 1 - r_k$ and $s \leq r_k$, the equality, $\mathrm{Vol}(\mathcal{W} \cap \{\bar{\boldsymbol{u}} + \boldsymbol{x} : x_1, \dots, x_k \geq 0\}) = \left(\frac{1}{2}\right)^k V_d(s)$ holds. On the other hand, the volume of $\mathcal{U}_k$ is equal to $\left(\frac{1}{2}\right)^k V_d(1)$, and thus, we have

$$\frac{\mathrm{Vol}(\mathcal{W} \cap V_k)}{\mathrm{Vol}(\mathcal{U}_k)} \geq \frac{\left(\frac{1}{2}\right)^k V_d(s)}{\left(\frac{1}{2}\right)^k V_d(1)} \geq q_1(\delta),$$

where the last inequality is the result of Lemma 4.3.

Next, we assume $1-r_k < \|\bar{\boldsymbol{u}}\| \leq 1$ and $k < d$. Let $\boldsymbol{e}_i$ $(i = 1, \dots, d)$ be unit coordinate vectors. If there exists $\boldsymbol{u} \in \mathcal{W}$ such as $\boldsymbol{e}_i^\top \boldsymbol{u} < 0$, we say, $\boldsymbol{e}_i$ cuts $\mathcal{W}$. We reorder unit coordinate vectors from $\boldsymbol{e}_1$ to $\boldsymbol{e}_k$ such that $\boldsymbol{e}_1, \dots, \boldsymbol{e}_\ell$ cut $\mathcal{W}$, and $\boldsymbol{e}_{\ell+1}, \dots, \boldsymbol{e}_k$ do not cut $\mathcal{W}$. For $\ell + 1 \leq i \leq k$, equality $\mathrm{Vol}(\mathcal{W} \cap V_{i-1}) = \mathrm{Vol}(\mathcal{W} \cap V_i)$ holds clearly. In what follows, we show that inequality $\mathrm{Vol}(\mathcal{W} \cap V_i) \geq \frac{1}{2}\mathrm{Vol}(\mathcal{W} \cap V_{i-1})$ holds for any $i = 1, \dots, \ell$. The key of the proof is to show that there exists a hyperplane that divides $\mathcal{W} \cap V_{i-1}$ into two subsets with same volume and $\mathcal{W} \cap V_i$ includes one of them. Let i be a positive integer less than or equal to ℓ, and Π_i be the orthogonal projection matrix onto the subspace spanned by $\{\boldsymbol{e}_i, \dots, \boldsymbol{e}_d\}$, that is, $\Pi_i \boldsymbol{u} = (0, \dots, 0, u_i, \dots, u_d)^\top$ for all $\boldsymbol{u} = (u_1, \dots, u_d)^\top \in \mathbb{R}^d$. For any $j = 1, \dots, i$, $0 \leq \bar{u}_j < s \leq r_k$ is satisfied, because $\boldsymbol{e}_1, \dots, \boldsymbol{e}_i$ cut $\mathcal{W}$. From the assumption $\|\bar{\boldsymbol{u}}\|^2 > (1 - r_k)^2$, we have inequality $\|\Pi_i \bar{\boldsymbol{u}}\|^2 > r_k^2$, since $\|\Pi_i \bar{\boldsymbol{u}}\|^2 > (1 - r_k)^2 - \bar{u}_1^2 \cdots - \bar{u}_{i-1}^2 > (1 - r_k)^2 - (k - 1)r_k^2 = r_k^2$. Thus, for any $\boldsymbol{u}$ such as $\|\boldsymbol{u} - \bar{\boldsymbol{u}}\|^2 \leq s^2 \, (\leq r_k^2)$, we have

$$\boldsymbol{u}^\top \Pi_i \bar{\boldsymbol{u}} > 0,$$

because $r_k^2 \geq \|\boldsymbol{u} - \bar{\boldsymbol{u}}\|^2 \geq \|\Pi_i(\boldsymbol{u} - \bar{\boldsymbol{u}})\|^2 \geq \|\Pi_i \bar{\boldsymbol{u}}\|^2 - 2\boldsymbol{u}^\top \Pi_i \bar{\boldsymbol{u}} > r_k^2 - 2\boldsymbol{u}^\top \Pi_i \bar{\boldsymbol{u}}$. Let us define the vector $\boldsymbol{b}$ as

$$\boldsymbol{b} = \boldsymbol{e}_i - \frac{\bar{u}_i}{\|\Pi_i \bar{\boldsymbol{u}}\|^2} \Pi_i \bar{\boldsymbol{u}},$$

where $\boldsymbol{b}$ is well-defined because of $\|\Pi_i \bar{\boldsymbol{u}}\|^2 > r_k^2 > 0$. The vector $\boldsymbol{b}$ is not zero vector, because inequalities, $0 \leq \bar{u}_i < s \leq r_k$ and $\|\Pi_i \bar{\boldsymbol{u}}\| > r_k$, assure that the norm of $\frac{\bar{u}_i}{\|\Pi_i \bar{\boldsymbol{u}}\|^2} \Pi_i \bar{\boldsymbol{u}}$ is strictly less than one. Note that the inequality, $\boldsymbol{e}_i^\top \boldsymbol{b} = 1 - \frac{\bar{u}_i^2}{\|\Pi_i \bar{\boldsymbol{u}}\|^2} \geq 0$ holds. In the following way, we find that $\mathcal{W} \cap V_{i-1}$ is symmetric with respect to the hyperplane defined by $\{\boldsymbol{u} \in \mathbb{R}^d : \boldsymbol{b}^\top \boldsymbol{u} = 0\}$. We have the equalities, $\boldsymbol{b}^\top \boldsymbol{e}_j = 0$ $(j = 1, \dots, i-1)$ and $\boldsymbol{b}^\top \bar{\boldsymbol{u}} = 0$. Note that $\boldsymbol{b}^\top \Pi_i \bar{\boldsymbol{u}} = 0$ is also satisfied. Let the symmetric transformation matrix T be $T = I_d - \frac{2}{\|\boldsymbol{b}\|^2} \boldsymbol{b}\boldsymbol{b}^\top$, where I_d is $d \times d$ identity matrix. Using equalities, $T\bar{\boldsymbol{u}} = \bar{\boldsymbol{u}}, T\boldsymbol{e}_j = \boldsymbol{e}_j$ $(j = 1, \dots, i-1)$, and $\|T\boldsymbol{u}\| = \|\boldsymbol{u}\|$ $(\forall \boldsymbol{u} \in \mathbb{R}^d)$, we can verify $\{T\boldsymbol{v} \mid \boldsymbol{v} \in \mathcal{W} \cap V_{i-1}\} = \mathcal{W} \cap V_{i-1}$, by confirming each condition of $\mathcal{W} \cap V_{i-1}$. Now, we give a proof of the incursion relation,

$$\mathcal{W} \cap V_{i-1} \cap \{\boldsymbol{u} : \boldsymbol{b}^\top \boldsymbol{u} \geq 0\} \subset \mathcal{W} \cap V_i.$$

First, we define an orthogonal decomposition of $\boldsymbol{u} \in \mathbb{R}^d$. Let U_1 be the subspace spanned by $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{i-1}\}$, and U_2' be the subspace spanned by $\{\boldsymbol{e}_i, \dots, \boldsymbol{e}_d\}$. Since U_2' involves $\boldsymbol{b}$ and $\Pi_i \bar{\boldsymbol{u}}$, the subspace U_2' is decomposed into two subspaces, one is the subspace spanned by $\{\boldsymbol{b}, \Pi_i \bar{\boldsymbol{u}}\}$ and the other is its orthogonal complement U_2 in U_2'. As the result, any vector $\boldsymbol{u} \in \mathcal{W} \cap V_{i-1} \cap \{\boldsymbol{u} : \boldsymbol{b}^\top \boldsymbol{u} \geq 0\}$ has the orthogonal decomposition,

$$\boldsymbol{u} = \alpha\, \boldsymbol{b} + \beta\, \Pi_i \bar{\boldsymbol{u}} + \boldsymbol{u}_1 + \boldsymbol{u}_2,$$

where $\alpha, \beta \in \mathbb{R}$, $\boldsymbol{u}_1 \in U_1$ and $\boldsymbol{u}_2 \in U_2$. Note that the equality $\boldsymbol{e}_i^\top \boldsymbol{u}_2 = 0$ is valid, since the definition of $\boldsymbol{b}$ implies that $\boldsymbol{e}_i$ is represented by linear combination of $\boldsymbol{b}$ and $\Pi_i \bar{\boldsymbol{u}}$. Recall that for $\boldsymbol{u} \in \mathcal{W} \cap V_{i-1} \cap \{\boldsymbol{u} : \boldsymbol{b}^\top \boldsymbol{u} \geq 0\}$, we have inequality, $\boldsymbol{u}^\top \Pi_i \bar{\boldsymbol{u}} > 0$, in addition to $\boldsymbol{b}^\top \boldsymbol{u} \geq 0$. Moreover, we have $\boldsymbol{b}^\top \Pi_i \bar{\boldsymbol{u}} = 0$ and $\boldsymbol{e}_i^\top \boldsymbol{b} \geq 0$. From $\boldsymbol{b}^\top \boldsymbol{u} \geq 0$ and $\boldsymbol{u}^\top \Pi_i \bar{\boldsymbol{u}} > 0$, we obtain $\alpha \geq 0$ and $\beta > 0$. Thus, the inequality, $\boldsymbol{e}_i^\top \boldsymbol{u} = \alpha\, \boldsymbol{e}_i^\top \boldsymbol{b} + \beta\, \boldsymbol{e}_i^\top \Pi_i \bar{\boldsymbol{u}} \geq 0$, holds, because of $\alpha \geq 0$, $\boldsymbol{e}_i^\top \boldsymbol{b} \geq 0$, $\beta > 0$ and $\boldsymbol{e}_i^\top \Pi_i \bar{\boldsymbol{u}} = \bar{u}_i \geq 0$. Therefore, $\boldsymbol{u} \in \mathcal{W} \cap V_i$ is valid. From the incursion relation, $\mathcal{W} \cap V_{i-1} \cap \{\boldsymbol{u} : \boldsymbol{b}^\top \boldsymbol{u} \geq 0\} \subset \mathcal{W} \cap V_i$, we have the inequality of volume formula, $\mathrm{Vol}(\mathcal{W} \cap V_i) \geq \frac{1}{2}\mathrm{Vol}(\mathcal{W} \cap V_{i-1})$, since the hyperplane defined by $\boldsymbol{b}^\top \boldsymbol{u} = 0$ separates $\mathcal{W} \cap V_{i-1}$ into two regions with the same volumes. As the conclusion, we obtain

$$\mathrm{Vol}(\mathcal{W} \cap V_k) = \cdots = \mathrm{Vol}(\mathcal{W} \cap V_\ell) \geq \frac{1}{2}\mathrm{Vol}(\mathcal{W} \cap V_{\ell-1}) \geq \cdots \geq \left(\frac{1}{2}\right)^\ell \mathrm{Vol}(\mathcal{W}) \geq \left(\frac{1}{2}\right)^k \mathrm{Vol}(\mathcal{W})$$

and thus, $\frac{\mathrm{Vol}(\mathcal{W} \cap V_k)}{\mathrm{Vol}(\mathcal{U}_k)} \geq q_1(\delta)$, in the same way as the case of $\|\bar{\boldsymbol{u}}\| \leq 1 - r_k$.

When $1 - r_k < \|\bar{\boldsymbol{u}}\| \leq 1$ and $k = d$, that is, nonnegative constraints are imposed for all components of $\boldsymbol{u}$, there exists an index i such that the region $\mathcal{W} \cap V_d$ is included in $\{\boldsymbol{u} : \boldsymbol{e}_i^\top \boldsymbol{u} \geq 0\}$. If there is no such index, inequality $0 \leq \bar{u}_i < s \leq r_d$, holds for all $i = 1, \ldots, d$, and this contradicts $1 - r_d < \|\bar{\boldsymbol{u}}\|$. Without loss of generality, the nonnegative constraint, $u_i \geq 0$, can be deleted, when we evaluate the volume of $\mathcal{W} \cap V_d$. As the result, only $d - 1$ nonnegative constraints are essential, and thus, the discussion under the assumption $k < d$ holds even for $k = d$. ■

5 Numerical Simulations

5.1 Test Problem

We consider a robust linear program:

$$(P) \ \Big| \ \min_{\boldsymbol{x}} \max_{\boldsymbol{u}} \ \boldsymbol{q}(\boldsymbol{u})^\top \boldsymbol{x} \ \text{ s.t. } \ \boldsymbol{u} \in \mathcal{U}, \ \boldsymbol{x} \in \mathcal{X},$$

where $\boldsymbol{q}(\boldsymbol{u}) := \boldsymbol{q}_0 + \sum_{j=1}^d u_j \boldsymbol{q}_j$, $\mathcal{U} := \{\boldsymbol{u} \in \mathbb{R}^d : \|\boldsymbol{u}\| \leq 1\}$ and $\mathcal{X} := \{\boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x}\| \leq 1\}$. This problem can be described as a second order cone program, which is solvable via interior point methods. We measure the difference between the optimal value of (P) and that of its sampled problem (P_N) experimentally and theoretically. Theorem 4.1 guarantees with probability at least $1 - \eta$ that the difference of values $\mathrm{opt}(P) - \mathrm{opt}(P_N)$ is within $[0, q^{-1}(\epsilon)]$ for $\epsilon \in (0, q(B))$, where $q(\delta)$ is a uniform lower bound of tail probability for $0 \leq \delta \leq B$. Several techniques to construct uniform lower bound of tail probability is presented in previous sections. Especially, Theorem 3.2 replaces the confidence probability $1-\eta$ with $(1-\eta)\times(1-\beta)$, and provides a general way to construct a uniform lower bound $q_0(\delta)$ as

$$
\begin{aligned}
&q_0(\delta) = \frac{1}{(\widehat{w}+\tau)\mathrm{Vol}(\mathcal{C})} \min \left\{ \frac{A_d\left(\sin^{-1}\frac{\alpha^* r}{R-\alpha^* r}\right)}{A_d(\pi)} V_d\left(\frac{\delta}{L}\right),\ D_d\left(\frac{\delta}{L}, r\right) \right\}, \\
&0 \le \delta \le B_0 := rL\left(\sqrt{1-\alpha^{*2}} + \frac{2\sqrt{1-\alpha^*}}{2-\alpha^*}\right),
\end{aligned}
$$

where α^* is given by 6. Note that $\tau > 0$ is arbitrary, $\widehat{w}$ is defined as (9) and $\mathcal{C}$ is hypercube including $\mathcal{U}$. On the other hand, for the above problem (P) with the ellipsoidal uncertainty set $\mathcal{U}$, it is possible to utilize easily achievable function that is provided as

$$
q_1(\delta) = \frac{1}{V_d(1)} D_d\left(\frac{\delta}{L}, 1\right), \qquad 0 \le \delta \le B_1 := 2L.
$$

To form $q_1(\delta)$, Lipschitz constant L and the dimension d of $\boldsymbol{u}$ are necessary. Additionally, the definition of $q_0(\delta)$ requires the diameter R of $\mathcal{U}$, the radius r of inscribed hypersphere in $\mathcal{U}$, and an upper bound $(\widehat{w}+\tau)\mathrm{Vol}(\mathcal{C})$ for the volume $\mathrm{Vol}(\mathcal{U})$.

Now, we evaluate the probabilistic theoretical error in these two ways: one is $q_1^{-1}(\epsilon)$ devised for the the ellipsoidal uncertainty set $\mathcal{U}$, and the other is $q_0^{-1}(\epsilon)$ under the assumption that the shape of $\mathcal{U}$ is not exactly known, that is, R, r and $\mathrm{Vol}(\mathcal{U})$ are unknown.

Lipschitz constant L: Lipschitz constant L of the objective function $\boldsymbol{q}(\boldsymbol{u})^\top \boldsymbol{x}$ can be estimated as the square root of the maximum eigenvalue of matrix $\sum_{i=1}^d \boldsymbol{q}_i \boldsymbol{q}_i^\top$, that is, $\sqrt{\sigma_{max}(\sum_{i=1}^d \boldsymbol{q}_i \boldsymbol{q}_i^\top)}$, since the inequality

$$
\mid \boldsymbol{q}(\boldsymbol{u})^\top \boldsymbol{x} - \boldsymbol{q}(\boldsymbol{v})^\top \boldsymbol{x} \mid = \mid \boldsymbol{d}(\boldsymbol{x})^\top (\boldsymbol{u}-\boldsymbol{v}) \mid \le \max_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{d}(\boldsymbol{x})\| \times \|\boldsymbol{u}-\boldsymbol{v}\|
$$

holds for $\boldsymbol{d}(\boldsymbol{x}) := (\boldsymbol{q}_1^\top \boldsymbol{x}, \dots, \boldsymbol{q}_d^\top \boldsymbol{x})^\top$.

Diameter R of $\mathcal{U}$: A hypercube including $\mathcal{U}$ is obtained as $[-1,1]^d$ via $2d$ quadratic programs (7). Then, the length of diagonal line of the hypercube, $2\sqrt{d}$, provides an upper bound for the diameter R of $\mathcal{U}$. Therefore, we regard $2\sqrt{d}$ as diameter R.

Radius r of inscribed hypersphere: The radius $r=1$ is achieved by solving the optimization problem, which consists of the objective function r and a positive semi-definite constraint (8) induced from $\mathcal{U} = \{\boldsymbol{u}\in\mathbb{R}^d : \|\boldsymbol{u}\| \le 1\}$.

Volume $\mathrm{Vol}(\mathcal{U})$ of $\mathcal{U}$: $\mathrm{Vol}(\mathcal{U})$ is estimated approximately by Monte-Carlo methods. Since a hypercube including $\mathrm{Vol}(\mathcal{U})$ is already obtained as $\mathcal{C} = [-1,1]^d$, $\widehat{w}$ is available with (9) and $\mathrm{Vol}(\mathcal{C}) = 2^d$.

Table 1: Theoretically estimated error for $\mathrm{opt}(P) - \mathrm{opt}(P_N)$ $(\eta = 0.01)$

		$q_0^{-1}(\epsilon)$	R.err-0	$q_1^{-1}(\epsilon)$	R.err-1 [%]
$N(\epsilon, \eta) = 2.50 \times 10^3$	$\epsilon = 4.88 \times 10^{-2}$	–	–	1.98	5.67 %
$N(\epsilon, \eta) = 2.50 \times 10^4$	$\epsilon = 7.14 \times 10^{-3}$	4.04	11.58 %	1.01	2.89 %
$N(\epsilon, \eta) = 4.50 \times 10^4$	$\epsilon = 4.30 \times 10^{-3}$	3.42	9.80 %	0.84	2.41 %
$N(\epsilon, \eta) = 6.89 \times 10^7$	$\epsilon = 5.69 \times 10^{-6}$	0.37	1.06 %	0.09	0.25 %

Legend: '–' denotes that ϵ is out of range.

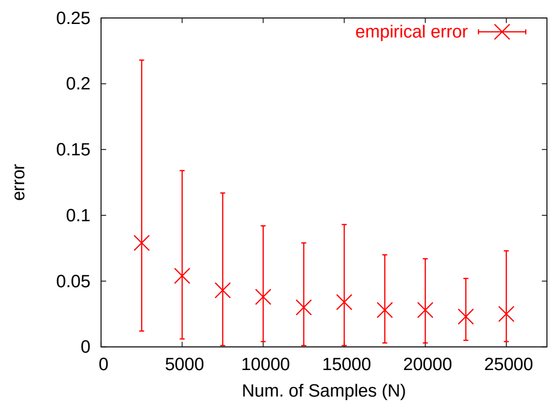

Figure 4: Empirical error $\mathrm{opt}(P) - \mathrm{opt}(P_N)$

5.2 A-priori Assessment

We constructed (P) with the dimension $m = 15$ of $\boldsymbol{x}$ and $d = 3$ of $\boldsymbol{u}$. The linear objective function consists of $\boldsymbol{q}_0 = (10, \dots, 10)^\top$ and $\boldsymbol{q}_j \in [-1, 1]^m$, $(j = 1, 2, 3)$, which are generated randomly. Then, Lipschitz constant $L = 4.01$. With the parameter setting such as $\alpha = 0.01$ and $\beta = 0.01$, we obtain an upper bound $(\widehat{w} + \alpha)\mathrm{Vol}(\mathcal{C}) = 4.23$ for $\mathrm{Vol}(\mathcal{U}) = 4.19$. Then, $B_0 = 6.13$ and $q_0(B_0) = 2.49 \times 10^{-2}$ are calculated.

Table 1 indicates the relation between the number of random samples $N(\epsilon, \eta)$ and the theoretical error measures (theoretical error $q_0^{-1}(\epsilon)$, $q_1^{-1}(\epsilon)$ and their theoretical relative errors) under the above parameter setting. The theoretical relative error is defined by R.err-$i := \frac{q_i^{-1}(\epsilon)}{|\mathrm{opt}(P)|} \times 100$ $(i = 0, 1)$. When the function q_1 is utilized to evaluate the error $\mathrm{opt}(P) - \mathrm{opt}(P_N)$ for an optimal solution of (P_N) with $N = 2.50 \times 10^4$ random samples, it is guaranteed with probability at least $1 - \eta (= 0.99)$ that the error is within 1.01, which corresponds to 2.89% theoretical relative error. The estimation accuracy, however, deteriorates with the use of generally constructed function q_0. Indeed, the evaluation of $q_0^{-1}(\epsilon)$ is guaranteed only in the small range $\epsilon \in (0,\ 2.49 \times 10^{-2})$, while $\epsilon \in (0, 1)$ for q_1. Moreover, the difference of estimations $q_0^{-1}(\epsilon)$ and $q_1^{-1}(\epsilon)$ is large with fixed $N(\epsilon, \eta)$.

Table 2: A-posteriori assessments for $\hat{\boldsymbol{x}}_{N_1}$ and $\hat{\boldsymbol{x}}_{N_2}$ $\quad (\eta = 0.01)$

	$\hat{\boldsymbol{x}}_{N_1}$ ($N_1 = 2500$)					
	δ_0	A.err-0	R.err-0	δ_1	A.err-1	R.err-1
$M = 8.68 \times 10^3$	1.607	1.659	4.75%	0.392	0.444	1.27 %
$M = 1.09 \times 10^6$	0.321	0.413	1.18%	0.077	0.169	0.48 %
$M = 1.36 \times 10^8$	0.064	0.164	0.47%	0.015	0.115	0.33 %
	$\hat{\boldsymbol{x}}_{N_2}$ ($N_2 = 45000$)					
	δ_0	A.err-0	R.err-0	δ_1	A.err-1	R.err-1
$M = 8.68 \times 10^3$	1.607	1.607	4.61 %	0.392	0.392	1.12 %
$M = 1.09 \times 10^6$	0.321	0.328	0.94 %	0.077	0.084	0.24 %
$M = 1.36 \times 10^8$	0.064	0.077	0.22 %	0.015	0.028	0.08 %

On the other hand, from an empirical point of view, less number N of samples may be required to attain an approximated solution $\hat{\boldsymbol{x}}_N$ with the empirical error $\mathrm{opt}(P) - \mathrm{opt}(P_N) \leq 1.01$. To show this, for each N, 100 different sets of random samples are drawn and sampled problems (P_N) constructed from the random samples are solved. Figure 4 shows the maximum, average and minimum values among 100 optimal values of (P_N) for each N. Certainly, the empirical error $\mathrm{opt}(P) - \mathrm{opt}(P_N)$ achieved with $N = 2.50 \times 10^4$ is far less than the theoretical error $q_1^{-1}(7.14 \times 10^{-3}) = 1.01$.

5.3 A-posteriori Assessment

Next, a-posteriori assessments are carried out for $\hat{\boldsymbol{x}}_{N_i}$ $(i = 1, 2)$, which are obtained via sampled problems (P_{N_i}) with $N_1 = 2500$ and $N_2 = 45000$, respectively. In a-posteriori assessment, we can utilize $\bar{q}(\delta, \hat{\boldsymbol{x}}_N)$, defined for the solution $\hat{\boldsymbol{x}}_N$, instead of uniform lower bound. For the concerned problem (P), $q_0(\delta)$ and $q_1(\delta)$ are reformed with Lipschitz constant $L = \|\boldsymbol{d}(\hat{\boldsymbol{x}}_N)\|$, where $\boldsymbol{d}(\boldsymbol{x}) := (\boldsymbol{q}_1^\top \boldsymbol{x}, \ldots, \boldsymbol{q}_d^\top \boldsymbol{x})^\top$. $L = 3.79$ is evaluated for $\hat{\boldsymbol{x}}_{N_i}$ $(i = 1, 2)$.

Now we compute function values and obtain $\max_{\boldsymbol{v} \in \mathcal{U}^{(M)}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{v})$ for some fixed $M > 0$. As a theoretical error measure, we define

$$\text{A.err-}i := \max_{\boldsymbol{v} \in \mathcal{U}^{(N)} \cup \mathcal{U}^{(M)}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{v}) - \max_{\boldsymbol{u} \in \mathcal{U}^{(N)}} f(\hat{\boldsymbol{x}}_N, \boldsymbol{u}) + \delta_i$$

for δ_i satisfying $M = \lceil \frac{\ln \eta}{\ln(1 - q_i(\delta_i))} \rceil$ with fixed M and $\eta = 0.01$. Theorem 4.2 guarantees with probability at least $1 - \eta$ that the error $\mathrm{opt}(P) - \mathrm{opt}(P_N)$ is less than A.err-i whenever $\delta_i \in (0, B_0]$ holds.

Table 2 shows theoretical error measures obtained via a-posteriori assessments for the solutions $\hat{\boldsymbol{x}}_{N_1}$ and $\hat{\boldsymbol{x}}_{N_2}$. R.err-$i$ is evaluated by $\frac{\text{A.err-}i}{\mathrm{opt}(P)}$ $(i = 1, 2)$. A-posteriori assessment with

$M = 1.36 \times 10^8$ ensures that these solutions $\hat{\boldsymbol{x}}_{N_i}$ $(i = 1, 2)$ approximate an optimal solution of (P) fairly well. Indeed, for $\hat{\boldsymbol{x}}_{N_1}$ with $N_1 = 2500$, a-posteriori assessment with function $q_1(\delta)$ guarantees that the relative error is within 0.33% with probability $1 - \eta$ $(= 0.99)$, while 5.67% relative error is guaranteed via a-priori assessment. When generally formulated function $q_0(\delta)$ is utilized for a-posteriori assessment, 0.47% and 0.22% relative errors are estimated for $\hat{\boldsymbol{x}}_{N_1}$ and $\hat{\boldsymbol{x}}_{N_2}$, respectively. Furthermore, as an advantage of a-posteriori assessment, we mention that the computational tasks necessary for a-posteriori assessment are just function evaluations and the assessment does not require computation for optimization. Therefore, even if the number M is large, computation of a-posteriori assessment is still tractable.

For achieving a nice approximation of an optimal solution of (P), it might be a clever way to solve (P_N) with appropriately large N and check the accuracy of the obtained solution $\hat{\boldsymbol{x}_N}$ via a-posteriori assessment. If the solution is sufficiently accurate, it can be accepted as almost optimal solution of (P).

6 Concluding Remarks

Sampling convex programs are applied to solve uncertain convex programs. Calafiore and Campi [4] have proposed sufficient number of random samples to achieve small violation probability, and along this line, we stochastically evaluate the worst-case violation. To derive an upper bound of the worst-case violation, uniform lower bound of tail probability $q(\delta)$ has important role. For general uncertainty set $\mathcal{U}$, we construct the function $q(\delta)$, and show a simple way of estimating $q(\delta)$. Using such $q(\delta)$, we give the relation between violation probability and the worst-case violation, which provides an upper bound of joint probability such that both violation probability and the worst-case violation take small values. The uniform lower bound $q(\delta)$ is also useful for a-posteriori assessment, which derives a reliable upper bound of the worst-case violation at an optimal solution of sampled convex program. We apply our results on the worst-case violation to min-max optimization problems, and derive upper bounds for optimal values.

Simple numerical simulations are also shown. The number of random samples tends to be large to make the worst-case violation enough small before solving optimization problems. However, throughout our numerical simulations, it is confirmed that a-posteriori assessment is effective to estimate the worst-case violation at an optimal solution of sampled convex program. Indeed, the a-posteriori assessment is carried out with high accuracy without heavy computational costs.

When uncertainty set is nonconvex, or more complicated shape, the uniform lower bound $q_0(\delta)$ is not so tight for tail probability. As the result, the number of random samples tends to be large, and the optimization problem will be intractable. Thus, as a future work, it is important to propose a way of reducing the number of randomly sampled constraint functions.

References

[1] A. Ben-Tal and A. Nemirovski: Robust convex optimization, *Mathematics of Operations Research*, 23 (4): 769–805 (1998).

[2] A. Ben-Tal and A. Nemirovski: Robust solutions of uncertain linear programs, *Math. Oper. Res.,* 25 (1): 1–13 (1999).

[3] Boyd, S., and Vandenberghe, L., *Convex Optimization* Cambridge Univ. Press, (2003).

[4] G. Calafiore and M.C. Campi: A new bound on the generalization rate of sampled convex programs, *43rd IEEE Conference on Decision and Control (CDC04)*, **5**, 5328-5333 (2004).

[5] G. Calafiore and M.C. Campi: The Scenario Approach to Robust Control Design, *IEEE Transactions on Automatic Control*, to appear.

[6] G. Calafiore and M.C. Campi: Uncertain convex programs: randomized solutions and confidence levels, *Mathematical Programming*, **102**, 25-46 (2005).

[7] G. Calafiore and L. E Ghaoui, Distributionally Robust Chance-Constrained Linear Programs with Applications, *Journal of Optimization Theory and Applications*, to appear.

[8] A. Charnes and W.W. Cooper: Chance constrained programming, *Manag. Sci.*, **6**, 73-79 (1959).

[9] L. Devroye and G. Lugosi: *Combinatorial Methods in Density Estimation Springer*, (2001).

[10] El Ghaoui and L., H. Lebret: Robust solutions to least-squares problems with uncertain data, *SIAM J. Matrix Anal. Appl.* **18**, 1035–1064 (1997).

[11] El Ghaoui, L., H. Lebret: Robust solutions to uncertain semidefinite programs, *SIAM J. Optim.* **9**, 33-52 (1998).

[12] D. Goldfarb and G. Iyengar: Robust convex quadratically constrained programs, *Mathematical Programming*, **97**, 495-515 (2003).

[13] D. Goldfarb and G. Iyengar: Robust portfolio selection problems, *Mathematics of Operations Research*, **28**, 1-38 (2003).

[14] P. Kouvelis and G. Yu: *Robust Discrete Optimization and its Applications*, Kluwer Academic Publishers, (1997).

[15] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret: Applications of second-order cone programming, Linear Algebra and its Applications, **284**, 193-228 (1998).

[16] A. Prékopa: *Stochastic Programming*, Kluwer Academic Publishers, (1995).

[17] R. Tempo, G. Calafiore, and F. Dabbene: *Randomized Algorithms for Analysis and Control of Uncertain Systems*, pSpringer-Verlag London Limited, (2005).

[18] S. Uryasev (editor): *Probabilistic Constrained Optimization: Methodology and Applications*, New York, (2000).

[19] L. Vandenberghe and S. Boyd: Semidefinite programming, SIAM Rev. **38**, 49-96 (1996).