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

Takafumi Kanamori[†] and Akiko Takeda[‡]

Abstract. Uncertain programs have been developed to deal with optimization problems including inexact data, *i.e.,* uncertainty. A deterministic approach called robust optimization is commonly applied to solve these problems. Recently, Calafiore and Campi have proposed a randomized approach based on sampling of constraints, where the number of samples is determined so that only small portion of original constraints is violated at the randomized solution. Our main concern is not only the probability of violation, but also the degree of violation *i.e.,* the worst-case violation. We derive an upper bound of the worst-case violation for the sampled convex programs and consider the relation between the probability of violation and worst-case violation. The probability of violation and the degree of violation are simultaneously bounded by small values, when the number of random samples is sufficiently large. Our method is applicable to not only a bounded uncertainty set but also an unbounded one such as Gaussian uncertain variables.

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

† Department of Computer Science and Mathematical Informatics, Nagoya University, Furocho, Chikusa-ku, Nagoya-shi, Aichi 464-8603, Japan. *kanamori@is.nagoya-u.ac.jp*

‡ Department of Administration Engineering, Keio University, 3-14-1 Hiyoshi, Kouhoku-ku, Yokohama, Kanagawa 223-8522, Japan. *takeda@ae.keio.ac.jp*

1 Introduction

Uncertain programs have been developed to deal with optimization problems including inexact data, *i.e.,* uncertainty. In realistic decision-making environments, the exact forms of objective functions or constraint functions are generally not known, since optimization problems are constructed based on observed data for which contamination is inevitable. Some information on the future may be required to formalize exact cost functions or constraints. The uncertain convex program (UCP) is formalized as a convex program under uncertain constraints. Without loss of generality, an objective function is assumed to be linear without uncertainty parameters, and only constraints have 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 expressed as $f(\boldsymbol{x}, \boldsymbol{u}) \leq 0$, where $\boldsymbol{x}$ is optimization variable and f is convex in $\boldsymbol{x}$. We do not know which uncertainty parameter will be realized; therefore, we need to deal with all the 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 standard approaches to UCP, see [1, 2, 10–13]. In the RCP approach, one looks for a solution that is feasible under all possible realizations of uncertainty parameters. This implies that the worst-case constraints are considered. The feasible region for RCP is described as $\{\boldsymbol{x} \mid f(\boldsymbol{x}, \boldsymbol{u}) \leq 0, \ \forall \boldsymbol{u} \in \mathcal{U}\}$. In some applications, the worst-case scenario needs to be considered, if the violation of constraints causes significant detriments. From the viewpoint of computation, although RCP is still a convex optimization problem, it is generally difficult to solve it numerically, because RCP may have infinitely many constraints. If the constraint function f and uncertainty set $\mathcal{U}$ satisfy some nice properties, RCP can be reduced to a tractable problem. For example, robust linear programs and robust second order cone programs result in second order cone programs and semidefinite programs, respectively [15].

Sampled convex program (SCP) is a practical alternative to RCP for problems with uncertainty. The purpose of SCP is to find an approximate solution for RCP, which satisfies almost all uncertainty constraints, while retaining the tractability. In SCP, a probability distribution P is defined on the uncertainty set $\mathcal{U}$, and the set of constraints for SCP is defined by random samples from P. Let random samples, $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N$, be N independent and identically distributed from P over $\mathcal{U}$. Then, the feasible region for SCP is defined as $\{\boldsymbol{x} \mid f(\boldsymbol{x}, \boldsymbol{u}_i) \leq 0, \ i = 1, \ldots, N\}$, which depends on the realization of random samples. SCP is a convex optimization problem with finite number of constraints, and therefore, it is tractable for a wide range of UCPs. This is an advantage of SCP over RCP, although resulting solutions do not necessarily satisfy all the constraints. Random sampling is a widely used technique in practical situations.

Estimation of the number of random samples N is important to guarantee that the resulting solution violates only a small portion of constraints. Calafiore and Campi [4, 6] and Campi and Garatti [8] have defined *violation probability* for randomized solutions, and proposed some practical lower bounds for the number of random samples to achieve a small violation probability. If the number of random samples N is determined properly according to the criterion proposed by Calafiore and Campi, the obtained solution satisfies almost all the constraints with high probability. Although the fact that the violation probability is equal to zero does not necessarily denote that all the constraints are satisfied, solutions with small violation probability are acceptable in many practical situations.

Chance constrained program (CCP) is one of the classical approaches for solving optimization problems with uncertainty [9], which also uses probability distributions over the uncertainty set $\mathcal{U}$. In general, the constraint set for CCP is nonconvex, and therefore, it is difficult to obtain exact solutions. Some researchers have studied the relation between CCP and SCP, see [6, 7].

In this study, we investigate the statistical properties of the values of constraint functions at the optimal solutions of SCP. Calafiore and Campi have mainly focused on percentage of uncertain constraints violated at an optimal solution of SCP. It is also important to assess the extent to which the solution violates each constraint function. In many applications, the degree of violation of a solution is measured with respect to the value of each constraint function, and plays a significant role in assessing the validity of the approximated solution. Therefore, we focus on the values of constraint functions for UCP.

The degree of violation at a given solution $\boldsymbol{x}$ for SCP is governed by the worst-case violation, which is defined as $\max_{\boldsymbol{u}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{u})$. Our main concern is to obtain an upper bound of the worst-case violation in a solution of SCP. When the uncertainty set $\mathcal{U}$ is unbounded, the value of $\max_{\boldsymbol{u}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{u})$ can be infinite. In such a case, we consider $\max_{\boldsymbol{u}\in\mathcal{A}} f(\boldsymbol{x},\boldsymbol{u})$ for a compact set $\mathcal{A}\subset\mathcal{U}$ instead of $\max_{\boldsymbol{u}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{u})$, and as the result $\max_{\boldsymbol{u}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{u})$ can be assessed up to a certain probability. To assess the worst-case violation, we need to evaluate the tail probability of $f(\boldsymbol{x},\boldsymbol{u})$, *i.e.,* the probability that $f(\boldsymbol{x},\boldsymbol{u})$ has a greater value than $\max_{\boldsymbol{u}\in\mathcal{U}} f(\boldsymbol{x},\boldsymbol{u})-\delta$ for some $\delta>0$. Under some conditions, a uniform lower bound of the tail probability is derived, which is applied to evaluate the worst-case violation. Moreover, we study the relation between the violation probability defined by Calafiore and Campi, and the worst-case violation. The worst-case violation and violation probability are assured to take small values with high probability in case of a large number of random samples. By exploiting the uniform lower bound of the tail probability, we can estimate the number of random samples required to obtain a sufficiently accurate solution before solving SCP. This is a priori evaluation. Min-max problems are common applications of UCP, and our results can be applied to determine the optimal value of min-max problems.

A criterion for a posteriori assessment is also proposed. After obtaining an optimal solution of SCP, one can make a posteriori assessment of the worst-case violation in the solution with small computation cost. The assessment is useful to evaluate the worst-case violation with high accuracy. In general, the uniform lower bound of the tail probability is not sufficiently tight in a priori evaluation, and therefore, the number of samples required to achieve high accuracy becomes quite large. Although it is intractable to solve SCP with such a large amount of constraint functions, a posteriori assessment at a given solution is still tractable, because in the phase of a posteriori assessment, we require to compute the values of constraint functions and not solve the optimization problem.

The paper is organized as follows. In Section 2, we introduce some results of sampled convex programs proposed by Calafiore and Campi [5] and Campi and Garatti [8]. Then, we define violation probability, the worst-case violation, and uniform lower bound of tail probability. In Section 3, illustrative examples are presented before discussing the details. In Section 4, the probabilistic assessment of the worst-case violation is presented. Here, we assume a condition that there exists a uniform lower bound of the tail probability. In Section 5, uniform lower bounds are derived under several conditions. Results of the numerical studies are presented in Section 6. Section 7 is devoted to concluding remarks. Long proofs are provided in the Appendix.

2 Worst-case Violation of Sampled Convex Programs

A general robust convex program is described as

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

where $f(\boldsymbol{x},\boldsymbol{u})$ is a convex function in $\boldsymbol{x}$, and $\mathcal{X}$ is a convex subset in $\mathbb{R}^m$. Throughout the paper, we assume the feasibility of (RCP). Calafiore and Campi [4,5], and Campi and Garatti [8] have considered sampled convex program (SCP) as an approximation of (RCP),

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

where $U_i,\ i=1,\dots,N$, are N independent identically distributed random variables on $\mathcal{U}$. Let P be the probability distribution of these random variables. The fundamental issue concerning randomized optimization problems is the number of samples needed to achieve the required accuracy of the solution. The feasible region of (SCP_N) is regarded as an approximation for the original robust optimization problem (RCP), which consists of infinitely many constraints. Note that the feasible region of (SCP_N) depends on random variables, and therefore, probabilistic discussions are useful to deal with convergence properties of optimal solutions.

In general, the optimal solution of (SCP_N) is not necessarily feasible under the constraints of (RCP). The ratio of violated constraints in the solution is measured by violation probability [4] defined below.

Definition 2.1. *Let Q be a probability distribution of a $\mathcal{U}$-valued random variable U'. Violation probability at $\boldsymbol{x}\in\mathcal{X}$ with respect to Q is defined by*

$$V(\boldsymbol{x}) = Q\{0 < f(\boldsymbol{x},U')\},$$

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

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

In the original definition of violation probability [4], the distribution for sampling P coincides with that for assessment Q. In this study, we also examine the case that P is not necessarily equal to Q. Throughout the paper, let P be the conditional probability determined by Q, *i.e.,* we assume the following.

Assumption 2.2. *Let $\mathcal{A}$ be a measurable subset of $\mathcal{U}$ such that $Q(\mathcal{A}) > 0$. The distribution P for sampling $U_1, \ldots, U_N$ is the conditional probability of Q on $\mathcal{A}$ defined as*

$$P(B) = Q(B|\mathcal{A}) = \frac{Q(B \cap \mathcal{A})}{Q(\mathcal{A})}$$

for any measurable subset $B \subset \mathcal{U}$.

Note that $P = Q$ holds if $\mathcal{A} = \mathcal{U}$. All the propositions in this paper are established under Assumption 2.2. In a later section, as a practical situation, we investigate the case where $\mathcal{A}$ is bounded while $\mathcal{U}$ is unbounded.

Calafiore and Campi [4, 5] and Campi and Garatti [8] have proposed that a solution of (SCP_N) fails to satisfy only a small portion of the original constraints, when sufficient number of samples are considered.

Theorem 2.3 (Campi and Garatti [8]). *Suppose that $\mathcal{A} = \mathcal{U}$ and that (SCP_N) is feasible. For a level parameter $\varepsilon \in (0, 1)$ and a confidence parameter $\eta \in (0, 1)$ let $N(\varepsilon, \eta)$ be*

$$N(\varepsilon, \eta) = \min\left\{N \in \mathbb{N} \mid \sum_{i=0}^{m-1} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i} \leq \eta\right\},$$

where m is the dimension of $\mathcal{X}$. The solution of (SCP_N) is denoted as $\widehat{\boldsymbol{x}}_N$. Then, for any positive integer $N \geq N(\varepsilon, \eta)$, the inequality

$$P^N\{V(\widehat{\boldsymbol{x}}_N) > \varepsilon\} \leq \eta$$

holds true, where $P^N\{\cdots\}$ denotes the probability over N independent random samples, $U_1, \ldots, U_N$.

Note that $N(\varepsilon, \eta)$ can be estimated before solving the optimization problems.

In Theorem 2.3, while the probability of violation is considered, the value of $f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u})$ is not considered. In this study, we consider the value of constraint function $f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u})$ on a subset of $\mathcal{U}$. The degree of violation on a subset of $\mathcal{U}$ is measured by the worst-case violation.

Definition 2.4. *Let $\mathcal{A}$ be a subset of $\mathcal{U}$. When $\max_{\boldsymbol{u} \in \mathcal{A}} f(\boldsymbol{x}, \boldsymbol{u}) > 0$,* the worst-case violation *of $\boldsymbol{x}$ over $\mathcal{A}$ is defined as $\max_{\boldsymbol{u} \in \mathcal{A}} f(\boldsymbol{x}, \boldsymbol{u})$.*

Here, we study the worst-case violation over a subset of the uncertainty set $\mathcal{U}$, because the worst-case violation over $\mathcal{U}$ is typically infinite. If the worst-case violation over the subset $\mathcal{A}$ is upper bounded, we can derive a probabilistic upper bound of the constraint functions of the set

$\mathcal{U}$. Although the worst-case violation over a subset $\mathcal{A}$ at the optimal solution $\widehat{\boldsymbol{x}}_N$ is generally greater than zero, it attains a non-positive value as N tends to infinity under conditions such as feasibility of (RCP).

We introduce the tail probability and the uniform lower bound, which are crucial concepts in the study.

Definition 2.5. *Let $\mathcal{A}$ be a subset of $\mathcal{U}$. The tail probability of the worst-case violation at $\boldsymbol{x}$ over $\mathcal{A}$ is defined as*

$$p(\delta, \boldsymbol{x}) = P\big\{ \max_{\boldsymbol{u}\in\mathcal{A}} f(\boldsymbol{x}, \boldsymbol{u}) - \delta < f(\boldsymbol{x}, U) \big\}, \quad U \sim P.$$

If there exists a strictly increasing function $q(\delta)$ for $\delta \in [0, B]$ satisfying

$$0 \le q(\delta) \le p(\delta, \boldsymbol{x}), \quad 0 \le \delta \le B, \ ^\forall \boldsymbol{x} \in \mathcal{X},$$

the function $q(\delta)$ is called the uniform lower bound of the tail probability of $\mathcal{A}$.

Although both P and $p(\delta, \boldsymbol{x})$ depend on $\mathcal{A}$, we drop the subscript to avoid any confusion.

If $\max_{\boldsymbol{u}\in\mathcal{A}} f(\boldsymbol{x}, \boldsymbol{u})$ is infinite, the tail probability $p(\delta, \boldsymbol{x})$ is equal to zero for any $\delta \ge 0$ and $\boldsymbol{x} \in \mathcal{X}$, and a uniform lower bound of the tail probability does not exist. Finding a subset $\mathcal{A}$ that induces a non-trivial tail probability is the key to derive an upper bound of the worst-case violation. In Section 5, we derive $q(\delta)$ for a Lipschitz continuous constraint $f(\boldsymbol{x}, \boldsymbol{u})$ on a bounded set $\mathcal{A}$.

3 Illustrative Examples

Before discussing the details, we show some numerical examples. Here, it is assumed that the constraint function $f(\boldsymbol{x}, \boldsymbol{u})$ is Lipschitz continuous on $\mathcal{X} \times \mathcal{U}$, and that the Lipschitz constant is less than or equal to one. Under the assumption, there exists a uniform lower bound, as shown in Section 5.

When the uncertainty set $\mathcal{U}$ is convex and compact, we can derive an upper bound of the worst-case violation over $\mathcal{U}$, *i.e.,* it holds the inequality of the form,

$$P^N\big\{V(\widehat{\boldsymbol{x}}_N) \le \varepsilon, \ \max_{\boldsymbol{u}\in\mathcal{U}} f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u}) \le q^{-1}(\varepsilon)\big\} \ge 1 - \eta, \quad N \ge N(\varepsilon, \eta)$$

as shown in Theorem 4.3. Here, the subset $\mathcal{A}$ is equal to $\mathcal{U}$, and therefore, the violation probability is assessed under the distribution P. Let $\mathcal{U}$ be a unit hypersphere $\{\boldsymbol{u} \in \mathbb{R}^d \mid \|\boldsymbol{u}\| \le 1\}$, and P be the truncated normal distribution on $\mathcal{U}$, which is derived from the normal distribution with mean $\mathbf{0}$ and identity matrix I_d as a variance-covariance matrix. The values of ε, $q_1^{-1}(\varepsilon)$, and $N(\varepsilon, \eta)$ are listed in Table 1, when $\eta = 0.01$, $d = 3, 10$, or 20, and the dimension of $\mathcal{X}$ is 10.

On the other hand, when the uncertainty set $\mathcal{U}$ is not compact, the worst-case violation over $\mathcal{U}$ may be infinite. Even in such a situation, we can derive a meaningful upper bound for the worst-case violation up to a certain probability. We consider the case of $\mathcal{U} = \mathbb{R}^d$. The distribution Q on $\mathcal{U}$ is the normal distribution with mean $\mathbf{0}$ and identity matrix I_d as

Table 1: Probabilistic assessment of the violation probability and the worst-case violation is expressed as $P^N\{V(\widehat{\boldsymbol{x}}_N) \leq \varepsilon,\ \max_{\boldsymbol{u}\in\mathcal{U}} f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u}) \leq q_1^{-1}(\varepsilon)\} \geq 0.99$ for $N \geq N(\varepsilon, \eta)$ ($\eta = 0.01$, $d = 3, 10$ or 20, $m = 10$) with the uncertainty set $\mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^d \mid \|\boldsymbol{u}\| \leq 1\}$.

	$q_1^{-1}(\varepsilon)$			
ε	$d = 3$	$d = 10$	$d = 20$	$N(\varepsilon, \eta)$
0.001	0.135237	0.584043	0.832902	18779
0.003	0.195319	0.661465	0.893968	6257
0.005	0.231805	0.701803	0.924900	3752
0.007	0.259594	0.730140	0.946302	2679
0.009	0.282539	0.752290	0.962893	2083

a variance-covariance matrix. Let $\mathcal{A}$ be a hypersphere centered at $\mathbf{0} \in \mathbb{R}^d$ with radius R_α, *i.e.,* $\mathcal{A} = \{\boldsymbol{u} \in \mathbb{R}^d \mid \|\boldsymbol{u}\| \leq R_\alpha\}$, where $\|\boldsymbol{u}\|$ is the Euclidean norm of $\boldsymbol{u}$, and R_α is determined such that $Q(\mathcal{A}) = 1 - \alpha$. The probability distribution P is the truncated normal distribution due to Assumption 2.2. As shown in Theorem 4.4, the inequality

$$P^N\left\{V(\widehat{\boldsymbol{x}}_N) \leq \alpha + \varepsilon - \alpha\varepsilon,\ Q(f(\widehat{\boldsymbol{x}}_N, U') > q^{-1}(\varepsilon)) \leq \alpha\right\} \geq 1 - 2\eta, \quad N \geq N(\varepsilon, \eta)$$

holds. Function q is a uniform lower bound of tail probability on $\mathcal{A}$, and some examples of q are illustrated in Section 5. The above inequality assures the following: for most optimal solutions of randomly generated problems (SCP_N), the violation probability is small, and the worst-case violation is at most $q^{-1}(\varepsilon)$. Table 2 shows the values of ε, $\alpha + \varepsilon - \alpha\varepsilon$, α, $q^{-1}(\varepsilon)$, and $N(\varepsilon, \eta)$, where $\eta = 0.005$, $d = 3, 10$, or 20, the dimension of $\mathcal{X}$ is 10, and $q = q_1$ defined in Lemma 5.2.

Previous studies considered only the violation probability $V(\widehat{x}_N)$ under $\mathcal{U} = \mathcal{A}$ [5, 6, 8]. On the other hand, our approach provides a probabilistic assessment of the worst-case violation.

4 Probabilistic Assessment

We study the upper bound of the worst-case violation. First, we consider the case of $\mathcal{A} = \mathcal{U}$, then extend the result to the case where $\mathcal{A} \subset \mathcal{U}$. We assume the existence of the uniform lower bound.

Assumption 4.1. *There exists a uniform lower bound $q(\delta)$ of the tail probability $p(\delta, \boldsymbol{x})$ on $\mathcal{A}$ for $0 \leq \delta \leq B$.*

In Section 5, we will describe some sufficient conditions for the existence of a uniform lower bound. In this section, we discuss a probabilistic assessment of the worst-case violation under the Assumption 4.1. The latter half of this section is devoted to a posteriori assessment of the worst-case violation. Once an optimal solution of (SCP_N) is obtained, probabilistic assessment of the worst-case violation of the solution is simply performed by random sampling on the uncertainty set.

Table 2: Probabilistic assessment of the violation probability and the worst-case violation is expressed as $P^N\left\{V(\widehat{\boldsymbol{x}}_N) \le \alpha+\varepsilon-\alpha\varepsilon,\ Q(f(\widehat{\boldsymbol{x}}_N, U') > q_1^{-1}(\varepsilon)) \le \alpha\right\} \ge 0.99$ for $N \ge N(\varepsilon,\eta)$ ($\eta = 0.005$, $d = 3, 10$ or 20, $m = 10$) with the uncertainty set $\mathcal{U} = \mathbb{R}^d$.

			$q_1^{-1}(\varepsilon)$			
ε	α	$\alpha+\varepsilon-\alpha\varepsilon$	$d=3$	$d=10$	$d=20$	$N(\varepsilon,\eta)$
0.001	0.01	0.01099	0.893106	2.929546	4.792834	19993
0.003	0.01	0.01297	1.155058	3.217771	5.084373	6661
0.005	0.01	0.01495	1.294734	3.364612	5.231991	3995
0.007	0.01	0.01693	1.393591	3.466617	5.334244	2852
0.009	0.01	0.01891	1.471216	3.545797	5.413517	2217
0.001	0.00901	0.01	0.914078	2.952241	4.815027	19993
0.003	0.00702	0.01	1.235271	3.296353	5.160170	6661
0.005	0.00503	0.01	1.457293	3.517713	5.378586	3995
0.007	0.00302	0.01	1.681087	3.730363	5.585356	2852
0.009	0.00101	0.01	2.019347	4.036630	5.877676	2217

4.1 A priori Assessment of the Worst-case Violation

The following lemma explains the relation between the worst-case violation and the uniform lower bound of the tail probability.

Lemma 4.2. *Under the Assumption 4.1 if* $P\{0 < f(\boldsymbol{x}, U)\} \le \varepsilon$ *holds for* $\varepsilon \in (0, q(B))$*, the inequality*

$$\max_{\boldsymbol{u}\in\mathcal{A}} f(\boldsymbol{x},\boldsymbol{u}) \le q^{-1}(\varepsilon)$$

is satisfied, where q^{-1} *is the inverse function of* q*.*

Proof. If the inequality $\max_{\boldsymbol{u}\in\mathcal{A}} f(\boldsymbol{x},\boldsymbol{u}) > B$ holds, we obtain

$$\varepsilon < q(B) \le P\big\{\max_{\boldsymbol{u}\in\mathcal{A}} f(\boldsymbol{x},\boldsymbol{u}) - B < f(\boldsymbol{x},U)\big\} \le P\big\{0 < f(\boldsymbol{x},U)\big\} \le \varepsilon,$$

which is a contradiction. Therefore, $\max_{\boldsymbol{u}\in\mathcal{A}} f(\boldsymbol{x},\boldsymbol{u}) \le B$ holds. If $\max_{\boldsymbol{u}\in\mathcal{A}} f(\boldsymbol{x},\boldsymbol{u}) \le 0$, the inequality of the lemma holds. On the other hand, applying the assumption 4.1, we obtain

$$q\big(\max_{\boldsymbol{u}\in\mathcal{A}} f(\boldsymbol{x},\boldsymbol{u})\big) \ \le\ p\big(\max_{\boldsymbol{u}\in\mathcal{A}} f(\boldsymbol{x},\boldsymbol{u}),\, \boldsymbol{x}\big) \ =\ P\{0 < f(\boldsymbol{x},U)\} \ \le\ \varepsilon.$$

■

We derive a probabilistic inequality for an upper bound of the worst-case violation.

Theorem 4.3. *Suppose that* $\mathcal{A} = \mathcal{U}$*, and Assumption 4.1 holds. For* $\varepsilon \in (0, q(B))$ *and* $\eta \in (0,1)$*, we assume* $N \ge N(\varepsilon,\eta)$*. Then, for an optimal solution* $\widehat{\boldsymbol{x}}_N$ *of* (SCP_N) *the inequality of the joint probability,*

$$P^N\big\{V(\widehat{\boldsymbol{x}}_N) \le \varepsilon,\ \max_{\boldsymbol{u}\in\mathcal{U}} f(\widehat{\boldsymbol{x}}_N,\boldsymbol{u}) \le q^{-1}(\varepsilon)\big\} \ge 1-\eta$$

holds.

Proof. It holds that $Q = P$ because $\mathcal{A} = \mathcal{U}$. The joint probability is lower bounded as

$$P^N\{\max_{\boldsymbol{u}\in\mathcal{U}} f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u}) \le q^{-1}(\varepsilon) \mid V(\widehat{\boldsymbol{x}}_N) \le \varepsilon\} P^N\{V(\widehat{\boldsymbol{x}}_N) \le \varepsilon\} = P^N\{V(\widehat{\boldsymbol{x}}_N) \le \varepsilon\} \ge 1 - \eta.$$

The first equality is the direct conclusion of Lemma 4.2, and the inequality is a result of Theorem 2.3. ∎

Theorem 4.3 is applicable for a priori assessment of the optimal solution in (SCP_N). In Section 5, we show a specific form of the uniform lower bound in the case that the uncertainty set $\mathcal{U}$ is compact, *e.g.* a hypersphere. Even if there does not exist any uniform lower bound of tail probability on $\mathcal{U}$, a uniform bound on a subset $\mathcal{A}$ can exist. The worst-case violation over $\mathcal{A}$ is upper bounded using the uniform lower bound.

Theorem 4.4. *Let $\alpha = 1 - Q(\mathcal{A}) \in (0, 1)$. Under the Assumption 4.1, the following inequalities hold.*

1. *For $0 < \varepsilon < q(B)$, $0 < \eta < 1$, and $N \ge N(\varepsilon, \eta)$,*

$$P^N\{Q(f(\widehat{\boldsymbol{x}}_N, U') \le q^{-1}(\varepsilon)) \ge 1 - \alpha\} \ge 1 - \eta, \qquad U' \sim Q. \tag{1}$$

2. *For $0 < \varepsilon < q(B)$, $0 < \eta < 1/2$, and $N \ge N(\varepsilon, \eta)$,*

$$P^N\{Q(f(\widehat{\boldsymbol{x}}_N, U') \le 0) \ge (1-\varepsilon)(1-\alpha),\ Q(f(\widehat{\boldsymbol{x}}_N, U') \le q^{-1}(\varepsilon)) \ge 1 - \alpha\} \ \ge\ 1 - 2\eta. \tag{2}$$

Proof. First, we prove the inequality (1). Let a subset E be

$$E = \{\boldsymbol{u} \in \mathcal{U} \mid f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u}) \le q^{-1}(\varepsilon)\} \subset \mathcal{U},$$

which depends on the solution $\widehat{\boldsymbol{x}}_N$. For $0 < \delta < 1$, the probability $P^N\{Q(E) < \delta\}$ is upper bounded as

$$\begin{aligned}
& P^N\{Q(E) < \delta\} \\
= \ & P^N\{(1-\alpha)P(E) + Q(E \cap \mathcal{A}^c) < \delta\} \\
\le \ & P^N\{P(E) < \frac{\delta}{1-\alpha}\} \\
= \ & P^N\{P(E) < \frac{\delta}{1-\alpha} \mid P(0 < f(\widehat{x}_N, U)) \le \varepsilon\} P^N\{P(0 < f(\widehat{x}_N, U)) \le \varepsilon\} \\
+ \ & P^N\{P(E) < \frac{\delta}{1-\alpha} \mid P(0 < f(\widehat{x}_N, U)) > \varepsilon\} P^N\{P(0 < f(\widehat{x}_N, U)) > \varepsilon\} \\
\le \ & P^N\{P(E) < \frac{\delta}{1-\alpha} \mid P(0 < f(\widehat{x}_N, U)) \le \varepsilon\} P^N\{P(0 < f(\widehat{x}_N, U)) \le \varepsilon\} + \eta.
\end{aligned}$$

The last inequality is obtained by applying $P^N\{P(0 < f(\widehat{x}_N, U)) > \varepsilon\} \le \eta$ for $N \ge N(\varepsilon, \eta)$. Note that $P(E) = 1$ if $P(0 < f(\widehat{x}_N, U)) \le \varepsilon$ by virtue of Lemma 4.2, and therefore, the equality

$$P^N\{P(E) < \frac{\delta}{1-\alpha} \mid P(0 < f(\widehat{x}_N, U)) \le \varepsilon\} = \begin{cases} 0 & \delta \le 1 - \alpha \\ 1 & \delta > 1 - \alpha \end{cases}$$

holds. Eventually, we obtain the inequality $P^N\{Q(E) < 1-\alpha\} \leq \eta$ by substituting $\delta = 1-\alpha$. As a result, the inequality

$$P^N\{Q(E) \geq 1-\alpha\} \geq 1-\eta.$$

is valid.

Next, we prove the inequality (2). If $P(0 < f(\widehat{\boldsymbol{x}}_N, U)) \leq \varepsilon$, then the inequalities

$$\begin{aligned} Q(0 < f(\widehat{\boldsymbol{x}}_N, U')) &= Q(0 < f(\widehat{\boldsymbol{x}}_N, U')|U' \in \mathcal{A})Q(U' \in \mathcal{A}) + Q(0 < f(\widehat{\boldsymbol{x}}_N, U')|U' \notin \mathcal{A})Q(U' \notin \mathcal{A}) \\ &\leq (1-\alpha)Q(0 < f(\widehat{\boldsymbol{x}}_N, U') \mid U' \in \mathcal{A}) + \alpha \\ &\leq (1-\alpha)\varepsilon + \alpha \end{aligned}$$

holds, because $Q(0 < f(\widehat{\boldsymbol{x}}_N, U') \,|\, U' \in \mathcal{A}) = P(0 < f(\widehat{\boldsymbol{x}}_N, U))$. Theorem 2.3 and the above inequality imply

$$\begin{aligned} 1-\eta &\leq P^N\big\{P(0 < f(\widehat{\boldsymbol{x}}_N, U)) \leq \epsilon\big\} \\ &\leq P^N\big\{Q(0 < f(\widehat{\boldsymbol{x}}_N, U')) \leq (1-\alpha)\varepsilon + \alpha\big\}. \end{aligned} \quad (3)$$

From (1) and (3), we have

$$\begin{aligned} &P^N\big\{Q(f(\widehat{\boldsymbol{x}}_N, U') \leq q^{-1}(\varepsilon)) < 1-\alpha\big\} \;\leq\; \eta, \\ &P^N\big\{Q(f(\widehat{\boldsymbol{x}}_N, U') \leq 0) < (1-\varepsilon)(1-\alpha)\big\} \;=\; P^N\big\{Q(0 < f(\widehat{\boldsymbol{x}}_N, U')) > (1-\alpha)\varepsilon + \alpha\big\} \;\leq\; \eta. \end{aligned}$$

Therefore, the inequality

$$\begin{aligned} &P^N\big\{Q(f(\widehat{\boldsymbol{x}}_N, U') \leq 0) \geq (1-\varepsilon)(1-\alpha),\; Q(f(\widehat{\boldsymbol{x}}_N, U') \leq q^{-1}(\varepsilon)) \geq 1-\alpha\big\} \\ &= 1 - P^N\big\{Q(f(\widehat{\boldsymbol{x}}_N, U') \leq 0) < (1-\varepsilon)(1-\alpha) \text{ or } Q(f(\widehat{\boldsymbol{x}}_N, U') \leq q^{-1}(\varepsilon)) < 1-\alpha\big\} \\ &\geq 1 - P^N\big\{Q(f(\widehat{\boldsymbol{x}}_N, U') \leq 0) < (1-\varepsilon)(1-\alpha)\big\} - P^N\big\{Q(f(\widehat{\boldsymbol{x}}_N, U') \leq q^{-1}(\varepsilon)) < 1-\alpha\big\} \\ &= 1-2\eta \end{aligned}$$

holds. ■

When $\mathcal{A} = \mathcal{U}$ holds with Assumption 4.1, we obtain $\alpha = 0$ and the inequality (2) is reduced to

$$P^N\big\{V(\widehat{\boldsymbol{x}}_N) \leq \varepsilon, \operatorname*{ess\,sup}_{\boldsymbol{u} \in \mathcal{U}} f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u}) \leq q^{-1}(\varepsilon)\big\} \;\geq\; 1-2\eta,$$

which is slightly weaker than that in Theorem 4.3.

4.2 A posteriori Assessments

Next, we conduct a posteriori assessment. Once an optimal solution $\widehat{\boldsymbol{x}}_N$ of (SCP_N) is obtained, one can make a posteriori assessment of the constraint function $f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u})$ by using additional random samples. We consider two cases: one in which the sampling distribution for a posteriori assessment is P, and the other in which it is Q.

First, suppose that additional samples, $\tilde{U}_1, \dots, \tilde{U}_M$, are independently and identically generated from the distribution P. Under the Assumption 4.1, a probabilistic upper bound of $\max_{\boldsymbol{u} \in \mathcal{A}} f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u})$ is expressed as

$$P^M\big\{ \max_{\boldsymbol{u} \in \mathcal{A}} f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u}) < \max_{i=1,\dots,M} f(\widehat{\boldsymbol{x}}_N, \tilde{U}_i) + \delta \big\} \ge 1 - \eta, \quad {}^\forall \delta \in (0, B), \quad {}^\forall \eta \in (0,1),$$

for $M \ge \ln \eta / \ln(1 - q(\delta))$. Here, $P^M\{\cdots\}$ denotes the probability distribution of the independent random variables, $\tilde{U}_1, \dots, \tilde{U}_M$. The above inequality is proved by using the Binomial formula. See Theorem 3.1 in Tempo et al. [18] for details. Note that a posteriori assessment does not require any computation for optimization, and therefore, even if the number of random samples M is large, computation of a posteriori assessment is still tractable.

On the other hand, additional samples distributed from Q enable the assessment of the probability distribution of $f(\widehat{\boldsymbol{x}}_N, U')$. Let $\tilde{U}_1, \dots, \tilde{U}_{M'}$ be independent random variables distributed from Q. Moreover, let us define Q_γ and $\widehat{Q}_\gamma$ as

$$Q_\gamma = Q(f(\widehat{\boldsymbol{x}}_N, U') \le \gamma), \quad \widehat{Q}_\gamma = \frac{1}{M'} \sum_{i=1}^{M'} I[\, f(\widehat{\boldsymbol{x}}_N, \tilde{U}_i) \le \gamma \,],$$

for a fixed $\widehat{\boldsymbol{x}}_N$, where $I[\cdot]$ is the indicator function. Then, it holds the Dvoretzky-Kiefer-Wolfowitz inequality holds [16],

$$Q^{M'}\big\{ \sup_\gamma \big| Q_\gamma - \widehat{Q}_\gamma \big| \le \varepsilon \big\} \ge 1 - \eta, \quad M' \ge \frac{1}{2\varepsilon^2} \log \frac{2}{\eta}. \tag{4}$$

Therefore, one can estimate $Q(f(\widehat{\boldsymbol{x}}_N, U') \le 0)$ and $Q(f(\widehat{\boldsymbol{x}}_N, U') \le q^{-1}(\varepsilon))$ by $\widehat{Q}_\gamma$ with $\gamma = 0$ and $\gamma = q^{-1}(\varepsilon)$, respectively, with a high degree of accuracy.

5 Uniform Lower Bound of Tail Probability

As shown in Section 4, it is important to derive a uniform lower bound of tail probability for assessing the worst-case violation. In this section, we derive the uniform lower bound for some uncertainty sets. First, common uncertainty sets such as a hypersphere are considered, and then, more general uncertainty sets are investigated. In the latter case, the so-called star shaped region is treated. Here, we assume the Lipschitz continuity for the constraint function.

Assumption 5.1. *The constraint function $f(\boldsymbol{x}, \boldsymbol{u})$ satisfies*

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

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

The tail probability has a lower bound under the Assumption 5.1. Let $\bar{\boldsymbol{u}}$ be the maximum solution of $\max_{\boldsymbol{u} \in \mathcal{A}} f(\boldsymbol{x}, \boldsymbol{u})$ for a fixed $\boldsymbol{x}$; then, $\|\boldsymbol{u} - \bar{\boldsymbol{u}}\| < \delta / L$ implies $f(\boldsymbol{x}, \bar{\boldsymbol{u}}) - f(\boldsymbol{x}, \boldsymbol{u}) < \delta$. As a result, the tail probability on $\mathcal{A}$ is lower bounded as

$$p(\delta, \boldsymbol{x}) = P\{f(\boldsymbol{x}, \bar{\boldsymbol{u}}) - f(\boldsymbol{x}, U) < \delta\} \ge P\big\{\|U - \bar{\boldsymbol{u}}\| < \delta / L\big\}.$$

Therefore, we derive a lower bound of $P\{\|U - \bar{\boldsymbol{u}}\| < \delta/L\}$ as a uniform lower bound of tail probability. In what follows, we consider only compact subsets and therefore, the existence of $\bar{\boldsymbol{u}}$ is assured.

5.1 Uniform Lower Bound on Simple Uncertainty Set

A uniform lower bound of the tail probability on $\mathcal{A}$ is derived, when the subset $\mathcal{A}$ is an ellipsoidal uncertainty set [1,2] or a norm-constrained uncertainty set [13].

Lemma 5.2. *Let $\mathcal{A}$ be an ellipsoidal set defined as $\mathcal{A} = \{\boldsymbol{u} \in \mathbb{R}^d \mid \|\boldsymbol{u}\| \le R\}$. We assume that the probability density of P with respect to the Lebesgue measure is represented as $h(\|\boldsymbol{u}\|)$ on $\mathcal{A}$ in which h is a monotonically non-increasing function. Let $h(r)$ be positive for $0 \le r < R$. Then, the function q_1 defined as*

$$q_1(\delta) = \int I[\boldsymbol{u} \in \mathcal{B}_{\delta,R}]\ h(\|\boldsymbol{u}\|)\, d\boldsymbol{u}, \quad 0 \le \delta \le 2LR,$$
$$\mathcal{B}_{\delta,R} = \left\{\boldsymbol{u} \in \mathbb{R}^d \mid \|\boldsymbol{u} - R\boldsymbol{e}_1\| \le \delta/L,\ \|\boldsymbol{u}\| \le R\right\}, \quad \boldsymbol{e}_1 = (1, 0, \dots, 0) \in \mathbb{R}^d,$$

is a uniform lower bound of the tail probability on $\mathcal{A}$.

In Theorems 4.3 and 4.4, the range of ε is restricted to the interval $(0, q(B))$, where B is the upper bound of δ. In Lemma 5.2, the upper bound of ε is equal to $q_1(2LR) = 1$, and therefore, ε can take any value in the range $(0, 1)$.

Proof. We derive a lower bound of $P\{\|U - \bar{\boldsymbol{u}}\| < \delta/L\}$, which does not depend on $\bar{\boldsymbol{u}} \in \mathcal{A}$. Since the probability density $h(\|\boldsymbol{u}\|)$ is rotationally invariant without loss of generality $\bar{\boldsymbol{u}} = (\bar{u}_1, 0, \dots, 0),\ 0 \le \bar{u}_1 \le R$ holds. Moreover, one can prove that the probability $P\{\|U - \bar{\boldsymbol{u}}\| < \delta/L\}$ is a decreasing function in terms of $\bar{u}_1$, and therefore, $\bar{\boldsymbol{u}} = R\boldsymbol{e}_1$ attains the minimum value. When $\bar{\boldsymbol{u}} = R\boldsymbol{e}_1$, q_1 is a strictly increasing function of δ in the range $0 \le \delta \le 2LR$. ■

Numerical examples in Section 3 are derived using Lemma 5.2. Let the uncertainty set $\mathcal{U}$ be $\mathbb{R}^d$, and $\mathcal{A}$ be $\{\boldsymbol{u} \in \mathbb{R}^d \mid \|\boldsymbol{u}\| \le R\}$. The probability distribution Q is the normal distribution $N(\boldsymbol{0}, I_d)$, and P is the truncated normal distribution on $\mathcal{A}$. Note that the density function of P satisfies the condition in Lemma 5.2. When $Q(\mathcal{A}) = 1 - \alpha$, then $R = \sqrt{\chi^2_{d,\alpha}}$, where $\chi^2_{d,\alpha}$ is the upper α-percentile of the chi-square distribution with d degrees of freedom. When P is the truncated normal distribution on $\mathcal{A}$, the integration in q_1 is reduced to one dimensional integration, which can be easily computed. Theorems 4.4 and 4.3 with $q = q_1$ provide the numerical results shown in Table 2 and Table 1, respectively.

In general, however, q_1 does not have an analytic representation, and may involve high dimensional integration. A simple method for approximating q_1 is Monte-Carlo sampling, which is expressed as

$$\widehat{q}_1(\delta) = \frac{1}{M}\sum_{i=1}^{M} I[\ \bar{U}_i \in \mathcal{B}_{\delta,R}\ ],$$

where $\bar{U}_1, \ldots, \bar{U}_M$ are independently and identically distributed from P. In addition, approximation error by random sampling can be incorporated into a priori assessment with minor modification. In fact, the VC dimension (see [19]) of the class of subsets $\{\mathcal{B}_{\delta,R} \mid \delta \in [0, 2LR]\}$ is one, and therefore the uniform convergence of $\widehat{q}_1$,

$$P^M\big\{\sup_{\delta} |q_1(\delta) - \widehat{q}_1(\delta)| > \varepsilon'\big\} < \eta' = \frac{2}{eM} e^{-\varepsilon'^2 M}$$

holds. Therefore, we obtain

$$\begin{aligned} & P^{N+M}\big\{Q(f(\widehat{\boldsymbol{x}}_N, U') \le 0) \ge (1-\varepsilon)(1-\alpha),\ Q(f(\widehat{\boldsymbol{x}}_N, U') \le \widehat{q}_1^{-1}(\varepsilon + \varepsilon')) \ge 1-\alpha\big\} \\ \ge\ & 1 - 2\eta - \eta' \end{aligned}$$

where P^{N+M} is the probability distribution of $M+N$ independent random variables on $\mathcal{A}$, which govern the behavior of $\widehat{\boldsymbol{x}}_N$ and $\widehat{q}_1$.

Lemma 5.2 is extended to the case of the norm-constrained uncertainty set.

Lemma 5.3. *Let $\mathcal{A}_k$ be the norm-constrained uncertainty set,*

$$\mathcal{A}_k = \big\{\boldsymbol{u} = \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix} \mid \|\boldsymbol{u}\| \le R,\ \boldsymbol{v}_1 \ge \boldsymbol{0},\ \boldsymbol{v}_1 \in \mathbb{R}^k,\ \boldsymbol{v}_2 \in \mathbb{R}^{d-k}\big\}.$$

We assume that the probability density of P is represented as $h(\|\boldsymbol{u}\|)I[\boldsymbol{u} \in \mathcal{A}_k]$, and h satisfies the same condition in Lemma 5.2. Then, function q_2, defined as

$$q_2(\delta) = \frac{h(R)}{2^k} \int I[\boldsymbol{u} \in \mathcal{B}_{\delta,R}] d\boldsymbol{u}, \quad 0 \le \delta \le \frac{LR}{1+\sqrt{k}},$$

is a uniform lower bound of the tail probability on $\mathcal{A}_k$.

The proof is provided in Appendix A. When non-negative constraints on $\boldsymbol{v}$ do not exist, *i.e.,* $k = 0$, we obtain $q_2(\delta) = h(R)\int I[\boldsymbol{u} \in \mathcal{B}_{\delta,R}]d\boldsymbol{u}$ for $0 \le \delta \le LR$. This is a lower bound of q_1.

In the assessment of the worst-case violation in Theorem 4.3 or 4.4, the upper bound of the violation probability ε should be less than $q(B)$. The uniform lower bound in Lemma 5.2 satisfies $q_1(2LR) = 1$, and therefore ε is not restricted. On the other hand, the upper bound of ε introduced from q_2 is expressed as

$$q_2(LR/(1+\sqrt{k})) = \frac{h(R)}{2^k} \int I[\ \|\boldsymbol{u} - R\boldsymbol{e}_1\| \le R/(1+\sqrt{k}),\ \|\boldsymbol{u}\| \le R\ ]\ d\boldsymbol{u}.$$

In general, the upper bound decreases as the dimension of $\mathcal{A}_k$ becomes large. Table 3 explains the relation between the dimension of the uncertainty set d and the upper bound $q_2(LR/(1+\sqrt{k}))$. The value of k for $\mathcal{A}_k$ is set to 0, 1, 5, or d. In the upper part of Table 3, the probability distribution Q is a truncated normal distribution on $\big\{\boldsymbol{u} = \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix} \mid \boldsymbol{v}_1 \ge \boldsymbol{0},\ \boldsymbol{v}_1 \in \mathbb{R}^k,\ \boldsymbol{v}_2 \in \mathbb{R}^{d-k}\big\}$, which is defined by the normal distribution with mean $\boldsymbol{0}$ and variance-covariance matrix I_d. The probability P is the conditional probability of Q over $\mathcal{A}_k$. On the other hand, in the lower part of Table 3, the probability distribution Q is set to the uniform distribution on $\mathcal{A}_k$,

Table 3: The values of $q_2(LR/(1+\sqrt{k}))$ are shown for each k to the dimension of uncertainty set d. In the upper part of the table, P and Q are the truncated normal distributions. In the lower part of the table, P and Q are the same uniform distribution on $\mathcal{A}_k$.

P, Q: truncated normal distribution ($\alpha = 0.01$)				
d	$k=0$	$k=1$	$k=5$	$k=d$
3	1.1034×10^{-2}	1.7930×10^{-3}	$-$	7.4690×10^{-4}
10	1.3231×10^{-3}	3.9634×10^{-6}	4.3669×10^{-8}	3.8802×10^{-9}
20	1.6009×10^{-4}	1.5109×10^{-9}	1.7512×10^{-13}	6.4643×10^{-18}

P, Q: uniform distribution ($P = Q$)				
d	$k=0$	$k=1$	$k=5$	$k=d$
3	3.1250×10^{-1}	5.0781×10^{-2}	$-$	2.1154×10^{-2}
10	8.1864×10^{-2}	2.4522×10^{-4}	2.7019×10^{-6}	2.4007×10^{-7}
20	1.5118×10^{-2}	1.4268×10^{-7}	1.6537×10^{-11}	6.1045×10^{-16}

and $\mathcal{U} = \mathcal{A}_k$ is assumed. Table 3 indicates that when the dimension d and the value of k are large, the upper bound of ε is small. As a result, the number of samples $N(\varepsilon, \eta)$ becomes extremely large. When we assess the worst-case violation over $\mathcal{U}$, it is practical to denote $\mathcal{A}$ as a hypersphere and $q_1(\delta)$ as the uniform lower bound.

5.2 Uniform Lower Bound on Star-Shaped Region

Next, we consider uniform lower bounds of tail probability for more general uncertainty sets. We assume that the set $\mathcal{A}$ is a type of a star-shaped region.

Assumption 5.4. *Let $\mathcal{A}$ be a compact set in $\mathbb{R}^d$. Suppose that there exists a d-dimensional hypersphere S in $\mathcal{A}$ that satisfies the condition,*

$$\mathrm{conv}(\{\boldsymbol{u}\} \cup S) \subset \mathcal{A}, \quad \textit{for any } \boldsymbol{u} \in \mathcal{A}, \tag{5}$$

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

When S is reduced to a point, condition (5) denotes that $\mathcal{A}$ is a star-shaped region. In this setting, the radius should be positive. When the radius of S tends to zero, the uniform lower bound derived from the argument below tends to zero, and as the result, the number of samples $N(\varepsilon, \eta)$ tends to infinity.

Lemma 5.5. *Suppose that Assumptions 5.1 and 5.4 hold, and the probability density of P is larger than or equal to $C > 0$ on $\mathcal{A}$. Let r and R be the radii of S in Assumption 5.4, and the diameter of $\mathcal{A}$ is defined by $R = \sup_{\boldsymbol{u},\boldsymbol{v} \in \mathcal{A}} \|\boldsymbol{u} - \boldsymbol{v}\|$. Then, the following function q_3 is a uniform*

Table 4: Probabilistic assessment of the violation probability and the worst-case violation is expressed as $P^N\{V(\widehat{\boldsymbol{x}}_N) \le \varepsilon,\ \max_{\boldsymbol{u}\in\mathcal{U}} f(\widehat{\boldsymbol{x}}_N, \boldsymbol{u}) \le q_3^{-1}(\varepsilon)\} \ge 0.99$ for $N \ge N(\varepsilon, \eta)$ ($\eta = 0.01$, $d = 3, 10$ or 20, $m = 10$) with the uncertainty set $\mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^d \mid \|\boldsymbol{u}\| \le 1\}$.

	$q_3^{-1}(\varepsilon)$			
ε	$d=3$	$d=10$	$d=20$	$N(\varepsilon,\eta)$
0.001	0.3051801	1.041995	1.277298	18779
0.003	0.4401498	1.163049	-	6257
0.005	0.5218502	1.223978	-	3752
0.007	0.5837831	1.265858	-	2679
0.009	0.6348016	1.298067	-	2083

lower bound of the tail probability on $\mathcal{A}$,

$$q_3(\delta) = C \min\left\{\frac{1}{4\pi}\mathrm{Beta}\left(\frac{\beta^2}{(R/r-\beta)^2}; \frac{d-1}{2}, \frac{1}{2}\right)\int I[\ \|\boldsymbol{u}\| \le \delta/L\ ]\ d\boldsymbol{u},\ \int I[\ \boldsymbol{u} \in \mathcal{B}_{\delta,r}\ ]\ d\boldsymbol{u}\right\}, \tag{6}$$

$$0 \le \delta \le rL\left(\sqrt{1-\beta^2} + \frac{2\sqrt{1-\beta}}{2-\beta}\right),$$

where β *is a parameter in* $(0,1)$ *and* $\mathrm{Beta}(x; a, b)$ *is the incomplete beta function defined as*

$$\mathrm{Beta}(x; a, b) = \int_0^x y^{a-1}(1-y)^{b-1}dy, \quad 0 \le x \le 1.$$

The proof is provided in Appendix B. Although there may exist a tighter uniform lower bound than $q_3(\delta)$, the computation of the bound will be more complicated.

The preferable β in (6), denoted by β^*, is determined as follows. When we apply q_3 to Theorem 4.3 or 4.4, the range of ε is $(0, q_3(B))$, where $B = rL\left(\sqrt{1-\beta^2} + \frac{2\sqrt{1-\beta}}{2-\beta}\right)$. Since larger $q_3(B)$ is better for practical use, β^* can be determined by the maximizer of $q_3(B)$ subject to $\beta \in (0,1)$, and the widest region of ε is obtained as $0 \le \varepsilon \le \varepsilon^*$. Note that β^* depends only on the dimension d and the ratio R/r. For a given d and R/r, numerical computation of β^* is easy.

The uniform lower bound q_3 in Lemma 5.5 is used for the probabilistic assessment of the violation probability and the worst-case violation. The value of β in q_3 is determined as mentioned above. The setup of probability distributions is the same as that of the examples illustrated in Section 3, and the results are displayed in Tables 4 and 5. Blank spaces in the tables mean that ε exceeds the upper bound ε^*. In general, q_3 is more loosely bound than q_1, and therefore, $q_3^{-1}(\varepsilon)$ is greater than $q_1^{-1}(\varepsilon)$.

When we use the uniform lower bound q_3 in Lemma 5.5 for assessing the worst-case violation, we need to estimate the parameters such as r, R, L, and C. In Appendix C, a simple approach of estimating these parameters is proposed.

Table 5: Probabilistic assessment of the violation probability and the worst-case violation is expressed as $P^N\left\{V(\widehat{\boldsymbol{x}}_N) \le \alpha+\varepsilon-\alpha\varepsilon,\ Q(f(\widehat{\boldsymbol{x}}_N,U') > q_3^{-1}(\varepsilon)) \le \alpha\right\} \ge 0.99$ for $N \ge N(\varepsilon,\eta)$ ($\eta = 0.005$, $d = 2$ or 3, $m = 10$) with the uncertainty set $\mathcal{U} = \mathbb{R}^d$.

ε	α	$\alpha+\varepsilon-\alpha\varepsilon$	$q_3^{-1}(\varepsilon)$ $d=2$	$q_3^{-1}(\varepsilon)$ $d=3$	$N(\varepsilon,\eta)$
0.001	0.01	0.01099	1.448107	3.143777	19993
0.003	0.01	0.01297	2.508223	4.534130	6661
0.005	0.01	0.01495	3.238118	-	3995
0.007	0.01	0.01693	3.831383	-	2852
0.009	0.01	0.01891	4.344404	-	2217
0.001	0.009	0.01	1.525685	3.264246	19993
0.003	0.007	0.01	2.993419	-	6661
0.005	0.005	0.01	4.567909	-	3995

6 Numerical Simulations

We perform the probabilistic assessment presented in the previous section for a maximum Sharpe ratio problem with the robust deviation criterion. We obtain a lower bound by (SCP_N) for the application problem, and then, perform probabilistic assessments for (SCP_N). All computations are conducted using an Opteron 850 (2.4 GHz), 8 GB of physical memory, and 1 MB of L2 cache size with SuSE Linux Enterprise Server 9.

6.1 Problem Setting

For a given expected return vector $\boldsymbol{\mu}$ and a positive definite covariance matrix $\boldsymbol{V}$, a maximum Sharpe ratio problem finds a portfolio that maximizes the Sharpe ratio, which is defined as

$$\max_{\boldsymbol{x}\in X} \frac{\boldsymbol{\mu}^\top \boldsymbol{x} - r_f}{\sqrt{\boldsymbol{x}^\top \boldsymbol{V}\boldsymbol{x}}},$$

where $r_f \ge 0$ is the expected return of a riskless asset. The numerator is the expected excess return on the portfolio, *i.e.*, the excess of return over the risk-free rate r_f, while the denominator is the standard deviation of the return. The Sharpe ratio is a measure to evaluate a portfolio.

The robust deviation Sharpe ratio problem, proposed by Krishnamurthy [14] minimizes the maximum deviation of the Sharpe ratio from the maximum ratio obtained for all possible realizations of the expected return vector $\boldsymbol{\mu}$. The robust deviation counterpart with uncertain expected return vector $\boldsymbol{\mu} \in \mathcal{A}_0 \subset \mathbb{R}^d$ is expressed as follows.

$$(RCP) \quad \left| \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{\mu}\in\mathcal{A}_0} \left\{h^*(\boldsymbol{\mu}) - h(\boldsymbol{x},\boldsymbol{\mu})\right\}, \right. \tag{7}$$

where $\mathcal{X} := \{\boldsymbol{x} \in \mathbb{R}^m : \boldsymbol{x} \ge \boldsymbol{0}, \boldsymbol{e}^\top \boldsymbol{x} = 1\}$, $\boldsymbol{e}$ denotes all-one vector $(1,\dots,1)^\top \in \mathbb{R}^m$,

$$h(\boldsymbol{x},\boldsymbol{\mu}) = \frac{(\boldsymbol{\mu} - r_f\boldsymbol{e})^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{V}\boldsymbol{x}}}, \quad \text{and} \quad h^*(\boldsymbol{\mu}) = \max_{\boldsymbol{x}\in X} \frac{(\boldsymbol{\mu} - r_f\boldsymbol{e})^\top \boldsymbol{x}}{\sqrt{\boldsymbol{x}^\top \boldsymbol{V}\boldsymbol{x}}}.$$

In terms of dimension, $m = d$ holds. The problem (7) is also described as

$$(RCP) \quad \left| \min_{\boldsymbol{x}\in\mathcal{X}, t\in\mathbb{R}} t \quad \text{s.t.} \quad h^*(\boldsymbol{\mu}) - h(\boldsymbol{x}, \boldsymbol{\mu}) - t \le 0, \quad {}^\forall\boldsymbol{\mu} \in \mathcal{A}_0. \right. \tag{8}$$

Applying the techniques of [13] and [14], problem (8), including fractional constraints, is equivalently transformed into

$$\begin{array}{ll} \min\limits_{\boldsymbol{x},t} & t \\ \text{s.t.} & h^*(\boldsymbol{\mu}) - (\boldsymbol{\mu} - r_f\boldsymbol{e})^\top \boldsymbol{x} - t \le 0, \quad {}^\forall\boldsymbol{\mu} \in \mathcal{A}_0 \\ & \boldsymbol{x}^\top \boldsymbol{V}\boldsymbol{x} \le 1, \ \boldsymbol{x} \ge \boldsymbol{0}, \end{array} \tag{9}$$

since the function of Sharpe ratio is homogeneous with respect to a portfolio $\boldsymbol{x}$, and the optimality is achieved at $\boldsymbol{x}^\top\boldsymbol{V}\boldsymbol{x} = 1$. By dividing an optimal solution $\boldsymbol{x}^*$ of (9) by $\boldsymbol{e}^\top\boldsymbol{x}^*$, we obtain an optimal solution of (RCP) with the same optimal value. As an uncertainty set of $\boldsymbol{\mu}$, Krishnamurthy [14] assumed a polytopic uncertainty set $\mathcal{A}_0 = \text{conv}\{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_\ell\}$, which is a convex hull of ℓ vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_\ell$. Then, an optimal solution of the robust deviation Sharpe ratio problem can be achieved by solving $(\ell + 1)$ second-order cone programming problems. Takeda et al. [17] considered the ellipsoidal uncertainty set for $\boldsymbol{\mu}$ and proposed another sampling approach based on the probabilistic assessment technique used in this study.

We consider the problem in which the number of assets ranges from 3 to 20. Let the sets $\mathcal{A}_0$ and $\mathcal{A}$ be

$$\begin{aligned} \mathcal{A}_0 &= \{\boldsymbol{\mu}(\boldsymbol{u}) \in \mathbb{R}^d \mid \mu_j(\boldsymbol{u}) = \mu_{0j}(1 + 0.3u_j), \ j = 1, \ldots, d, \ \boldsymbol{u} \in \mathcal{A}\}, \\ \mathcal{A} &= \{\boldsymbol{u} \in \mathbb{R}^d \mid \|\boldsymbol{u}\| \le 1\}, \end{aligned}$$

where $\boldsymbol{\mu}_0^\top = (\mu_{01}, \ldots, \mu_{0d}) = (0.6, 0.58, 0.56, \ldots, 0.62 - 0.02d) \in \mathbb{R}^d$. The matrix $\boldsymbol{V}$ is defined as follows: 1000 samples from $[\boldsymbol{0}, 2\boldsymbol{\mu}_0]$ are randomly generated, and a positive definite covariance matrix $\boldsymbol{V}$ is constructed from those samples. Let the risk-free rate r_f be 0.10.

Here, we deal with the following two settings of the uncertain set $\mathcal{U}$.

(Case 1: compact uncertainty set) The uncertainty set is defined as $\mathcal{U} = \mathcal{A}$. The probability distribution Q on $\mathcal{U}$ is the truncated normal distribution derived from the multinomial normal distribution, $N_d(\boldsymbol{0}, (\chi^2_{d,\alpha})^{-1}\boldsymbol{I}_d)$, where $\chi^2_{d,\alpha}$ is the upper α-percentile of the chi-square distribution with d degrees of freedom. The value of α is set to 0.01. The sampling distribution is the conditional probability P on $\mathcal{A}$, which is the same as Q.

(Case 2: unbounded uncertainty set) The uncertainty set is defined as $\mathcal{U} = \mathbb{R}^d$, and the distribution Q on $\mathcal{U}$ is a multinomial normal distribution, $N_d(\boldsymbol{0}, (\chi^2_{d,\alpha})^{-1}\boldsymbol{I}_d)$. Note that $Q(\mathcal{A}) = 1-\alpha$ holds. The value of α is set to 0.01. The distribution P on $\mathcal{A}$ is the truncated normal distribution derived from Q.

In cases 1 and 2, the set $\mathcal{A}$ and the conditional probability P are the same, and the distribution Q is different. The set $\mathcal{A}$ and the conditional probability P satisfy the conditions in Lemma 5.2, and therefore, the function q_1 is available as the uniform lower bound of tail probability.

6.2 Probabilistic Assessment of (SCP_N)

Now, we obtain an approximate solution for the robust deviation optimization problem, and perform a probabilistic assessment of the solution. Note that in the case of min-max optimization problems, the worst-case violation provides an assessment of the optimal value of (RCP). By denoting the objective function of (RCP) as $g(\boldsymbol{x}, \boldsymbol{u}) = h^*(\boldsymbol{\mu}(\boldsymbol{u})) - h(\boldsymbol{x}, \boldsymbol{\mu}(\boldsymbol{u}))$ and using random samples U_i, $i = 1, \ldots, N$, which are independently and identically generated from the distribution P on $\mathcal{A}$, we define a sampled convex program of the min-max problem as

$$(SCP_N) \quad \Big| \min_{\boldsymbol{x} \in \mathcal{X},\, t \in R} t \quad \text{s.t.} \;\; g(\boldsymbol{x}, U_i) - t \le 0 \quad i = 1, \ldots, N.$$

Let optimal solutions of (RCP) and (SCP_N) be $(\boldsymbol{x}^*, t^*)$ and $(\widehat{\boldsymbol{x}}_N, \widehat{t}_N)$, respectively. Theorem 4.3 implies that for $N \ge N(\varepsilon, \eta)$,

$$\max_{\boldsymbol{u} \in \mathcal{A}} g(\widehat{\boldsymbol{x}}_N, \boldsymbol{u}) - \widehat{t}_N = \max_{\boldsymbol{u} \in \mathcal{A}} g(\widehat{\boldsymbol{x}}_N, \boldsymbol{u}) - \max_{i=1,\ldots,N} g(\widehat{\boldsymbol{x}}_N, U_i) \le q_1^{-1}(\varepsilon)$$

holds in high probability. By noting that $\max_{\boldsymbol{u} \in \mathcal{A}} g(\boldsymbol{x}^*, \boldsymbol{u}) \le \max_{\boldsymbol{u} \in \mathcal{A}} g(\widehat{\boldsymbol{x}}_N, \boldsymbol{u})$, the inequality:

$$0 \le \max_{\boldsymbol{u} \in \mathcal{A}} g(\boldsymbol{x}^*, \boldsymbol{u}) - \max_{i=1,\ldots,N} g(\widehat{\boldsymbol{x}}_N, U_i) \le q_1^{-1}(\varepsilon)$$

follows, and therefore, the optimal value of (RCP) is shown to be in

$$[\max_{i=1,\ldots,N} g(\widehat{\boldsymbol{x}}_N, U_i), \max_{i=1,\ldots,N} g(\widehat{\boldsymbol{x}}_N, U_i) + q_1^{-1}(\varepsilon)]$$

in high probability.

The sample size $N(\varepsilon, \eta)$ given by a priori assessment becomes extremely large for small error $q_1^{-1}(\varepsilon)$, and solving (SCP_N) with $N(\varepsilon, \eta)$ constraints becomes difficult. Tables 6 and 7 show the relation among ε, $q_1^{-1}(\varepsilon)$, and $N(\varepsilon, \eta)$ for cases 1 and 2, respectively. The violation probability $V(\widehat{\boldsymbol{x}}_N, \widehat{t}_N)$ is obtained by computing $Q(g(\widehat{\boldsymbol{x}}_N, U') - \widehat{t}_N > 0)$. Note that the values of $q_1^{-1}(\varepsilon)$ are the same as in Tables 6 and 7, since $\mathcal{A}$ and P are the same in cases 1 and 2. When the dimension of uncertainty d is high, the upper bound $q_1^{-1}(\varepsilon)$ cannot be improved even if $N(\varepsilon, \eta)$ becomes large. In numerical simulations, we construct a relaxation problem $(SCP_{\bar{N}})$ with $\bar{N} = 1000$ and perform a posteriori assessment for the solution $(\widehat{\boldsymbol{x}}_{\bar{N}}, \widehat{t}_{\bar{N}})$.

A posteriori assessment of the worst-case violation: Next, we perform a posteriori assessment of the worst-case violation:

$$\max_{\boldsymbol{u} \in \mathcal{A}} \{g(\widehat{\boldsymbol{x}}_{\bar{N}}, \boldsymbol{u}) - \widehat{t}_{\bar{N}}\}, \tag{10}$$

evaluated at $(\widehat{\boldsymbol{x}}_{\bar{N}}, \widehat{t}_{\bar{N}})$ of $(SCP_{\bar{N}})$ with $\bar{N} = 1000$. The following a posteriori assessment is valid for both cases 1 and 2. Using additional samples $\tilde{U}_i$, $i = 1, \ldots, M = \lceil \ln \eta / \ln(1 - q_1(\delta)) \rceil$, which are independently and identically distributed from P, we determine the empirical worst-case violation:

$$\beta_M = \max_{i=1,\ldots,M} \{g(\widehat{\boldsymbol{x}}_{\bar{N}}, \tilde{U}_i) - \widehat{t}_{\bar{N}}\},$$

Table 6: A priori assessment of the violation probability and the worst-case violation is expressed as $P^N\{V(\widehat{\boldsymbol{x}}_N, \widehat{t}_N) \le \varepsilon,\ \max_{\boldsymbol{u}\in\mathcal{U}} g(\widehat{\boldsymbol{x}}_N, \boldsymbol{u}) - \widehat{t}_N \le q_1^{-1}(\varepsilon)\} \ge 0.99$ for $N \ge N(\varepsilon, \eta)$ ($\eta = 0.01$, $d = 3, 10$ or 20, $m = d$) under the case 1 uncertainty.

	$d=3$		$d=10$		$d=20$	
ε	$q_1^{-1}(\varepsilon)$	$N(\varepsilon,\eta)$	$q_1^{-1}(\varepsilon)$	$N(\varepsilon,\eta)$	$q_1^{-1}(\varepsilon)$	$N(\varepsilon,\eta)$
0.00001	0.078215	840592	0.442139	1878307	0.691440	3184531
0.0002	0.184064	42027	0.566877	93911	0.798570	159221
0.0004	0.220570	21012	0.600514	46953	0.826838	79608
0.0006	0.244358	14007	0.621205	31301	0.844151	53070
0.0008	0.262405	10505	0.636390	23475	0.856809	39801

Table 7: A priori assessment of the violation probability and the worst-case violation is expressed as $P^N\{V(\widehat{\boldsymbol{x}}_N, \widehat{t}_N) \le \alpha + \varepsilon - \alpha\varepsilon,\ Q(g(\widehat{\boldsymbol{x}}_N, U') - \widehat{t}_N > q_1^{-1}(\varepsilon)) \le \alpha\} \ge 0.99$ for $N \ge N(\varepsilon, \eta)$ ($\eta = 0.005$, $d = 3, 10$ or 20, $m = d$) under the case 2 uncertainty.

			$d=3$		$d=10$		$d=20$	
ε	α	$\alpha+\varepsilon-\alpha\varepsilon$	$q_1^{-1}(\varepsilon)$	$N(\varepsilon,\eta)$	$q_1^{-1}(\varepsilon)$	$N(\varepsilon,\eta)$	$q_1^{-1}(\varepsilon)$	$N(\varepsilon,\eta)$
0.00001	0.01	0.010010	0.078215	927376	0.442139	1999837	0.691440	3338291
0.0002	0.01	0.010198	0.184064	46366	0.566877	99987	0.798570	166908
0.0004	0.01	0.010396	0.220570	23181	0.600514	49991	0.826838	83451
0.0006	0.01	0.010594	0.244358	15453	0.621205	33326	0.844151	55632
0.0008	0.01	0.010792	0.262405	11589	0.636390	24993	0.856809	41722

and obtain the upper bound for (10) as $\beta_M + \delta$. In numerical studies, we solve the problem ($SCP_{\bar{N}}$) with a reasonably large number of examples $\bar{N}$, and determine β_M for a posteriori assessment.

A posteriori assessment was performed for the robust deviation Sharpe ratio problem with $d = 3$. Figure 1 shows the number of samples M and upper bounds $\beta_M + \delta$ for parameter values δ. The left figure shows that an exponentially large number of samples is required to perform a posteriori evaluation. In a robust deviation problem, the computation cost rapidly increases as the problem size increases, because we need to solve an optimization problem to compute $h^*(\boldsymbol{\mu})$ even in the phase of a posteriori assessment. As shown in Table 8, small value of δ and large value of d result in a huge number of samples for a posteriori assessment. Therefore, for a larger-sized problem, estimation of the probability distribution of $g(\widehat{\boldsymbol{x}}_{\bar{N}}, U) - \widehat{t}_{\bar{N}}$ may be useful, because the number of function evaluation M' in (4) does not depend on the dimension d.

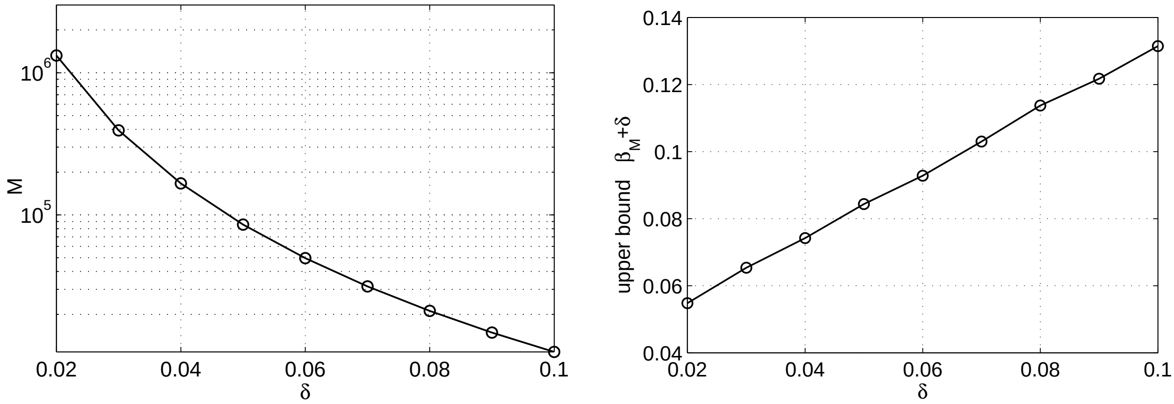

Figure 1: The left figure shows the number of samples $M = \lceil \ln \eta / \ln(1 - q_1(\delta)) \rceil$ for a-posteriori assessment of a solution of $(SCP_{\bar{N}})$ with $\bar{N} = 1000$. The right one shows upper bounds $\beta_M + \delta$ for the worst-case violation such that $P^M \{ \max_{\boldsymbol{u} \in \mathcal{A}} g(\widehat{\boldsymbol{x}}_{\bar{N}}, \boldsymbol{u}) - \widehat{t}_{\bar{N}} < \beta_M + \delta \} \geq 0.99$.

Table 8: The relation between δ and M for a posteriori assessment of the worst-case violation via $M = \lceil \ln \eta / \ln(1 - q_1(\delta)) \rceil$ with $\eta = 0.01$.

$d = 3$		$d = 10$		$d = 20$	
δ	M	δ	M	δ	M
0.2090	1243	0.2133	113791629	0.2217	1.296240×10^{15}
0.4180	167	0.4266	148122	0.4434	2049393902
0.6269	53	0.6399	3640	0.6651	1198309
0.8359	24	0.8531	311	0.8868	8910
1.0449	13	1.0664	54	1.1085	303

A posteriori assessment for the distribution Q_γ: We calculate the empirical distribution function:

$$\widehat{Q}_\gamma = \frac{1}{M'} \sum_{i=1}^{M'} I[\, g(\widehat{\boldsymbol{x}}_{\bar{N}}, \tilde{U}_i) - \widehat{t}_{\bar{N}} \leq \gamma \,],$$

using random samples $\tilde{U}_i$ from the probability distribution Q on $\mathcal{U}$. Remember that the distribution Q is different in cases 1 and 2. If the sample size is $M' \geq \frac{1}{2\varepsilon^2} \ln \frac{2}{0.01}$, then

$$Q^{M'} \{ \sup_\gamma |Q_\gamma - \widehat{Q}_\gamma| \leq \varepsilon \} \geq 0.99 \tag{11}$$

holds for $Q_\gamma = Q(g(\widehat{\boldsymbol{x}}_{\bar{N}}, U') - \widehat{t}_{\bar{N}} \leq \gamma)$. It implies that for any $\gamma \in \mathbb{R}$, Q_γ uniformly exists in the interval $[\widehat{Q}_\gamma - \varepsilon, \widehat{Q}_\gamma + \varepsilon]$ with a high probability. The necessary sample size M' for (11) is determined by ε, and therefore, is not related to the problem size such as d. Table 9 displays the relation between ε and M'. For the robust deviation problem with $d = 3$ and $d = 20$, the empirical distribution $\widehat{Q}_\gamma$ is determined, as shown in Figure 2. In addition, upper bounds $\widehat{Q}_\gamma + 0.002$ of Q_γ are shown for two cases. The figures indicate that the probability $\widehat{Q}_\gamma$ of case

Table 9: Relation between ε and M' for a posteriori assessment of the probability distribution of the worst-case violation via $M' = \lceil \frac{1}{2\varepsilon^2} \ln(1/\eta) \rceil$ with $\eta = 0.01$.

ε	M'
0.002	662290
0.004	165573
0.006	73588
0.008	41394
0.010	26492

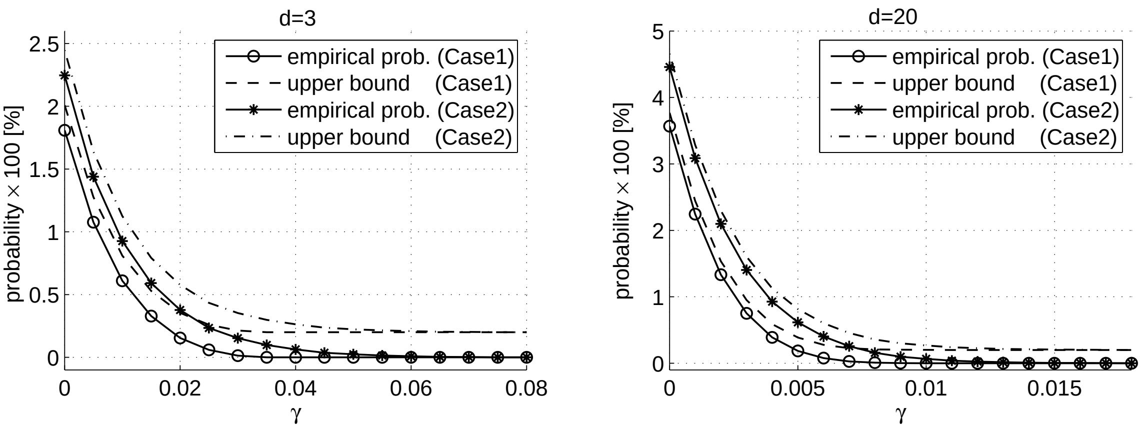

Figure 2: The left figure is for a $d = 3$ dimensional problem $(SCP_{\bar{N}})$ (its optimal value is 0.045), and the right figure for a $d = 20$ dimensional problem (its optimal value is 0.021). The solid line shows empirical probability $\widehat{Q}_\gamma$, and the dotted lines show a theoretical upper bound $\widehat{Q}_\gamma + 0.002$ of Q_γ under the sample size $M' = 662290$ for a posteriori assessment.

2 is likely to be higher than that of case 1, which is reasonable, since in case 2, vector $\boldsymbol{u} \in \mathcal{U}$ is ignored for the construction of $(SCP_{\bar{N}})$ with probability $\alpha = 0.01$.

7 Concluding Remarks

Sampled convex programs are applied to solve uncertain convex programs. Calafiore and Campi [4] and Campi and Garatti [8] proposed the required number of random samples for achieving a small violation probability. Following the same approach, we have performed a probabilistic assessment of the worst-case violation.

The uniform lower bound of tail probability $q(\delta) \leq p(\delta, \boldsymbol{x})$ is significant to derive an upper bound of the worst-case violation. If there exists a uniform lower bound, one can make a priori and a posteriori assessments of the worst-case violation. Some uniform lower bounds are derived for uncertainty sets, such as an ellipsoidal set, a norm-constraint uncertainty set, and a star-shaped region. By following our method, it is possible to assess the worst-case violation up to a certain probability using the subset $\mathcal{A}$ of the uncertain set $\mathcal{U}$. As a result, we can deal with

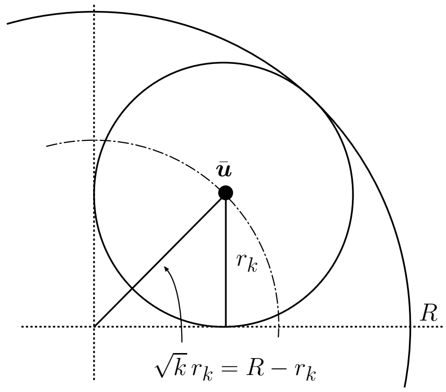

Figure 3: If $\|\bar{\boldsymbol{u}}\| \le R - r_k$, the hypersphere with radius r_k centered at $\bar{\boldsymbol{u}}$ is included in the hypersphere with radius R.

unbounded uncertain variables such as the Gaussian distribution.

In numerical studies, the proposed method is applied to solve robust deviation optimization problems. Although the number of samples for SCP is less than the required number given by $N(\varepsilon, \delta)$, a posteriori assessment shows that the worst-case violation is not large. Although in many practical applications, the required number $N(\varepsilon, \delta)$ may be extremely large, the accuracy of the solution is found to be sufficiently high by a posteriori assessment.

As a future study, we need to develop a technique to drop the sampled constraints in order to reduce the computation cost. At the same time, it is important to derive tighter uniform lower bounds of tail probability.

A Proof of Lemma 5.3

In the following proof, $\mathrm{Vol}(S)$ denotes the volume of the set S with respect to the Lebesgue measure.

Proof. Let r_k be $R/(\sqrt{k}+1)$, and s be δ/L for the sake of simplicity. We assume $s \le r_k$. For any $\bar{\boldsymbol{u}} \in \mathcal{A}_k$, we estimate an uniform lower bound of the volume of $\{\boldsymbol{u} \in \mathcal{A}_k \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| \le s\}$. This set is described as $\mathcal{W} \cap V_k$, where $\mathcal{W} = \{\boldsymbol{u} \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| \le s,\ \|\boldsymbol{u}\| \le R\}$ and $V_k = \{\boldsymbol{u} \in \mathbb{R}^d \mid u_i \ge 0,\ \ i = 1, \dots, k\}$. The set $\mathcal{W}$ is considered in Lemma 5.2, while the range $[0, r_k L]$ is smaller than $[0, 2LR]$. The proof is divided into two parts. First, the case of $\|\bar{\boldsymbol{u}}\| \le R - r_k$ is studied, and then, the case of $R - r_k < \|\bar{\boldsymbol{u}}\| \le R$ is considered. Geometrical meaning of these conditions is illustrated in Figure 3.

First, we prove $P\{\|U - \bar{\boldsymbol{u}}\| \le \delta/L\} \ge q_2(\delta)$ under the condition of $\|\bar{\boldsymbol{u}}\| \le R - r_k$. Let $\boldsymbol{u}$ be an element of $\mathcal{W} \cap \{\bar{\boldsymbol{u}} + \boldsymbol{x} \mid x_1, \dots, x_k \ge 0\}$; then, $u_i \ge \bar{u}_i$ holds for $i = 1, \dots, k$. Since $\bar{\boldsymbol{u}}$ lies in $\mathcal{A}_k$, we have inequalities $\bar{u}_i \ge 0, i = 1, \dots, k$, and therefore, $u_i \ge \bar{u}_i \ge 0, i = 1, \dots, k$ are valid. This indicates that $\boldsymbol{u} \in \mathcal{W} \cap V_k$. Therefore, it holds that $\mathrm{Vol}(\mathcal{W} \cap V_k) \ge \mathrm{Vol}(\mathcal{W} \cap \{\bar{\boldsymbol{u}} + \boldsymbol{x} \mid x_1, \dots, x_k \ge 0\})$. When$\|\bar{\boldsymbol{u}}\| \le R - r_k$ and $s \le r_k$, the inclusion relation, $\{\boldsymbol{u} \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| \le s\} \subset \{\boldsymbol{u} \mid \|\boldsymbol{u}\| \le R\}$, holds. Thus, we have $\mathcal{W} = \{\boldsymbol{u} \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| \leq s\}$. These results imply that

$$\begin{aligned}
P\{\|U - \bar{\boldsymbol{u}}\| \leq \delta/L\} &= \int I[\boldsymbol{u} \in \mathcal{W} \cap V_k\ ]\ h(\|\boldsymbol{u}\|)\ d\boldsymbol{u} \\
&\geq h(R)\mathrm{Vol}(\mathcal{W} \cap V_k) \\
&\geq h(R)\mathrm{Vol}(\mathcal{W} \cap \{\bar{\boldsymbol{u}} + \boldsymbol{x} \mid x_1, \ldots, x_k \geq 0\}) \\
&= \frac{h(R)}{2^k} \int I[\ \boldsymbol{u} \in \mathcal{W}\ ]\ d\boldsymbol{u} \qquad (\because\ \mathcal{W} = \{\boldsymbol{u} \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| \leq s\}) \\
&\geq \frac{h(R)}{2^k} \int I[\ \boldsymbol{u} \in \mathcal{B}_{\delta,R}\ ]\ d\boldsymbol{u}.
\end{aligned}$$

The last inequality is derived from Lemma 5.2 with P as the uniform distribution.

Next, we assume $R - r_k < \|\bar{\boldsymbol{u}}\| \leq R$ and $k < d$. Let $\boldsymbol{e}_i$ $(i = 1, \ldots, d)$ be unit coordinate vectors. If there exists $\boldsymbol{u} \in \mathcal{W}$ such that $\boldsymbol{e}_i^\top \boldsymbol{u} < 0$, then we say that $\boldsymbol{e}_i$ cuts $\mathcal{W}$. We reorder unit coordinate vectors from $\boldsymbol{e}_1$ to $\boldsymbol{e}_k$ such that $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_\ell$ cut $\mathcal{W}$, and $\boldsymbol{e}_{\ell+1}, \ldots, \boldsymbol{e}_k$ do not cut $\mathcal{W}$. For $\ell + 1 \leq i \leq k$, equality $\mathrm{Vol}(\mathcal{W} \cap V_{i-1}) = \mathrm{Vol}(\mathcal{W} \cap V_i)$ clearly holds. 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 $i = 1, \ldots, \ell$. 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 the same volume and that $\mathcal{W} \cap V_i$ includes one of them. Let i be a positive integer less than or equal to ℓ, and Π_i be an orthogonal projection matrix onto the subspace spanned by $\{\boldsymbol{e}_i, \ldots, \boldsymbol{e}_d\}$, *i.e.*, $\Pi_i \boldsymbol{u} = (0, \ldots, 0, u_i, \ldots, u_d)^\top$ for all $\boldsymbol{u} = (u_1, \ldots, u_d)^\top \in \mathbb{R}^d$. When $\boldsymbol{e}_j$ cuts $\mathcal{W}$, the j-th element of $\bar{\boldsymbol{u}} - s\boldsymbol{e}_j$ should be negative. Hence, for any $j = 1, \ldots, i$, the inequalities $0 \leq \bar{u}_j < s \leq r_k$ are satisfied. From the assumption $\|\bar{\boldsymbol{u}}\|^2 > (R - r_k)^2$, we have inequality $\|\Pi_i \bar{\boldsymbol{u}}\|^2 > r_k^2$, since $\|\Pi_i \bar{\boldsymbol{u}}\|^2 > (R - r_k)^2 - \bar{u}_1^2 \cdots - \bar{u}_{i-1}^2 > (R - r_k)^2 - (k-1)r_k^2 = r_k^2$. Therefore, 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 a 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 $\{\boldsymbol{u} \in \mathbb{R}^d \mid \boldsymbol{b}^\top \boldsymbol{u} = 0\}$. We have the equalities, $\boldsymbol{b}^\top \boldsymbol{e}_j = 0$ $(j = 1, \ldots, i-1)$ 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 defined as $T = I_d - \frac{2}{\|\boldsymbol{b}\|^2} \boldsymbol{b}\boldsymbol{b}^\top$, where I_d is a $d \times d$ identity matrix. Using equalities, $T\bar{\boldsymbol{u}} = \bar{\boldsymbol{u}}, T\boldsymbol{e}_j = \boldsymbol{e}_j$ $(j = 1, \ldots, i-1)$, 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 present a proof of the incursion relation,

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

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 divided 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 a result, any vector $\boldsymbol{u} \in \mathcal{W} \cap V_{i-1} \cap \{\boldsymbol{u} \mid \boldsymbol{b}^\top \boldsymbol{u} \geq 0\}$ 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 a linear combination of $\boldsymbol{b}$ and $\Pi_i \bar{\boldsymbol{u}}$. Recall that for $\boldsymbol{u} \in \mathcal{W} \cap V_{i-1} \cap \{\boldsymbol{u} \mid \boldsymbol{b}^\top \boldsymbol{u} \geq 0\}$, we have the 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 $\beta > 0$. Therefore, 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 $\alpha \geq 0$, $\boldsymbol{e}_i^\top \boldsymbol{b} \geq 0$, $\beta > 0$, and $\boldsymbol{e}_i^\top \Pi_i \bar{\boldsymbol{u}} = \bar{u}_i \geq 0$. As a result, $\boldsymbol{u} \in \mathcal{W} \cap V_i$ is confirmed to be valid. From the incursion relation $\mathcal{W} \cap V_{i-1} \cap \{\boldsymbol{u} \mid \boldsymbol{b}^\top \boldsymbol{u} \geq 0\} \subset \mathcal{W} \cap V_i$, we have the inequality of volume formula $\mathrm{Vol}(\mathcal{W} \cap V_i) \geq \mathrm{Vol}(\mathcal{W} \cap V_{i-1})/2$, 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 a result, 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 therefore,

$$P\{\|U - \bar{\boldsymbol{u}}\| \leq \delta/L\} \geq h(R)\mathrm{Vol}(\mathcal{W} \cap V_k) \geq \frac{h(R)}{2^k} \int I[\, \boldsymbol{u} \in \mathcal{W}\,]\, d\boldsymbol{u} \geq \frac{h(R)}{2^k} \int I[\, \boldsymbol{u} \in \mathcal{B}_{\delta,R}\,]\, d\boldsymbol{u}.$$

When $R - r_k < \|\bar{\boldsymbol{u}}\| \leq 1$ and $k = d$, *i.e.,* non-negative constraints are imposed for all components of $\boldsymbol{u}$, there exists an index i such that the region $\mathcal{W}$ is included in $\{\boldsymbol{u} \mid \boldsymbol{e}_i^\top \boldsymbol{u} \geq 0\}$. If such an index does not exist, inequality $0 \leq \bar{u}_i < s \leq r_d$ holds for all $i = 1, \dots, d$, which contradicts $R - r_d < \|\bar{\boldsymbol{u}}\|$. Without loss of generality, the non-negative constraint $u_i \geq 0$ can be deleted, when we calculate the volume of $\mathcal{W} \cap V_d$. As a result, only $d-1$ non-negative constraints are essential, and therefore, the discussion under the assumption $k < d$ holds even for $k = d$. ■

B Proof of Lemma 5.5

Proof. $\bar{\boldsymbol{u}}$ is a maximum solution of $\max_{\boldsymbol{u} \in \mathcal{A}} f(\boldsymbol{x}, \boldsymbol{u})$ with a fixed $\boldsymbol{x}$. A uniform lower bound of $P\{\|U - \bar{\boldsymbol{u}}\| \leq \delta/L\}$ is constructed. Let $C > 0$ be a lower bound of the probability density of P on $\mathcal{A}$. Since the probability distribution is positive on $\mathcal{A}$, it holds that

$$P\big\{\|U - \bar{\boldsymbol{u}}\| < \delta/L\big\} \geq C \int I[\, \boldsymbol{u} \in \mathcal{A},\, \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\,] d\boldsymbol{u}.$$

We derive a lower bound of the above integral. First, we introduce some notations. Let r be the radius of S in Assumption 5.4, and R be the diameter of $\mathcal{A}$ defined by $\sup_{\boldsymbol{u}, \boldsymbol{v} \in \mathcal{A}} \|\boldsymbol{u} - \boldsymbol{v}\|$. The hypersphere of radius w at center $\boldsymbol{y} \in \mathbb{R}^d$ is denoted by $S(\boldsymbol{y}, w)$, *i.e.,* $S(\boldsymbol{y}, w) = \big\{\boldsymbol{x} \in$

$\mathbb{R}^d \mid \|\boldsymbol{x} - \boldsymbol{y}\| \le w\}$. We also define $A_d(\theta)$ for $\theta \in [0, \pi]$ as the surface area of a part of a d-dimensional unit hypersphere defined by

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

In other words, $A_d(\theta)$ is a solid angle in $\mathbb{R}^d$. In Figure 4 (a), $A_d(\theta)$ for $d = 2$ is illustrated by a bold curve.

Assumption 5.4 assures that there exists a hypersphere such that $S(\boldsymbol{c}, r) \subset \mathcal{A}$. Therefore, it holds that

$$\mathrm{Vol}(\{\boldsymbol{u} \in \mathcal{A} \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}) \ \ge\ \mathrm{Vol}(\{\boldsymbol{u} \in S(\boldsymbol{c}, r) \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}). \tag{12}$$

$\mathrm{Vol}(B)$ denotes the volume of the set B with respect to the Lebesgue measure. The right-hand side of (12) is the volume of the intersection of two hyperspheres. Suppose that the radii of these hyperspheres are fixed, and that $\bar{\boldsymbol{u}} \in S(\boldsymbol{c}, r)$. Then, the volume of the intersection attains a minimum value if $\bar{\boldsymbol{u}}$ lies on the boundary of $S(\boldsymbol{c}, r)$. As a result, the inequality

$$\mathrm{Vol}(\{\boldsymbol{u} \in \mathcal{A} \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}) \ge \int I[\boldsymbol{u} \in \mathcal{B}_{\delta,r}]\, d\boldsymbol{u}, \quad 0 \le \delta \le 2Lr,$$
$$\mathcal{B}_{\delta,r} = \{\boldsymbol{u} \in \mathbb{R}^d \mid \|\boldsymbol{u} - r\boldsymbol{e}_1\| \le \delta/L,\ \|\boldsymbol{u}\| \le r\}, \quad \boldsymbol{e}_1 = (1, 0, \dots, 0) \in \mathbb{R}^d$$

holds on the condition $\bar{\boldsymbol{u}} \in S(\boldsymbol{c}, r)$.

Next, we consider the case of $\bar{\boldsymbol{u}} \notin S(\boldsymbol{c}, r)$. Under this condition, there exists a hypersphere $S(\boldsymbol{d}_\beta, \beta r) \subset S(\boldsymbol{c}, r)$ for $0 < \beta < 1$, where $\boldsymbol{d}_\beta$ is defined as

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

Let us define two subsets K_1 and K_2 as

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

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

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

The meaning of $\ell(\beta)$ is illustrated in Figure 4 (b). Note that for $0 \le \ell \le \ell(\beta)$, the inclusion relation $S(\bar{\boldsymbol{u}}, \ell) \cap K_2 \subset K_1 \subset \mathcal{A}$ holds. Therefore, for $0 \le \delta/L \le \ell(\beta)$, it holds that

$$\begin{aligned}\mathrm{Vol}(\{\boldsymbol{u} \in \mathcal{A} \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}) \ &\ge\ \mathrm{Vol}(\{\boldsymbol{u} \in S(\bar{\boldsymbol{u}}, \ell) \cap K_2 \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}) \\ &=\ \frac{A_d(\phi)}{A_d(\pi)}\mathrm{Vol}(S(\bar{\boldsymbol{u}}, \delta/L)),\end{aligned}$$

where ϕ is defined as $\sin\phi = \beta r/\|\bar{\boldsymbol{u}} - \boldsymbol{d}_\beta\|$. Since $\|\bar{\boldsymbol{u}} - \boldsymbol{d}_\beta\| + \beta r \le R$ is satisfied, ϕ is lower bounded by $\sin^{-1}(\beta r/(R - \beta r))$. As a result, a lower bound of the volume of $\{\boldsymbol{u} \in \mathcal{A} \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}$ for $0 \le \delta/L \le \ell(\beta)$ is given by

$$\mathrm{Vol}(\{\boldsymbol{u} \in \mathcal{A} \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\}) \ \ge\ \frac{A_d(\phi)}{A_d(\pi)}\mathrm{Vol}(S(\bar{\boldsymbol{u}}, \delta/L)) \ \ge\ \frac{A_d(\sin^{-1}\frac{\beta r}{R - \beta r})}{A_d(\pi)}\mathrm{Vol}(S(\bar{\boldsymbol{u}}, \delta/L)).$$

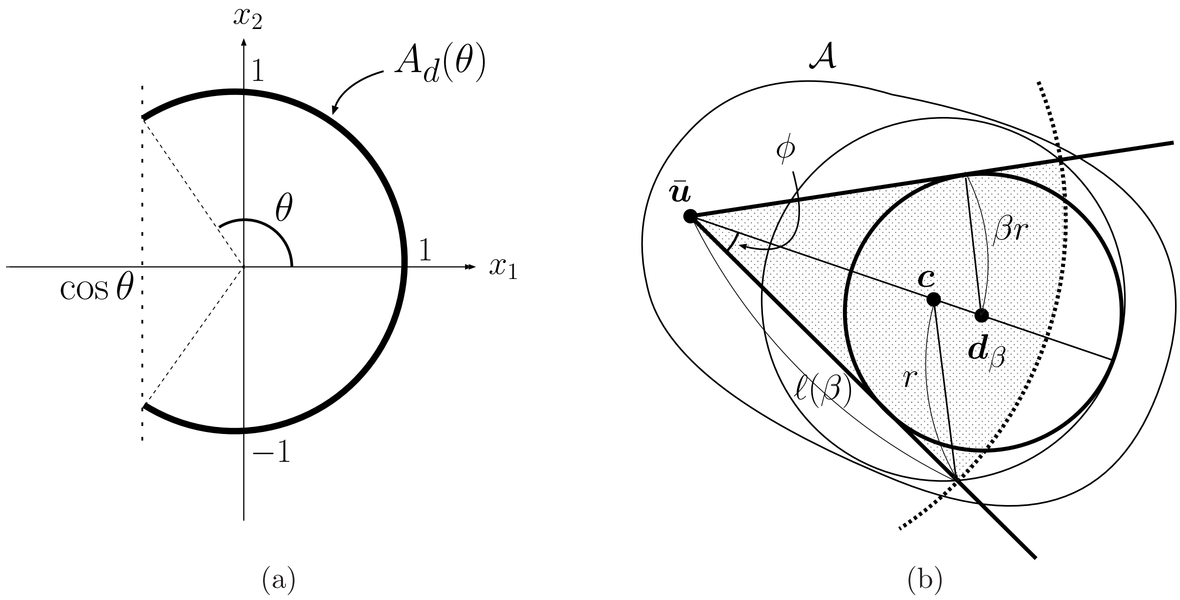

Figure 4: (a) Surface area $A_d(\theta)$ of a part of the unit hypersphere $\{\boldsymbol{x} = (x_1, \cdots, x_d) \in \mathbb{R}^d \;\; \|\boldsymbol{x}\| = 1,\ \cos\theta \le x_1\}$. In the figure, surface area for $d = 2$, which denotes the length of the thick arch, is shown. (b) Definitions of $S(\boldsymbol{c}, r)$, $S(\boldsymbol{d}_\beta, \beta r)$, ϕ, and $\ell(\beta)$ are illustrated. The shaded set denotes $S(\bar{\boldsymbol{u}}, \ell(\beta)) \cap K_2$.

Therefore, we obtain

$$P\{\boldsymbol{u} \in \mathcal{A} \mid \|\bar{\boldsymbol{u}} - \boldsymbol{u}\| < \delta/L\} \ge C\, \frac{A_d\big(\sin^{-1} \frac{\beta r}{R - \beta r}\big)}{A_d(\pi)}\, \mathrm{Vol}(S(\mathbf{0}, \delta/L)) \tag{13}$$

for $0 \le \delta/L \le \ell(\beta)$. Substituting

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

into the cosine formula

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

we obtain

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

For $\|\bar{\boldsymbol{u}} - \boldsymbol{c}\| \ge r$, the first term of $\ell(\beta)$ is an increasing function of $\|\bar{\boldsymbol{u}} - \boldsymbol{c}\|$. Therefore, we have the inequality,

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

Therefore, inequality (13) holds for $0 \le \delta/L \le r\left(\sqrt{1-\beta^2}+\frac{2\sqrt{1-\beta}}{2-\beta}\right)$. It can be seen that $0 \le \sqrt{1-\beta^2}+\frac{2\sqrt{1-\beta}}{2-\beta} \le 2$ for $0 \le \beta \le 1$.

In summary, for any $\beta \in (0,1)$, it holds that

$$q_3(\delta) = C \min\left\{ \frac{A_d(\sin^{-1}\frac{\beta r}{R-\beta r})}{A_d(\pi)} \mathrm{Vol}(S(\mathbf{0},\delta/L)),\ \int I[\ \boldsymbol{u} \in \mathcal{B}_{\delta,r}\ ]\ d\boldsymbol{u} \right\} \ \le\ p(\delta,\boldsymbol{x}),$$

$$0 \le \delta \le rL\left(\sqrt{1-\beta^2}+\frac{2\sqrt{1-\beta}}{2-\beta}\right).$$

It is clear that the lower bound is a strictly increasing function with respect to δ. The solid angle $A_d(\theta)$ is represented by the incomplete beta function. ■

C Estimation of parameters in q_3

A unified way of identifying C and L may not exist, when C is a lower bound of the probability density of P on $\mathcal{A}$ and L is the Lipschitz constant of the constraint function $f(\boldsymbol{x},\boldsymbol{u})$. The Lipschitz constant for a specific function f is described below.

Lipschitz constant L: We consider a quadratic constraint function $f(\boldsymbol{x},\boldsymbol{u})$ in $\boldsymbol{x}$ which is linearly perturbed in terms of $\boldsymbol{u}$, *i.e.*, $f(\boldsymbol{x},\boldsymbol{u}) := \boldsymbol{x}^\top \boldsymbol{Q}(\boldsymbol{u})\boldsymbol{x} + \boldsymbol{q}(\boldsymbol{u})^\top \boldsymbol{x} + \gamma(\boldsymbol{u})$, 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

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

It is possible to roughly estimate 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 the diameter r_x of $\mathcal{X}$ such that $\max\limits_{\boldsymbol{x}_1,\boldsymbol{x}_2\in\mathcal{X}} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| \le r_x$.

For some uncertainty sets, it is not so difficult to find the inscribed hypersphere and the diameter.

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

$$\left|\ \max\ \boldsymbol{e}_i^\top \boldsymbol{u} \quad \text{s.t. } \boldsymbol{u} \in \mathcal{A}\ ,\right. \tag{14}$$

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 the inscribed hypersphere of $\mathcal{A}$ with some examples. For the first example, we assume that uncertainty set $\mathcal{A}$ is described as

$$\mathcal{A} = \left\{ \boldsymbol{u} \in \mathbb{R}^d \mid \boldsymbol{a}_i^T \boldsymbol{u} \le b_i,\ i = 1, \dots, p \right\}.$$

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

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

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

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

holds. As a result, the maximum radius of the inscribed hypersphere is given by the optimal value of the linear program,

$$\left| \begin{array}{l} \max_{r,\boldsymbol{c}} \ r \\ \quad \text{s.t.} \ \ r\,\|\boldsymbol{a}_i\| + \boldsymbol{a}_i^T \boldsymbol{c} \le b_i, \quad i = 1, \dots, p, \\ \qquad\qquad\qquad\qquad -r \le 0. \end{array} \right.$$

In the second example, $\mathcal{A}$ is specified by quadratic constraints,

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

where A_j denotes a positive definite matrix, and the maximum radius r is determined such that $\mathcal{B} \subset \mathcal{A}$. We first determine 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

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

which is equivalent to the condition that there exists $\lambda_j \ge 0$ such that

$$\begin{pmatrix} -\lambda_j - e_j - \boldsymbol{d}_j^T A_j^{-1} \boldsymbol{d}_j & \boldsymbol{0}^T & (\boldsymbol{c} + A_j^{-1}\boldsymbol{d}_j)^T \\ \boldsymbol{0} & \lambda_j I & rI \\ \boldsymbol{c} + A_j^{-1}\boldsymbol{d}_j & rI & A_j^{-1} \end{pmatrix} \succeq 0. \tag{15}$$

For the details of the derivation, refer to [3]. Therefore, for the uncertainty set described by quadratic functions and linear functions, the maximum radius of inscribed hypersphere can be determined by semi-definite programs.

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

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. *Mathematics of Operations Research*, 25(1):1–13, 1999.

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

[4] G. Calafiore and M. C. Campi. A new bound on the generalization rate of sampled convex programs. In *IEEE Conference on Decision and Control*, volume 5, pages 5328–5333, 2004.

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

[6] G. Calafiore and M. C. Campi:. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[7] G. Calafiore and L. El Ghaoui. Distributionally robust chance-constrained linear programs with applications. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.

[8] M. C. Campi and S. Garatti. The exact feasibility of randomized solutions of robust convex programs. *Optimization-online*, 2008. available at http://www.optimization-online.org/DB_HTML/2007/07/1727.html.

[9] A. Charnes and W. W. Cooper. Chance constrained programming. *Management Science*, 6:73–79, 1959.

[10] L. El Ghaoui and L. H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18:1035–1064, 1997.

[11] L. El Ghaoui and L. H. Lebret. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization*, 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] V. Krishnamurthy. Robust optimization in finance. Technical report, Second Summer Paper for the Doctoral Program (Carnegie Mellon University, supervised by Tütüncü R.H.), 2004.

[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] P. Massart. The tight constant in the dvoretzky-kiefer-wolfowitz inequality. *Annals of Probability*, 18(3):1269–1283, 1990.

[17] A. Takeda, S. Taguchi, and T. Tanaka. Relaxation methods for robust deviation and relative robust decision models. Technical report, Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2006.

[18] R. Tempo, E. W. Bai, and F. Dabbene. Probabilistic robustness analysis: Explicit bounds for the minimum number of samples. *Systems & Control Letters*, 30:237–242, 1997.

[19] V. N. Vapnik. *Statistical Learning Theory.* Wiley, 1998.