Concentration of Measure for Chance-Constrained Optimization

Sadegh Soudjani * **Rupak Majumdar** **

* *School of Computing, Newcastle University, United Kingdom*
** *Max Planck Institute for Software Systems, Germany*

Abstract: Chance-constrained optimization problems optimize a cost function in the presence of probabilistic constraints. They are convex in very special cases and, in practice, they are solved using approximation techniques. In this paper, we study approximation of chance constraints for the class of probability distributions that satisfy a concentration of measure property. We show that using concentration of measure, we can transform chance constraints to constraints on expectations, which can then be solved based on scenario optimization. Our approach depends solely on the concentration of measure property of the uncertainty and does not require the objective or constraint functions to be convex. We also give bounds on the required number of scenarios for achieving a certain confidence. We demonstrate our approach on a non-convex chanced-constrained optimization, and benchmark our technique against alternative approaches in the literature on chance-constrained LQG problem.

Keywords: Chance-Constrained Optimization, Non-Convex Scenario Program, Concentration of Measure, Stochastic Optimization, Randomized Optimization, Linear Quadratic Gaussian

1. INTRODUCTION

Chance-constrained programming (CCP) (Prékopa (1995)) is an important technique to optimize a cost function in the presence of random parameters. It arises in many problems in engineering and finance, where a full description of the system and the effect of all factors may not be available. For example, in controller synthesis problems it is desirable to pick the "best" control input among the set of possible valid inputs in the presence of disturbances for which we may only have a stochastic model. This problem is cast as an optimization question which minimizes an objective function modeling the system performance, while ensuring that the system constraints are met "as much as possible." That is, while constraints may be violated under rare, unexpected, events, by modeling or approximating the distribution of the random parameter, it makes sense to call decisions feasible (in a stochastic sense) whenever they are feasible with high probability. CCP is in contrast to the robust approach, which bounds the worst-case range of disturbances, and which can consequently be extremely conservative.

Unfortunately, chance-constrained optimization problems have feasible domains which are in general non-convex. Thus, other than a few restricted cases, numerical methods must be employed for obtaining a solution. A widely used numerical method is *randomized optimization* (Campi and Garatti (2008); Calafiore (2010)). In this approach, one constructs a scenario program (SP) by taking samples from the uncertain variables and requires that the (chance) constraints hold for all observed values of samples. If the objective and constraint functions are convex, the required number of samples can be selected to guarantee that with certain confidence the solution of SP is feasible for the original CCP. The main advantage of convex SP is that no knowledge of properties of uncertainty is required but the guarantee holds under the special condition that objective and constraint functions are convex with respect to the decision variables. Moreover, increasing number of samples reduces in general the chance of getting a feasible SP and decreases the (sub)optimal performance. In recent work, Grammatico et al. (2016) address random non-convex programs by solving multiple SPs with different convex objective functions, but they restrict the constraint function to be separable non-convex.

In this paper, we show that if the uncertainty distribution satisfies a *concentration of measure* property, then one can significantly generalize the scenario-based approach. Concentration of measure phenomenon roughly states that, if a set in a probability space has measure at least one half, "most" of the points in the probability space are "close" to the set, and if a function on the probability space is regular enough, the chance that this function deviates too much from its expectation (or median) is very small.

In particular, given a CCP, we construct a SP which takes samples from the uncertain parameters, but, instead of forcing the constraint to hold for all individual samples, requires that the constraint is satisfied *in average* for the samples with a predefined margin from its boundary. With an appropriate choice for the margin, which depends on the concentration of measure property, we can relate the feasible solutions of the original CCP with that of the SP. Our approach does *not* assume convexity of the objective or constraint functions, but only Lipschitz continuity of the constraint function w.r.t. the uncertainty parameter, which is a reasonable assumption and less restrictive. Thus, we show that concentration of measure can be used to significantly expand the scope of randomized optimization.

Concentration of measure is a powerful property of distributions, and has been used in many problems in combinatorics (Alon and Spencer (2016)), the analysis of randomized algorithms (Barvinok (1997)), and for discrete optimization (Dubhashi and Panconesi (2009)). They include "classical" Chernoff bounds for sums of independent Bernoulli variables and Gaussian concentration inequalities, to more advanced Poincaré, log-Sobolev, and Talagrand inequalities. In the general form, the inequalities take the form

$$\mathbb{P}(|f-\mathbb{E}(f)|>t)\le c\cdot e^{-c't}$$

bounding probability of f deviating from its expected value. While classical results show the existence of such constants c and c', for practical applications, we are also interested in optimizing the constants: the smaller c and larger c', the less conservative solution obtained from SP. Thus, we have revisited proofs of concentration of measure (Ledoux (1999)) with an attempt to improve the constants.

The rest of this paper is organized as follows. In Section 2 we define chance-constrained programs and the concentration of measure property for the uncertainty. In Section 3 we discuss the construction of a scenario program and the selection of the number of scenarios, and connect these to the original CCP through concentration of measure. Section 4 gives an overview of the concentration phenomenon together with improved bounds. We use a non-convex CCP as a running example and in Section 5 we compare our approach with alternative techniques from literature on an LQG problem.

2. PRELIMINARIES: CCP

Consider a random variable $\delta\in\Omega\subseteq\mathbb{R}^p$ defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$. With respect to this random variable, we define the Chance-Constrained Program

$$CCP:\begin{cases}\min\limits_{x\in\mathcal{X}} J(x)\\ s.t.\ \mathbb{P}\left(\{\delta\in\Omega\,|\,g(x,\delta)\le 0\}\right)\ge 1-\epsilon.\end{cases}\tag{1}$$

where $x\in\mathcal{X}$ is the decision variable belonging to the compact admissible set $\mathcal{X}\subset\mathbb{R}^n$, $J:\mathbb{R}^n\to\mathbb{R}$ is a lower semi-continuous cost function, $g:\mathbb{R}^n\times\mathbb{R}^p\to\mathbb{R}$ is the constraint function, and $\epsilon\in(0,1)$ is constraint-violation tolerance. We assume that for any fixed $\bar{x}\in\mathcal{X}$, the mapping $\delta\mapsto g(\bar{x},\delta)$ is measurable and for any fixed $\bar{\delta}\in\Omega$, the mapping $x\mapsto g(x,\bar{\delta})$ is lower semi-continuous.

Theorem 1. The CCP (1) is well-defined and attains a solution if it is feasible.

Proof of Theorem 1 relies on reverse Fatou's lemma (Royden (1988)) and is omitted here due to space limitation. Note that here we do not put any assumption on the convexity of the function $g(\cdot,\delta)$ or on the cost function $J(\cdot)$, which is a common practice in the scenario approach for optimization (cf. Campi and Garatti (2011); Grammatico et al. (2016)).

In this paper, we make the following assumption on the probability space $(\Omega,\mathcal{F},\mathbb{P})$.

Assumption 2. (Concentration of Measure). Probability space $(\Omega,\mathcal{F},\mathbb{P})$ satisfies the inequality

$$\mathbb{P}\left(|f(\delta)-\mathbb{E}\left(f(\delta)\right)|\le t\right)\ge 1-h(t),\quad\forall t\ge 0,\tag{2}$$

for any function $f:\Omega\to\mathbb{R}$ that belongs to a class of functions $\mathcal{C}$ such that $g(x,\cdot)\in\mathcal{C}$ for any $x\in\mathcal{X}$, and where $h:\mathbb{R}^{\ge 0}\to[0,1]$ is a monotonically decreasing function.

We use Assumption 2 in the next section to construct a scenario program for finding a (possibly sub-optimal) solution of CCP (1). We discuss in Section 4 how this assumption holds for many well-known distributions. Note that monotonicity of $h(\cdot)$ is not restrictive since we can always replace $h(\cdot)$ with another monotonically decreasing function $\bar{h}(\cdot)$. Inequality (2) is still valid with $\bar{h}(\cdot)$ if $\bar{h}(\cdot)\ge h(\cdot)$.

The following non-convex CCP is used for demonstrating our approach.

Example 3. Consider uniformly distributed random variable $\delta\sim U[0,1]$ and the following CCP

$$\begin{cases}\min\ x_2\\ s.t.\ x_1\in[0,1],\ \ x_2\in\mathbb{R}\\ \qquad\mathbb{P}\left(\{\delta\in\Omega\,:\,x_2\ge 1-|x_1-\delta|\}\right)\ge 1-\epsilon.\end{cases}\tag{3}$$

Since any $(x_1,x_2)\in\mathbb{R}^2$ with $x_2\ge 1$ is feasible and with $x_2<0$ is infeasible, we can put $\mathcal{X}=[0,1]\times[0,1]$. In this example $g(x,\delta)=1-x_2-|x_1-\delta|$ with is obviously non-convex on $\mathcal{X}$ for any $\delta\in[0,1]$. Feasible domain of this CCP is non-convex. The exact optimal value of CCP (3) will be $1-\epsilon$ which is obtained at $x=(0,1-\epsilon)$ or $x=(1,1-\epsilon)$.

3. FROM CCP TO SP

We define the following Expected-Constrained Program (ECP) that is tightly connected to CCP (1) under Assumption (2), for each choice of the parameter $\beta\ge 0$:

$$\mathrm{ECP}(\beta):\begin{cases}\min\limits_{x\in\mathcal{X}} J(x)\\ s.t.\ \mathbb{E}g(x,\delta)+\beta+h^{-1}(\epsilon)\le 0.\end{cases}\tag{4}$$

Proposition 4. The feasible domain of CCP (1) includes the feasible domain of Expected-Constrained Program (4) for all values of $\beta\ge 0$.

The expectation operator in (4) still prevents us to efficiently compute a solution. Therefore, we define the following Scenario Program (SP) that replaces the expectation with its empirical mean,

$$\mathrm{SP}(\gamma):\begin{cases}\min\limits_{x\in\mathcal{X},\gamma_i\in\mathbb{R}} J(x)\\ s.t.\quad g(x,\delta^i)\le\gamma_i,\quad i=1,\ldots,N,\\ \qquad\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\gamma_i+\gamma+h^{-1}(\epsilon)\le 0,\end{cases}\tag{5}$$

where δ^i, $i=1,2,\ldots,N$, each defined on the probability space $(\Omega,\mathcal{F},\mathbb{P})$, are independent and $\gamma>0$ is a positive parameter.

We make the following assumption to formally connect the two optimizations (4) and (5).

Assumption 5. Variance of $g(x,\delta)$ is bounded: there is a constant M_v such that

$$\mathbb{E}\left[(g(x,\delta)-\mathbb{E}(g(x,\delta)))^2\right]\le M_v,\quad\forall x\in\mathcal{X}.$$

This assumption already holds if $g(x,\delta)$ is bounded.

Theorem 6. Under Assumption 5, any feasible solution of $\mathrm{ECP}(\beta)$ in (4) is a feasible solution of $\mathrm{SP}(\gamma)$ in (5) with probability

$$1-\frac{M_v}{N(\beta-\gamma)^2},$$

for any $\beta>\gamma\ge 0$ and any number of samples N.

Corollary 7. Under Assumption 5, if $\mathrm{ECP}(\beta)$ in (4) is feasible, then $\mathrm{SP}(\gamma)$ in (5) is also feasible with confidence $1-\alpha\in(0,1)$ when $\beta>\gamma\ge 0$ and number of samples N satisfy the inequality $N\ge M_v/[\alpha(\beta-\gamma)^2]$.

The above theorem relates feasibility of ECP (4) to that of SP (5). In practice it is desirable to have confidence on the solution of SP (5) being a (possibly sub-optimal) solution for CCP (1). In order to provide such a confidence, we require one of the following technical assumptions.

Assumption 8. ECP(β_0) (4) is feasible for some $\beta_0 > 0$.

Assumption 9. SP (5) is feasible for any $N \in \mathbb{N}$ and any choice of samples $\delta^i \in \Omega$.

In case of Assumption 8 we get a lower bound on probability of having a feasible SP (5) according to Theorem 6. In case of Assumption 9, SP (5) always has an optimal value since $\mathcal{X}$ is compact, and its optimal value can in principle be written as a function of samples $\delta^i \in \Omega$. Note that these assumptions are less restrictive than the one posed in (Campi and Garatti (2008)) as we do not require non-empty interior for feasible domain of SP (5).

Proposition 10. Under Assumptions 5 and 8, if SP(γ) in (5) is feasible for samples δ^i, then ECP(0) in (4) is also feasible with confidence $1-\alpha$, where number of samples N satisfies

$$N \geq \frac{M_v}{\alpha}\left[\frac{1}{\gamma^2} + \frac{1}{(\beta_0-\gamma)^2}\right]. \tag{6}$$

Moreover, having concentration inequality (2), CCP (1) will be also feasible with confidence $1-\alpha$.

Remark 11. This theorem can give a tradeoff between optimality and number of samples for achieving a confidence. We solve SP(γ) with $\gamma = \beta_0/2$ and the associated number of samples in (6). If the optimal solution happens to be at the boundaries of optimization constraint, we can improve the solution at the cost of larger computational effort by reducing γ and increasing number of samples.

Proposition 12. Under Assumptions 5 and 9, any solution of SP(γ) in (5) is a feasible solution of ECP(0) in (4) (and hence a feasible solution of CCP (1) having Assumption 2) with confidence $1-\alpha$ if $\gamma > 0$ and number of samples N satisfy $N \geq M_v/(\alpha\gamma^2)$.

Example 13. Uniformly distributed random variable $\delta \sim U[0,1]$ satisfies the inequality

$$\mathbb{P}\left(\delta \,:\, g(x,\delta) - \mathbb{E}\left(g(x,\delta)\right) \leq t\right) \geq \min\{2t, 1\},$$

for all $t \geq 0$ and $x \in \mathbb{R}^2$. This enables us to have the same results for uniform distribution with $h(t) = \max\{1-2t, 0\}$. The ECP will be

$$\begin{cases} \min \ x_2 \\ s.t. \quad x_1, x_2 \in [0,1] \\ \qquad 1 - x_2 - \mathbb{E}|x_1 - \delta| + \beta + (1-\epsilon)/2 \leq 0. \end{cases} \tag{7}$$

The feasible domain of ECP (7) is $x_2 \geq 1 + x_1 - x_1^2 + \beta - \epsilon/2$. As theoretically shown, this domain is a subset of feasible domain of CCP (3). This gives the sub-optimal value $(1-\epsilon/2+\beta)$ for CCP (3) at exactly the same optimal points of CCP (3).

Now let us examine the following SP without the need for explicit computation of $\mathbb{E}|x_1 - \delta|$,

$$\begin{cases} \min \ x_2 \\ s.t. \quad x_1, x_2 \in [0,1], \ \gamma_i \in \mathbb{R} \\ \qquad 1 - x_2 - |x_1 - \delta^i| \leq \gamma_i, \quad i = 1, \ldots, N \\ \qquad \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \gamma_i + \gamma + (1-\epsilon)/2 \leq 0. \end{cases} \tag{8}$$

Solution of this optimization is

$$x_2^* = 1 - \epsilon/2 + \gamma + \left|\frac{\sum_i \delta^i}{N} - \frac{1}{2}\right|,$$

if $x_2^* \leq 1$, otherwise it is infeasible. As we see, $\gamma > 0$ can be selected sufficiently small and N sufficiently large such that the optimization is feasible. The optimal value is taken at one of the two boundary points $x_1 = 0$ or $x_1 = 1$ depending on the term inside absolute value.

In order to theoretically relate (8) and (3), we require to check the imposed assumptions. We verify Assumption 8 by trying a single point in (7). Taking $x_1 = 0$ reveals that (7) is feasible for $\beta_0 \leq \epsilon/2$. According to Proposition 10, we have

$$x_2^* = 1 - \epsilon/2 + \beta_0/2 + \left|\frac{\sum_i \delta^i}{N} - \frac{1}{2}\right|,$$

as a sub-optimal solution for CCP (3) with confidence $(1-\alpha)$ if we take $N \geq 8/(3\alpha\beta_0^2)$. This solution is very close to $1-\epsilon/2$ when we take the limit $\beta_0 \to 0$ and $N \to +\infty$. On the other hand, if we try $x_1 = 0.5$, we get that $\beta_0 \leq \epsilon/2 - 1/4$, which puts the constraint $\epsilon > 0.5$ for drawing the same conclusion.

We can make the choice of x for verifying Assumption 8 more intelligent. Intuitively, we take N_0 samples and minimize the empirical mean of $g(x,\delta)$. Algorithm 1 presents the combined approach of verifying Assumption 8 and finding a feasible solution for CCP (1). In this algorithm **do-until** loop tries to find $\hat{x}$ for verifying Assumption 8. In step 6 number of samples is selected according to the outcome of this loop and then SP (5) is solved. Note that **do-until** loop terminates with probability one if Assumption 8 holds. The choice of N_0 can also be made adaptive w.r.t. the outcome of optimization is step 3.

Algorithm 1 Computing a (possibly sub-optimal) solution for CCP (1).

input: CCP (1) with J, g, ϵ, Confidence $1-\alpha$, Number of samples N_0, constant M_v and function h
1: **do:**
2: take N_0 independent samples δ^i
3: Compute $(\hat{x}, \hat{\gamma}, \hat{\gamma}_i) = \arg\max\{\gamma \,|\, x \in \mathcal{X}, \gamma, \gamma_i \in \mathbb{R}\}$ under constraints in (5)
4: Compute $\beta_0 = -\mathbb{E}(g(\hat{x},\delta)) - h^{-1}(\epsilon)$
5: **until** $\beta_0 > 0$
6: Select number of samples $N \geq 8M_v/(\alpha\beta_0^2)$
7: Take N independent samples δ^i
8: Solve SP (5) with $\gamma = \beta_0/2$ to get x^* if it is feasible
output: Feasible point x^* for CCP (1) with confidence $1-\alpha$

4. CONCENTRATION OF MEASURE

Assumption 2 has a central role in our approach but establishing inequality (2) is generally difficult for multivariate probability distributions. In this section, we discuss *concentration of measure* that enables us to identify classes of distributions that satisfy this assumption. The most important feature of this phenomenon is that it is dimension free, and thus extends the results from one dimension to product probability spaces.

Concentration of measure phenomenon can be explained in terms of sets. It roughly states that, if a set A in probability space Ω has measure at least one half, "most" of the points of Ω are "close" to A. However, we are more often interested in functions rather than sets in order to satisfy Assumption 2 with inequalities of the form (2). Concentration of measure phenomenon has also an interpretation in terms of functions that if a function f on a metric space (Y, d) equipped with a probability measure μ is sufficiently regular, it is very concentrated around its median (hence around its mean).

In the following we discuss concentration of measure inequalities based on Poincaré and log-Sobolev inequalities. We also discuss the modified log-Sobolev inequality for exponential distributions. In measure theory, the focus of these results (e.g., Ledoux (2005)) is mainly on proving the existence of function $h(\cdot)$ in (2). We have improved the concentration of measure results founded on these inequalities and have classified the well-known distributions that satisfy one of these inequalities.

Let $(Y, \mathcal{B}, \mu)$ be a probability space. We denote by $\mathbb{E}$ integration w.r.t. μ, and by $(L^p, \|\cdot\|_\infty)$ the Lebesgue spaces over $(Y, \mathcal{B}, \mu)$. We further denote the variance of any function $f \in L^2$ by

$$\mathsf{Var}(f) := \mathbb{E}\left[(f - \mathbb{E}(f))^2\right] = \mathbb{E}(f^2) - (\mathbb{E}(f))^2. \quad (9)$$

If f is a non-negative function on Y, we define the entropy of f w.r.t. μ as

$$\mathsf{Ent}(f) := \mathbb{E}(f \log f) - \mathbb{E}(f) \log \mathbb{E}(f). \quad (10)$$

In order to have a bounded quantity for $\mathsf{Ent}(f)$ we restrict this definition to functions that $\mathbb{E}(f \log f) < \infty$ with the convention $0 \log 0 = 0$. Note that $\mathsf{Ent}(f) \geq 0$ and that $\mathsf{Ent}(af) = a\mathsf{Ent}(f)$ for any $a \geq 0$. Using Jensen's inequality one can easily show that $\mathsf{Ent}(f) \geq 0$. On some subset $\mathcal{A} \subseteq L^2$ of measurable functions $f : Y \to \mathbb{R}$, consider now a map, or energy, $\mathcal{E} : \mathcal{A} \to \mathbb{R}_{\geq 0}$.

Definition 14. We say that μ satisfies a spectral gap or *Poincaré inequality* w.r.t. the pair $(\mathcal{A}, \mathcal{E})$ if there exists $C > 0$ such that

$$\mathsf{Var}(f) \leq C\mathcal{E}(f), \quad \forall f \in \mathcal{A}. \quad (11)$$

Definition 15. We say that μ satisfies a *logarithmic Sobolev inequality* w.r.t. the pair $(\mathcal{A}, \mathcal{E})$ if there exists $C > 0$ such that

$$\mathsf{Ent}(f^2) \leq C\mathcal{E}(f), \quad (12)$$

for every $f \in \mathcal{A}$ with $\mathbb{E}(f^2 \log f^2) < \infty$.

Having these abstract definitions we assume Y is a metric space (Y, d) equipped with its Borel sigma algebra $\mathcal{B}$. The energy functional $\mathcal{E}$ can be selected as

$$\mathcal{E}(f) := \mathbb{E}\left(|\nabla f|^2\right) = \int_Y |\nabla f|^2(y) d\mu(y), \quad (13)$$

where $|\nabla f|$ is the abstract length of the gradient of f,

$$|\nabla f|(y) := \limsup_{d(y', y) \to 0} \frac{|f(y') - f(y)|}{d(y', y)}.$$

The subset $\mathcal{A} \subseteq L^2$ will also be

$$\mathcal{A} := \{f : Y \to \mathbb{R},\ f \text{ Lipschitz}\}, \quad (14)$$

the set of all Lipschitz functions on Y, which implies $|\nabla f|(y) \leq \|f\|_{Lip}$ for any $f \in \mathcal{A}$.

Theorem 16. Let (Y, d, μ) be a metric probability space, energy functional (13), and class $\mathcal{A}$ in (14). If μ satisfies the Poincaré inequality in (11) with constant C w.r.t. $(\mathcal{A}, \mathcal{E})$, then we have the exponential concentration

$$\mu\left(|f - \mathbb{E}(f)| > t\right) \leq 6.8 e^{-\lambda_0 t}, \quad \forall f \in \mathcal{A},\ \forall t > 0, \quad (15)$$

with $\lambda_0 = \frac{1}{\rho}\sqrt{\frac{2}{C}}$, and $\rho = \|f\|_{Lip}$.

This inequality is proved in (Ledoux (1999)) with focus on the existence of exponential bound $\kappa e^{-\lambda_0 t}$ and the bound $240 e^{-\lambda_0 t}$ is provided in (Naor (2008)). Theorem 16 improves this bound by reducing it to $6.8 e^{-\lambda_0 t}$.

Theorem 17. (Herbst's Theorem). Let (Y, d, μ) be a metric probability space, energy functional (13), and class $\mathcal{A}$ in (14). If μ satisfies the log-Sobolev inequality in (12) with constant C w.r.t. $(\mathcal{A}, \mathcal{E})$, then we have concentration inequality

$$\mu\left(|f - \mathbb{E}(f)| > t\right) \leq 2 e^{-\lambda_1 t^2}, \quad \forall f \in \mathcal{A},\ \forall t > 0, \quad (16)$$

where $\lambda_1 = 1/(\rho^2 C)$ and $\rho = \|f\|_{Lip}$.

One important feature of both variance and entropy defined in (9)-(10) is their product property (Ledoux (1999)). Assume we are given probability spaces $(Y_i, \mathcal{B}_i, \mu_i)$, $i = 1, 2, \ldots, n$. Denote their product probability space by $(Y, \mathcal{B}, \mu)$, where $\mu := \mu_1 \otimes \cdots \otimes \mu_n$, $Y := Y_1 \times \cdots Y_n$ and $\mathcal{B}$ is the product sigma algebra. Given a function $f : Y \to \mathbb{R}$ on the product space, we define functions $f_i : Y_i \to \mathbb{R}$, with

$$f_i(y_i) = f(y_1, \ldots, y_{i-1}, y_i, y_{i+1}, \ldots, y_n),$$

with $y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n$ being fixed. Under appropriate integrability conditions, we have the following inequalities for variance and entropy

$$\mathsf{Var}_\mu(f) \leq \sum_{i=1}^n \mathbb{E}_\mu\left(\mathsf{Var}_{\mu_i}(f_i)\right)$$

$$\mathsf{Ent}_\mu(f) \leq \sum_{i=1}^n \mathbb{E}_\mu\left(\mathsf{Ent}_{\mu_i}(f_i)\right),$$

where $\mathsf{Var}, \mathsf{Ent}$, and $\mathbb{E}$ are computed w.r.t. the measure written as subscript. This product property tells us that in order to establish a Poincaré or logarithmic Sobolev inequality in product spaces, it will be enough to deal with dimension one (Talagrand (1995)).

In the following we discuss some well-known probability measures that satisfy concentration of measure inequality.

4.1 Gaussian concentration inequality

Suppose μ is the standard Gaussian measure on $\mathbb{R}$. The Logarithmic Sobolev inequality $\mathsf{Ent}_\mu(\phi^2) \leq 2\mathbb{E}_\mu\left(\phi'^2\right)$ holds for any smooth function $\phi : \mathbb{R} \to \mathbb{R}$. By the product property of entropy, the multivariate Gaussian measure on $\mathbb{R}^n$ satisfies (12) with $C = 2$. Then the concentration inequality (16) holds for Gaussian measure with $\lambda_1 = 1/2\rho^2$.

4.2 Exponential concentration inequality

Consider the exponential measure $d\mu(y) = \frac{1}{2} e^{-|y|} dy$ w.r.t. Lebesgue measure on $\mathbb{R}$. Then, $\mathsf{Var}_\mu(\phi) \leq 4\mathbb{E}_\mu\left(\phi'^2\right)$ for any smooth ϕ. By the product property of variance, product of exponential measures on $\mathbb{R}^n$, denoted by ν_n, satisfies (11) with $C = 4$. Then the concentration inequality (15) holds for multivariate exponential measure with $\lambda_0 = 1/\rho\sqrt{2}$. This bound can be further improved for small values of t as

$$\nu_n\left(|f - \mathbb{E}(f)| > t\right) \leq 2 \exp\left(-\min\left(\frac{\kappa_1 t}{\beta}, \frac{\kappa_2 t^2}{\rho^2}\right)\right), \quad (17)$$

for every $t \geq 0$, with $\kappa_1 = \frac{1}{4}$, $\kappa_2 = \frac{1}{16}$, $\rho = \|f\|_{Lip}$, and $\beta \geq 0$ satisfying

$$|f(y) - f(y')| \leq \beta \sum_{i=1}^n |y_i - y_i'|, \quad \forall y, y' \in \mathbb{R}^n.$$

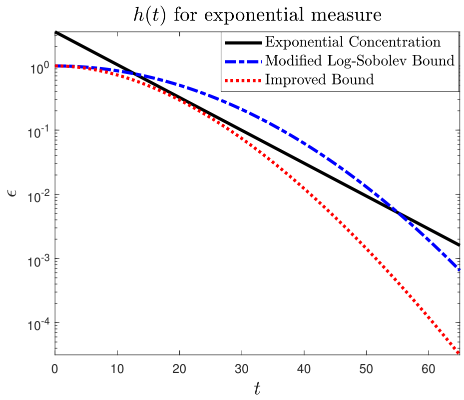

Fig. 1. Comparing $\epsilon = h(t)$ form three concentration inequalities of exponential measure. Plot is in logarithmic scale. Black curve is the bound in (15) for exponential measure, blue curve is the bound from modified log-Sobolev inequality (17) and red in the improved bound in (18).

The proof of (17) is presented in (Bobkov and Ledoux (1997)) and is based on a modified version of log-Sobolev inequality (12). Since we are interested in the tightest possible bound in (17), we improve the above bound in the following theorem.

Theorem 18. Constants κ_1, κ_2 in (17) can be selected freely with the condition that $\kappa_1, \kappa_2 \in (0,1)$, $(1-\kappa_1)^2 = 8\kappa_2$. Bound (17) can also be improved by $2\exp\left(-u^2(t)\right)$ with

$$u(t) := \frac{\rho\sqrt{2}}{\beta}\left[\sqrt{1+\frac{\beta t}{2\rho^2}}-1\right]. \tag{18}$$

Figure 1 compares three upper bounds (15), (17), and (18) (for one-side concentration, i.e., without factor 2) for values $\rho = 6$ and $\beta = 1$. In this case (18) gives the smallest value of $t = h^{-1}(\epsilon)$ for all values of ϵ.

Remark 19. If a function f is almost everywhere differentiable, constants ρ and β can be computed as

$$\sum_{i=1}^{n} |\partial_i f|^2 \leq \rho^2 \quad \text{and} \quad \max_{1\leq i\leq n} |\partial_i f| \leq \beta \quad \text{a.e.}\ .$$

4.3 Strong Log-Concave Measures

A measure μ on $\mathbb{R}^n$ is strongly log-concave if $d\mu(y) = \mu(y)dy$ with

$$\mu(y) = h(y)c\gamma(cy), \quad c > 0,$$

where $\ln h(\cdot)$ is concave and $\gamma(\cdot)$ is density function of standard Gaussian measure. Strong log-concavity is preserved under convolution and marginalization (Saumard and Wellner (2014)). A sufficient condition of being strong log-concave is that $(-\log\mu)''(y) \geq \lambda\mathbb{I}_n$ for some $\lambda > 0$ for all $y \in \mathbb{R}^n$. Under this sufficient condition, the measure satisfies both (15) with $C = 1/\lambda$ and (12) with $C = \lambda$. Thus concentration inequalities (17) and (16) are true for this class of measures. Some examples of strong log-concave densities are γ the Gaussian density and

- h the logistic density $h(y) = e^y/(1+e^y)^2$;
- h the Gumbel density $h(y) = \exp(y - e^y)$;
- or $\mu(y) = zh(y)h(-y)$ with h being the Gumbel density and z a normalizing constant.

A review of log-concave and strong log-concave measures can be found in (Saumard and Wellner (2014)).

4.4 Measures with Bounded Support

A function f on $\mathbb{R}^n$ is called *separately convex* if it is convex in each coordinate, i.e., convex in the directions of coordinate axes. For instance function $f(y_1, y_2) = y_1 y_2$ is separately convex but it is not convex. Another example is $f(y) = y^T Q y$ with a symmetric matrix Q. This function is convex only if Q is positive semi-definite, but for being separately convex we only need to ensure that the diagonal element of Q are non-negative.

Let f be a separately convex Lipschitz function on $\mathbb{R}^n$ with Lipschitz constant $\|f\|_{Lip} \leq 1$. Then, for every product probability $\mathbb{P}$ on $[0,1]^n$, and every $t \geq 0$,

$$\mathbb{P}\left(f \geq \mathbb{E}(f) + t\right) \leq e^{-t^2/2}.$$

This result enables us to solve CCPs where uncertainty has bounded support and $g(x,\delta)$ is separately convex w.r.t δ.

5. CASE STUDY: LQG PROBLEM

We demonstrate the effectiveness of our approach on the optimization required in chance-constrained Linear Quadratic Gaussian (LQG) problem. Consider the linear time-invariant dynamical system

$$x_{k+1} = Ax_k + Bu_k + F\delta_k, \quad k = 0,1,2,\ldots \tag{19}$$

where $x_k \in \mathbb{R}^d$ is the state, u_k is the control input, and A, B, F are matrices with appropriate dimensions. $\{\delta_k\}_k$ is a sequence of independent standard Gaussian random vectors with $\delta_k \in \mathbb{R}^p$. Control input u_k takes values in a compact set $U \subset \mathbb{R}^m$. The objective of LQG problem is to optimize the cost function

$$\mathcal{J}(x_0, u) := \mathbb{E}\left[\sum_{k=1}^{L}(x_k^T Q_k x_k + u_k^T R_k u_k) + x_L^T Q_L x_L\right],$$

over the sequence of control inputs $u = [u_0^T, \ldots, u_{L-1}^T]^T$. At the same time, we would like to keep the sequence of states inside a safe region C with high probability:

$$\mathbb{P}\left([x_1^T, \ldots, x_L^T]^T \in C | x_0\right) \geq 1 - \epsilon. \tag{20}$$

After expanding the linear dynamics (19) and substituting them in both the objective function $\mathcal{J}(x_0, u)$ and constraint (20) we need to solve an optimization of the form

$$\min_u (u^T \bar{Q} u + 2x_0^T \bar{R} u + c) \tag{21}$$
$$u \in U^{mL},\ \mathbb{P}\left((\bar{A}x_0 + \bar{B}u + \bar{F}\delta) \in C | x_0\right) \geq 1 - \epsilon,$$

with $\delta = [\delta_0^T, \ldots, \delta_{L-1}^T]^T \in \mathbb{R}^{pL}$ and appropriately defined matrices $\bar{Q}, \bar{R}, \bar{A}, \bar{B}, \bar{F}, c$.

This problem is studied in (Hokayem et al. (2013)) addressing closed-loop policies. To keep the presentation focused we only consider open-loop policies with the understanding that our approach is applicable also to closed-loop policies. We select the following numerical values for the system dynamics

$$A = \begin{bmatrix} -1 & 10 \\ 0 & -2 \end{bmatrix},\ B = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ F = \sigma\begin{bmatrix} 1 \\ 1 \end{bmatrix},\ x_0 = \begin{bmatrix} 0.3 \\ 0.3 \end{bmatrix},$$

with $\sigma := 0.1$. We also consider the objective function with horizon $L = 5$ and matrices $Q_k = R_k = \mathbb{I}_2$ for all k. For the sake of comparison, we take C as in (Hokayem et al. (2013)) to be an ℓ_2-ball $C = \{\xi \in \mathbb{R}^{dL},\ \|\xi\|_2 \leq r\}$. In the following we apply our approach to this problem and compare it against (Hokayem et al. (2013)) and the scenario approach of (Grammatico et al. (2016)).

Our approach. We have Lipschitz continuous function

$$g(u,\delta) = \|\bar{A}x_0 + \bar{B}u + \bar{F}\delta\|_2 - r,$$

with Lipschitz constant $\rho := \|\bar{F}\|_2$. Due to the Gaussian measure $h(t) = \exp(-t^2/2\rho^2)$ thus $h^{-1}(\epsilon) = \rho\sqrt{2\ln 1/\epsilon}$. Variance of g is bounded by $\varrho + 2\sqrt{\varrho}\|\bar{A}x_0 + \bar{B}u\|_2$ with $\varrho := Tr(\bar{F}^T\bar{F})$ thus Assumption 5 holds.

$$\begin{cases} \min\limits_{u,\gamma_i}(u^T\bar{Q}u + 2x_0^T\bar{R}u + c) \\ \text{s.t.} \quad \|\bar{A}x_0 + \bar{B}u + \bar{F}\delta^i\|_2 \le \gamma_i, \quad i = 1,2,\dots,N \\ \qquad \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\gamma_i + \gamma + \rho\sqrt{2\ln 1/\epsilon} \le r. \end{cases}$$

The safe set is considered to be the 10-dim ℓ_2-ball with radius $r = 64$ and threshold 0.95 ($\epsilon = 0.05$). Note that the system without input goes out of this ball in expectation, so it is not a large ball for these dynamics. We also consider the input space $U = [-5, 5]$. By taking input that minimizes $\|\bar{A}x_0 + \bar{B}u\|_2$, the constraint is feasible with $\beta_0 = 2.32$. We take $\gamma = 1.16$ and number of samples N according to Theorem 6 for confidence $1 - \alpha = 0.99$.

Approach of (Hokayem et al. (2013)) relies on inequalities that hold only for Gaussian measure and are dimension dependent. Chance constraint of (1) in conservatively replaced by

$$\|\bar{A}x_0 + \bar{B}u\|_2 + \|\bar{F}\|_2\sqrt{\frac{Lp}{1-\eta}} \le r, \quad \eta := 2\sqrt{\frac{\ln 1/\epsilon}{Lp}},$$

where η has to be inside the open interval $(0, 1)$. As it is also mentioned in (Hokayem et al. (2013)), this puts a lower bound on the probability threshold ϵ that can be achieved. In this case the constraint is infeasible for all values of threshold $\epsilon \in (0, 0.7)$, which makes the approach impractical.

Scenario approach of (Grammatico et al. (2016)). Since the constraint of this case study can be transformed into a convex constraint and the objective function is already convex, the approach of (Grammatico et al. (2016)) is applicable, which gives the following optimization

$$\begin{cases} \min\limits_{u}(u^T\bar{Q}u + 2x_0^T\bar{R}u + c) \\ \text{s.t.} \quad \|\bar{A}x_0 + \bar{B}u + \bar{F}\delta^i\|_2^2 \le r^2, \quad i = 1,2,\dots,N. \end{cases}$$

Number of samples $N = 345$ is selected according to (Grammatico et al., 2016, Eqn. (5)) that depends on the required confidence α, probability threshold ϵ, and dimension of decision variables.

Comparison. Approach of (Hokayem et al. (2013)) is infeasible for values of threshold $\epsilon \in (0, 0.7)$. We run our scenario program and that of (Grammatico et al. (2016)) 500 times. Our SP is feasible in all runs and the optimal value has mean 364.4 and standard deviation 9×10^{-10}. SP of (Grammatico et al. (2016)) is infeasible in 25 runs (5% of the cases) and the optimal values in feasible runs have mean 368.6 and standard deviation 23.7. Note that here we have not used sampling and discarding (Campi and Garatti (2011)), which will improve the optimal values. The main weakness will be getting infeasible SP. This results in violation of *recursive feasibility* (Morari et al. (2014)), which is a standing assumption naturally made when applying scenario optimization in model predictive control framework. Note that by increasing number of samples, probability of getting an infeasible SP will decrease in our approach, while this probability will increase in SP of (Grammatico et al. (2016)) due to adding more constraints.

6. CONCLUSION

In this work we proposed a scenario program for solving chance-constrained optimizations. Our approach does not require convexity of the objective or constraint function but relies on knowing that uncertainty has concentration of measure property. Instead of satisfying constraints for all observed samples of uncertainty, we allow violation of constraints but we require that in average the value of constraints be away from their boundaries. Concentration of measure enables us to specify how much it should be away in order to guarantee having a feasible solution for the original optimization with certain confidence. We benchmarked our technique against approaches from literature on LQG control problem.

REFERENCES

Alon, N. and Spencer, J. (2016). *The Probabilistic Method*. Wiley, 4th edition.

Barvinok, A. (1997). Measure concentration in optimization. *Mathematical Programming*, 79(1), 33–53.

Bobkov, S. and Ledoux, M. (1997). Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probability Theory and Related Fields*, 107, 383–400.

Calafiore, G.C. (2010). Random convex programs. *SIAM Journal on Optimization*, 20(6), 3427–3464.

Campi, M.C. and Garatti, S. (2011). A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality. *Journal of Optimization Theory and Applications*, 148(2), 257–280.

Campi, M.C. and Garatti, S. (2008). The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3), 1211–1230.

Dubhashi, D. and Panconesi, A. (2009). *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, New York, NY, USA.

Grammatico, S., Zhang, X., Margellos, K., Goulart, P., and Lygeros, J. (2016). A scenario approach for non-convex control design. *IEEE Transactions on Automatic Control*, 61(2), 334–345.

Hokayem, P., Chatterjee, D., and Lygeros, J. (2013). Chance-constrained LQG with bounded control policies. In *52nd IEEE Conference on Decision and Control*, 2471–2476.

Ledoux, M. (2005). *The Concentration of Measure Phenomenon*. Mathematical surveys and monographs. American Mathematical Society.

Ledoux, M. (1999). Concentration of measure and logarithmic Sobolev inequalities. *Sminaire de probabilits de Strasbourg*, 33, 120–216.

Morari, M., Fagiano, L., Schildbach, G., and Frei, C. (2014). The scenario approach for stochastic model predictive control with bounds on closed-loop constraint violations. *Automatica*, 50(12), 3009–3018.

Naor, A. (2008). Lecture notes on concentration of measure.

Prékopa, A. (1995). *Stochastic Programming*. Mathematics and Its Applications. Springer Netherlands.

Royden, H. (1988). *Real Analysis*. Mathematics and statistics. Macmillan.

Saumard, A. and Wellner, J.A. (2014). Log-concavity and strong log-concavity: A review. *Statist. Surv.*, 8, 45–114.

Talagrand, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, 81(1), 73–205.