

RESEARCH

Open Access



# A smooth approximation approach for optimization with probabilistic constraints based on sigmoid function

Yong H. Ren<sup>1\*</sup> , Ying Xiong<sup>1</sup>, Yu H. Yan<sup>1</sup> and Jian Gu<sup>2\*</sup>

\*Correspondence:  
ryhong@163.com;  
gujian@dloou.edu.cn

<sup>1</sup>School of Mathematics, Liaoning Normal University, Dalian, P.R. China

<sup>2</sup>School of Information Engineering, Dalian Ocean University, Dalian, P.R. China

## Abstract

Many practical problems, such as computer science, communications network, product design, system control, statistics and finance, etc., can be formulated as a probabilistic constrained optimization problem (PCOP) which is challenging to solve since it is usually nonconvex and nonsmooth. Effective methods for the probabilistic constrained optimization problem mostly focus on approximation techniques, such as convex approximation, D.C. (difference of two convex functions) approximation, and so on. This paper aims at studying a smooth approximation approach. A smooth approximation to the probabilistic constraint function based on a sigmoid function is analyzed. Equivalence of PCOP and the corresponding approximation problem are shown under some appropriate assumptions. Sequential convex approximation (SCA) algorithm is implemented to solve the smooth approximation problem. Numerical results suggest that the smooth approximation approach proposed is effective for optimization problems with probabilistic constraints.

**Keywords:** Probabilistic constrained optimization problem; Sigmoid function; Smooth approximation approach; Sequential convex approximation

## 1 Introduction

It is well known that the probabilistic constrained optimization has been applied extensively to practical problems, such as computer science, communications network, product design, system control, statistics and finance, etc. For instance, a manufacturer needs to maximize its profit while satisfying demands from customers to make decisions on production and inventory. However, there is often uncertainty in the demands of the customers. To handle the uncertainty in the constraints, a natural approach is to require that all constraints be satisfied with a given high probability e.g. 90%. The resulted optimization problem is called a probabilistic constrained optimization problem (PCOP).

Consider the probabilistic constrained optimization problem as follows:

$$\begin{aligned} & \min_{x \in X} h(x) \\ & \text{s.t. } \Pr\{c_1(x, \xi) \leq 0, c_2(x, \xi) \leq 0, \dots, c_m(x, \xi) \leq 0\} \geq 1 - \alpha, \end{aligned} \tag{PCOP}$$

© The Author(s) 2022. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

where  $x$  is a  $d$ -dimensional vector of decision variables,  $\xi$  is a  $k$ -dimensional vector of uncertain parameters, and support of  $\xi$ , denoted as  $\Xi$ , is a closed subset of  $\mathfrak{R}^k$ ,  $X$  is a subset of  $\mathfrak{R}^d$ ,  $h : \mathfrak{R}^d \rightarrow \mathfrak{R}$  and  $c_i : \mathfrak{R}^d \times \Xi \rightarrow \mathfrak{R}$ ,  $i = 1, \dots, m$ , are real-valued functions, and functions  $h$  and  $c_i$ ,  $i = 1, \dots, m$ , are convex and continuously differentiable in  $x$  for every  $\xi \in \Xi$ .

Let  $c(x, \xi) = \max\{c_1(x, \xi), \dots, c_m(x, \xi)\}$ . Note that  $c(x, \xi)$  is a convex function of  $x$  since  $c_i(x, \xi)$ ,  $i = 1, \dots, m$ , are all convex in  $x$ . Let

$$p(x) = 1 - \Pr\{c_1(x, \xi) \leq 0, c_2(x, \xi) \leq 0, \dots, c_m(x, \xi) \leq 0\} = \Pr\{c(x, \xi) > 0\}.$$

Then  $p(x)$  is the probability that at least one constraint is violated. Furthermore, the probabilistic constrained optimization problem (PCOP) is reformulated as follows:

$$\begin{aligned} \min_{x \in X} \quad & h(x) \\ \text{s.t.} \quad & p(x) \leq \alpha. \end{aligned}$$

The literature on PCOPs can be dated back to Charnes et al. [1], who first considered a single probabilistic constrained optimization problem (SPCOP), and Miller and Wagner [2], who first considered a joint probabilistic constrained optimization problem (JPCOP). Since then, probabilistic constrained optimization has been studied extensively. PCOPs are generally challenging to solve. There are three major difficulties in solving a PCOP. First, the probabilistic constraint does not necessarily preserve the convexity. The set defined by the probabilistic constraint may not be convex even if all  $c_i$  are convex. Second, it is difficult to evaluate the probabilistic constraint function, that is to say, there is no closed-form expression of the probabilistic constraint generally. Third, the probabilistic constraint does not necessarily preserve the smoothness.

Numerical methods for PCOP have attracted great attention in the field of optimization (see [3–7]).

Convex conservative approximations have been proposed e.g. the quadratic approximation [8], the Bernstein approximation [9], and the conditional value-at-risk (CVaR) approximation [10]. These methods seek to find a convex subset of the (possible nonconvex) feasible set and find the optimal solution in the subset. To solve JPCOPs, probabilistic inequalities (e.g. Bonferroni’s inequality) have to be used to break a joint probabilistic constraint into multiple single probabilistic constraints, which often makes approximations even more conservative.

Monte Carlo simulation is often used to evaluate  $p(x)$  when its closed-form is not available. Luedtke and Ahmed [11] studied the sample average approximation of the PCOPs. Moreover, the scenario approach is introduced by Calafiore and Campi [12, 13] and De Farias and Van Roy [14]), which solves the following problem:

$$\begin{aligned} \min_{x \in X} \quad & h(x) \\ \text{s.t.} \quad & c_i(x, \xi_l) \leq 0, \quad i = 1, \dots, m, l = 1, \dots, n, \end{aligned}$$

where  $\xi_1, \xi_2, \dots, \xi_n$  are independent observations of  $\xi$  which are often generated from a Monte Carlo simulation. The critical issue is how to determine the sample size  $n$  to ensure the probability requirement. The new problem under the scenario approach is easier

to solve because  $c_i$  is convex, smooth, and easy to evaluate. Yang et al. [15] presented a scenario-based method to solve the chance constrained optimization for the nonconvex program.

Hong et al. [16] proposed an  $\varepsilon$ -approximation approach, which reformulated a PCOP into a D.C. (difference of convex) program, which is solved by an  $\varepsilon$ -approximation together with a sequence convex approximation (SCA) algorithm. Under some technical conditions, the solutions of the sequences of approximations converge to the set of Karush–Kuhn–Tucker (KKT) points of the PCOP. Based on the  $\varepsilon$ -approximation approach, Hu et al. [17] proposed a smooth Monte Carlo (SMC) approach. Shan et al. [18] proposed a class of smoothing functions to approximate the joint probabilistic constraint function, based on which smooth optimization problems are constructed to approximate PCOP. They show that the solutions of a sequence of smoothing approximations converge to a KKT point of PCOP under a certain asymptotic regime.

Owing to the nonsmoothness of the probabilistic constraint, we establish a smooth approximation technique based on sigmoid function, which has many desired properties. Moreover, the sequential convex approximation (SCA) algorithm is implemented to solve the smooth approximate problem. In each iteration of the SCA algorithm, a gradient-based Monte Carlo method is applied to solve a convex stochastic program. We prove that, under some assumptions, a KKT point of PCOP can be obtained by solving the sequence of convex programs.

The rest of this paper is organized as follows. In Sect. 2, we provide the smooth technique based on sigmoid function to approximate problem (PCOP) and show the desired convergence properties. Sequential convex approximation algorithm is discussed to solve the smoothing approximation problem. The numerical results are reported in Sect. 3, and some conclusions are presented in Sect. 4.

## 2 Materials and methods

### 2.1 Sigmoid approximation to PCOP

In this section, we use a sigmoid function to approximate a probabilistic constraint function, and obtain an approximation problem based on the sigmoid function.

#### 2.1.1 Sigmoid function

Probabilistic constrained optimization problem (PCOP) can be reformulated as follows:

$$\begin{aligned} & \lim_{x \in X} h(x) \\ \text{s.t. } & p(x) \leq \alpha, \end{aligned} \tag{P}$$

where  $p(x) = \Pr\{c(x, \xi) > 0\} = E[\mathbf{1}_{(0,+\infty)}(c(x, \xi))]$  and

$$\mathbf{1}_{(0,+\infty)}(z) := \begin{cases} 1 & \text{if } z \in (0, +\infty), \\ 0 & \text{if } z \notin (0, +\infty). \end{cases}$$

Consider the sigmoid function

$$\psi(z, \mu) = \frac{1}{1 + e^{-\mu^{-1}z}},$$

where  $\mu > 0$  is a parameter and  $\psi : \mathfrak{R}^2 \rightarrow \mathfrak{R}$  is a real-valued function. It is obvious that  $\psi(z, \mu) \in (0, 1)$  for any  $z \in \mathfrak{R}$ .

**Proposition 2.1** *For  $\mu > 0$ , the sigmoid function has the following properties.*

- (i) For any  $z \in \mathfrak{R} \setminus \{0\}$ ,  $\lim_{\mu \searrow 0} \psi(z, \mu) = \mathbf{1}_{(0, +\infty)}(z)$ .
- (ii)  $\psi(z, \mu)$  is nondecreasing in  $z$ .
- (iii)  $\psi(z, \mu)$  is monotone increasing in  $\mu$  when  $z < 0$ ,  $\psi(z, \mu)$  is monotone decreasing in  $\mu$  when  $z > 0$ .
- (iv)  $\psi(z, \mu)$  is smooth with respect to  $z$ .

*Proof* (i) When  $z > 0$ , we have

$$\lim_{\mu \searrow 0} \psi(z, \mu) = \lim_{\mu \searrow 0} \frac{1}{1 + e^{-\mu^{-1}z}} = 1.$$

When  $z < 0$ , we have

$$\lim_{\mu \searrow 0} \psi(z, \mu) = \lim_{\mu \searrow 0} \frac{1}{1 + e^{-\mu^{-1}z}} = 0.$$

- (ii) Differentiating the function  $\psi(z, \mu)$  with respect to  $z$  yields

$$\frac{\partial}{\partial z} \psi(z, \mu) = \frac{e^{-\mu^{-1}z}}{\mu(1 + e^{-\mu^{-1}z})^2} > 0.$$

- (iii) Differentiating the function  $\psi(z, \mu)$  with respect to  $\mu$ , we obtain

$$\frac{\partial}{\partial \mu} \psi(z, \mu) = \frac{-ze^{-\mu^{-1}z}}{(\mu + \mu e^{-\mu^{-1}z})^2}.$$

Obviously,  $\frac{\partial}{\partial \mu} \psi(z, \mu) < 0$ , when  $z > 0$ ;  $\frac{\partial}{\partial \mu} \psi(z, \mu) > 0$ , when  $z < 0$ .

- (iv) The result is obvious. □

We can observe that the sigmoid function  $\psi(z, \mu)$  approximates the characteristic function  $\mathbf{1}_{(0, +\infty)}(z)$  when parameter  $\mu > 0$  is small enough (see Fig. 1).

### 2.1.2 Sigmoid approximation

Let

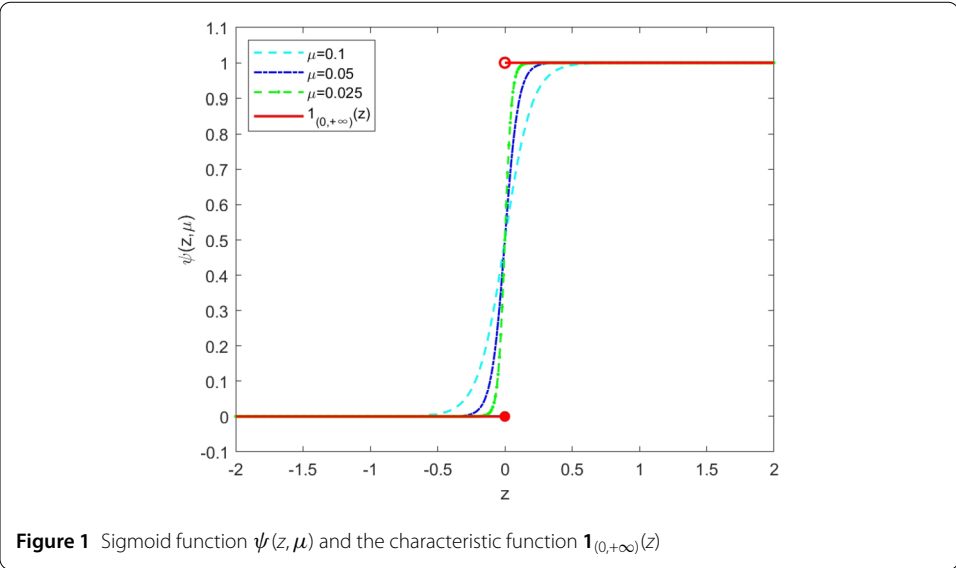
$$\tilde{p}(x, \mu) := E[\psi(z, \mu)] = E[\psi(c(x, \xi), \mu)],$$

then we have  $\tilde{p}(x) = \lim_{\mu \searrow 0} \tilde{p}(x, \mu)$ . The approximation problem is built as follows:

$$\begin{aligned} \min_{x \in X} \quad & h(x) \\ \text{s.t.} \quad & \tilde{p}(x) \leq \alpha. \end{aligned} \tag{\tilde{P}}$$

Let  $\tilde{\Omega}$  and  $\Omega$  denote the feasible sets of problem  $(\tilde{P})$  and problem  $(P)$ , respectively.

Now we make the following assumptions.



**Figure 1** Sigmoid function  $\psi(z, \mu)$  and the characteristic function  $\mathbf{1}_{(0,+\infty)}(z)$

**Assumption 1** The set  $X$  is a compact and convex subset of  $\mathfrak{N}^d$ , and the support of  $\xi$ , denoted as  $\Xi$ , is a closed subset of  $\mathfrak{N}^k$ . For any  $\xi \in \Xi$ ,  $h(x)$  and  $c_i(x, \xi)$ ,  $i = 1, \dots, m$ , are continuously differentiable and convex in  $x$  for any  $x \in \mathcal{O}$ , where  $\mathcal{O}$  is a bounded open set such that  $X \subset \mathcal{O}$ .

**Assumption 2** Function  $c(x, \cdot)$  is measurable for every  $x \in \mathfrak{N}^d$  and  $c(\cdot, \xi)$  is continuous for a.e.  $\xi \in \Xi$ .

**Assumption 3** For every  $\bar{x} \in X$ , the set  $\{\xi \in \Xi : c(\bar{x}, \xi) = 0\}$  has P-measure zero i.e.  $c(\bar{x}, \xi) \neq 0$ , w.p.1.

**Assumption 4** For any  $x \in \mathcal{O}$ ,  $\Pr\{c_i(x, \xi) = c_j(x, \xi)\} = 0$ ,  $i \neq j$  i.e.  $c(x, \xi)$  is differentiable with respect to  $x$  w.p.1.

The following theorem suggests the equivalence between problem  $(\tilde{P})$  and problem  $(P)$ .

**Theorem 2.2** Assume that Assumptions 1–4 hold. Then, for any  $x \in X$  and  $\mu > 0$ ,

$$\lim_{\mu \searrow 0} \tilde{p}(x, \mu) = p(x).$$

*Proof* Let  $z = c(x, \xi)$ , it follows from Proposition 2.1 and the Lebesgue control convergence theorem that

$$\begin{aligned} \tilde{p}(x) &= \lim_{\mu \searrow 0} \tilde{p}(x, \mu) = \lim_{\mu \searrow 0} E[\psi(c(x, \xi), \mu)] \\ &= \lim_{\mu \searrow 0} \int_{\Xi} \psi(c(x, \xi), \mu) dp(\xi) \\ &= \int_{\Xi} \lim_{\mu \searrow 0} \psi(c(x, \xi), \mu) dp(\xi) \end{aligned}$$

$$\begin{aligned}
 &= \int_{\Xi} \mathbf{1}_{(0,+\infty)}(c(x, \xi)) dp(\xi) \\
 &= E[\mathbf{1}_{(0,+\infty)}(c(x, \xi))] \\
 &= p(x). \tag*{$\square$}
 \end{aligned}$$

An important question for sigmoid approximation is how to choose parameter  $\mu$ . Consider the parameterized approximation of problem  $(\tilde{P})$  as follows:

$$\begin{aligned}
 &\min_{x \in X} h(x) \\
 &\text{s.t. } \tilde{p}(x, \mu) \leq \alpha. \tag*{$(\tilde{P}_\mu)$}
 \end{aligned}$$

Denote  $\Omega_\mu := \{x \in X : \tilde{p}(x, \mu) \leq \alpha\}$  as the feasible set of problem  $(\tilde{P}_\mu)$ . Let  $S_\mu$  and  $v_\mu$  be the set of optimal solutions and the optimal value of problem  $(\tilde{P}_\mu)$ , respectively. The following theorem describes the properties of problem  $(\tilde{P}_\mu)$ .

**Assumption 5** Let  $\Omega = \{x \in X : p(x) \leq \alpha\}$  and  $\Omega^I = \{x \in X : p(x) < \alpha\}$ . Then  $\Omega = \text{cl}\Omega^I$ .

**Theorem 2.3** Suppose that Assumptions 1 to 5 are satisfied. Then  $\lim_{\mu \searrow 0} \Omega_\mu = \Omega$ .

*Proof* It suffices to prove that the inclusions

$$\limsup_{\mu \searrow 0} \Omega_\mu \subset \Omega \subset \liminf_{\mu \searrow 0} \Omega_\mu$$

are valid.

Given any  $\bar{x} \in \limsup_{\mu \searrow 0} \Omega_\mu$ , there exist  $\mu_k \rightarrow 0$  and  $x^k \in \Omega_{\mu_k}$  such that  $x^k \rightarrow \bar{x}$ , which means that the inequality

$$\tilde{p}(x^k, \mu_k) \leq \alpha$$

holds. Taking  $k \rightarrow +\infty$ , we have

$$\lim_{\mu_k \rightarrow 0, x^k \rightarrow \bar{x}} \tilde{p}(x^k, \mu_k) \leq \alpha.$$

Owing to the continuity of function  $\tilde{p}(x, \mu)$  in  $x$ , it follows that  $p(\bar{x}) \leq \alpha$ , which means  $\bar{x} \in \Omega$ .

Consequently, inclusion  $\limsup_{\mu \searrow 0} \Omega_\mu \subset \Omega$  holds.

For any  $\bar{x} \in \Omega^I$ , we have  $p(\bar{x}) < \alpha$ . It follows from Theorem 2.2 that

$$\lim_{\mu \searrow 0} \tilde{p}(\bar{x}, \mu) < \alpha.$$

Consider  $\epsilon = \frac{1}{2}(\alpha - p(\bar{x})) > 0$ , given any  $\mu_k \searrow 0$ , there exists  $N \in \mathcal{N}_\infty$  for  $k \in N$ , one has

$$\tilde{p}(\bar{x}, \mu_k) \leq p(\bar{x}) + \epsilon = \frac{1}{2}(\alpha + p(\bar{x})) < \alpha$$

i.e.  $\bar{x} \in \text{int}(\Omega_{\mu_k})$ . So there exist  $N \supset N_1 \in \mathcal{N}_\infty$  and  $x^k \in \Omega_{\mu_k}$ ,  $k \in N_1$  such that  $x^k \xrightarrow[N_1]{} \bar{x}$ , which implies that the inclusion  $\Omega^I \subset \liminf_{\mu \searrow 0} \Omega_\mu$  holds. Since the set  $\liminf_{\mu \searrow 0} \Omega_\mu$  is closed, it follows by Assumption 5 that

$$\Omega = \text{cl}\Omega^I \subset \liminf_{\mu \searrow 0} \Omega_\mu.$$

As a result,  $\lim_{\mu \searrow 0} \Omega_\mu = \Omega$  holds. □

For sets  $A, B \subset \mathfrak{R}^d$ , let  $d(x, A) = \inf_{x' \in A} \|x - x'\|$  denote the distance from  $x \in \mathfrak{R}^d$  to  $A$ , and  $D(A, B) = \sup_{x \in A} d(x, B)$  denote the deviation of the set  $A$  from the set  $B$ , see [19]. Note that it is a measure of the distance between two sets. Let  $S$  and  $v$  denote the set of optimal solutions and optimal value of problem (P), respectively. It is easy to prove the following theorem.

**Theorem 2.4** *Suppose that Assumptions 1 to 5 are satisfied. Then, for any  $x \in X$ ,  $\lim_{\mu \searrow 0} v_\mu = v$  and  $\lim_{\mu \searrow 0} D(S_\mu, S) = 0$ .*

*Proof* Define

$$\bar{h}(x) = h(x) + I_\Omega(x), \quad \bar{h}_\mu(x) = h(x) + I_{\Omega_\mu}(x),$$

where

$$I_A(x) = \begin{cases} 0 & \text{if } x \in A, \\ +\infty & \text{if } x \notin A. \end{cases}$$

In view of Theorem 2.3 and [20, Proposition 7.4(f)], we have that  $I_{\Omega_\mu}(\cdot)$  epi-converges to  $I_\Omega(\cdot)$  as  $\mu \searrow 0$ . Since  $h(\cdot)$  is continuous, it follows that  $\bar{h}_\mu(\cdot)$  epi-converges to  $\bar{h}(\cdot)$  as  $\mu \searrow 0$ . Note that  $\Omega_\mu$  and  $\Omega$  are compact, we have that  $\bar{h}_\mu(\cdot)$  and  $\bar{h}(\cdot)$  are lower semi-continuous and proper. Then, by [20, Theorem 7.33], we have  $v_\mu \rightarrow v$  and

$$\limsup_{\mu \searrow 0} S_\mu \subset S \text{ (as } c(x, \xi) < 0 \text{)} \quad \text{or} \quad \liminf_{\mu \searrow 0} S_\mu \supset S \text{ (as } c(x, \xi) > 0 \text{)}.$$

Since  $S_\mu$  and  $S$  are subsets of the compact set  $X$ , they are uniformly compact. According to [20, Example 4.13], we have that  $\lim_{\mu \searrow 0} D(S_\mu, S) = 0$ . This concludes the proof of the theorem. □

Theorem 2.4 shows that the optimal solutions of problem  $(\tilde{P}_\mu)$  provide good approximation to those of problem (P) when  $\mu \searrow 0$ . However, problem  $(\tilde{P}_\mu)$  is generally a nonconvex problem. Therefore, finding an optimal solution to the problem may be difficult. We often only find KKT points of problem  $(\tilde{P}_\mu)$ . In the rest of this subsection, we analyze the convergence of the KKT points of problem  $(\tilde{P}_\mu)$  to those of problem (P) as  $\mu \searrow 0$ .

Let  $\Lambda$  and  $\Lambda_\mu$  denote the sets of KKT pairs of problem (P) and  $(\tilde{P}_\mu)$ , respectively, as follows:

$$\Lambda = \left\{ \begin{array}{l} (x, \lambda) \in \Omega \times \mathfrak{R}_+ : 0 \in \nabla h(x) + \lambda \partial p(x) + N_X(x) \\ \lambda [p(x) - \alpha] = 0 \end{array} \right\},$$

$$\Lambda_\mu = \left\{ (x, \lambda) \in \Omega_\mu \times \mathfrak{R}_+ : \begin{aligned} &0 \in \nabla h(x) + \lambda \nabla_x \tilde{p}(x, \mu) + N_X(x) \\ &\lambda [\tilde{p}(x, \mu) - \alpha] = 0 \end{aligned} \right\},$$

where  $\partial p(x)$  represents the sub-differential of function  $p$  at  $x$  and  $N_X(x)$  denotes the normal cone to  $X$  at  $x$ .

**Theorem 2.5** *Suppose that Assumptions 1 to 5 are satisfied. Then  $\limsup_{\mu \searrow 0} \Lambda_\mu \subset \Lambda$ .*

*Proof* For any  $(x, \lambda) \in \limsup_{\mu \searrow 0} \Lambda_\mu$ , there exist  $\mu_k \searrow 0$  and  $(x^k, \lambda^k) \in \Lambda_{\mu_k}$  such that  $(x^k, \lambda^k) \rightarrow (x, \lambda)$ . The inclusion  $(x^k, \lambda^k) \in \Lambda_{\mu_k}$  means

$$\begin{cases} 0 \in \nabla h(x^k) + \lambda^k \nabla_x \tilde{p}(x^k, \mu_k) + N_X(x^k), \\ \lambda^k [\tilde{p}(x^k, \mu_k) - \alpha] = 0, \end{cases}$$

which are proved from the following four parts.

Part I, by Assumption 1, we have  $h(x)$  is continuously differentiable. Then  $\nabla h(x)$  is continuously differentiable. So  $\nabla h(x^k) \rightarrow \nabla h(x)$  as  $x^k \rightarrow x$  holds.

Part II, by Theorem 2.2, we know that  $\lim_{\mu \searrow 0} \tilde{p}(x, \mu) = p(x)$ . Then  $\tilde{p}(x^k, \mu_k) \rightarrow p(x)$  as  $x^k \rightarrow x, \mu_k \searrow 0$ .

Part III, by Assumption 4, we obtain by calculation that, for  $\mu > 0$ ,

$$\nabla_x \tilde{p}(x, \mu) = \nabla_x E[\psi(c(x, \xi), \mu)] = E[\nabla_x \psi(c(x, \xi), \mu)]$$

and

$$\lim_{\mu \searrow 0} \nabla_x \psi(c(x, \xi), \mu) = \lim_{\mu \searrow 0} \frac{\mu^{-1} e^{-\mu^{-1} c(x, \xi)}}{(1 + e^{-\mu^{-1} c(x, \xi)})^2} \nabla_x c(x, \xi) = \begin{cases} \mathbf{0} & c(x, \xi) \neq 0, \\ +\infty & c(x, \xi) = 0. \end{cases}$$

Thus, it follows from Assumption 3 that

$$\lim_{\mu \searrow 0} \nabla_x \tilde{p}(x, \mu) = \mathbf{0},$$

Consequently,  $\lim_{\mu \searrow 0} \nabla_x \tilde{p}(x, \mu) \in \partial p(x)$ . Since  $\nabla_x \tilde{p}(x, \mu)$  is continuous,

$$\lim_{x^k \rightarrow x, \mu_k \searrow 0} \nabla_x \tilde{p}(x^k, \mu_k) \in \partial p(x).$$

Part IV, by Theorem 2.3, we can obtain  $\lim_{\mu \searrow 0} \Omega_\mu = \Omega$ , since  $x^k \in \Omega_{\mu_k}$  and  $x^k \rightarrow x$ . Therefore  $x \in \Omega$ . By [20, Proposition 6.6], we have

$$\limsup_{x^k \rightarrow x} N_X(x^k) = N_X(x),$$

where  $x, x^k \in X$ . In conclusion,  $(x, \lambda) \in \Lambda$ , so  $\limsup_{\mu \searrow 0} \Lambda_\mu \subset \Lambda$  holds. This concludes the proof. □

Theorem 2.4 and 2.5 ensure that problem  $(\tilde{P}_\mu)$  can approximate problem (P) very well. Therefore, we can solve problem  $(\tilde{P}_\mu)$  instead of problem (P) provided that the smoothing parameter  $\mu$  is sufficiently close to 0.



We see that problem  $(\tilde{P}_\mu)$  is also a smooth approximation of the original PCOP. Similarly, as problem  $(\tilde{P}_\mu)$  may be nonconvex, its optimal solutions may not be guaranteed by the optimization procedures such as the SCA algorithm that will be introduced in Sect. 2.2. We observe that problem  $(\tilde{P}_\mu)$  directly approaches the original PCOP as  $\mu \searrow 0$ . However, the probability function  $\Pr\{c(x, \xi) > 0\}$  in the PCOP is in general nonsmooth, nonconvex, and also may not be locally Lipschitz continuous. Consequently, none of its gradient, subdifferential in convex context, and Clarke’s generalized gradient [21] are available, which makes the conventional KKT conditions for smooth optimization, the subdifferential conditions for nonsmooth convex optimization, and the generalized gradient conditions for locally Lipschitz continuous optimization not applicable to the PCOP. In this paper, we try to give a depiction of the possible optimality conditions for the PCOP and show the convergence of the stationary points for problem  $(\tilde{P}_\mu)$  in the proposed new context. In the rest of the paper we shall focus on the computational and implementation issues of the proposed smoothing approach.

### 2.2 Sequential convex approximation method

In this section, we applied sequential convex approximation (SCA) constructed in [16] to solve the approximation problem  $(\tilde{P}_\mu)$ .

Let

$$H_1(x, \xi, \mu) := \frac{1}{1 + e^{-\mu^{-1}c(x, \xi)}} + \frac{1}{8}\mu^{-2}c_+(x, \xi)^2$$

and

$$H_2(x, \xi, \mu) := \frac{1}{8}\mu^{-2}c_+(x, \xi)^2,$$

where  $c_+(x, \xi) = \max\{0, c(x, \xi)\}$ . Then, for any  $\mu > 0$ , both  $H_1(x, \xi, \mu)$  and  $H_2(x, \xi, \mu)$  are convex in  $x$ .

In fact, let  $z = c(x, \xi)$ , we can obtain that

$$\frac{\partial}{\partial z}H_2(z, \mu) = \frac{1}{4}\mu^{-2}z_+ \geq 0.$$

Thus  $H_2(z, \mu)$  is monotone increasing. It follows from [22, Proposition 2.1.7] that  $H_2(x, \xi, \mu)$  is convex in  $x$ . Similarly,  $\psi(x, \xi, \mu)$  is also convex in  $x$  by [22, Proposition 2.1.1],  $H_1(x, \xi, \mu)$  is also convex in  $x$ . And for any  $\mu > 0$ , both  $H_1(x, \xi, \mu)$  and  $H_2(x, \xi, \mu)$  are continuously differentiable in  $x$  for every  $\xi \in \Xi$ . Denote

$$g_1(x, \mu) = \Psi_1(x, \mu) - \alpha, \quad g_2(x, \mu) = \Psi_2(x, \mu),$$

where  $\Psi_1(x, \mu) = E[H_1(x, \xi, \mu)]$ ,  $\Psi_2(x, \mu) = E[H_2(x, \xi, \mu)]$ .

Problem  $(\tilde{P}_\mu)$  can be reformulated as the following DC program:

$$\begin{aligned} \min_{x \in X} \quad & h(x) \\ \text{s.t.} \quad & g(x, \mu) := g_1(x, \mu) - g_2(x, \mu) \leq 0. \end{aligned} \tag{P_\mu}$$

Let  $Z(\mu)$  be the feasible set of problem  $(P_\mu)$  and

$$Z(\mu)_y = \{x \in X : g_1(x, \mu) - [g_2(y, \mu) + \nabla_x g_2(y, \mu)^T(x - y)] \leq 0\}$$

for any  $y \in Z(\mu)$ . Note that

$$g_2(y, \mu) + \nabla_x g_2(y, \mu)^T(x - y)$$

is the first order Taylor expansion at any point  $y \in Z(\mu)$  to approximate  $g_2(x, \mu)$ . Since  $g_2(x, \mu)$  is convex in  $x$ , we have

$$g_2(x, \mu) \geq g_2(y, \mu) + \nabla_x g_2(y, \mu)^T(x - y), \quad x \in X,$$

which implies that

$$g_1(x, \mu) - g_2(x, \mu) \leq g_1(x, \mu) - [g_2(y, \mu) + \nabla_x g_2(y, \mu)^T(x - y)].$$

Then  $Z(\mu) \supset Z(\mu)_y$  for any  $y \in Z(\mu)$ . Moreover, since

$$g_1(x, \mu) - [g_2(y, \mu) + \nabla_x g_2(y, \mu)^T(x - y)]$$

is a smooth convex function of  $x$ , we have  $Z(\mu)_y$  is a convex subset of  $Z(\mu)$ .

For any  $y \in Z(\mu)$ , define problem as follows:

$$\begin{aligned} \min \quad & h(x) \\ \text{s.t.} \quad & x \in Z(\mu)_y. \end{aligned} \tag{CP(\mu, y)}$$

Then  $(CP(\mu, y))$  is a convex smooth approximation of problem  $(P_\mu)$ . It leads us to using the following algorithm to solve problem  $(P_\mu)$ .

**Algorithm SCA** ([16])

- step 0 Given an initial point  $x_0 \in Z(\mu)$  and set  $k = 0$ ;
- step 1 Stop if  $x^k$  satisfies the KKT condition of problem  $(P_\mu)$ ;
- step 2 Solve  $CP(\mu, x^k)$  to obtain its optimal solution  $x^{k+1}$ ;
- step 3 Set  $k = k + 1$  and return to step 1.

Algorithm SCA is easy to implement, since we only need to solve the convex optimization problem  $CP(\mu, x^k)$  in each iteration. It also has some desired properties, which we summarize in the following theorems without proofs. Note that we say Slater’s condition holds at  $y \in Z(\mu)$  if  $\text{int}Z(\mu)_y \neq \emptyset$ , where  $\text{int}A$  denotes the interior of a set  $A$ .

**Theorem 2.6** For problem  $(P_\mu)$ , let  $\{y_k\} \subset X$  be a sequence converging to  $\bar{y} \in Z(\mu)$  at which Slater’s condition holds. Then  $\lim_{k \rightarrow +\infty} Z(\mu)_{y_k} = Z(\mu)_{\bar{y}}$ .

**Theorem 2.7** Suppose that the Slater’s condition holds and  $\{x^k\}$  is a sequence of solutions generated by Algorithm SCA for problem  $(P_\mu)$  starting from  $x_0 \in Z(\mu)$ . Then,

- (i)  $\{x^k\} \subset Z(\mu)$  and  $\{h(x^k)\}$  is a convergent nonincreasing sequence;
- (ii) if  $x^{k+1} = x^k$ , then  $x^k$  is a KKT point of problem  $(P_\mu)$ ;
- (iii) suppose that  $\bar{x}$  is a cluster point of  $\{x^k\}$ , then  $\bar{x}$  is a KKT point of problem  $(P_\mu)$ .

Theorem 2.7 shows that we make improvement at each iteration and the sequence of objective values converges to a certain value, and if the algorithm terminates after a finite number of iterations, we actually reach a KKT point. It also ensures that all limit points of the sequence of solutions generated are KKT points. Therefore, it demonstrates that the algorithm has the desired convergence property. If problem  $(P_\mu)$  has only a KKT point that is better than the initial solution or only has a single KKT point, Algorithm SCA guarantees to converge to a global optimal solution.

### 3 Results and discussion

In this section we implement Algorithm SCA in MATLAB and use the inherent function *fmincon* to solve the convex sub-problem in each iteration. All the programs are run on a laptop with Intel(R) Core(TM) i5-8265U CPU @ 1.60 GHz 1.80 GHz and 8.00 GB of RAM.

*Example 3.1* Consider the following problem:

$$\begin{aligned} \min_{x \geq 0} \quad & - \sum_{j=1}^{10} x_j \\ \text{s.t.} \quad & \Pr \left\{ \sum_{j=1}^{10} \xi_{ij}^2 x_j^2 \leq 100, i = 1, \dots, 10 \right\} \geq 1 - \alpha, \end{aligned} \tag{1}$$

where  $x = (x_1, \dots, x_{10})^T \in \mathfrak{R}^{10}$ ,  $\xi_i = (\xi_{i1}, \dots, \xi_{i10})^T \in \mathfrak{R}^{10}$  for  $i = 1, \dots, 10$  are random variables. Assume that the random variables  $\xi_{ij}$ ,  $i = 1, \dots, 10, j = 1, \dots, 10$  are independent and identically distributed (i.i.d.) standard normal random variables.

It is easy to know that the optimal solution  $x^*$  of problem (1) is

$$x_1^* = x_2^* = \dots = x_{10}^* = \left[ \frac{100}{F_{\chi_{10}^2}^{-1}((1 - \alpha)^{\frac{1}{10}})} \right]^{\frac{1}{2}}, \tag{2}$$

where  $F_{\chi_{10}^2}^{-1}$  is the inverse chi-squared distribution function with 10 degrees of freedom.

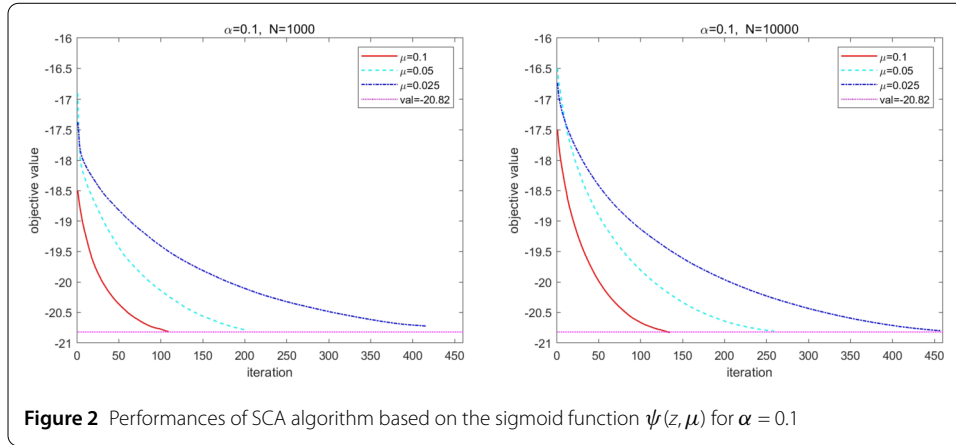
By calculating according to Eq. (2), we can obtain that the optimal solution and optimal value of problem (1) are  $(2.082, 2.082, \dots, 2.082)^T$  and  $-20.82$  respectively as  $\alpha = 0.1$ , and  $(1.995, 1.995, \dots, 1.995)^T$  and  $-19.95$  respectively as  $\alpha = 0.05$ .

Denote  $c_i(x, \xi) = \sum_{j=1}^{10} \xi_{ij}^2 x_j^2 - 100$ ,  $i = 1, \dots, 10$ , then problem (1) is a JPCOP as defined in problem (PCOP). For different  $\alpha$ , we run the Algorithm SCA a great number of replications for sample size  $N = 1000$  and  $N = 10,000$  respectively. The algorithm is stopped if  $\|x^{k+1} - x^k\|^2 \leq 10^{-6}$ . Numerical results are rounded to two decimal places.

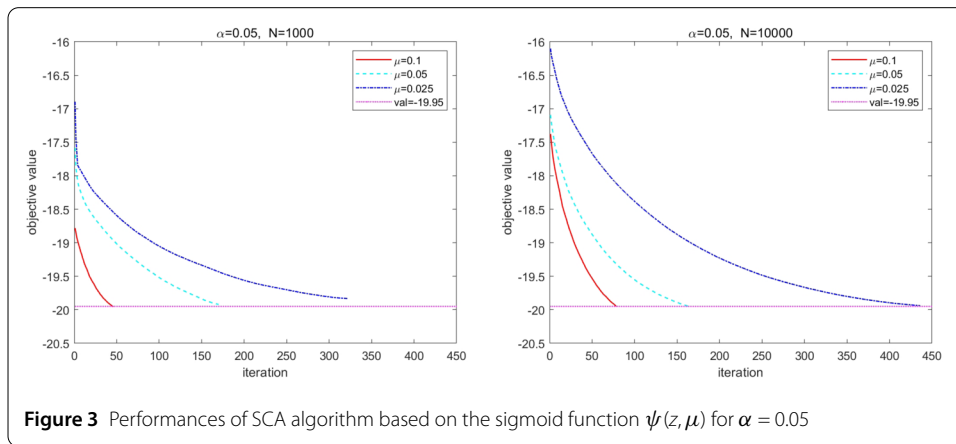
Table 1 reports typical performances for different scalar  $\mu = 0.1, 0.05, 0.025$  as  $\alpha = 0.1, 0.05$  respectively. It is obvious that the performances of the algorithm are very stable for parameter  $\mu$  small enough. It always converges to similar objective values. And the smaller the confidence  $\alpha$  the faster the convergence under the same conditions.

**Table 1** Numerical results for different  $\alpha$  and  $\mu$

$\alpha$	$\mu$	Number of iterations		Final residue		Objective value	
		$N = 1000$	$N = 10,000$	$N = 1000$	$N = 10,000$	$N = 1000$	$N = 10,000$
0.1	0.1	109	135	$1.95e-10$	$5.14e-09$	-20.81	-20.82
	0.05	201	259	$9.69e-07$	$8.67e-08$	-20.79	-20.81
	0.025	416	458	$9.99e-08$	$6.99e-08$	-20.72	-20.80
0.05	0.1	46	79	$8.87e-06$	$8.10e-06$	-19.95	-19.95
	0.05	170	166	$8.47e-07$	$8.22e-07$	-19.92	-19.95
	0.025	322	436	$8.39e-08$	$6.79e-08$	-19.83	-19.94



**Figure 2** Performances of SCA algorithm based on the sigmoid function  $\psi(z, \mu)$  for  $\alpha = 0.1$



**Figure 3** Performances of SCA algorithm based on the sigmoid function  $\psi(z, \mu)$  for  $\alpha = 0.05$

Figure 2 and Fig. 3 illustrate the objective values of all iterations for different parameters  $\mu$  as  $\alpha = 0.1$  and  $\alpha = 0.05$  respectively. The algorithm starts with the objective function dropping rapidly and then gradually slows down.

By comparison, it is found that when the sample size is sufficiently large, the approximation problem has desirable convergence. Furthermore, numerical results suggest that the smooth approximation approach based on a sigmoid function for solving optimization with probabilistic constraints is implementable.

#### 4 Conclusions

We used a sigmoid function to approximate the characteristic function and built the corresponding smooth approximation problem. Sequential convex approximation (SCA) al-

gorithm was introduced to solve the smooth approximation problem. Numerical results showed that the smooth approximation approach for optimization with probabilistic constraints based on the sigmoid function was implementable and guaranteed desired convergence properties under certain conditions. In the future, we will further explore a class of smooth approximation methods for solving probabilistic constrained optimization problems and compare different approximation algorithms numerically.

#### Acknowledgements

We would like to thank for the constructive feedback provided by the reviewers.

#### Funding

Our work is supported by National Natural Science Foundation of China under Grant(12171219, 11801054); Liaoning Province Department of Education Scientific Research General Project (LJ2019005).

#### Availability of data and materials

Not applicable.

#### Declarations

##### Competing interests

The authors declare that they have no competing interests.

##### Authors' contributions

YR performed the analysis with constructive discussions and wrote the manuscript. YX and YY performed the experiments. JG contributed significantly to data analysis. All authors have read and approved the manuscript.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 30 November 2021 Accepted: 22 March 2022 Published online: 08 April 2022

#### References

1. Charnes, A., Cooper, W.W., Symonds, H.: Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Manag. Sci.* **4**, 235–263 (1958)
2. Miller, L.B., Wagner, H.M.: Chance constrained programming with joint constraints. *Oper. Res.* **13**, 930–945 (1965)
3. Som, A., Some, K., Compaore, A., Some, B.: Exponential penalty function with MOMA-plus for the multiobjective optimization problems. *Appl. Anal. Optim.* **5**, 323–334 (2021)
4. Zhao, J., Bin, M., Liu, Z.: A class of nonlinear differential optimization problems in finite dimensional spaces. *Appl. Anal. Optim.* **5**, 145–156 (2021)
5. Zhao, X., Kobis, M.A., Yao, Y., Yao, J.C.: A projected subgradient method for nondifferentiable quasiconvex multiobjective optimization problems. *J. Optim. Theory Appl.* **190**, 82–107 (2021)
6. Zhang, J., Xu, H.F., Zhang, L.W.: Quantitative stability analysis of stochastic quasi-variational inequality problems and applications. *Math. Program.* **165**, 433–470 (2017)
7. Zhang, J., Xu, H.F., Zhang, L.W.: Quantitative stability analysis for distributionally robust optimization with moment constraints. *SIAM J. Optim.* **26**, 1855–1882 (2016)
8. Ben-Tal, A., Nemirovski, A.: Robust solutions of linear programming problems contaminated with uncertain data. *Math. Program.* **88**, 411–424 (2000)
9. Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. *SIAM J. Optim.* **17**, 347–375 (2006)
10. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. *J. Risk* **2**, 21–41 (2000)
11. Luedtke, J., Ahmed, S.: A sample approximation approach for optimization with probabilistic constraints. *SIAM J. Optim.* **19**, 674–699 (2008)
12. Calafiore, G., Campi, M.C.: Uncertain convex programs: randomized solutions and confidence levels. *Math. Program.* **102**, 25–46 (2005)
13. Calafiore, G., Campi, M.C.: The scenario approach to robust control design. *IEEE Trans. Autom. Control* **51**, 742–753 (2006)
14. Farias, D.P.D., Roy, B.V.: On constraint sampling in the linear programming approach to approximate dynamic programming. *Math. Oper. Res.* **29**, 462–478 (2004)
15. Yang, Y., Sutanto, C.: Chance-constrained optimization for nonconvex programs using scenario-based methods. *ISA Trans.* **90**, 157–168 (2019)
16. Hong, L.J., Yang, Y., Zhang, L.W.: Sequential convex approximations to joint chance constrained programs: a Monte Carlo approach. *Oper. Res.* **59**, 617–630 (2011)
17. Hu, Z., Hong, L.J., Zhang, L.W.: A smooth Monte Carlo approach to joint chance-constrained programs. *IIE Trans.* **45**, 716–735 (2013)
18. Shan, F., Zhang, L.W., Xiao, X.T.: A smoothing function approach to joint chance-constrained programs. *J. Optim. Theory Appl.* **163**, 181–199 (2014)

19. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia (2009)
20. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer, Berlin (1998)
21. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
22. Hiriart-Urruty, J.B., Lemaréchal, C.: Fundamentals of Convex Analysis. Springer, Berlin (2001)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---