Approximating inequality systems within probability functions: studying implications for problems and consistency of first-order information

Wim van Ackooij[1*], Diego Morales-Poblete[2], Pedro Pérez-Aros[2,3], David Villacís[4]

[1]OSIRIS, EDF Lab Paris-Saclay, 7 Boulevard Gaspard Monge, Palaiseau, 91120, France.

[2]Departamento de Ingeniería Matemática, Universidad de Chile, Beauchef 850, Santiago, Chile.

[3]Centro de Modelamiento Matemático, Universidad de Chile, Beauchef 850, Santiago, Chile.

[4]Departamento de Métodos Cuantitativos, Universidad Loyola Andalucía, Av. de las Universidades 2, Seville, 41704, Spain.

*Corresponding author(s). E-mail(s): wim.van-ackooij@edf.fr;

Contributing authors: dhmorales@dim.uchile.cl; pperez@dim.uchile.cl; davillacis@uloyola.es;

Abstract

In this work, we are concerned with the study of optimization problems featuring so-called probability or chance constraints. Probability constraints measure the level of satisfaction of an underlying random inequality system and ensure that this level is high enough. Such an underlying inequality system could be expressed by an abstractly known or perhaps costly to evaluate function. While perhaps then it is possible to theoretically investigate first-order properties of the probability function or perhaps even make numerical use of them, the computational cost could be important. We suggest an inner approximation framework, thus providing a range of approximate probability functions. We establish that this sequence converges hypographically and continuously to the nominal probability function. Furthermore, we examine the convergence of (sub-)gradients, provide a suitable formula for the (sub-)gradients of the approximate probability functions, and prove that under mild assumptions, this sequence of subgradients converges to a subgradient of the original probability function. We also examine the consistency of optimization problems wherein the probability constraint is replaced by its approximation. Finally, we illustrate our results with numerical applications and propose several algorithms based on our findings.

Keywords: Stochastic programming, Probability functions, Epigraphical convergence, Variational Analysis, Nonsmooth Analysis

MSC Classification: 90C15 , 65K10 , 49J52 , 49J53

1 Introduction

Chance-constrained optimization is a mathematical and computational framework used in decision-making under uncertainty. It is particularly useful in situations where the decision-maker must account for the likelihood of uncertain events or parameters affecting the outcome of an optimization problem. Formally, the optimization problem is presented as

$$\min f(x) \text{ s.t. } \varphi(x) \geq p, \tag{1}$$

where φ is the so-called probability function defined as

$$\varphi(x) := \mathbb{P}(\omega \in \Omega : g(x, \xi(\omega)) \leq 0), \tag{2}$$

where the function $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ represents an inequality for the random vector $\xi : \Omega \to \mathbb{R}^m$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Here, let us notice that the function g is not required to be smooth, so the framework of (2) covers the case of joint probability functions by passing to the pointwise maximum of the functions that satisfy the system of inequalities, and of course using the same idea more general settings as the probabilistic/robust chance-constrained programming (see, e.g., [1–4]), or even optimal control problems with probabilistic constraints (see, e.g, [5–7]).

In chance-constrained optimization, the goal is to find an optimal solution while ensuring that certain constraints are satisfied with a specified probability or confidence level, that is, a vector x is feasible for problem (1) if and only if the random inequality $g(x, \xi) \leq 0$ is satisfied with at least probability p. These constraints represent the uncertain factors in the problem. The approach involves incorporating probabilistic models or distributions for these uncertainties and then formulating constraints to ensure they are met with a desired level of reliability. This model finds applications in various fields, such as finance, engineering, logistics, and energy management, where random variables or uncertain parameters influence outcomes. By quantifying and managing the risk associated with these uncertainties, chance-constrained optimization helps decision-makers make more robust and reliable choices in complex and uncertain environments.

The numerical implementation of chance-constrained optimization problems presents several technical challenges. One is the need to accurately model and characterize the underlying uncertainty in the problem. This requires statistical methods and data collection, as well as a good understanding of the probabilistic distributions involved. Additionally, solving chance-constrained optimization problems typically involves handling nonlinear optimization algorithms, which could be computationally intensive and prone to convergence issues, especially when dealing with high-dimensional or non-convex problems. These numerical methods, as any, also raise the question of balancing the trade-off between computational efficiency and the desired confidence level in constraint satisfaction. We refer to [8] for a precise treatment of multiple methods capable of handling probability constraints. Implementing chance-constrained optimization effectively demands an understanding of probability theory, optimization algorithms, and numerical methods to address these technical difficulties and obtain reliable solutions. This is achievable while keeping the probability functions as is, which can be observed from successful applications such as those considered in [9–12].

Numerical algorithms handling probability functions, as is, typically require first-order information. Obtaining first-order information has been extensively studied as testified by the classic references: [12–20]. Unlike using probability functions, as is, it is popular to deal with approximations of various kinds. A popular instantiation of this idea involves replacing the probability function with an "easier-to-compute" function for the purpose of optimization algorithms. For instance, the probability function can be written as an expectation, and then the characteristic function can be replaced with a smooth parameterized approximation. These and other related ideas can be found in works such as [21–25]. One of the main challenges of such an approach is ensuring proper convergence of the approximated solutions and optimal values to the intended targets of the original chance-constrained optimization model. Evidently, when the

parameters reach their limiting values, the numerical properties of the approximations can become unstable. Therefore, a trade-off between precision and computability must be established. Other approaches abandon altogether the idea of convergence. This evidently raises the question of the relation between the approximated problem and the original one, notably in terms of feasibility. Inner approximations, such as "safe-tractable" ones, replacing, for instance, the probability function by CVaR, in contrast, may be so conservative as to have an empty feasible set.

In this work, however, we focus on modifying the underlying random system $g(x,\xi) \leq 0$ by a sequence of approximations $g_k(x,\xi) \leq 0$. Specifically, we will show that this inner modification leads to approximations of the original probability function by a sequence of probability functions that inherit some of the convergence properties of the sequence g_k to g. This approach maintains the approximation problem within the framework of a chance-constrained problem. Consequently, the approximate optimal solution can be interpreted as a vector that satisfies the random inequality $g_k(x,\xi) \leq 0$ with a high probability level, for a random constraint $g_k(x,\xi) \leq 0$ that approximates the original and is defined by the user. Therefore, the interpretation is simpler. We also show that this approximation could have a computational advantage in terms of evaluation cost, over evaluating the probability function itself.

This paper is organized as follows. In section 2, we lay down the basic notation used, provide insights into the generalized spherical radial decomposition, and discuss the suggested approximation scheme as well as some basic related facts. Section 3 provides results into the differentiability of the approximate probability functions and discusses the consistency of the first-order information. The implications of replacing a probability function with its approximant in a given problem are considered in section 4, where consistency under classic assumptions is shown. Finally, section 5 features numerical experiments showing how the suggested framework can be put to work.

2 Notation, background, inner approximations

In this section, we present some classical notions used in convex and variational analysis, along with some preliminary results that constitute the basis of our investigation.

2.1 Notions and notation from variational analysis

Throughout this work, we will mostly use standard notation. Indeed, let $\overline{\mathbb{R}} := [-\infty, \infty]$ be the extended real line, and $\mathbb{R}^n$ be the Euclidean space with the norm $\|\cdot\|$. If $x \in \mathbb{R}^n$ and $r > 0$, we use $\mathbb{B}_r(x)$ to denote the closed ball centered at x with radius r, and $\mathbb{B}$ to denote the closed unit ball. For a set $\mathcal{O} \subset \mathbb{R}^n$, $x \overset{\mathcal{O}}{\to} \bar{x}$ indicates that x converges to $\bar{x}$ while staying within $\mathcal{O}$. The characteristic function of a set $\mathcal{O}$ is denoted by $\mathbb{1}_{\mathcal{O}}$, i.e., $\mathbb{1}_{\mathcal{O}} = 1$ if $x \in \mathcal{O}$ and zero otherwise. Furthermore, for a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, we denote *Fenchel's conjugate* as $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$,

$$f^*(v) = \sup_x \{\langle v, x\rangle - f(x)\}.$$

Let $\{C_k\}_{k\in\mathbb{N}}$ be a sequence of sets of $\mathbb{R}^n$. For this sequence, we may define its outer and inner limit sets, respectively, as follows:

$$\limsup_{k\to\infty} C_k := \{x : \exists N \in \mathcal{N}_\infty^\#, \exists x_k \in C_k (k \in N) \quad with \quad x_k \to x\},$$
$$\liminf_{k\to\infty} C_k := \{x : \exists N \in \mathcal{N}_\infty, \exists x_k \in C_k (k \in N) \quad with \quad x_k \to x\},$$

where $\mathcal{N}_\infty = \{N \subset \mathbb{N} : \mathbb{N}\backslash N \text{ is finite}\}$ and $\mathcal{N}_\infty^\# = \{N \subset \mathbb{N} : N \text{ is infinite}\}$. If the outer and the inner limit sets are equal, the *Painlevé-Kuratowski* limit is said to exist and is defined as:

$$\lim_{k\to\infty} C_k := \liminf_{k\to\infty} C_k = \limsup_{k\to\infty} C_k.$$

If we consider a sequence of functions $\{f_k\}_{k\in\mathbb{N}}$ defined on $\mathbb{R}^n$, we say that it *epigraphically converges* to f, denoted by $f_k \xrightarrow{e} f$, if their corresponding epigraphs, epi f_k converge in the *Painlevé-Kuratowski* sense to epi f. This property can be specified alternatively as shown in the following proposition (see, e.g., [26]).

Proposition 1. *Let $\{f_k\}_{k\in\mathbb{N}}$ be a sequence of functions on $\mathbb{R}^n$, then, it epigraphically converges to f if and only if the following two conditions hold:*

$$\liminf_{k\to\infty} f_k(x_k) \geq f(x) \text{ for any sequence } x_k \to x, \text{ and}$$
$$\limsup_{k\to\infty} f_k(x_k) \leq f(x) \text{ for some sequence } x_k \to x.$$

Furthermore, we say that f_k *hypographically converges* to f if and only if $-f_k$ epigraphically converges to $-f$.

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a given lower semicontinuous function and let a point $\bar{x}$ where $f(\bar{x})$ is finite be given. We say that a vector $v \in \mathbb{R}^n$ is a *regular* subgradient of f at $\bar{x}$, written as $v \in \hat{\partial} f(\bar{x})$, if

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x}\rangle + o(\|x - \bar{x}\|), \text{ for all } x \in \mathbb{R}^n.$$

Moreover, we say v is a *basic* subgradient of f at $\bar{x}$, written as $v \in \partial f(\bar{x})$, if there are sequences $x_k \to \bar{x}$ with $f(x_k) \to f(\bar{x})$ and $v_k \in \hat{\partial} f(x_k)$ with $v_k \to v$. For convenience, we set $\partial f(x) := \emptyset =: \hat{\partial} f(x)$ for $|f(x)| = +\infty$. The regular subgradient is also referred to as the *Fréchet* subgradient, whereas the "basic" subgradient also goes by the name of *limiting* or *Kruger-Mordukhovich* subgradient.

When f is locally Lipschitzian at $\bar{x}$, we define the *convexified* or *Clarke* subdifferential as

$$\overline{\partial} f(\bar{x}) := \operatorname{co}\{\partial f(\bar{x})\}.$$

The latter can also be defined as the convex hull of all possible limits along which f is differentiable, remembering Rademacher's theorem.

A function f that is locally Lipschitz at $\bar{x}$ is said to be *lower regular* at $\bar{x}$ provided that $\partial f(\bar{x}) = \hat{\partial} f(\bar{x})$. In such a case, all the previous subdifferentials coincide, i.e., $\overline{\partial} f(\bar{x}) = \hat{\partial} f(\bar{x})$. Moreover, the function f is said to be *lower regular* on U provided that it is *lower regular* at any point $x \in U$. It is easy to see that if f is *Clarke-regular* at $\bar{x}$ (see [27, Definition 2.3.4] or [26, Definition 7.25]), then it is lower-regular at $\bar{x}$.

When dealing with a sequence of functions $f_k : \mathbb{R}^n \to \overline{\mathbb{R}}$, it is important to consider the property of subgradient consistency. This concept or similar notions have been employed by various authors in different contexts (see, e.g., [28–31]).

Definition 1. *For a sequence of functions $f_k : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ locally Lipschitz around $\bar{x}$, we say that the sequence f_k is (basic-)subgradient consistent at $\bar{x}$ if:*

$$\operatorname{co}\left(\limsup_{\substack{x\to\bar{x}\\ k\to\infty}} \partial f_k(x)\right) = \overline{\partial} f(\bar{x}), \tag{3}$$

where

$$\limsup_{\substack{x\to\bar{x}\\ k\to\infty}} \partial f_k(x) := \{x^* : \exists x_k \to \bar{x}, x_k^* \to x^* \text{ with } x_k^* \in \partial f_k(x_k)\}.$$

Finally, if the functions f_k are differentiable and the previous property holds, we refer to the sequence as gradient consistent for short.

The following proposition shows that the subgradient consistency does not depend on the subgradient (regular, basic, or convexified) used in the left-hand side of (3).

Definition 2. *We say that a sequence of functions $f_k : \mathbb{R}^n \to \overline{\mathbb{R}}$ is uniformly Lipschitz around x if there exists a constant $L > 0$ and a neighbourhood U of x such that*

$$|f_k(y) - f_k(z)| \leq L\|y - z\|, \quad \text{for all } y, z \in U \text{ and all } k \in \mathbb{N}.$$

Proposition 2. *Let $f_k : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a sequence of functions that is uniformly Lipschitz around x. Then, the following holds:*

$$\operatorname{co}\left(\limsup_{\substack{x\to\bar{x}\\k\to\infty}} \hat{\partial} f_k(x)\right) = \operatorname{co}\left(\limsup_{\substack{x\to\bar{x}\\k\to\infty}} \partial f_k(x)\right) = \operatorname{co}\left(\limsup_{\substack{x\to\bar{x}\\k\to\infty}} \overline{\partial} f_k(x)\right).$$

Proof. Let us notice that by using a diagonal argument, we have

$$\limsup_{\substack{x\to\bar{x}\\k\to\infty}} \hat{\partial} f_k(x) = \limsup_{\substack{x\to\bar{x}\\k\to\infty}} \partial f_k(x).$$

Second, it is easy to see that

$$\operatorname{co}\left(\limsup_{\substack{x\to\bar{x}\\k\to\infty}} \partial f_k(x)\right) \subseteq \operatorname{co}\left(\limsup_{\substack{x\to\bar{x}\\k\to\infty}} \overline{\partial} f_k(x)\right).$$

Now, it is enough to prove:

$$\limsup_{\substack{x\to\bar{x}\\k\to\infty}} \overline{\partial} f_k(x) \subseteq \operatorname{co}\left(\limsup_{\substack{x\to\bar{x}\\k\to\infty}} \partial f_k(x)\right).$$

Let us consider a point $x^* \in \limsup_{\substack{x\to\bar{x}\\k\to\infty}} \overline{\partial} f_k(x)$. It follows that there is a sequence $x^*_\nu \in \operatorname{co}\left(\partial f_{\gamma(\nu)}(x_\nu)\right)$ for some sequence $x_\nu \to \bar{x}$ such that $x^*_\nu \to x^*$, for some increasing function $\gamma\colon \mathbb{N} \to \mathbb{N}$. Now, we can, as a result of Carathéodory's theorem, express

$$x^*_\nu = \sum_{i=1}^{n+1} \lambda^i_\nu x^*_{\nu,i} \text{ with } x^*_{\nu,i} \in \partial f_{\gamma(\nu)}(x_\nu),\ \sum_{i=1}^{n+1} \lambda^i_\nu = 1 \text{ and } \lambda^i_\nu \geq 0.$$

Now, we can assume that for all $i = 1, \ldots, n+1$ the sequence (λ^i_ν) converges to some λ_i. Moreover, due to the fact that the functions are Lipschitz continuous around $\bar{x}$ (all the functions with the same constant), we have that there exists a constant $M > 0$ such that $\|x^*_{\nu,i}\| \leq M$ for all $i = 1, \ldots, n+1$ and all ν. Therefore, we can assume that $(x^*_{\nu,i})$ also converges, let us say $x^*_{\nu,i} \to x^*_i$. Hence, $x^*_i \in \limsup_{\substack{x\to\bar{x}\\k\to\infty}} \partial f_k(x)$, consequently we conclude that $x^* \in \operatorname{co}\left(\limsup_{\substack{x\to\bar{x}\\k\to\infty}} \partial f_k(x)\right)$, and that concludes the proof. □ □

2.2 Probability functions and Spherical radial-like decomposition

Let $\xi : \Omega \to \mathbb{R}^m$ be an m-dimensional random vector, which possesses a continuous density with respect to the Lebesgue measure that we denote with f_ξ. Furthermore, let us consider a continuous function $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and let us denote the probability function φ, defined in (2). The subsequent proposition provides adequate conditions for guaranteeing the continuity of the probability function φ with respect to the data function g. The result is classic and can be found in [32, Proposition 2.2] (see also [33–35])

Proposition 3. *Let us consider the probability functions φ defined in (2), for a continuous function $g : \mathbb{R}^n \to \mathbb{R}$. Then, the probability function φ is an upper semicontinuous function on $\mathbb{R}^n$. Moreover, if the set $\{z : g(x, z) = 0\}$ has null measure, then φ is continuous at x.*

When the function g is convex with respect to its second argument, that is, $g(x,\cdot)$ is convex for all $x \in \mathbb{R}^n$, the so-called *spherical radial-like decomposition*, introduced in [36], allows us to rewrite the above probability for a general density function f_ξ as:

$$\varphi(x) = \int\limits_{v\in\mathbb{S}^{m-1}} e(x,v)d\mu_\zeta(v), \tag{4}$$

where $e : \mathbb{R}^n \times \mathbb{S}^{m-1} \to [0,+\infty]$ is the *radial probability-like* function given by

$$e(x,v) = \frac{2\pi^{\frac{m}{2}}|\det(L)|}{\Gamma(\frac{m}{2})} \int\limits_{\{r\geq 0: g(x,rLv)\leq 0\}} r^{m-1} f_\xi(rLv)dr.$$

The matrix L is an arbitrary non-singular matrix. In order to simplify the notation, let us define the *density-like function* θ:

$$\theta(r,v) := \frac{2\pi^{\frac{m}{2}}|\det(L)|}{\Gamma(\frac{m}{2})} r^{m-1} f_\xi(rLv). \tag{5}$$

In the context of the spherical radial-like decomposition (4), it is crucial to recognize that the matrix L assumes an abstract role of a non-singular matrix and can be arbitrarily chosen to enhance computations of the probability function. However, notable advantages arise when dealing with an elliptically symmetric distribution ξ. In such instances, it is common to let the matrix L result from the Cholesky decomposition of the covariance-like matrix associated with ξ. This specific utilization of the matrix L facilitates the representation of the *density-like function* θ, as defined in (5), in a more convenient form that does not rely on the variable v. To elaborate further, in cases where ξ is a Gaussian random vector, the function θ simplifies to a Chi-distribution with m degrees of freedom (see, e.g., [36] and the references therein).

2.3 Approximating functions from below

Let a mapping $h\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be given but arbitrary and let us consider an arbitrary sequence of functions $h_k\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ which converges monotonically nondecreasing to h as $k \to \infty$, i.e., $h_k(x) \nearrow h(x)$, as $k \to \infty$ for each $x \in \mathbb{R}^n$.

The next proposition establishes two further properties of approximating in this way:

Proposition 4. *Let $h_k\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be a sequence of lower semicontinuous functions that converges monotonically nondecreasing to $h\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ as $k \to \infty$. Then, the following are true:*

a) epi-convergence holds, i.e., $h_k \xrightarrow{e} h$ and h is in particular lower semicontinuous.

b) if the mapping h is moreover upper semicontinuous at x, then continuous convergence holds at x, i.e., for all $x_k \to x$

$$\lim_{k\to\infty} h_k(x_k) = h(x).$$

Proof. Let us first show item a). As a result of monotonicity, we have for all $x \in \mathbb{R}^n$ and $k_1 \leq k_2$, that $h_{k_1}(x) \leq h_{k_2}(x) \leq h(x)$. This means that the sequence of epigraphs decreases in the following way:

$$\operatorname{epi} h_{k_1} \supseteq \operatorname{epi} h_{k_2} \supseteq \operatorname{epi} h.$$

Therefore, by [26, Exercise 4.3], $\lim_{k\to\infty} \operatorname{epi} h_k = \bigcap_{k\in\mathbb{N}} \operatorname{cl}\operatorname{epi} h_k$ and the limit is closed. Since each function h_k is lower semicontinuous, $\operatorname{cl}\operatorname{epi} h_k = \operatorname{epi} h_k$. As a result, it remains to show:

$$\operatorname{epi} h \supseteq \bigcap_{k\in\mathbb{N}} \operatorname{epi} h_k.$$

To prove this, let us take $(x, \lambda) \in \bigcap_{k\in\mathbb{N}} \operatorname{epi} h_k$, this means that for all k, $h_k(x) \leq \lambda$. Since h_k converges monotonically, we have

$$\lim_{k\to\infty} h_k(x) = h(x)$$

and so we can conclude that $\lambda \geq h(x)$, thus yielding the result. Note that this implies in particular that $\operatorname{epi} h$ is closed as a Painlevé-Kuratowski limit, i.e., h is lower semicontinuous (but this also follows directly from [26, Proposition 1.26]). We can now focus on item b). As a result of item a) being shown and Proposition 1, we have for any sequence $(x_k)_{k\in\mathbb{N}}$ that converges to x, that:

$$\liminf_{k\to\infty} h_k(x_k) \geq h(x). \tag{6}$$

Furthermore, by monotonicity, i.e., $h_k(x_k) \leq h(x_k)$ for all k and taking the limit, using upper semicontinuity of h, on both sides, we get $h(x) \geq \limsup_{k\to\infty} h_k(x_k)$. We must thus have equality of both limits with moreover $h(x)$, i.e., item b). □ □

2.4 Approximating the probability functions from within

We now return to the probability function of (2) and assume, given somehow, a sequence of functions g_k monotonically nondecreasing towards g. To give an example, since g is assumed convex in the second argument, one can think of x-parameterized cutting plane models of some kind. Several concrete examples will be explored in section 5. With the help of this sequence g_k, we can define an approximated probability function $\varphi_k : \mathbb{R}^n \to [0, 1]$:

$$\varphi_k(x) := \mathbb{P}(\omega \in \Omega : g_k(x, \xi(\omega)) \leq 0), \tag{7}$$

where ξ has a continuous density with respect to the Lebesgue measure.

Theorem 5. *Let us consider a family of continuous $g_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ such that $g_k \nearrow g$ as $k \to \infty$. Moreover, considering the probability functions φ and φ_k defined in (2) and (7), respectively. Then, the following properties hold:*

a) Monotonicity: for all $k_1 > k_2$, $\varphi_{k_1}(x) \leq \varphi_{k_2}(x)$, for all $x \in \mathbb{R}^n$
b) Pointwise convergence: $\varphi_k(x) \searrow \varphi(x)$, for all $x \in \mathbb{R}^n$.
c) For any sequence $x_k \to x$ in $\mathbb{R}^n$

$$\limsup_{k\to\infty} \varphi_k(x_k) \leq \varphi(x).$$

Proof. We will begin the proof by demonstrating the monotonicity of the sequence φ_k from item a). Suppose $k_1 > k_2$. Then, by the monotonicity of the sequence $(g_k)_k$, we obtain the following inclusions:

$$\{\omega \in \Omega : g(x, \xi(\omega)) \leq 0\} \subseteq \{\omega \in \Omega : g_{k_1}(x, \xi(\omega)) \leq 0\} \subseteq \{\omega \in \Omega : g_{k_2}(x, \xi(\omega)) \leq 0\}, \text{ for all } x \in \mathbb{R}^n.$$

In particular, it implies that $\varphi(x) \leq \varphi_{k_1}(x) \leq \varphi_{k_2}(x)$ for all x, thus establishing the monotonicity of the sequence. With $\varphi_k(x)$, being a decreasing sequence bounded from below, one immediately gets: $\lim_{k\to\infty} \varphi_k(x) \geq \varphi(x)$ and this for arbitrary $x \in \mathbb{R}^n$. We will now show item c), which we can immediately use to conclude the proof of b).

To this end, let x be fixed but arbitrary and $x_k \to x$ an arbitrary sequence. We now define the sets:

$$B_k = \{z \in \mathbb{R}^m : g_k(x_k, z) > 0\}, \text{ and } B = \{z \in \mathbb{R}^m : g(x, z) > 0\}.$$

For an arbitrary but fixed $z \in \mathbb{R}^m$, using Proposition 4 and Proposition 1, we get:

$$\liminf_{k\to\infty} g_k(x_k, z) \geq g(x, z).$$

This implies that $\liminf_{k\to\infty} \mathbb{1}_{B_k}(z) \geq \mathbb{1}_B(z)$ and since z was arbitrary, this holds for all z. Additionally, since ξ has a continuous density and using Fatou's Lemma, we obtain

$$\liminf_{k\to\infty}(1-\varphi_k(x_k)) = \liminf_{k\to\infty} \mathbb{P}(\xi^{-1}(B_k)) \geq \mathbb{P}(\xi^{-1}(B)) = 1-\varphi(x).$$

Therefore, $\limsup_{k\to\infty} \varphi_k(x_k) \leq \varphi(x)$, which is item c). □ □

The approximating functions φ_k actually hypographically converge to φ as we have just shown. Let us make this fully apparent in the following result.

Corollary 6. *In the setting of Theorem 5 the sequence φ_k given in (7) hypographically converges to the probability function φ. Moreover, if in addition the set $\{z \in \mathbb{R}^m \ : g(x,z) = 0\}$ is a Lebesgue null-set, then φ_k will converge continuously to φ at x.*

Proof. As a result of Proposition 3 each probability function φ_k is upper semicontinuous. Moreover, as a result of Theorem 5 b), $-\varphi_k$ converges monotonically nondecreasing to $-\varphi$. At this stage, we may invoke Proposition 4 a) to conclude on the hypographical convergence.

In order to address the second point, let us observe that under the given additional assumption that the set $\{z \in \mathbb{R}^m \ : g(x,z) = 0\}$ is a Lebesgue null-set; the function φ is continuous at x (see Proposition 3). Now, item b) of Proposition 4 may be used to conclude. □ □

Remark 1. *The given Lebesgue null set regularity conditions are very mild. They hold, for instance, whenever g_k, g are convex in the second argument and allow for a Slater point, i.e., the existence of some z with $g_k(x,z) < 0$ and $g(x,z) < 0$ respectively (see [37, Theorem 1]).*

3 Generalized differentiation

3.1 Generalized differentiation of the approximate probability functions

In this section, we present sufficient conditions under which the approximate probability functions of (7) are differentiable in a generalized sense. The derived formulæ will be made precise in the current setting but rely on past results for the most part. These formulæ will be used in the next section to obtain consistency of gradients. In order to obtain a more precise yet useful formula, we will assume that the functions g_k have a specific structure.

Formally, we will assume that the functions g_k are represented as the maximum of a finite family of C^1 functions which is convex with respect to its second argument, that is, there exists $g_{k,i} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ with $i \in T_k$ with $\#T_k < +\infty$ such that

$$g_k(x,z) := \max\{g_{k,i}(x,z) : i \in T_k\} \tag{8}$$

for continuous differentiable functions $g_{k,i}$ such that $g_{k,i}(x,\cdot)$ is convex for all $x \in \mathbb{R}^n$.

Although we have assumed a finite index set, it is worth noting that the literature contains broader cases, such as structures defined by the maximum of nonsmooth functions or the supremum of arbitrary families of functions; however, these are beyond the scope of the current manuscript. For further details, we direct the reader to [4, 38].

3.1.1 Resolvants, directions, and controlling growth

We will be interested in studying the differentiability of φ_k around a trial point $\bar{x}$ previously set aside. The results given below will involve a neighbourhood U of $\bar{x}$, occasionally appropriately shrunken. Our first assumption is that zero is a Slater point for the constrained system $g_k(\bar{x},\cdot) \leq 0$, that is,

$$g_k(\bar{x},0) < 0. \tag{9}$$

We observe that the neighbourhood U can be taken such that this also holds for $x \in U$. Furthermore as a result of convexity in the second argument and should ξ be elliptically symmetric with mean vector 0 (which can be assumed without loss of generality), this condition holds automatically at any $\bar{x}$ with $\varphi_k(\bar{x}) > \frac{1}{2}$ (see [39, Corollary 2.1], and [40, Proposition 3.11]).

For $x \in U$ and $i \in T_k$, we define the sets of finite and infinite directions with respect to $g_{k,i}$ as

$$F_{k,i}(x) := \{v \in \mathbb{S}^{m-1} \ : \exists r \geq 0 : g_{k,i}(x, rLv) = 0\}, \text{ and } I_{k,i}(x) := \mathbb{S}^{m-1} \backslash F_{k,i}(x), \tag{10}$$

respectively. The finite and infinite directions with respect to g_k are defined as $F_k(x) = \cup_{i=1}^{s} F_{k,i}(x)$ and $I_k(x) = \cap_{i=1}^{s} I_{k,i}(x)$.

Now, for each $i \in T_k$, let us introduce the radial function associated with $g_{k,i}$ as the function $\rho_{k,i} : U \times \mathbb{S}^{m-1} \to [0, +\infty]$ given by

$$\rho_{k,i}(x, v) := \sup_{r \geq 0} \{r \ : \ g_{k,i}(x, rLv) \leq 0\} \tag{11}$$

and the radial function associated with g_k is given by $\rho_k(x, v) := \min \{\rho_{k,i}(x, v) : i \in T_k\}$. Furthermore, following [36], let us introduce the following growth condition:

Definition 3 (Growth condition)**.** *Let $\eta_\xi : \mathbb{R}^m \to [0, +\infty)$ be a mapping such*

$$\lim_{\|z\| \to +\infty} \|z\|^m f_\xi(z) \eta_\xi(z) = 0, \tag{12}$$

where f_ξ is the density of the random vector ξ with respect to the Lebesgue measure. Furthermore, we say that the family of mappings $\{g_{k,i}\}_{i \in T_k}$ satisfies the growth condition at $\bar{x}$ if for some $\epsilon, l > 0$

$$\|\nabla_x g_{k,i}(x, z)\| \leq \eta_\xi(z) \text{ for all } x \in \mathbb{B}_\epsilon(\bar{x}); \ \ \forall z : \|z\| \geq l; \ \ \forall i \in T_k. \tag{13}$$

3.1.2 Formulæ for subdifferentials and gradients

With the preparatory notions and notation laid down, we can now present the main (sub-)differentiability result for the approximate probability functions.

Theorem 7. *Let $\bar{x} \in \mathbb{R}^n$ be given such that*

- *$g_k(\bar{x}, 0) < 0$, and*
- *the family of mappings $\{g_{k,i}\}_{i \in T_k}$ satisfies the growth-condition at $\bar{x}$, given in Definition 3.*

Then the probability function (7) is locally Lipschitz on a neighbourhood U' of $\bar{x}$ and

$$\overline{\partial} \varphi_k(x) = - \int_{v \in F_k(x)} \bigcup_{i \in T_x^k(v)} \left\{ \frac{\theta(\rho_k(x, v), v) \nabla_x g_{k,i}\left(x, \rho_k\left(x, v\right) Lv\right)}{\left\langle \nabla_z g_{k,i}\left(x, \rho_k\left(x, v\right) Lv\right), Lv \right\rangle} \right\} d\mu_\zeta(v), \text{ for all } x \in U'. \tag{14}$$

where $T_x^k(v) := \{i \in T_k : \rho_{k,i}(x, v) = \rho_k(x, v)\}$. Moreover $-\varphi_k$ is lower-regular on U.

Proof. Let us split the proof into six claims

Claim 1: The inclusion $\subseteq$ holds in (14). Indeed, first let us notice that by [36, Corollary 3.2] the probability function (7) is locally Lipschitz at $\bar{x}$ and following inclusion holds

$$\partial \varphi_k(x) \subset - \int_{v \in F_k(x)} \bigcup_{i \in T_x^k(v)} \left\{ \frac{\theta(\rho_k(x, v), v) \nabla_x g_{k,i}\left(x, \rho_k\left(x, v\right) Lv\right)}{\left\langle \nabla_z g_{k,i}\left(x, \rho_k\left(x, v\right) Lv\right), Lv \right\rangle} \right\} d\mu_\zeta(v).$$

for all x in an appropriate neighbourhood of $\bar{x}$. Moreover, as a result of the convexity of the set-valued integrals under nonatomic measures (see, e.g., [41, Theorem 8.6.3]), the integral set in the right-hand side

of (14) is a convex and compact set (actually the integrably bounded condition required in [41, Theorem 8.6.3] follows from [36, Proposition 3.4]). This implies that (14) holds with inclusion.

Following [36, Lemma 3.2], we write the radial probability function associated with the function $g_{k,i}$ with $i \in T_k$ as

$$e_i(x,v) = \int_0^{\rho_{k,i}(x,v)} \theta(t,v)dt.$$

Claim 2: There exists an open neighbourhood U_i of $\bar{x}$ and $r_i > 0$ such that

$$|e_i(y_1,v) - e_i(y_2,v)| \leq r_i\|y_1 - y_2\|, \text{ for all } y_1, y_2 \in U_i \tag{15}$$

Using [36, Proposition 3.4] we have that the regular subdifferential of e_i, with respect to x, is bounded for all $(x,v) \in U_i \times \mathbb{S}^{m-1}$ by r_i, in some convex neighbourhood U_i. Then, the application of the mean value theorem (see, e.g., [42, Corollary 4.14]) implies that the function e_i must satisfy the Lipschitz condition (15).

Claim 3: There exists an open neighbourhood U_i of $\bar{x}$ such that $e_i : U_i \times \mathbb{S}^{m-1} \to \mathbb{R}$ is continuously differentiable with respect to x. Moreover, in this case

$$\nabla_x e_i(x,v) = \begin{cases} \frac{\theta(\rho_k(x,v),v)\nabla_x g_{k,i}(x,\rho_k(x,v)Lv)}{\langle \nabla_z g_{k,i}(x,\rho_k(x,v)Lv),Lv\rangle} & \text{if } v \in F_k(x), \\ 0 & \text{if } v \in I_k(x). \end{cases} \tag{16}$$

Indeed, let us denote

$$\Theta(s,v) := \int_0^s \theta(t,v)\,dt. \tag{17}$$

It follows from the continuity of θ that Θ is continuously differentiable with respect to s. On the one hand, by [36, Lemma 3.4], there exists a neighbourhood U_i' of $\bar{x}$ such that for all $x \in U_i'$ and $v \in \mathbb{S}^{m-1}$, the function ρ_i is continuously differentiable with respect to x at any $v \in F_k(x)$. Therefore, the function e_i, being the composition of Θ and $\rho_{k,i}$, is continuously differentiable with respect to x in the same neighbourhood. In this case

$$\nabla_x e_i(x,v) = \frac{\theta(\rho_k(x,v),v)\nabla_x g_{k,i}\left(x,\rho_{k,i}\left(x,v\right)Lv\right)}{\langle \nabla_z g_{k,i}\left(x,\rho_{k,i}\left(x,v\right)Lv\right),Lv\rangle}.$$

On the other hand, [36, Proposition 3.4] yields that $\partial_x e_i(x,v) \subseteq \{0\}$ for any $v \in I_k(x)$. This fact together with the Lipschitz continuity provided in (15) and [42, Theorem 4.17] implies that $e_i(\cdot,v)$ is (strictly) differentiable at x. Therefore, we have proved that the function $e_i(\cdot,v)$ is differentiable at any $x \in U_i$ and all $v \in \mathbb{S}^{m-1}$ (for some neighbourhood U_i of $\bar{x}$). Finally, [42, Thoerem 4.17] also shows that $e_i(\cdot,v)$ is C^1 around x whenever $v \in \mathbb{S}^{m-1}$ and $x \in U_i$, and that ends the proof of the claim.

Claim 4: There exists a neighbourhood U of $\bar{x}$ such that $e(x,v) = \min_{i\in T_k} e_i(x,v)$ for all $x \in U$ and all $v \in \mathbb{S}^{m-1}$.

Since the function Θ, defined in (17), is nondecreasing with respect to s, it is easy to see that

$$e(x,v) = \Theta(\rho(x,v),v) = \Theta(\min_{i\in T_k} \rho_{k,i}(x,v),v) = \min_{i\in T_k} \Theta(\rho_{k,i}(x,v),v) = \min_{i\in T_k} e_i(x,v),$$

whenever the functions ρ and $\rho_{k,i}$ are well-defined, which concludes the proof of the claim.

Claim 5: There exists a neighbourhood U of $\bar{x}$ such that for all $v \in \mathbb{S}^{m-1}$ and all $x \in U$ the function $e(\cdot,v)$ is Clarke-regular at x (see [27, Definition 2.3.4]).

Using Claim 4, we have that $-e(\cdot, v)$ is Clarke-regular at x, being the maximum of continuously differentiable functions (see, e.g., [27, Theorem 2.8.2]), and consequently by [27, Proposition 2.5.6] the function $e(\cdot, v)$ is Clarke-regular at x for all $x \in U$ and all $v \in \mathbb{S}^{m-1}$, for a suitable neighbourhood U of $\bar{x}$.

Claim 6: There exists a neighbourhood U of $\bar{x}$ such that for all $x \in U$ the function $-\varphi_k$ is Clarke-regular at x and the equality holds in (14).

Indeed, by [27, Theorem 2.7.2], we have that $-\varphi_k$ is Clarke regular, furthermore using [27, Proposition 2.3.1 and Theorem 2.7.2] we get

$$\overline{\partial}\varphi_k(x) = -\int\limits_{v \in F_k(x)} \overline{\partial}_x e(x, v) d\mu_\zeta(v), \text{ for all } x \in U. \tag{18}$$

Finally, we have that the calculus rules for the maximum (see, e.g., [27, Theorem 2.8.2]), the formula (16) and the convexity of the set-valued integral yield the equality in (14). □ □

A natural next question is whether or not there is some hope that the probability functions of (7) are actually continuously differentiable. This would clearly be the case if, locally, the right-hand side in equation (14) boils down to a singleton. This, in turn, happens whenever the active index set $T_x^k(v)$ is a singleton. Imposing that as an abstract condition is not very nice since one wonders how one can verify it concretely. Fortunately, a constraint qualification condition can be introduced, playing precisely the right role. Indeed, the authors of [43] introduced the *Rank-2-Constraint Qualification* (R2CQ). This condition inherently entails that the active index set $T_x^k(v)$ is μ_ζ-almost surely single-valued. This condition has been employed in several articles, as evidenced by references [36, 44, 45].

Definition 4. *Let us consider the following system of inequalities:*

$$h_j(x, z) \leq 0, \forall j \in J. \tag{19}$$

Then, for any $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$, we denote by

$$J(x, z) := \{j \in J | h_j(x, z) = 0\}$$

the active index set of h at (x, z). We say that the inequality system (19) satisfies the Rank-2-Constraint Qualification *(R2CQ) at (x, z) if*

$$\text{rank}\{\nabla_z h_i(x, z), \nabla_z h_j(x, z)\} = 2, \ \forall i, j \in J(x, z), i \neq j, \forall z \in \mathbb{R}^m \text{ with } \max_{j \in J} h_j(x, z) \leq 0. \tag{20}$$

Based on the given definition, we can observe certain similarities between the concept of R2CQ (Rank-2-Constraint Qualification) and *Linear Independence Constraint Qualification* (LICQ) in the context of nonlinear programming. However, the fundamental difference between these two concepts is that R2CQ compares the gradients of active constraints in pairs, whereas LICQ demands all gradients to be linearly independent. Thus, we can infer that LICQ is a stronger condition than R2CQ, as LICQ implies R2CQ. It therefore follows immediately from an "optimizer's" viewpoint that this condition is very reasonable for the underlying system $g_k(x, .) \leq 0$.

The next corollary shows that, under mild assumptions, the smoothness of the regularization g_k is inherited (locally) by the probability function.

Corollary 8. *Under the same assumptions of Theorem 7, let us suppose that R2CQ holds at $\bar{x}$ with respect to the system $g_{k,i}(x, z) \leq 0$, $i \in T_k$. Then, the probability function φ_k, defined in (7), is Fréchet differentiable at $\bar{x}$ and the following gradient formula holds*

$$\nabla\varphi_k(\bar{x}) = -\int\limits_{v \in F_k(\bar{x})} \frac{\theta(\rho_k(\bar{x}, v), v) \nabla_x g_{k,i_v}\left(\bar{x}, \rho_k\left(\bar{x}, v\right) Lv\right)}{\left\langle \nabla_z g_{k,i_v}\left(\bar{x}, \rho_k\left(\bar{x}, v\right) Lv\right), Lv \right\rangle} d\mu_\zeta(v), \tag{21}$$

where i_v is any selection of active indexes $T_{\bar{x}}^k(v)$, that is, $i_v \in T_{\bar{x}}^k(v)$. Moreover, if the R2CQ holds on a neighbourhood of $\bar{x}$. Then, the function φ_k is continuously differentiable around $\bar{x}$.

Proof. Indeed under (R2CQ), as a result of [43, Lemma 4.3], $\mu_\zeta(\{v \in \mathbb{S}^{m-1} : \#T_{\bar{x}}^k(v) > 1\}) = 0$. The arguments of [43, Theorem 4.1] - using, in fact, Propositions 2.1.2, 2.2.4 and its subsequent Corollary from [27] then allow us to conclude. □ □

We conclude this section by addressing a specific corollary that arises when the whole function g_k is continuous differentiability. In this setting, the differentiability of the data implies the corresponding differentiability of the probability functions at points where the data function g satisfies the Slater and growth conditions.

Corollary 9. *Under the same assumptions of Theorem 7, let us suppose the function g_k is continuously differentiable. Then, the probability function is continuously differentiable around $\bar{x}$ and the following gradient formula holds:*

$$\nabla\varphi_k(x) = -\int_{v\in F_k(x)} \frac{\theta(\rho_k(x,v),v)\nabla_x g_k\left(x,\rho_k\left(x,v\right)Lv\right)}{\langle\nabla_z g_k\left(x,\rho_k\left(x,v\right)Lv\right),Lv\rangle} d\mu_\zeta(v). \tag{22}$$

3.2 Consistency of the approximate gradients with the probability function

We start this section by providing the following (sub-)differential variational principle for the probability function φ using the inner regularized functions φ_k. The proof of this result follows directly from [38, Lemma 2.1], whose proof relies on standard techniques and variational principles used to approximate gradients by means of convergence of functions, such as Ekeland's or Borwein-Preiss' variational principle. Also, this approximation result can be obtained from the theory of epi-graphical convergence and uses nothing special from the setting of the probability function that we are developing here (see, e.g., [26] for further references).

Proposition 10. *Under the same assumptions of Theorem 5, we have that for every $x^* \in \hat{\partial}\varphi(\bar{x})$ and every $\epsilon > 0$, there exists $k \in \mathbb{N}$, $x_k \in \mathbb{R}^n$ and $x_k^* \in \partial\varphi_k(x_k)$ such that:*

$$\|\bar{x} - x_k\| + \|x^* - x_k^*\| + |\varphi(\bar{x}) - \varphi(x_k)| \leq \epsilon.$$

Particularly, we have that $\partial\varphi(\bar{x}) \subseteq \limsup_{x\to\bar{x},\, k\to\infty} \partial\varphi_k(x)$.

Proof. As a consequence of Theorem 5, we have that $\varphi(x) = \inf_{k\in\mathbb{N}} \varphi_k(x)$. Then, applying [38, Lemma 2.1] we get the result. □ □

We can now turn our attention to the convergence of the resolvant functions prior to examining subgradient consistency. To this end, let us recall that the mapping g is assumed convex in the second argument, that we have a monotonically non-decreasing sequence of continuous maps g_k, also convex in the second argument and converging pointwise to g. We also have a trial point $\bar{x}$ and (open) neighbourhood U of this point on which we assume

$$g(x,0) < 0,\ \forall x \in U. \tag{23}$$

We now define the resolvant functions, akin to what was done in (11), associated with g and the approximation g_k as follows. Let $\rho : U \times \mathbb{S}^{m-1} \to [0,+\infty]$ and $\rho_k : U \times \mathbb{S}^{m-1} \to [0,+\infty]$ be given by, respectively:

$$\rho(x,v) := \sup_{r\geq 0}\{r :\ g(x,rLv) \leq 0\}, \tag{24a}$$

$$\rho_k(x,v) := \sup_{r\geq 0}\{r :\ g_k(x,rLv) \leq 0\}. \tag{24b}$$

With this put in place, the following result guarantees the continuity of ρ_k, which is a key property for numerical computations.

Proposition 11. *Let us consider a sequence of functions g_k which converges continuously to g. Consider a point $x \in U$. Then, for every sequence $(x_k, v_k) \to (x, \bar{v})$ we have that $\rho_k(x_k, v_k) \to \rho(x, \bar{v})$.*

Proof. Consider a sequence $(x_k, v_k) \to (x, \bar{v})$ and a sequence $r_k = \rho_k(x_k, v_k)$. First, we notice that $\bar{r} := \limsup_k r_k > 0$, otherwise we arrive a contradiction with (23). Now, let us consider $r \in (0, \bar{r})$. By using the definition of the radial function, we have that $g_k(x_k, r_k L v_k) = 0$. We may of course assume k sufficiently large so that $x_k \in U$ and therefore we have $g_k(x_k, 0) < 0$. Now, using the fact that $r < \bar{r}$, we can consider a subsequence r_{k_j} such that

$$r < r_{k_j} \text{ for all } j \in \mathbb{N}.$$

The Slater condition ($g_{k_j}(x_{k_j}, 0) < 0$) along with with the convexity of $g_{k_j}(x_{k_j}, \cdot)$ implies that for all $j \in \mathbb{N}$, we have $g_{k_j}(x_{k_j}, rLv_{k_j}) < 0 = g_{k_j}(x_{k_j}, r_{k_j} L v_{k_j})$.
Therefore, due to the convergence of g_k to the function g and this just stated fact, we have that

$$0 = \lim_j g_{k_j}(x_{k_j}, r_{k_j} L v_{k_j}) \geq \lim_j g_{k_j}(x_{k_j}, rLv_{k_j}) = g(\bar{x}, rL\bar{v}),$$

which implies that $r \leq \rho(x, \bar{v})$. Since $r < \bar{r}$ was arbitrary, we get that $\bar{r} \leq \rho(x, \bar{v})$.
Now, if $\underline{r} := \liminf_k r_k = +\infty$, the proof is complete. Hence, we focus on the case that $\underline{r} < +\infty$. Let us pick $r \in \mathbb{R}$ such that $r > \underline{r}$. It follows that there exists a subsequence such that $r > r_{k_j}$ for all $j \in \mathbb{N}$. Again, it follows from convexity of g_{k_j} in the second argument and $g_k(x_k, 0) < 0$ that $g_{k_j}(x_{k_j}, rLv_{k_j}) > 0 = g_{k_j}(x_{k_j}, r_{k_j} L v_{k_j})$. Then, by the convergence of the functions g_k and the just stated fact, we get:

$$0 = \lim g_{k_j}(x_{k_j}, r_{k_j} L v_{k_j}) \leq \lim g_{k_j}(x_{k_j}, rLv_{k_j}) = g(\bar{x}, rL\bar{v}),$$

which implies that $r \geq \rho(x, \bar{v})$. Again, since r was arbitrary, we obtain that $\underline{r} \geq \rho(x, \bar{v})$. The last shows that

$$\limsup_k \rho_k(x_k, v_k) \leq \rho(x, \bar{v}) \leq \liminf_k \rho_k(x_k, v_k),$$

which ends the proof. □ □

Let us define the finite and infinite directions in $x \in \mathbb{R}^n$ with respect to g respectively, as

$$F(x) := \{v \in \mathbb{S}^{m-1} : \exists r > 0 : g(x, rLv) = 0\}, \text{ and } I(x) := \mathbb{S}^{m-1} \backslash F(x). \tag{25}$$

Definition 5. *We will say that the probability function φ, defined in (2), satisfies the spherical radial first order integral formula at $\bar{x}$ if the following holds:*

$$\overline{\partial}\varphi(\bar{x}) = -\int_{v \in F(x)} \left\{ \frac{\theta(\rho(\bar{x}, v), v) x^*}{\langle z^*, Lv \rangle} : (x^*, z^*) \in \partial g(\bar{x}, \rho(\bar{x}, v) Lv) \right\} d\mu_\zeta(v). \tag{26}$$

We recall that it has been shown that $\overline{\partial}\varphi$ is included in the right-hand side of (26) under the conditions of the current work. We refer the reader to [32, 36, 43]. Definition 5 thus only requests equality to hold rather than inclusion. Equality can be justified in some cases - along the lines of the argument of Theorem 7 above. We refer the reader to [5, 6].
In order to get the (sub-)gradient consistency of our approach, we require a uniform growth condition over the family of functions.

Definition 6 (Uniform growth condition)**.** *Let η_ξ be a continuous mapping satisfying (12). We say that the family of mappings $\{g_{k,i}\}_{i\in T_k, k\in\mathbb{N}}$ satisfies the uniform growth condition at $\bar{x}$ if for some $\epsilon > 0$, it holds*

$$\|\nabla_x g_{k,i}(x,z)\| \leq \eta_\xi(z) \text{ for all } x \in \mathbb{B}_\epsilon(\bar{x}) \ \ \forall z \in \mathbb{R}^n; \ \ \forall i \in T_k, \ \ \forall k \in \mathbb{N}. \tag{27}$$

We can now establish the main result of this section, the consistency of the approximation in terms of gradients:

Theorem 12. *Consider a sequence of functions g_k, given in the form of (8), such that the uniform growth condition holds at $\bar{x}$. Moreover, suppose that:*

- *The functions g_k converge monotonically to g*
- *The sequence g_k satisfies the gradient consistency property on $\mathbb{R}^n \times \mathbb{R}^m$.*
- *g is locally Lipschitz and lower-regular on $\mathbb{R}^n \times \mathbb{R}^m$*
- *$g(\bar{x}, 0) < 0$*

Then

$$\limsup_{\substack{x\to\bar{x}\\ k\to\infty}} \partial\varphi_k(x) \subseteq -\int_{v\in F(\bar{x})} \left\{ \frac{\theta(\rho(\bar{x},v),v)x^*}{\langle z^*, Lv\rangle} : (x^*,z^*) \in \partial g(\bar{x}, \rho_k(\bar{x},v)Lv) \right\} d\mu_\zeta(v). \tag{28}$$

Furthermore, if the probability function φ, defined in (2), satisfies the spherical radial first order integral formula at $\bar{x}$ of Definition 5, then the sequence $\{\varphi_k\}_k$ is (basic-) subgradient consistent at $\bar{x}$ (see Definition 1).

Proof. Let U denote once again the neighbourhood of $\bar{x}$ such that $g(x,0) < 0$ for all $x \in U$. Let us consider a sequence $x_k \to \bar{x}$, $x_k^* \to x^*$ with $x_k \in U$ and $x_k^* \in \partial\varphi_k(x_k)$. For k large enough, we may employ Theorem 7, and therefore, we can assume that the subgradient formula (14) holds. Upon potentially discarding the beginning of the sequence, we may assume it holds for all k. As a result, it follows that there exists integrable functions $y_k : \mathbb{S}^{m-1} \to \mathbb{R}^n$ with $y_k(v) \in S_k(x_k, v)$ a.e. $v \in \mathbb{S}^{m-1}$ such that $x_k^* = \int_{\mathbb{S}^{m-1}} y_k(v) d\mu_\zeta(v)$, where the set-valued mapping $S_k : U \times \mathbb{S}^{m-1} \rightrightarrows \mathbb{R}^n$ is given by

$$S_k(x,v) := \begin{cases} \left\{ \frac{\theta(\rho_k(x,v),v)\nabla_x g_{k,i}(x,\rho_k(x,v)Lv)}{\langle \nabla_z g_{k,i}(x,\rho_k(x,v)Lv), Lv\rangle} : i \in T_x^k(v) \right\} & \text{if } v \in F_k(x) \\ \{0\} & \text{if } v \in I_k(x) \end{cases}.$$

Let us split the proof into four claims:

Claim 1: There exist $\epsilon > 0$ and $C > 0$ such that

$$\sup\{\|u^*\| : u^* \in S_k(u,v)\} \leq C, \text{ for all } u \in \mathbb{B}_\epsilon(\bar{x}), \ \forall v \in \mathbb{S}^{m-1}, \forall k \in \mathbb{N}. \tag{29}$$

Indeed, let $\epsilon > 0$ be given by the uniform growth condition (see Definition 6) and consider $(u,v) \in \mathbb{B}_\epsilon(\bar{x}) \times \mathbb{S}^{m-1}$, $k \in \mathbb{N}$, and $u^* \in S_k(u,v)$. Hence, if $v \in I_k(u)$, we get that $\|u^*\| = 0$, on the other hand, let us assume that $v \in F_k(u)$. Then, we can assume that u belongs to the neighbourhood, where the growth condition (13) holds. Moreover, in this case

$$u^* = \frac{\theta(\rho_k(x,v),v)\nabla_x g_{k,i}\left(x,\rho_k\left(x,v\right)Lv\right)}{\langle \nabla_z g_{k,i}\left(x,\rho_k\left(x,v\right)Lv\right), Lv\rangle} \text{ for some } i \in T_u^k(v).$$

Hence, using the convexity of the functions g (see, e.g., [36, Lemma 3.3]) we have that

$$\langle \nabla_z g_{k,i}\left(x,\rho_k\left(x,v\right)Lv\right), Lv\rangle \geq \frac{|g_k(u,0)|}{\rho_k(u,v)}.$$

Now, using the uniform growth condition (27), defining $z_{u,v} := \rho_k(u,v)Lv$, we get that given $l \geq 0$,

$$\|u^*\| \leq \begin{cases} c_0 \|z_{u,v}\|^m f_\xi(z_{u,v})\eta_\xi(z_{u,v}), & \text{if } \|z_{u,v}\| \geq l \\ c_0 \sup\{\|z\|^m f_\xi(z)\|\nabla_x g(x,z)\| : \|z\| \leq l\} & \text{if } \|z_{u,v}\| \leq l, \end{cases} \tag{30}$$

where $c_0 := \sup\{\frac{1}{\|Lv\|} : v \in \mathbb{S}^{m-1}\}\frac{2\pi^{\frac{m}{2}}|\det(L)|}{\Gamma(\frac{m}{2})}$. Now, as a result of (12), the first term tends to 0 as $l \to \infty$, and the second is bounded as the maximum of continuous maps on a compact. The existence of $C > 0$ has thus been shown.

Claim 2:

$$x^* \in \int_{\mathbb{S}^{m-1}} \limsup_{k\to+\infty, x'\to\bar{x}} S_k(x', v) d\mu_\zeta(v). \tag{31}$$

Indeed, using a Fatou-type theorem (see, e.g., [46, Corollary 4.1]), we obtain this formula.

Claim 3: We have that

$$\limsup_{k\to+\infty, x'\to\bar{x}} S_k(x', v) \subseteq \begin{cases} \left\{\frac{\theta(\rho(\bar{x},v),v)x^*}{\langle z^*, Lv\rangle} : (x^*, z^*) \in \partial g(\bar{x}, \rho(\bar{x},v)Lv)\right\} & \text{if } v \in F(x) \\ \{0\} & v \in I(x) \end{cases}$$

and consequently (28) holds.

Indeed, let us fix $v \in \mathbb{S}^{m-1}$. Consider $u_j \to \bar{x}$ and $u_j^* \in S_{k_j}(u_j, v)$ with $u_j^* \to u^*$.

First, if $v \in I_{k_j}(u_j)$ for infinitely many j, in fact j sufficiently large, we have, by Proposition 11 that $v \in I(\bar{x})$ and $\|u^*\| = 0$, so the result holds.

Therefore, in the rest of the proof, we can assume that $v \in F_{k_j}(u_j)$ for all $j \in \mathbb{N}$. Consequently, by definition of S_k, we have that

$$u_j^* = \frac{\theta(\rho_{k_j}(u_j,v),v)\nabla_x g_{k_j,i}\left(u_j, \rho_{k_j}\left(u_j,v\right)Lv\right)}{\left\langle \nabla_z g_{k_j,i}\left(u_j, \rho_{k_j}\left(u_j,v\right)Lv\right), Lv\right\rangle} \text{ for some } i \in T_{u_j}^{k_j}(v). \tag{32}$$

Let us define $z_j := \rho_{k_j}(u_j,v)Lv$. Consider two separate cases, first if $v \in I(\bar{x})$. Then, Proposition 11 implies that $\rho_{k_j}(u_j,v) \to +\infty$, this implies that $\|z_j\| \to \infty$. Hence, the growth condition (see (12)) and the estimation given in (30) for each j implies that there exists some $c_0 > 0$ such that

$$\|u^*\| = \lim_{j\to\infty} \|u_j^*\| \leq c_0 \lim_{j\to\infty} \|z_j\|^m f_\xi(z_j)\eta_\xi(z_j) = 0,$$

implying that the claim holds in this case. Second, if $v \in F(\bar{x})$, we can define $\bar{z} := \rho(x,v)Lv$ and $(w_j^*, z_j^*) := (\nabla_x g_{k_j,i}(u_j, z_j), \nabla_z g_{k_j,i}(u_j, z_j))$. Since, the index i belongs to $T_{u_j}^{k_j}(v)$ we have that

$$g_{k_j,i}(u_j, z_j) = g_{k_j}(u_j, z_j),$$

which by the subdifferential calculus rule of the maximum (see, e.g, [42, Theorem 4.10]), implies that $(w_j^*, z_j^*) \in \partial g_{k_j}(u_j, z_j)$. Now, it follows from Proposition 11 that $(u_j, z_j) \to (\bar{x}, \bar{z})$. Moreover, it follows from the convexity of the sequence of functions with respect to z that the sequence $(z_j^*)_j$ must be bounded. Moreover, due to the uniform growth condition

$$\|w_j^*\| \leq \sup\{\eta_\xi(z) : z \in \mathbb{B}_{r_0}(0)\}, \text{ with } r_0 := \sup\|z_j\| + 1.$$

Then, due to the continuity of η_ξ, we get that the sequence (w_j^*) also must be bounded. Then, taking a subsequence if necessary and by the gradient consistency assumption over g_k we have that $(w_j^*, z_j^*) \to (w^*, z^*) \in \overline{\partial} g(\bar{x}, \bar{z})$, and consequently by lower-regularity of g, we get that $(w^*, z^*) \in \partial g(\bar{x}, \bar{z})$, and that concludes the proof of the claim.

Claim 4: Under the additional assumption the sequence $(\varphi_k)_k$ satisfies the gradient consistency property at $\bar{x}$.

This follows simply by noticing that Proposition 10, (26), (28) and the convexity and closedness of the set-valued integral (see, e.g., [41, Theorem 8.6.3]) implies that

$$\overline{\partial}\varphi(\bar{x}) \subseteq \operatorname{co}\left(\limsup_{\substack{x\to\bar{x}\\ k\to\infty}} \partial\varphi_k(x)\right) \subseteq -\int_{v\in F(\bar{x})} \left\{\frac{\theta(\rho(\bar{x},v),v)x^*}{\langle z^*, Lv\rangle} : (x^*,z^*) \in \partial g(\bar{x},\rho(\bar{x},v)Lv)\right\} = \overline{\partial}\varphi(\bar{x}),$$

which ends the proof of the result. □ □

4 On the consistency of approximate chance-constrained problems

4.1 On convergence of problems

In this section, we will study, following classic arguments, the impact of replacing the probability function φ, defined in (2), with the regularization φ_k given in (7). Beyond just the regularization or approximation of the probability function, we also will consider the possibility of approximation of the objective function. To make this precise, consider a fixed reliability parameter $p \in (0,1]$. Furthermore, let us be given a lower semicontinuous (lsc) function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ playing the role of objective. We now address the following optimization problem:

$$\begin{array}{ll} \min & \psi(x) \\ \text{s.t} & x \in M := \{x \in \mathbb{R}^n : \varphi(x) \geq p\}, \end{array} \tag{P}$$

where φ is the probability function defined in (2). Furthermore, we consider the family of problems

$$\begin{array}{ll} \min & \psi_k(x) \\ \text{s.t} & x \in M_k := \{x \in \mathbb{R}^n : \varphi_k(x) \geq p_k\}, \end{array} \tag{P_k}$$

where φ_k is given in (7), ψ_k represents a generic regularization of the objective function, and $p_k \uparrow p$. Replacing the initial objective function ψ in problem (P), and the reliability parameter p with a function ψ_k in problem (P_k), and a reliability parameter p_k provides greater flexibility in the studied model, allowing, for instance, alternative algorithms in numerical applications. Let $v(P)$ and $v(P_k)$ denote the optimal values of problems (P) and (P_k), respectively.

The following proposition establishes the Painlevé-Kuratowski convergence of the feasible sets of problem (P_k) to the feasible set stipulated in the original optimization problem (P). The proof of the result below relies on classic lines of argumentation, as, for instance, encountered in [21, 26, 47] and many other works.

Proposition 13. *Under the same assumptions of Theorem 5, and considering $p \in (0,1]$ and $p_k \nearrow p$. Then, the sets M and M_k are closed and the sequence of sets $\{M_k\}_k$ converge to M in the Painlevé-Kuratowski sense.*

Proof. First, by the upper semicontinuity of φ and φ_k, which results from Proposition 3, we have that the sets M and M_k are closed. Now for the convergence of the sequence $\{M_k\}_k$ it is enough to use [26, Proposition 7.7] on the epi-convergence of $-\varphi_k$ to $-\varphi$. obtained by corollary 6. □ □

We can now present a standard "convergence" result of the approximate problems towards the original problem. The following theorem delineates this mathematical connection between problems (P) and (P_k).

Theorem 14. *Under the assumptions of Theorem 5, consider $p \in (0,1]$ and $p_k \nearrow p$. Furthermore, let us consider a sequence of lower semicontinuous functions $\psi_k : \mathbb{R}^n \to \overline{\mathbb{R}}$ converging monotonically nondecreasingly to ψ as $k \to \infty$. Assume that there exists some $k_0 \in \mathbb{N}$ and $x_0 \in M \cap \operatorname{dom}\psi$ such that $M_{k_0} \cap \{x : \psi_{k_0}(x) \leq \psi(x_0)\}$ is bounded. Then,*

a) $v(P), v(P_k)$ *are finite for all* $k \geq k_0$ *and* $v(P_k) \nearrow v(P)$ *as* $k \to \infty$.
b) Let (x_k) *be any sequence of (globally) optimal solutions for* (P_k)*, then this sequence* (x_k) *is bounded. Moreover, any cluster point* $\bar{x}$ *of this sequence is an optimal solution of problem* (P).
c) If problem (P) *has a unique solution, then every sequence of minimizers of* (P_k) *converges to that solution.*

Proof. First let us notice that for all $k_2 \geq k_1 \geq k_0$ the following inclusions hold

$$x_0 \in M \cap \{x : \psi(x) \leq \psi(x_0)\} \subset B_{k_2} \subseteq B_{k_1} \subseteq B_{k_0}, \tag{33}$$

where $B_k := M_k \cap \{x : \psi_k(x) \leq \psi(x_0)\}$. Now, by Proposition 13 the set M and M_k are closed, so the nonempty sets $M \cap \{x : \psi(x) \leq \psi(x_0)\}$ and B_k are compact. Since ψ and ψ_k are lower semicontinuous, we have that the problems (P) and (P_k) admit a solution. Therefore, $v(P), v(P_k) \in \mathbb{R}$ for all $k \geq k_0$.

Now, using the inclusion (33) we have that $v(P_{k_1}) \leq v(P_{k_2}) \leq v(P)$ for all $k_1 \leq k_2$ large enough, which implies that $\lim_{k\to\infty} v(P_k) \leq v(P)$.

Now, let x_k denote a global minimizer of problem (P_k). It follows by (33) that $x_k \in B_{k_0}$. As a result, there exists a converging subsequence, let us say $x_{k_j} \to \bar{x}$. Due to Proposition 13 the limit point belongs to M, and consequently $v(P) \leq \psi(\bar{x})$. Moreover, as a result of Proposition 4 $\psi_{k_j} \overset{e}{\to} \psi$, so we have

$$\psi(\bar{x}) \leq \liminf_{k\to\infty} \psi_{k_j}(x_{k_j}) = \lim_{k\to\infty} v(P_k) \leq v(P),$$

which shows that $v(P_k) \to v(P)$ as $k \to \infty$, and at the same time that $\bar{x}$ is a solution of problem (P).

To derive the final result, we observe that any accumulation point of the sequence must be the unique minimizer of problem (P). Therefore the sequence can not have other accumulation points, thereby completing the proof. □ □

Although the previous result is powerful, it unfortunately requires globally solving the approximate problems P_k. As a result, one may wonder when it is reasonable to assume that this can be done. The next section will discuss situations under which P_k is a problem from convex optimization, and such global optimality can be reasonable.

4.2 On globally solving the approximate problems: convexity

This section gathers sufficient conditions ensuring the convexity of the sequence of optimization problems (P_k). Although it is frequently stated that probability constraints are not convex, there are, in fact, many situations wherein convexity can be assured. This was recognized a long time ago by Prékopa in his seminal work, e.g., [48]. Although probability functions themselves can never be concave, there is a chance that the upper-level sets (M_k, M in our notation) are convex. This would require only much weaker quasi-concavity. Finally, it is worth noting that two kinds of convexity results exist: those ensuring convexity of M_k, M regardless of the safety levels p_k, p and those of the "eventual kind" requiring that p_k, p are sufficiently large. The second kind has been extensively explored in e.g., [49–51].

In order to present some conditions under which convexity of M_k can be present, let us recall some generalized notations of concavity. A nonnegative function f defined over a convex set $D \subset \mathbb{R}^n$ is said to be α-concave, where $\alpha \in [-\infty, \infty]$ provided that for all $x, y \in D$ and all $\lambda \in [0, 1]$, the following inequality is satisfied:

$$f(\lambda x + (1-\lambda)y) \geq m_\alpha(f(x), f(y), \lambda),$$

with the function $m_\alpha : \mathbb{R}_+ \times \mathbb{R}_+ \times [0,1] \to \mathbb{R}$ defined by

$$m_\alpha(a,b,\lambda) = \begin{cases} a^\lambda b^{1-\lambda} & \text{if } \alpha = 0, \\ \max\{a,b\} & \text{if } \alpha = \infty, \\ \min\{a,b\} & \text{if } \alpha = -\infty, \\ 0 & \text{if } ab = 0, \text{ and } \alpha \leq 0, \\ (\lambda a^\alpha + (1-\lambda)b^\alpha)^{1/\alpha} & \text{otherwise.} \end{cases}$$

We follow here the presentation of [52]. In the above definition, the case $\alpha = 0$ corresponds to log-concavity. Evidently, for $\alpha = 1$, the function is just concave. In particular all functions f that are α-concave, are also β-concave for $\beta \leq \alpha$ and thus in particular $-\infty$-concave, i.e., quasi-concave.

Furthermore, considering a random vector ξ, the probability distribution of ξ, that is, $\mathbb{P}_\xi(A) := \mathbb{P}(\xi \in A)$ is said to be α-concave provided that for any convex sets A and B in $\mathbb{R}^m$, along with any λ in the interval $[0,1]$ the following inequality holds

$$\mathbb{P}_\xi(\lambda A + (1-\lambda)B) \geq m_\alpha(\mathbb{P}_\xi(A), \mathbb{P}_\xi(B), \lambda).$$

As prominent examples, we can mention multivariate Gaussian random vectors or uniform random vectors over convex sets. We refer the reader to [52, 53] for further examples.

We can now establish Prékopa's classic result to our setting:

Corollary 15. *Suppose that ξ has an α-concave probability distribution and g_k and g are quasi-convex (jointly in both arguments). Then, for every $\lambda > 0$, and any $p \in (0,1)$ the functions φ_k and φ are α-concave on the sets $\{x \in \mathbb{R}^n : \exists z \in \mathbb{R}^m \; s.t \; g_k(x,z) \leq 0\}$ and $\{x \in \mathbb{R}^n : \exists z \in \mathbb{R}^m \; s.t \; g(x,z) \leq 0\}$, respectively. Consequently, for any $p, p_k \in (0,1]$ the sets M_k and M are convex. Moreover, suppose that the objective function ψ and ψ_k in the optimization problem (P) and (P_k) are convex and lower semicontinuous. Furthermore, assume that ψ is coercive and strictly convex. Then, any sequence of solutions of problems (P_k) converges to the unique minimizer of (P), provided that problem (P) is feasible.*

Proof. The α-concavity of the functions φ and φ_k follows from a direct application of [52, Theorem 4.39, p. 108]. In particular, φ and φ_k are quasi-concave; hence, the sets M and M_k, being upper-level sets of these functions, are convex. Now suppose that the objective function ψ in the optimization problem (P) is convex, coercive, lower semicontinuous and $B := M \cap \{x : \psi(x) \leq \psi(x_0)\} \neq \emptyset$, for some $x_0 \in M$ (recall that (P) is feasible). It follows from the coercivity, strict convexity of ψ, and convexity of M that the problem (P) has a unique solution. Now, let us define $B_k := M_k \cap \{x : \psi_k(x) \leq \psi(x_0)\}$, by the convergence of the involved functions ψ_k and φ_k to ψ and φ, respectively and (33) we have that B_k Painlevé-Kuratowski converges to B. Hence, by the coercivity of ψ, the set B is bounded, and by the convexity of B_k, we have that B_k must be bounded for large enough k. Then, Theorem 14 allows us to conclude. □ □

We can evidently derive similar results of convexity in the eventual case relaxing the requirement of joint quasi-convexity of g_k, g to, for instance, convexity in both arguments - but not jointly, e.g., see [51]. Alternatively, the separable case allows for generalized concavity in x through the choice or presence of a specific Copula linking the components of ξ. Here we refer to [50, 54]. The use of generalized concavity can also be generalized, and this is in both situations. Here we refer the reader to [55].

5 Putting the framework into practice: experiments and illustrations

In order to illustrate the developments in this section, let us begin by showing how the spherical-radial decomposition can be put to work in order to compute - with the same cost - the probability value and a

"gradient". To this end, let φ be as in (2), where ξ is a random vector with density f_ξ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a continuous function, which is convex in the second argument. We consider the scenario where ξ follows an elliptically symmetric distribution, characterized by a mean vector μ and covariance-like matrix Σ. At any evaluation point x that satisfies the Slater condition, meaning $g(x, \mu) < 0$, the density-like function θ remains independent of v and belongs to a well-defined class. In the specific case where ξ follows a multivariate Gaussian distribution, as discussed below, θ corresponds to a Chi density with m degrees of freedom.

Algorithm 1 is designed to compute the spherical-radial decomposition for φ, providing estimates for both the function's value and its gradient, as given by (4) and (22), respectively. While initially presented in a simple Monte-Carlo version, the algorithm can be enhanced with more sophisticated techniques along classic lines, such as those described in [56].

Algorithm 1: Computing the spherical-radial decomposition for φ

Data: $x \in \mathbb{R}^n$, with $g(x, \mu) < 0$, $N > 0$, $\xi \sim \mathcal{N}(\mu, \Sigma)$ and L such that $\Sigma = LL^\top$.
Result: $\varphi(x)$, $\nabla\varphi(x)$

Sample N points $\{v_1, v_2, \ldots, v_N\}$ distributed over the unit sphere $\mathbb{S}^{m-1}$, i.e.,

$$\mathcal{V} := \{v_j \in \mathbb{R}^m : \ \|v_j\|^2 = 1 \text{ for } j = 1, \ldots, N\}.$$

Initialize $val, s \leftarrow 0$
for $v \in \mathcal{V}$ **do**
  Find the zero, denoted as r^*, of the one-dimensional function:

$$g(x, r^* Lv + \mu) = 0.$$

  or establish $r^* = \infty$.
  Compute the Chi-probability according to $F_\chi(r^*)$ where F_χ represents the cumulative distribution function of the one-dimensional Chi-distribution with m degrees of freedom.
  $val \leftarrow val + F_\chi(r^*)$.
  if $r^* < \infty$ **then**
    $s \leftarrow s - f_\chi(r^*) \frac{\nabla_x g(x, \mu + r^* Lv)}{\langle \nabla_z g(x, \mu + r^* Lv), Lv \rangle}$ where f_χ represents the density function of the one-dimensional Chi-distribution with m degrees of freedom.
  end
end
Return $\frac{val}{N}, \frac{s}{N}$.

Some comments are in order:

- it is immediately clear how to update the scheme when ξ is another elliptically symmetric random vector: it suffices to replace the Chi distribution and density functions by an appropriate other choice, such as Fisher-Snedecor when ξ is multivariate Student.
- the scheme can be enhanced with an online estimation of the sample variance and stopped early if sufficiently small.
- detecting if $r^* = \infty$ is not hard, it suffices to verify if $g(x, \mu + \bar{r} Lv) < 0$ for say $\bar{r} = 50$, since for such $\bar{r}$: $F_\chi(\bar{r})$ is numerically 1 and $f_\chi(\bar{r})$ is numerically zero.
- computing r^* can be done with a (1 dimensional) dichotomy procedure or Newton-Rhapson approach if no analytical expression of r^* is available.

In this algorithm, when r^* is not available analytically, there is a "computationally costly part" in computing it. The cost is not really high, but since we must compute r^*, N times (per iteration of some

overall solver), gain in its evaluation can be shown to be beneficial overall. This topic is explored in the next section:

5.1 Functions known through oracles

In order to make it work, let $C_0 \subseteq C_1 \subseteq ... \subseteq C_k$, be increasing compact sets such that $\bigcup_{k\geq 0} C_k = \mathbb{R}^m$. We can evidently pick these sets such that for any decreasing sequence $\delta_k \downarrow 0$, $\mathbb{P}[\xi \in C_k] \geq 1 - \delta_k$, since the probability measure associated with ξ is a Borel probability measure and hence tight. Furthermore, let us be given a sequence $\varepsilon_k \downarrow 0$ as well as a bijection $f : \mathbb{N} \to \mathbb{N}^2$.

A convex mapping $h : \mathbb{R}^m \to \mathbb{R}$ can now be approximated using its cutting plane model. Indeed for any collection of points $z_1, ..., z_k$, with $s_j \in \partial h(z_j)$, $j = 1, ..., k$, it is well known that

$$\check{h}_k(z) = \max_{j=1,...,k} h(z_j) + \langle s_j, z - z_j \rangle ,$$

is polyhedral convex and $\check{h}_k(z) \leq h(z)$ for all $z \in \mathbb{R}^m$. Moreover evidently $h(z_j) = \check{h}_k(z_j)$ for $j = 1, ..., k$. Moreover for each $\varepsilon > 0$, some $\delta_k > 0$ can be found such that $h(z) - \check{h}_k(z) \leq \varepsilon$ for $z \in \bigcup_j \mathbb{B}(z_j, \delta_k)$. This follows from the continuity of both functions and them being equal in each of the points.

Our goal is for $\check{h}_k$ to converge monotonically nondecreasingly to h. It is ensured by the following proposition:

Proposition 16. *Assume that the points $Z_k = \{z_1, ..., z_k\}$ are generated as follows. For a given k, let $(k_1, k_2) = f(k)$ be available and let Z'_k be the collection of centers of the finite subcover of C_{k_1} covered by ε_{k_2} balls. Then $Z_k = Z_{k-1} \cup Z'_k$ and let $\check{h}_k$ be the cutting plane model with elements in the "bundle" Z_k. Then $\check{h}_k$ converges monotonically nondecreasingly to h.*

Proof. It holds trivially that $\check{h}_k \leq \check{h}_{k+1} \leq h$. It remains to show that for any $z \in \mathbb{R}^m$, $h(z) - \check{h}_k(z) \to 0$. To this end, let us pick $z \in \mathbb{R}^m$ arbitrary but fixed. By construction, there exists K such that for all $k \geq K$, $z \in C_k$. There also exists a sequence of ball centers z_{k_j} and $\varepsilon_{k_j} \downarrow 0$, such that $\left|z - z_{k_j}\right| \leq \varepsilon_{k_j}$. As a result of continuity of the function h, we have $h(z_{k_j}) \to h(z)$, and moreover for all $\ell \geq k_j$, $\check{h}_\ell(z_{k_j}) = \check{h}_{k_j}(z_{k_j}) = h(z_{k_j})$. Finally for any z' and such ℓ:

$$h(z') - \check{h}_\ell(z') \leq h(z') - \check{h}_{k_j}(z'),$$

together with the continuity of the cutting plane model, the result follows. □ □

With this put in place, we can now turn our attention to the advantage of moving to a cutting plane model for the mapping g. For instance, if $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is convex in both arguments, then with g_k a cutting-plane model for g, we have the added advantage of being able to compute the resolvant mapping explicitly. Indeed, in this case:

$$\rho_i(x, v) = \begin{cases} \frac{\langle z_i^*, z\rangle - \langle x_i^*, x - x_i\rangle - g(x_i, z_i)}{\langle z_i^*, v\rangle} & \text{if } \langle z_i^*, v\rangle > 0, \\ +\infty & \text{else.} \end{cases} \tag{34}$$

with $(x_i^*, z_i^*) \in \partial g(x_i, z_i)$ and

$$\rho(x, v) = \min_{i=1,...,k} \rho_i(x, v).$$

As a particular example of this situation, we can consider the function g to be defined as:

$$g(x, z) = \min_y \zeta(x, y, z), \tag{35}$$

where $\zeta : \mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^m \to \mathbb{R}$ is a continuous function that is convex with respect to (x, y, z). In this case it is well known that at some given $(\bar{x}, \bar{z})$, it suffices to identify the minimizing $\bar{y}$ and that then

$(x^*, z^*) \in \partial_{x,z}\zeta(\bar{x}, \bar{y}, \bar{z})$ is such that $(x^*, z^*) \in \partial g(\bar{x}, \bar{z})$. Indeed, the approximated model obtained by the cutting plane model for a sampling $(\bar{x}_i, \bar{z}_i)$ and its corresponding (x_i^*, z_i^*) read as follows:

$$\varphi_k(x) := \mathbb{P}\left(\xi \in \Omega : g_k(x, \xi) := \max_{i=1,\dots,k} \{\zeta(\bar{x}_i, \bar{y}_i, \bar{z}_i) + \langle x_i^*, x - \bar{x}_i\rangle + \langle z_i^*, z - \bar{z}_i\rangle\} \leq 0\right). \tag{36}$$

Example 1. *Let us consider the following example with $A \in \mathbb{R}^{n\times s}$, $B \in \mathbb{R}^{m\times s}$ and $c \in \mathbb{R}$, such that the function ζ, is defined as:*

$$\zeta(x, y, z) = \frac{1}{2}\|x - Ay\|^2 + \frac{1}{2}\|z - By\|^2 + \frac{1}{2}\|y\|^2 - c, \tag{37}$$

the domain of the function g is defined as $\mathrm{dom}\,(g) = \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^m : \|x\|_\infty \leq 1\}$. *We will randomly sample the domain of g using a uniform distribution in the unit ball. Let us recall that since we have an explicit formulation for the conjugate function, this can also be used to provide lower estimations of the function due to the Fenchel inequality between a convex function and its conjugate.*

To gain insight into the behavior of the cutting plane model defined in (36), we begin with a two-dimensional setting, specifically taking $n = m = s = 2$. For simplicity and ensure strong convexity with respect to the second variable, we consider $A = B = I$ and set $c = 1$, which guarantees the existence and uniqueness of the minimizer in (37).

Figure 1 illustrates the resulting function g, obtained by solving (35) in this setting. The function is evaluated on a 20×20 grid over x, with z fixed at zero. We provide both a three-dimensional surface plot and a two-dimensional cross-section at $x_2 = 0$ to show the accuracy and convergence properties of the cutting plane approximations as the number of approximation points k increases.

In particular, Figure 1a presents a 3D visualization of $g(x, 0)$. The colored surface corresponds to the true function, while the gray mesh beneath depicts the cutting plane model for $k = 4$. This comparison reveals how the model begins to capture the underlying geometry of g in three-dimensional space.

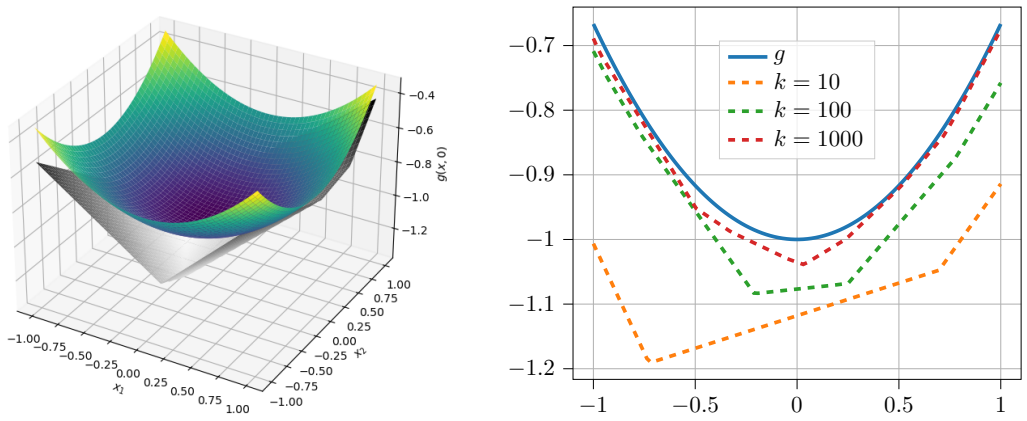
To further highlight convergence behavior, Figure 1b shows a 2D slice of the function at $x_2 = 0$. The solid blue curve represents the true function g, and the dashed lines show piecewise linear approximations corresponding to the cutting plane model for $k = 10$, 100 and 1000. As k increases, the approximation align more closely non decreasingly with the true function, demonstrating the progressive refinement and effectiveness of the model.

We now turn our attention to the numerical solution of problem (36), leveraging the cutting plane approximations of the function g introduced previously. Specifically, we approximate the associated probability function φ_k using the spherical radial decomposition method (Algorithm 1), evaluated over a 20×20 grid withing the domain.

Figure 2 illustrated the result of this approximation. In Figure 2a, we present the exact function φ, computed also using Algorithm 1 with a large number of samples ($N = 20000$), alongside its approximations φ_k, computed using the cutting plane model for $k = 20$ (blue), $k = 100$ (orange) and $k = 1000$ (green). As expected, increasing k leads to a more accurate approximation, with φ_k converging to the true function.

Figure 2b provides a two-dimensional cross-sectional view comparing the exact function and the approximations φ_k for different values of k. Additionally, we include the cross-section of the probability function obtained via Monte Carlo sampling, denoted as φ^{MC}, to highlight the smoothness and accuracy between the two approaches.

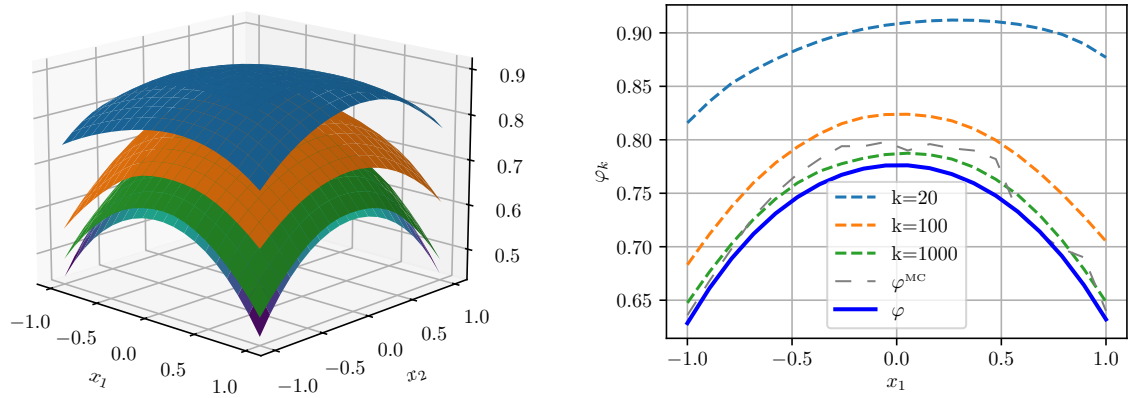
To further examine the behavior of the cutting plane method, we consider two variants based on the computation of the resolvent mapping ρ. The first, φ_k^{Nwt}, employs a Newton-type method as described in (24) and used in Step 4 of Algorithm 1. The second, φ_k^{Bld}, utilizes the closed-form expression given in (34). This distinction allows us to assess the trade-off between computational complexity and approximation fidelity in both implementations.



(a) Approximation

(b) Cross-section at $x_2 = 0$

Fig. 1: Approximation of the function $x \mapsto g(x, 0)$ in Example 1.
(a) Three-dimensional visualization of the function $g(x, 0)$ (colored surface) alongside its cutting-plane approximation for $k = 4$ (gray mesh). (b) Cross-sectional plot at $x_2 = 0$, comparing the true function g (solid blue line) with its cutting-plane approximations for increasing values of $k = 10$, 100, and 1000 (dashed lines), highlighting the improvement in accuracy as k grows.



(a) Approximation

(b) Cross-section at $x_2 = 0$

Fig. 2: Approximation of the function φ in Example 1.
(a) Exact probability function φ and its cutting-plane approximations φ_k for $k = 20$ (blue), $k = 100$ (orange), and $k = 1000$ (green), illustrating convergence toward the true model as k increases. (b) Two-dimensional cross-section of φ and φ_k for varying values of k, along with the Monte Carlo approximation, denoted by φ^{MC}, highlighting differences in smoothness and accuracy.

We begin our analysis by comparing the Monte Carlo (MC) approximation with the spherical radial decomposition (SRD) approach, as summarized in Table 1. Here, N denotes the number of samples used to compute the approximation, both for Monte Carlo method φ^{MC} and the SRD method φ^{SRD}. To assess accuracy, both approximations are compared against a high-fidelity reference solution φ, obtained via SRD using $N = 20000$ samples. The evaluation is performed over a 20×20 grid, where the error and computation time are computed at each grid point and aggregated to produce the relevant metrics.

As shown in Table 1, the SRD method demonstrates significantly greater accuracy than the MC approach. Notably, even with just $N = 100$ samples, SRD achieves an approximation error that is more than an order of magnitude smaller than that of MC with $N = 1000$. This suggests that Monte Carlo would

	$\lvert\varphi - \varphi_N^{\mathrm{MC}}\rvert$			t_N^{MC}	$\lvert\varphi - \varphi_N^{\mathrm{SRD}}\rvert$			t_N^{SRD}
N	mean ± std	min	max	mean ± std	mean ± std	min	max	mean ± std
100	0.097 ± 0.023	5.0e-03	0.171	0.122 ± 0.046 s	0.001 ± 0.001	2.9e-06	0.005	1.320 ± 0.132 s
250	0.040 ± 0.012	7.4e-04	0.065	0.140 ± 0.049 s	0.001 ± 0.001	1.6e-06	0.004	1.971 ± 0.131 s
500	0.017 ± 0.009	2.2e-04	0.044	0.175 ± 0.052 s	0.001 ± 0.000	5.0e-06	0.002	3.137 ± 0.169 s
1000	0.009 ± 0.006	2.5e-06	0.033	0.242 ± 0.071 s	0.001 ± 0.001	3.6e-06	0.004	5.398 ± 0.168 s

Table 1: Accuracy and computation time for φ^{MC} and φ^{SRD} approximations.

require at least two orders of magnitude more samples to achieve comparable accuracy. In addition to its higher precision, SRD also exhibits lower variance across the grid, indicating consistent performance and increased reliability of the approximation. Together, these observations highlight the superior sample efficiency and numerical stability of the SRD method compared to standard Monte Carlo sampling.

We now proceed to analyze the numerical performance of the cutting plane model. Tables 2 and 3 report, respectively, the approximation accuracy and computational efficiency of the probability function φ_k for various values of k.

Regarding the accuracy metrics (Table 2), the following quantities are displayed:

$\boldsymbol{k}$: Number of cutting planes used.
$\boldsymbol{|\varphi - \varphi_k^{MC}|}$: Absolute error between the true function φ and its Monte Carlo approximation.
$\boldsymbol{|\varphi - \varphi_k^{Nwt}|}$: Absolute error using the Newton-based cutting-plane approximation.
$\boldsymbol{|\varphi - \varphi_k^{Bdl}|}$: Absolute error using the closed-form resolvent mapping.

Table 2 reveals that, even at low sample sizes, the cutting plane method outperforms the MC approach when the number of cutting planes k is sufficiently large (highlighted in gray in the table). Interestingly, the threshold value of k required to see this improvement is smaller for lower sample sizes, indicating that the cutting plane model becomes effective more rapidly in lower-sample regimes. This performance gain is not only in terms of reduced mean error, but also in the lower variance and maximum error values across the grid.

Now, regarding the numerical performance of the cutting plane model in Table 3, reports the computation times associated with different approximation methods. Let us denote by φ_N^{SRD} the evaluation of the probability function using the SRD method with a given number of samples N. The table includes the following columns:

$\boldsymbol{t_k^{MC}}$: Time taken (in seconds) to evaluate φ_k^{MC}.
$\boldsymbol{t_k^{Nwt}}$: The computation time (in seconds) for calculating the probability function φ_k^{Nwt}.
$\boldsymbol{t_k^{Bdl}}$: The computation time (in seconds) for calculating the probability function φ_k^{Bld}.
$\boldsymbol{t_k^{Nwt}/t_k^{Bdl}}$: The speedup ratio of the computation time of φ_k^{Nwt} to $\varphi_k^{\mathbf{Bdl}}$.
$\boldsymbol{t_N^{SRD}/t_k^{Bdl}}$: The speedup ratio of the computation time of φ_N^{SRD} to φ_k^{Bld}.

Indeed, this table confirms the increasing computational cost with growing k. However, it also highlights a key advantage of our approach: approximations based on the cutting plane model, specially using the closed form for the resolvant φ_k^{Bld}, are consistently faster than the SRD-based computation of the probability function φ_N^{SRD}. This speedup is specially notable for smaller values of k. In contrast, the Newton-based approximation φ_k^{Nwt}, does not always exhibit this advantage; in fact, several instances where it is slower than φ_N^{SRD} are highlighted in gray in the table. Nevertheless, across all tested values of k, the closed-form approximation φ_k^{Bld} consistently yields the most efficient performance, often outperforming the Newton-based version φ_k^{Nwt} by a substancial margin.

Finally, an important remark in this section concerns the availability of gradient information provided when using SRD-based approximations. As outlined in Algorithm 1, gradient information is obtained at a negligible extra cost, since it is inherently computed during the evaluation of the probability function. In contrast, the MC-based approximation yields only function values and does not provide any gradient estimates. One could argue that finite differences could be employed to estimate gradient information, but the combination of MC estimations and finite differences is not good, e.g., [56]. On top this would

		$\lvert\varphi - \varphi_k^{\mathbf{MC}}\rvert$			$\lvert\varphi - \varphi_k^{\mathbf{Nwt}}\rvert$			$\lvert\varphi - \varphi_k^{\mathbf{Bdl}}\rvert$		
N	k	**mean ± std**	**min**	**max**	**mean ± std**	**min**	**max**	**mean ± std**	**min**	**max**
100	100	0.141 ± 0.022	9.3e-02	0.231	0.059 ± 0.015	4.4e-02	0.154	0.059 ± 0.015	4.4e-02	0.154
	200	0.128 ± 0.021	8.2e-02	0.211	0.038 ± 0.009	2.5e-02	0.084	0.038 ± 0.009	2.5e-02	0.084
	500	0.117 ± 0.019	6.6e-02	0.183	0.024 ± 0.006	1.3e-02	0.049	0.024 ± 0.006	1.3e-02	0.049
	1000	0.112 ± 0.020	5.5e-02	0.181	0.017 ± 0.005	9.2e-03	0.041	0.017 ± 0.005	9.2e-03	0.041
	2000	0.108 ± 0.021	1.5e-02	0.178	0.012 ± 0.003	6.7e-03	0.030	0.012 ± 0.003	6.7e-03	0.030
	5000	0.103 ± 0.022	5.0e-03	0.171	0.007 ± 0.002	3.3e-03	0.020	0.007 ± 0.002	3.3e-03	0.020
	10000	0.102 ± 0.022	5.0e-03	0.171	0.005 ± 0.002	2.7e-03	0.015	0.005 ± 0.002	2.7e-03	0.015
	20000	0.100 ± 0.022	1.5e-02	0.171	0.004 ± 0.001	1.6e-03	0.012	0.004 ± 0.001	1.6e-03	0.012
	50000	0.099 ± 0.022	5.3e-03	0.171	0.002 ± 0.001	4.1e-04	0.009	0.002 ± 0.001	4.1e-04	0.009
250	100	0.090 ± 0.019	5.9e-02	0.170	0.059 ± 0.015	4.5e-02	0.150	0.059 ± 0.015	4.5e-02	0.150
	200	0.073 ± 0.015	3.5e-02	0.120	0.039 ± 0.009	2.6e-02	0.083	0.039 ± 0.009	2.6e-02	0.083
	500	0.061 ± 0.013	3.1e-02	0.096	0.024 ± 0.006	1.3e-02	0.051	0.024 ± 0.006	1.3e-02	0.051
	1000	0.055 ± 0.012	2.7e-02	0.096	0.017 ± 0.005	9.6e-03	0.040	0.017 ± 0.005	9.6e-03	0.040
	2000	0.050 ± 0.012	1.1e-02	0.080	0.012 ± 0.003	7.0e-03	0.028	0.012 ± 0.003	7.0e-03	0.028
	5000	0.047 ± 0.012	3.0e-03	0.074	0.007 ± 0.002	3.4e-03	0.018	0.007 ± 0.002	3.4e-03	0.018
	10000	0.045 ± 0.012	3.0e-03	0.074	0.005 ± 0.002	2.8e-03	0.014	0.005 ± 0.002	2.8e-03	0.014
	20000	0.043 ± 0.012	3.3e-03	0.070	0.004 ± 0.001	1.7e-03	0.011	0.004 ± 0.001	1.7e-03	0.011
	50000	0.043 ± 0.012	7.4e-04	0.065	0.002 ± 0.001	5.7e-04	0.008	0.002 ± 0.001	5.7e-04	0.008
500	100	0.071 ± 0.014	4.5e-02	0.140	0.059 ± 0.014	4.5e-02	0.146	0.059 ± 0.014	4.5e-02	0.146
	200	0.053 ± 0.011	2.8e-02	0.092	0.039 ± 0.009	2.7e-02	0.082	0.039 ± 0.009	2.7e-02	0.082
	500	0.039 ± 0.009	2.0e-02	0.072	0.024 ± 0.006	1.4e-02	0.054	0.024 ± 0.006	1.4e-02	0.054
	1000	0.032 ± 0.009	8.5e-03	0.069	0.017 ± 0.005	9.9e-03	0.039	0.017 ± 0.005	9.9e-03	0.039
	2000	0.028 ± 0.009	7.3e-03	0.063	0.012 ± 0.003	7.2e-03	0.026	0.012 ± 0.003	7.2e-03	0.026
	5000	0.024 ± 0.009	1.3e-04	0.057	0.007 ± 0.002	3.7e-03	0.017	0.007 ± 0.002	3.7e-03	0.017
	10000	0.022 ± 0.009	2.2e-04	0.053	0.005 ± 0.002	2.9e-03	0.012	0.005 ± 0.002	2.9e-03	0.012
	20000	0.020 ± 0.009	2.2e-04	0.049	0.004 ± 0.001	1.8e-03	0.010	0.004 ± 0.001	1.8e-03	0.010
	50000	0.019 ± 0.009	2.2e-04	0.047	0.002 ± 0.001	6.8e-04	0.007	0.002 ± 0.001	6.8e-04	0.007
1000	100	0.059 ± 0.015	3.2e-02	0.131	0.060 ± 0.014	4.5e-02	0.147	0.060 ± 0.014	4.5e-02	0.147
	200	0.041 ± 0.014	1.1e-02	0.087	0.040 ± 0.009	2.8e-02	0.083	0.040 ± 0.009	2.8e-02	0.083
	500	0.026 ± 0.012	3.3e-03	0.075	0.025 ± 0.006	1.4e-02	0.055	0.025 ± 0.006	1.4e-02	0.055
	1000	0.019 ± 0.012	1.3e-06	0.057	0.017 ± 0.005	1.0e-02	0.040	0.017 ± 0.005	1.0e-02	0.040
	2000	0.014 ± 0.010	9.6e-06	0.050	0.012 ± 0.003	7.3e-03	0.024	0.012 ± 0.003	7.3e-03	0.024
	5000	0.011 ± 0.009	7.3e-08	0.038	0.007 ± 0.002	3.6e-03	0.017	0.007 ± 0.002	3.6e-03	0.017
	10000	0.010 ± 0.008	2.5e-05	0.037	0.005 ± 0.002	2.9e-03	0.013	0.005 ± 0.002	2.9e-03	0.013
	20000	0.010 ± 0.007	2.5e-05	0.035	0.004 ± 0.001	1.5e-03	0.011	0.004 ± 0.001	1.5e-03	0.011
	50000	0.009 ± 0.007	1.1e-04	0.035	0.002 ± 0.001	4.4e-04	0.008	0.002 ± 0.001	4.4e-04	0.008

Table 2: Accuracy of approximating the probability function φ.

require a very significant sample size N and $2n$ evaluations to get a poor estimate of a gradient. This limitation makes the Monte Carlo method unsuitable for use in any gradient-based optimization pipeline. This property is further explored in the next section.

5.2 Additional smoothing

Let us begin by introducing the smooth maximum function formally:

Definition 7. *Let $\alpha > 0$ be an arbitrary parameter. We define the smoothing function $\mathcal{M}_\alpha^n : \mathbb{R}^n \to \mathbb{R}$ as follows $\mathcal{M}_\alpha^n(x) = \alpha^{-1}\log(\langle u, \exp(\alpha x)\rangle)$, where $u \in \mathbb{R}^n$ is the all one vector.*

Then, we can derive the following properties easily:

Lemma 17. *The smoothing function is jointly convex, continuously differentiable and for all $x \in \mathbb{R}^n$ we have $\max_i x_i + \alpha^{-1}\log(n) \geq \mathcal{M}_\alpha^n(x) > \max_i x_i$. The function $\mathcal{M}_\alpha^n$ is strictly increasing, i.e., for any $x, y \in \mathbb{R}^n$ with $x_i < y_i$ for at least one $i = 1, ..., n$ and $x_i \leq y_i$ elsewhere, we have $\mathcal{M}_\alpha^n(x) < \mathcal{M}_\alpha^n(y)$. The mapping is also decreasing in α, i.e.,*

$$\mathcal{M}_r^n(x) < \mathcal{M}_\alpha^n(x), \forall r > \alpha > 0, x \in \mathbb{R}^n, \tag{38}$$

		t_k^{MC}	t_k^{Nwt}	t_k^{Bdl}		
N	k	mean ± std	mean ± std	mean ± std	$t_k^{\mathrm{Nwt}}/t_k^{\mathrm{Bdl}}$	$t_N^{\mathrm{SRD}}/t_k^{\mathrm{Bdl}}$
	100	0.073 ± 0.007 s	0.949 ± 0.108 s	0.778 ± 0.134 s	1.2	1.7
	200	0.064 ± 0.006 s	0.897 ± 0.095 s	0.681 ± 0.103 s	1.3	1.9
	500	0.056 ± 0.005 s	0.891 ± 0.095 s	0.681 ± 0.101 s	1.3	1.9
	1000	0.062 ± 0.005 s	0.917 ± 0.091 s	0.725 ± 0.093 s	1.3	1.8
100	2000	0.060 ± 0.005 s	1.087 ± 0.125 s	0.872 ± 0.135 s	1.2	1.5
	5000	0.069 ± 0.005 s	1.283 ± 0.159 s	1.038 ± 0.152 s	1.2	1.3
	10000	0.067 ± 0.004 s	1.204 ± 0.131 s	0.927 ± 0.130 s	1.3	1.4
	20000	0.127 ± 0.010 s	1.978 ± 0.259 s	1.409 ± 0.189 s	1.4	0.9
	50000	0.095 ± 0.005 s	2.125 ± 0.291 s	1.375 ± 0.176 s	1.5	1.0
	100	0.092 ± 0.007 s	1.363 ± 0.112 s	0.906 ± 0.109 s	1.5	2.2
	200	0.081 ± 0.006 s	1.356 ± 0.107 s	0.858 ± 0.091 s	1.6	2.3
	500	0.075 ± 0.007 s	1.370 ± 0.109 s	0.901 ± 0.112 s	1.5	2.2
	1000	0.078 ± 0.006 s	1.427 ± 0.098 s	0.985 ± 0.110 s	1.4	2.0
250	2000	0.078 ± 0.005 s	1.610 ± 0.138 s	1.162 ± 0.144 s	1.4	1.7
	5000	0.091 ± 0.007 s	1.839 ± 0.166 s	1.333 ± 0.167 s	1.4	1.5
	10000	0.090 ± 0.006 s	1.807 ± 0.142 s	1.241 ± 0.155 s	1.5	1.6
	20000	0.161 ± 0.011 s	2.526 ± 0.196 s	1.764 ± 0.197 s	1.4	1.1
	50000	0.145 ± 0.006 s	3.074 ± 0.213 s	1.975 ± 0.195 s	1.6	1.0
	100	0.138 ± 0.012 s	2.153 ± 0.138 s	1.224 ± 0.112 s	1.8	2.6
	200	0.124 ± 0.010 s	2.116 ± 0.130 s	1.177 ± 0.103 s	1.8	2.7
	500	0.112 ± 0.009 s	2.171 ± 0.117 s	1.242 ± 0.116 s	1.7	2.5
	1000	0.116 ± 0.008 s	2.265 ± 0.118 s	1.445 ± 0.123 s	1.6	2.2
500	2000	0.118 ± 0.009 s	2.501 ± 0.144 s	1.602 ± 0.161 s	1.6	2.0
	5000	0.140 ± 0.015 s	2.887 ± 0.179 s	1.784 ± 0.172 s	1.6	1.8
	10000	0.140 ± 0.012 s	2.943 ± 0.199 s	1.756 ± 0.167 s	1.7	1.8
	20000	0.243 ± 0.023 s	4.196 ± 0.312 s	2.372 ± 0.218 s	1.8	1.3
	50000	0.265 ± 0.021 s	5.223 ± 0.262 s	2.961 ± 0.200 s	1.8	1.1
	100	0.238 ± 0.034 s	3.536 ± 0.161 s	1.796 ± 0.129 s	2.0	3.0
	200	0.215 ± 0.025 s	3.604 ± 0.152 s	1.769 ± 0.118 s	2.0	3.1
	500	0.203 ± 0.022 s	3.790 ± 0.141 s	1.912 ± 0.139 s	2.0	2.8
	1000	0.201 ± 0.019 s	4.004 ± 0.150 s	2.304 ± 0.153 s	1.7	2.3
1000	2000	0.213 ± 0.024 s	4.299 ± 0.180 s	2.450 ± 0.167 s	1.8	2.2
	5000	0.247 ± 0.033 s	4.818 ± 0.190 s	2.715 ± 0.186 s	1.8	2.0
	10000	0.263 ± 0.037 s	5.040 ± 0.202 s	2.718 ± 0.169 s	1.9	2.0
	20000	0.400 ± 0.047 s	6.563 ± 0.239 s	3.575 ± 0.234 s	1.8	1.5
	50000	0.528 ± 0.057 s	9.631 ± 0.313 s	4.887 ± 0.232 s	2.0	1.1

Table 3: Computation time for the evaluation of the probability function φ.

and

$$\lim_{\alpha\to\infty} \mathcal{M}_\alpha^n(x) = \max_i x_i. \tag{39}$$

We observe that this is well known, and we refer the reader to, for instance, e.g., [57]. We can now combine the smoothing function with the cutting plane model to obtain a sequence of smooth approximants. The resulting function $g_k(x,z) = \mathcal{M}_\alpha^k(\ell_1(x,z),...,\ell_k(x,z))$, where $\ell_j(x,z)$ are the cutting plane linearizations of the original mapping g. The following corollary summarizes the findings:

Corollary 18. *Let the mapping $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be jointly convex and assume that its cutting plane model is constructed following Proposition 16. Then the following holds true:*

a) the smoothed cutting plane model converges monotonically nondecreasingly to g, i.e., $g_k(x,z) \nearrow g(x,z)$.

b) $g_k(x,z) \leq g(x,z) \leq g_k(x,z) + \alpha^{-1}\ln(k)$.

Example 2 (Smooth Chance-Constrained Problem)**.** *Let us consider the probability function defined using the function described in Example 1. Its associated chance-constrained test problem reads:*

$$\begin{aligned} \min_{x} \quad & \mu^T x \\ s.t. \quad & \varphi_k(x) := \mathbb{P}\left[\xi \in \Omega : g_k(x,\xi) = \mathcal{M}_\alpha^k(\ell_1(x,\xi), ..., \ell_k(x,\xi)) \leq 0\right] \geq p \\ & x \geq 0, \end{aligned} \tag{40}$$

where $x \in \mathbb{R}^n$ is the decision variable, g_k is built according to the cutting plane model and smoothed according to Definition 7, and $p \in [0,1]$ is the probability threshold.

One of the main advantages of working a smooth version g_k lies in the result presented in Corollary 9, which ensures the differentiability of the corresponding probability function φ_k when g_k is smooth. This is an important property that can be utilized numerically, as having access to both a value and gradient from the oracle enables us to make use of commonly used nonlinear optimization solver packages. For this experiment, we employ the `SciPy` implementation of the SLSQP algorithm to approximate the solution of the smooth chance-constrained problem (40).

Tables 4 and Table 5 report the solutions to the chance-constrained problem (40), obtained using the SRD-based approximation applied to the smooth cutting plane model (x_k^*), as well as and the solution derived from original function (35), evaluated with a fixed number of samples N. Alongside the solution x_k^*, the tables present the corresponding cost function value (fun), execution time, number of objective function evaluations (nfev), number of Jacobian evaluations (njev), and the error with respect to a high-accuracy reference solution (x^{SRD}) computed using $N = 10000$ samples. Results are reported for two probability threshold levels: p = 0.7 and p = 0.75, respectively.

The results from these tables demonstrate a clear trend: as k increases, the solution x_k^* converges towards the reference solution x^{SRD}, both in terms of the Euclidean distance and objective function value. Moreover, for moderate values of k, the cutting-plane-based approach yields a considerable speedup over the direct SRD method, especially for a smaller number of cutting planes k. This performance gain is reflected in significantly reduced computation times, while the number of solver iterations, objective function evaluations, and Jacobian evaluations remains comparable.

To conclude this analysis, Figure 3 presents two subplots corresponding to the threshold levels $p = 0.7$ and $p = 0.75$, respectively. Each plot illustrates the approximated solutions obtained from the cutting-plane-based problems, along with their associated feasible sets. The figure clearly demonstrates that as the number of cutting planes increases, the feasible sets (dashed lines) derived from the cutting plane approximation of the probability function progressively converge to those of the original problem. This convergence is also reflected in the solutions computed by the solver (colored dots), which approach the true solution as the approximation improves.

As a final remark, regarding the sampling methodology used for selecting the points at which the cutting-plane approximations are constructed in (36), the results confirm that increasing the number of cutting planes leads to more accurate solutions. Nevertheless, we believe that a more strategic or adaptive selection mechanism for the cutting planes could further enhance the quality of the approximation while reducing the number of required samples. Exploring such approaches represents a promising direction for future research.

Supplementary information. Not applicable

Acknowledgements. Research of Pedro Pérez-Aros was partially supported by Centro de Modelamiento Matemático (CMM), ACE210010 and FB210005, BASAL funds for center of excellence and ANID-Chile grant: MATH-AMSUD 23-MATH-09 and MATH-AMSUD 23-MATH-17, ECOS-ANID ECOS320027, Fondecyt Regular 1220886, Fondecyt Regular 1240335, Fondecyt Regular 1240120 and Exploración 13220097.

		x_k^*	$\|x_k^* - x^{\mathrm{SRD}}\|$	$f(x_k^*)$	time	speedup	nit	nfev	njev
N	k								
100	100	[0.7834 0.7573]	0.364953	-1.540673	0.824 s	2.243x	8	29	8
	200	[0.6857 0.7161]	0.267610	-1.401794	0.797 s	2.322x	9	30	9
	500	[0.6895 0.6481]	0.222597	-1.337666	0.720 s	2.568x	8	26	8
	1000	[0.6905 0.6272]	0.211082	-1.317738	0.789 s	2.345x	8	28	8
	2000	[0.6769 0.6101]	0.190405	-1.287041	0.738 s	2.504x	8	26	8
	5000	[0.6407 0.6232]	0.169176	-1.263909	0.763 s	2.424x	7	22	7
	10000	[0.6441 0.6159]	0.167003	-1.260012	0.931 s	1.986x	9	31	9
	20000	[0.6521 0.5973]	0.162603	-1.249465	0.997 s	1.855x	7	25	7
	50000	[0.6493 0.5938]	0.158281	-1.243036	1.572 s	1.176x	7	22	7
	SRD	[0.6385 0.5905]	0.147470	-1.229051	1.849 s	-	8	25	8
250	100	[0.7878 0.6224]	0.295077	-1.410204	1.372 s	3.408x	7	24	7
	200	[0.7108 0.5613]	0.202447	-1.272122	1.331 s	3.512x	7	23	7
	500	[0.6743 0.5437]	0.162855	-1.217962	1.406 s	3.325x	7	25	7
	1000	[0.6688 0.5279]	0.154843	-1.196643	1.439 s	3.250x	8	26	8
	2000	[0.6653 0.5023]	0.150539	-1.167606	1.552 s	3.013x	8	27	8
	5000	[0.6310 0.5106]	0.116021	-1.141602	1.437 s	3.254x	7	23	7
	10000	[0.6273 0.5083]	0.112328	-1.135550	2.049 s	2.281x	8	32	8
	20000	[0.6180 0.5061]	0.103090	-1.124072	2.008 s	2.328x	7	24	7
	50000	[0.6171 0.5009]	0.102574	-1.117968	3.955 s	1.182x	8	25	8
	SRD	[0.6061 0.4990]	0.091830	-1.105100	4.675 s	-	8	25	8
500	100	[0.6991 0.6773]	0.248792	-1.376424	2.414 s	3.543x	7	25	7
	200	[0.5878 0.6597]	0.166428	-1.247468	2.164 s	3.952x	7	23	7
	500	[0.5890 0.5956]	0.113103	-1.184524	2.519 s	3.395x	8	27	8
	1000	[0.5926 0.5672]	0.096420	-1.159787	2.479 s	3.450x	8	26	8
	2000	[0.6021 0.5321]	0.089941	-1.134260	2.403 s	3.559x	7	24	7
	5000	[0.5665 0.5432]	0.061270	-1.109678	2.566 s	3.333x	8	25	8
	10000	[0.5639 0.5362]	0.055506	-1.100119	3.002 s	2.849x	8	27	8
	20000	[0.5460 0.5455]	0.047203	-1.091594	3.649 s	2.344x	8	25	8
	50000	[0.5497 0.5367]	0.043809	-1.086419	10.483 s	0.816x	8	32	8
	SRD	[0.5447 0.5281]	0.034850	-1.072867	8.552 s	-	7	23	7
1000	100	[0.6891 0.6706]	0.236834	-1.359644	4.256 s	4.498x	7	23	7
	200	[0.5712 0.6569]	0.157243	-1.228063	4.887 s	3.917x	8	26	8
	500	[0.5724 0.5915]	0.099705	-1.163930	4.744 s	4.035x	7	22	7
	1000	[0.5828 0.5540]	0.080833	-1.136799	4.055 s	4.721x	6	20	6
	2000	[0.5881 0.5254]	0.074701	-1.113466	4.762 s	4.020x	7	25	7
	5000	[0.5524 0.5363]	0.045771	-1.088732	5.019 s	3.814x	8	25	8
	10000	[0.5511 0.5272]	0.040014	-1.078288	5.495 s	3.484x	8	26	8
	20000	[0.5311 0.5402]	0.034221	-1.071303	7.338 s	2.609x	8	25	8
	50000	[0.5359 0.5299]	0.028865	-1.065793	15.286 s	1.252x	8	26	8
	SRD	[0.5317 0.5213]	0.020182	-1.052997	19.143 s	-	8	26	8

Table 4: Solutions of the smooth chance-constrained problem with $p = 0.7$.

Declarations

Conflict of interest/Competing interests

There are no conflict of interests

Data availability

The used data can be made available upon request to the authors

Code availability

The computer codes can be made available upon request to the authors.

N	k	x_k^*	$\|x_k^* - x^{\mathrm{SRD}}\|$	$f(x_k^*)$	time	speedup	nit	nfev	njev
100	100	[0.6929 0.5702]	0.470093	-1.263082	0.685 s	2.939x	8	29	8
	200	[0.5103 0.5440]	0.316205	-1.054355	0.624 s	3.223x	8	27	8
	500	[0.5225 0.4662]	0.271085	-0.988707	0.666 s	3.019x	8	27	8
	1000	[0.5162 0.4461]	0.254163	-0.962276	0.626 s	3.216x	8	26	8
	2000	[0.5171 0.4049]	0.234162	-0.922041	0.685 s	2.938x	8	27	8
	5000	[0.4902 0.3893]	0.203102	-0.879493	0.777 s	2.590x	8	27	8
	10000	[0.4790 0.3852]	0.191183	-0.864107	0.848 s	2.374x	9	30	9
	20000	[0.4591 0.3954]	0.178656	-0.854492	1.248 s	1.612x	10	34	10
	50000	[0.4359 0.4097]	0.168081	-0.845554	2.335 s	0.862x	9	30	9
	SRD	[0.4375 0.3859]	0.155325	-0.823441	2.012 s	-	8	26	8
250	100	[0.6415 0.4889]	0.383319	-1.130355	1.228 s	4.493x	7	23	7
	200	[0.5066 0.4241]	0.234138	-0.930671	1.350 s	4.087x	8	26	8
	500	[0.5142 0.3684]	0.217761	-0.882597	1.378 s	4.006x	8	26	8
	1000	[0.4965 0.3468]	0.194969	-0.843346	1.875 s	2.944x	10	34	10
	2000	[0.5000 0.3132]	0.193581	-0.813277	1.791 s	3.081x	8	29	8
	5000	[0.4569 0.3048]	0.150178	-0.761712	1.621 s	3.406x	8	26	8
	10000	[0.4447 0.2971]	0.138022	-0.741787	2.827 s	1.953x	10	34	10
	20000	[0.4109 0.3152]	0.104951	-0.726114	3.137 s	1.760x	10	32	10
	50000	[0.4045 0.3087]	0.097904	-0.713184	5.811 s	0.950x	9	29	9
	SRD	[0.4024 0.2913]	0.096258	-0.693694	5.519 s	-	8	27	8
500	100	[0.5386 0.5523]	0.341129	-1.090902	2.406 s	6.192x	8	27	8
	200	[0.4062 0.5096]	0.230107	-0.915757	2.281 s	6.531x	8	26	8
	500	[0.4302 0.4218]	0.171945	-0.851965	2.277 s	6.544x	8	26	8
	1000	[0.4263 0.3814]	0.143504	-0.807750	2.568 s	5.802x	9	29	9
	2000	[0.4235 0.3501]	0.126193	-0.773580	2.621 s	5.684x	9	29	9
	5000	[0.3813 0.3415]	0.084289	-0.722741	2.806 s	5.311x	9	29	9
	10000	[0.3677 0.3321]	0.067907	-0.699756	3.497 s	4.260x	9	31	9
	20000	[0.3481 0.3427]	0.057935	-0.690756	4.712 s	3.162x	9	31	9
	50000	[0.3459 0.3328]	0.049813	-0.678773	8.716 s	1.709x	9	28	9
	SRD	[0.3373 0.3199]	0.035414	-0.657272	14.899 s	-	8	37	8
1000	100	[0.5221 0.5563]	0.333109	-1.078302	4.614 s	4.590x	8	26	8
	200	[0.4021 0.5009]	0.220489	-0.902979	5.277 s	4.013x	9	30	9
	500	[0.4323 0.4037]	0.161505	-0.835983	4.520 s	4.685x	8	25	8
	1000	[0.4133 0.3726]	0.127712	-0.785819	4.826 s	4.388x	8	26	8
	2000	[0.4077 0.3474]	0.110648	-0.755093	5.132 s	4.127x	8	25	8
	5000	[0.3719 0.3345]	0.072775	-0.706392	7.260 s	2.917x	9	30	9
	10000	[0.3604 0.3211]	0.056928	-0.681524	8.749 s	2.421x	10	32	10
	20000	[0.3374 0.3341]	0.044382	-0.671575	8.427 s	2.513x	9	29	9
	50000	[0.3326 0.3264]	0.035512	-0.659026	18.104 s	1.170x	9	29	9
	SRD	[0.3239 0.3130]	0.020358	-0.636937	21.178 s	-	9	29	9

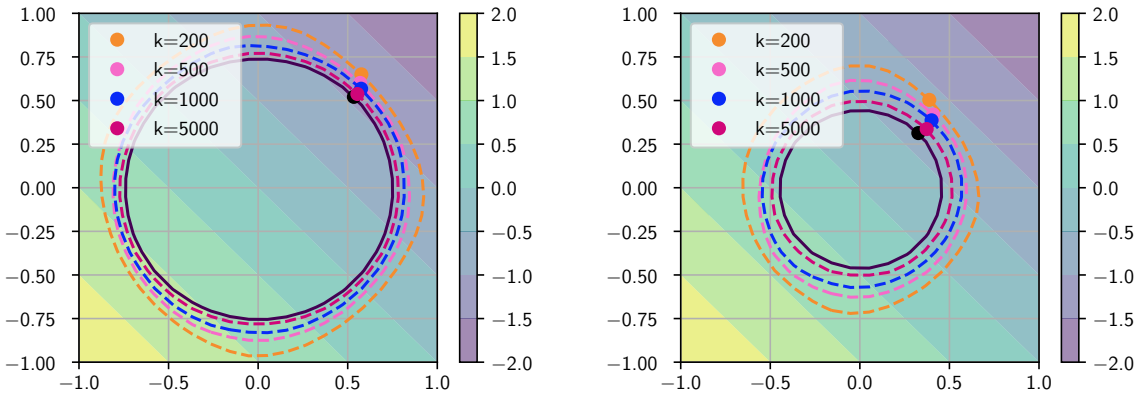
Table 5: Solutions of the smooth chance-constrained problem with $p = 0.75$.

References

[1] González Gradón, T., Heitsch, H., Henrion, R.: A joint model of probabilistic/robust constraints for gas transport management in stationary networks. Computational Management Science **14**, 443–460 (2017) https://doi.org/10.1007/s10287-017-0284-7

[2] Berthold, H., Heitsch, H., Henrion, R., Schwientek, J.: On the algorithmic solution of optimization problems subject to probabilistic/robust (probust) constraints. Math. Methods Oper. Res. **96**(1), 1–37 (2022) https://doi.org/10.1007/s00186-021-00764-8

[3] Adelhütte, D., Aßmann, D., Gonzàlez-Gradòn, T., Gugat, M., Heitsch, H., Henrion, R., Liers, F., Nitsche, S., Schultz, R., Stingl, M., Wintergerst, D.: Joint model of probabilistic (probust) constraints with application to gas network optimization. Vietnam Journal of Mathematics **49**(4), 1097–1130 (2021) https://doi.org/10.1007/s10013-020-00434-y



(a) $p = 0.7$ (b) $p = 0.75$

Fig. 3: Graphical comparison of the solution to the chance-constrained problem
The figure shows the obtained solutions to the chance-constrained problem using the exact probability constraint φ (black dot) and its smooth approximations φ_k, computed using the spherical radial decomposition method for various values of k. Subplots (a) and (b) correspond to threshold levels $p = 0.7$ and $p = 0.75$, respectively. The plots illustrate the convergence of the feasible sets and solutions as the number of cutting planes increases.

[4] van Ackooij, W., Henrion, R., Pérez-Aros, P.: Generalized gradients for probabilistic/robust (probust) constraints. Optimization **69**(7-8), 1451–1479 (2020) https://doi.org/10.1080/02331934.2019.1576670

[5] Geiersbach, C., Henrion, R., Pérez-Aros, P.: Numerical solution of an optimal control problem with probabilistic and almost sure state constraints (2023)

[6] Geiersbach, C., Henrion, R.: Optimality conditions in control problems with random state constraints in probabilistic or almost-sure form (2024)

[7] van Ackooij, W., Henrion, R., Zidani, H.: Pontryagin's principle for some probabilistic control problems. Applied Mathematics and Optimization **90**, 1–36 (2024) https://doi.org/10.1007/s00245-024-10151-4

[8] van Ackooij, W., de Oliveira, W.: Methods of Nonsmooth Optimization in Stochastic Programming. International Series in Operations Research & Management Science, p. 574. Springer, ??? (2025). https://doi.org/10.1007/978-3-031-84836-0 . https://link.springer.com/book/9783031848360

[9] Bremer, I., Henrion, R., Möller, A.: Probabilistic constraints via SQP solver: Application to a renewable energy management problem. Computational Management Science **12**, 435–459 (2015)

[10] van Ackooij, W., Sagastizábal, C.: Constrained bundle methods for upper inexact oracles with application to joint chance constrained energy problems. SIAM Journal on Optimization **24**(2), 733–765 (2014)

[11] van Ackooij, W., Finardi, E.C., Matiussi Ramalho, G.: An exact solution method for the hydrothermal unit commitment under wind power uncertainty with joint probability constraints. IEEE Transactions on Power Systems **33**(6), 6487–6500 (2018) https://doi.org/10.1109/TPWRS.2018.2848594

[12] Royset, J.O., Polak, E.: Extensions of stochastic optimization results to problems with system failure

probability functions. Journal of Optimization Theory and Applications **133**(1), 1–18 (2007) https://doi.org/10.1007/s10957-007-9178-0

[13] Garnier, J., Omrane, A., Rouchdy, Y.: Asymptotic formulas for the derivatives of probability functions and their Monte Carlo estimations. European Journal of Operations Research **198**, 848–858 (2009) https://doi.org/10.1016/j.ejor.2008.09.026

[14] Kibzun, A.I., Uryas'ev, S.: Differentiability of probability function. Stoch. Anal. Appl. **16**, 1101–1128 (1998) https://doi.org/10.1080/07362999808809581

[15] Marti, K.: Differentiation of probability functions : The transformation method. Computers and Mathematics with Applications **30**, 361–382 (1995) https://doi.org/10.1016/0898-1221(95)00113-1

[16] Pflug, G., Weisshaupt, H.: Probability gradient estimation by set-valued calculus and applications in network design. SIAM J. Optim. **15**, 898–914 (2005) https://doi.org/10.1137/S1052623403431639

[17] Raik, E.: The differentiability in the parameter of the probability function and optimization of the probability function via the stochastic pseudogradient method (russian). Izvestiya Akad. Nayk Est. SSR, Phis. Math. **24**(1), 3–6 (1975)

[18] Royset, J.O., Polak, E.: Implementable algorithm for stochastic optimization using sample average approximations. Journal of Optimization Theory and Applications **122**(1), 157–184 (2004) https://doi.org/10.1023/B:JOTA.0000041734.06199.71

[19] Uryas'ev, S.: Derivatives of probability functions and integrals over sets given by inequalities. Journal of Computational and Applied Mathematics **56**(1-2), 197–223 (1994) https://doi.org/10.1016/0377-0427(94)90388-3

[20] Uryas'ev, S.: Derivatives of probability functions and some applications. Annals of Operations Research **56**, 287–311 (1995) https://doi.org/10.1007/BF02031712

[21] Hong, L.J., Yang, Y., Zhang, L.: Sequential convex approximations to joint chance constrained programed: A monte carlo approach. Operations Research **3**(59), 617–630 (2011)

[22] Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. SIAM Journal of Optimization **17**(4), 969–996 (2006)

[23] Geletu, A., Hoffmann, A., Klöppel, M., Li, P.: An inner-outer approximation approach to chance constrained optimization. SIAM Journal on Optimization **27**(3), 1834–1857 (2017)

[24] Shan, F., Zhang, L.W., Xiao, X.: A smoothing function approach to joint chance constrained programs. J. Optim Theory Appl **163**, 181–199 (2014)

[25] Peña-Ordieres, A., Luedtke, J.R., Wächter, A.: Solving chance-constrained problems via a smooth sample-based nonlinear approximation. SIAM Journal on Optimization **30**(3), 2221–2250 (2020)

[26] Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis, 3rd edn. Grundlehren der mathematischen Wissenschaften, vol. 317, p. 734. Springer, Berlin (2009). https://doi.org/10.1007/978-3-642-02431-3

[27] Clarke, F.H.: Optimization and Nonsmooth Analysis. Classics in Applied Mathematics, p. 320. Society for Industrial and Applied Mathematics, Philadelphia (1987). https://doi.org/10.1137/1.9781611971309

[28] Burke, J.V., Hoheisel, T., Kanzow, C.: Gradient consistency for integral-convolution smoothing

functions. Set-Valued Var. Anal. **21**(2), 359–376 (2013) https://doi.org/10.1007/s11228-013-0235-6

[29] Chen, X.: Smoothing methods for nonsmooth, nonconvex minimization. Math. Program. **134**(1), 71–99 (2012) https://doi.org/10.1007/s10107-012-0569-0

[30] Borges, P., Sagastizábal, C., Solodov, M.: A regularized smoothing method for fully parameterized convex problems with applications to convex and nonconvex two-stage stochastic programming. Math. Program. **189**(1-2), 117–149 (2021) https://doi.org/10.1007/s10107-020-01582-2

[31] Pérez-Aros, P.: Ergodic approach to robust optimization and infinite programming problems. Set-Valued Var. Anal. **29**(2), 409–423 (2021) https://doi.org/10.1007/s11228-020-00567-9

[32] van Ackooij, W.: A discussion of probability functions and constraints from a variational perspective. Set-Valued and Variational Analysis **28**(4), 585–609 (2020) https://doi.org/10.1007/s11228-020-00552-2

[33] Guo, S., Xu, H., Zhang, L.: Convergence analysis for mathematical programs with distributionally robust chance constraint. SIAM Journal on Optimization **27**(2), 784–816 (2017) https://doi.org/10.1137/15M1036592

[34] Römisch, W., Schultz, R.: Stability analysis for stochastic programs. Annals of Operations Research **30**, 241–266 (1991)

[35] Henrion, R.: Optimierungsprobleme mit wahrscheinlichkeitsrestriktionen: Modelle, struktur, numerik. Lecture Notes, 1–53 (2016)

[36] van Ackooij, W., Pérez-Aros, P.: Generalized differentiation of probability functions: parameter dependent sets given by intersections of convex sets and complements of convex sets. Applied Mathematics & Optimization **85**(2), 1–39 (2022)

[37] Lang, R.: A note on the measurability of convex sets. Archiv des Mathematik **47**(1), 90–92 (1986)

[38] van Ackooij, W., Pérez-Aros, P.: Generalized differentiation of probability functions acting on an infinite system of constraints. SIAM Journal on Optimization **29**(3), 2179–2210 (2019)

[39] van Ackooij, W., Malick, J.: Decomposition algorithm for large-scale two-stage unit-commitment. Annals of Operations Research **238**(1), 587–613 (2016) https://doi.org/10.1007/s10479-015-2029-8

[40] van Ackooij, W., Henrion, R.: Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions. SIAM Journal on Optimization **24**(4), 1864–1889 (2014) https://doi.org/10.1137/130922689

[41] Aubin, J.-P., Frankowska, H.: Set-valued Analysis. Modern Birkhäuser Classics, p. 461. Birkhäuser Boston, Inc., Boston, MA, Boston (2009). https://doi.org/10.1007/978-0-8176-4848-0 . Reprint of the 1990 edition [MR1048347]. https://doi.org/10.1007/978-0-8176-4848-0

[42] Mordukhovich, B.S.: Variational Analysis and Applications, p. 622. Springer, Swiss (2018). https://doi.org/10.1007/978-3-319-92775-6

[43] van Ackooij, W., Henrion, R.: (Sub-) Gradient formulae for probability functions of random inequality systems under Gaussian distribution. SIAM Journal on Uncertainty Quantification **5**(1), 63–87 (2017) https://doi.org/10.1137/16M1061308

[44] van Ackooij, W., Javal, P., Pérez-Aros, P.: Derivatives of probability functions acting on parameter

dependent unions of polyhedra. Set-Valued and Variational Analysis **30**(2), 487–519 (2022) https://doi.org/10.1007/s11228-021-00598-w

[45] van Ackooij, W., Aleksovska, I., Zuniga, M.M.: (sub-)differentiability of probability functions with elliptical distributions. Set Valued and Variational Analysis **26**(4), 887–910 (2018) https://doi.org/10.1007/s11228-017-0454-3

[46] Balder, E.J., Sambucini, A.R.: Fatou's lemma for multifunctions with unbounded values in a dual space. J. Convex Anal. **12**(2), 383–395 (2005)

[47] Attouch, H.: Variational Convergence for Functions and Operators. Applicable Mathematics Series, p. 423. Pitman (Advanced Publishing Program), Boston, MA, ??? (1984)

[48] Prékopa, A.: Logarithmic concave measures with applications to stochastic programming. Acta Scientiarium Mathematicarum (Szeged) **32**, 301–316 (1971)

[49] Henrion, R., Strugarek, C.: Convexity of chance constraints with independent random variables. Computational Optimization and Applications **41**, 263–276 (2008)

[50] van Ackooij, W.: Eventual convexity of chance constrained feasible sets. Optimization (A Journal of Math. Programming and Operations Research) **64**(5), 1263–1284 (2015) https://doi.org/10.1080/02331934.2013.855211

[51] van Ackooij, W., Malick, J.: Eventual convexity of probability constraints with elliptical distributions. Mathematical Programming **175**(1), 1–27 (2019) https://doi.org/10.1007/s10107-018-1230-3

[52] Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming. Modeling and Theory. MPS-SIAM series on optimization, vol. 9. SIAM and MPS, Philadelphia, ??? (2009). https://doi.org/10.1137/1.9780898718751

[53] Prékopa, A.: Stochastic Programming. Kluwer, Dordrecht, ??? (1995). https://doi.org/10.1007/978-94-017-3087-7

[54] Henrion, R., Strugarek, C.: Convexity of chance constraints with dependent random variables: the use of copulae. In: Bertocchi, M., Consigli, G., Dempster, M.A.H. (eds.) Stochastic Optimization Methods in Finance and Energy: New Financial Products and Energy Market Strategies. International Series in Operations Research and Management Science, pp. 427–439. Springer, ??? (2011)

[55] Laguel, Y., van Ackooij, W., Malick, J., Matiussi Ramalho, G.: On the convexity of level-sets of probability functions. Journal of Convex Analysis **29**(2), 1–32 (2022)

[56] Heitsch, H.: On probabilistic capacity maximization in a stationary gas network. Optimization **69**(3), 575–604 (2020) https://doi.org/10.1080/02331934.2019.1625353

[57] Shan, F., Xiao, X.T., Zhang, L.W.: Convergence analysis on a smoothing approach to joint chance constrained programs. Optimization **65**(12), 2171–2193 (2016)