SIAM J. OPTIM.
Vol. 24, No. 4, pp. 1864–1889

GRADIENT FORMULAE FOR NONLINEAR PROBABILISTIC CONSTRAINTS WITH GAUSSIAN AND GAUSSIAN-LIKE DISTRIBUTIONS*

WIM VAN ACKOOIJ† AND RENÉ HENRION‡

Abstract. Probabilistic constraints represent a major model of stochastic optimization. A possible approach for solving probabilistically constrained optimization problems consists in applying nonlinear programming methods. To do so, one has to provide sufficiently precise approximations for values and gradients of probability functions. For linear probabilistic constraints under Gaussian distribution this can be done successfully by analytically reducing these values and gradients to values of Gaussian distribution functions and computing the latter, for instance, by Genz's code. For nonlinear models one may fall back on the spherical-radial decomposition of Gaussian random vectors and apply, for instance, Deák's sampling scheme for the uniform distribution on the sphere in order to compute values of corresponding probability functions. The present paper demonstrates how the same sampling scheme can be used to simultaneously compute gradients of these probability functions. More precisely, we prove a formula representing these gradients in the Gaussian case as a certain integral over the sphere again. The result is also extended to alternative distributions with an emphasis on the multivariate Student's (or t-) distribution.

Key words. stochastic optimization, probabilistic constraints, chance constraints, gradients of probability functions

AMS subject classifications. 49M37, 65K05, 90C15

DOI. 10.1137/130922689

1. Introduction.

A probabilistic constraint is an inequality of the type

$$\mathbb{P}\left(g(x,\xi) \leq 0\right) \geq p, \tag{1.1}$$

where g is a mapping defining a (random) inequality system and ξ is an s-dimensional random vector defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The constraint (1.1) expresses the fact that a decision vector x is feasible if and only if the random inequality system $g(x,\xi) \leq 0$ is satisfied with probability at least $p \in [0,1]$. Probabilistic constraints are important for engineering problems involving uncertain data. Applications can be found in water management, telecommunications, electricity network expansion, mineral blending, chemical engineering, etc. For a comprehensive overview of the theory, numerics, and applications of probabilistic constraints, we refer the reader to, e.g., [28, 29, 31]. Initiated by Charnes and Cooper [6] and pioneered by Prékopa (e.g., by his celebrated log-concavity theorem [27]), the analysis of probabilistic constraints has attracted much attention in recent years with a focus on algorithmic approaches. Without providing an exhaustive list here, we refer the reader to models such as robust optimization [2], a penalty approach [12], p-efficient points [9, 10], scenario approximation [5], sample average approximation [25], and convex approximation [23].

*Received by the editors May 28, 2013; accepted for publication (in revised form) July 29, 2014; published electronically October 30, 2014.
http://www.siam.org/journals/siopt/24-4/92268.html
†EDF R&D, OSIRIS, F-92141 Clamart Cedex, France, and Ecole Centrale Paris, F-92295 Châtenay-Malabry, France (wim.van-ackooij@edf.fr).
‡Weierstrass Institute Berlin, 10117 Berlin, Germany (henrion@wias-berlin.de). This author's work was supported by the DFG Research Center MATHEON "Mathematics for Key Technologies" in Berlin.

The present paper is motivated by the traditional nonlinear programming approach to the solution of probabilistically constrained optimization problems: from a formal viewpoint, (1.1) is a conventional inequality constraint $\varphi(x) \geq p$ with $\varphi(x) := \mathbb{P}(g(x,\xi) \leq 0)$. On the other hand, a major difficulty arises from the fact that, typically, no analytical expression is available for φ. All one can hope for, in general, are tools for numerically approximating φ. Beyond crude Monte Carlo estimation of the probability defining φ, there exist many more efficient approaches based, for instance, on graph-theoretical arguments [4], variance reduction [33], quasi–Monte Carlo (QMC) techniques, or sparse grid numerical integration [14]. It seems, however, that such approaches are most successful when exploiting the special model structure (i.e., the mapping g and the distribution of ξ). For instance, in the special case of separable constraints $g(x,\xi) = \xi - x$, and of ξ having a regular Gaussian distribution (such that φ reduces to a multivariate Gaussian distribution function), one may employ an efficient code by Genz [15, 16] which is based on a numerical integration scheme combining separation and reordering of variables with randomized QMC. A similar technique has been proposed for the multivariate Student's (or t-) distribution [16]. The numerical evaluation of other multivariate distribution functions such as Gamma or exponential distribution has been discussed, e.g., in [24, 32].

For an efficient solution of probabilistically constrained problems via numerical nonlinear optimization, it is evidently not sufficient to calculate just functional values of φ; one also has to have access to gradients of φ. The need to calculate gradients of probability functions has been recognized for a long time and has given rise to many papers on representing such gradients (e.g., [13, 19, 21, 26, 34]). In the separable case with Gaussian distribution mentioned above, it is well known [28, p. 203] that partial derivatives of φ can be reduced analytically to function values $\tilde{\varphi}$ of another Gaussian distribution with modified parameters. This allows one to employ the same efficient method (e.g., by Genz) available for values of Gaussian distribution functions in order to compute gradients simultaneously and to control the error for calculating $\nabla\varphi$ and φ simultaneously [17]. Interestingly, this special circumstance can be extended to more general models: it has been demonstrated in [18, 37, 38] that for general linear probabilistic constraints $\varphi(x) := \mathbb{P}(T(x)\xi \leq a(x)) \geq p$ under Gaussian distribution and with possibly singular matrix $T(x)$, the computation not only of φ (which is evident) but also of $\nabla\varphi$ can be analytically reduced to the computation of Gaussian distribution functions. Combining appropriately these ideas with Genz's code and an SQP solver, it is possible to solve corresponding optimization problems for Gaussian random vectors in dimension of up to a few hundred (where the dimension of the decision vector x is less influential). Applications to various problems of power management can be found, e.g., in [1, 18, 37, 38, 39].

When considering models which are nonlinear in ξ, a reduction to distribution functions is no longer possible. In this case, another approach, the so-called spherical-radial decomposition of Gaussian random vectors (see, e.g., [16]) appears to be promising for calculating both function values and gradients of φ. More precisely, let ξ be an m-dimensional random vector normally distributed according to $\xi \sim \mathcal{N}(0,R)$ for some correlation matrix R. Then, $\xi = \eta L\zeta$, where $R = LL^T$ is the Cholesky decomposition of R, η has a chi-distribution with m degrees of freedom, and ζ has a uniform distribution over the Euclidean unit sphere

$$\mathbb{S}^{m-1} := \left\{ z \in \mathbb{R}^m \,\middle|\, \sum_{i=1}^{m} z_i^2 = 1 \right\}$$

of $\mathbb{R}^m$. As a consequence, for any Lebesgue measurable set $M \subseteq \mathbb{R}^m$, its probability may be represented as

$$\mathbb{P}\left(\xi \in M\right) = \int_{v \in \mathbb{S}^{m-1}} \mu_\eta\left(\{r \geq 0 : rLv \cap M \neq \emptyset\}\right) d\mu_\zeta, \tag{1.2}$$

where μ_η and μ_ζ are the laws of η and ζ, respectively. More generally, one may approximate the integral

$$\int_{v \in \mathbb{S}^{m-1}} h(v) d\mu_\zeta \tag{1.3}$$

for any Lebesgue measurable function $h : \mathbb{S}^{m-1} \to \mathbb{R}$. For the special case $h(v) := \mu_\eta\left(\{r \geq 0 : rLv \cap M \neq \emptyset\}\right)$, we obtain the probability (1.2), which can be numerically computed by combining efficient sampling schemes on $\mathbb{S}^{m-1}$ with fast computation of the one-dimensional sets in braces for each fixed sampling point $v \in \mathbb{S}^{m-1}$. In this paper, we will show how—with another function $h(v)$—the same (efficient) sampling scheme of the unit sphere can be employed to simultaneously compute values and derivatives of the probability (1.2) with respect to an exterior parameter playing the role of a decision vector x in a probabilistic constraint.

To illustrate the potential advantage of spherical-radial decomposition (1.2) over a crude Monte Carlo approximation of $\mathbb{P}\left(\xi \in M\right)$, assume that there exists a mapping ρ such that

$$\{r \geq 0 : rLv \cap M \neq \emptyset\} = [0, \rho(v)] \quad \forall v \in \mathbb{S}^{m-1}. \tag{1.4}$$

Then, by (1.2), we may write the desired probability as an expected value of two different random variables:

$$p := \mathbb{P}\left(\xi \in M\right) = \mathbb{E}_\xi(\mathbb{I}_M(\xi)) = \mathbb{E}_\zeta(\mu_\eta([0, \rho(\zeta)])),$$

where $\mathbb{I}_A(z) = 1$ if $z \in A$ and $\mathbb{I}_A(z) = 0$ otherwise. This leaves us with two possibilities for approximating p: first, one may empirically approximate $\mathbb{I}_M(\xi)$ by sampling ξ (crude Monte Carlo), and, second, one may empirically approximate $\mu_\eta([0, \rho(\zeta)])$ by sampling ζ on the sphere. Now one can show (see the appendix) that

$$\mathrm{Var}_\zeta\left(\mu_\eta([0, \rho(\zeta)])\right) \leq \mathrm{Var}_\xi\left(\mathbb{I}_M(\xi)\right), \tag{1.5}$$

indicating that sampling by means of the spherical-radial decomposition should lead to a possibly strong reduction of variance compared to crude Monte Carlo sampling. For an early proposal of sampling the uniform distribution on the sphere, we refer the reader to Deák [7, 8]. Recently, much progress has been made toward optimal QMC sampling on the unit sphere, e.g., [3] (or on a Cartesian product of unit spheres [20]), which gives hope for even more efficient procedures. In [3], for instance, certain N-point configurations on the sphere are identified which belong to the class of so-called QMC designs and as such allow one to derive powerful estimates of the approximation error in terms of the sample size. Based on such efficient sampling schemes for the unit sphere and applying them simultaneously to values and gradients of probability functions, the results of this paper may serve as a basis for a numerical treatment of nonlinear convex probabilistic constraints with Gaussian and alternative distributions via nonlinear optimization.

In this paper we will consider probability functions $\varphi(x) := \mathbb{P}(g(x,\xi) \le 0)$, where ξ has a multivariate Gaussian distribution and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a continuously differentiable function convex in its second argument. As mentioned above, the primary goal of this paper is to provide a formula for $\nabla\varphi$ which allows one to make use of sampling schemes for the unit sphere in order to compute φ and $\nabla\varphi$ simultaneously. One may be tempted to simply differentiate with respect to x under the integral representing φ in terms of the spherical-radial decomposition (1.2) of ξ. A corresponding formula can be found, without justification, however, in [11, section 9.2]. A proof of this formula can be found in [30] under the implicit assumption that the set $\{\xi \mid g(\bar{x},\xi) \le 0\}$ is bounded at the point $\bar{x}$ of interest. Boundedness is also the key assumption for many general gradient formulae of probability functions (e.g., [26, 34, 35, 36]). On the other hand, boundedness is a very restrictive assumption in probabilistic programming. Even in the simple case of mappings $g(x,\xi) := \xi - x \le 0$, corresponding to the evaluation of multivariate distribution functions, it is violated. In order to illustrate the potential failure of differentiability of probability functions in the absence of boundedness, we start the paper with a short section presenting a counterexample suited to the framework considered here. In this example, all input data are nice (ξ has a regular Gaussian distribution, g is smooth and additionally convex in its second argument, and the inequality defined by g satisfies the Slater condition or, equivalently, the linear independence constraint qualification), yet the probability function fails to be differentiable due to the absence of compactness. In order to guarantee differentiability without boundedness, one usually has to provide additional arguments allowing for an application of Lebesgue's dominated convergence theorem. These frequently nonevident arguments are typically derived from the specific structure of a probability function (distribution of ξ, structure of the mapping g). Not surprisingly, finding these arguments in the context of our paper requires major effort. As a result, we present not just a formal technical but a *verifiable* condition ensuring differentiability also in the unbounded case. The validity of this condition is checked for several significant examples. In particular, a generalization of the obtained result to non-Gaussian distributions (such as χ^2 or multivariate Student's) is provided by exploiting the fact that certain important multivariate distributions can be led back to the Gaussian one by a nonlinear transformation not affecting the original structure imposed on our data.

2. Potential nondifferentiability of probability functions with nice data in the absence of boundedness.

We start this section by presenting a minimum requirement on the problem data in the spirit of this paper, ensuring at least continuity of the resulting probability function.

PROPOSITION 2.1. *Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be continuous and convex in the second argument. Let $\xi \sim \mathcal{N}(0,R)$ for some correlation matrix R. Define $\varphi(x) := \mathbb{P}(g(x,\xi) \le 0)$. Let $\bar{x} \in \mathbb{R}^n$ be arbitrary. If $g(\bar{x},0) < 0$, then φ is continuous in $\bar{x}$.*

Proof. Due to the existence of a Slater point, the set $\{z \in \mathbb{R}^m | g(\bar{x},z) = 0\}$ has Lebesgue measure zero as the boundary of the convex set $\{z \in \mathbb{R}^m | g(\bar{x},z) \le 0\}$. Since ξ has a density, it follows that $\mathbb{P}(g(\bar{x},\xi) = 0) = 0$. Along with the continuity of g, this entails the continuity of φ. ∎

In order to derive the differentiability of φ, it is not sufficient to add continuous differentiability of g to the assumptions of Proposition 2.1, as is shown by the following example. It follows that differentiability for φ requires the additional assumption of compactness of the set $\{z \in \mathbb{R}^m | g(\bar{x},z) \le 0\}$ or, in case one wants to admit noncompact sets, a kind of growth condition for the function $\|\nabla_x g(x,\cdot)\|$ in a neighborhood of $\bar{x}$, as will be presented in this paper.

PROPOSITION 2.2. *Let $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ be defined by*

$$g(x_1, x_2, z_1, z_2) := x_1^2 e^{h(z_1)} + x_2 z_2 - 1, \quad \text{where} \quad h(t) := -1 - 2\log(1 - \Phi(t))$$

and Φ is the cumulative distribution function of the one-dimensional standard Gaussian distribution. Let $\xi \sim \mathcal{N}(0, I_2)$ and $\bar{x} = (0, 1)$. Then, the following hold true:

1. *g is continuously differentiable.*
2. *g is convex in the second argument.*
3. *$g(\bar{x}, 0) = g(0, 1, 0, 0) < 0$.*
4. *φ is not differentiable at $\bar{x}$.*

Proof. Statement 1 is evident from Φ being smooth and satisfying $\Phi < 1$. The Slater condition, i.e., statement 3, is evident as well. As for statement 2, note first that $1 - \Phi$ is a log-concave function [28, Theorem 4.2.4] which implies that $-\log(1 - \Phi)$ and, hence, h are convex functions. As a composition with the exponential as an outer function—which is convex and increasing—e^h is convex too. It follows that g is convex in the second argument.

In order to show nondifferentiability of φ at $\bar{x}$, observe first that the partial function $\varphi(\cdot, 1)$ attains its global maximum at 0. Indeed, the implication

$$g(t, 1, z_1, z_2) \le 0 \Longrightarrow g(0, 1, z_1, z_2) \le 0,$$

which is valid for all t, z_1, z_2, yields that

$$\varphi(t, 1) = \mathbb{P}(g(t, 1, \xi_1, \xi_2) \le 0) \le \mathbb{P}(g(0, 1, \xi_1, \xi_2) \le 0) = \varphi(0, 1)$$

for all t. Now, if φ was differentiable at $\bar{x}$, then $\varphi'(\cdot, 1) = 0$. We will show that there exists some $\varepsilon > 0$ with

$$\varphi(0, 1) - \varphi(t, 1) \ge \varepsilon t \quad \forall t \in (0, 1), \tag{2.1}$$

which clearly contradicts $\varphi'(\cdot, 1) = 0$ and thus differentiability of φ. By definition of φ and g and by the assumption of $\xi \sim \mathcal{N}(0, I_2)$ (implying $\xi_1, \xi_2 \sim \mathcal{N}(0, 1)$), we have that

$$\begin{aligned}
\varphi(0, 1) - \varphi(t, 1) &= \mathbb{P}(\xi_2 \le 1) - \mathbb{P}\left(\xi_2 \le 1 - t^2 e^{h(\xi_1)}\right) \\
&= \Phi(1) - \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{1 - t^2 e^{h(z_1)}} e^{-(z_1^2 + z_2^2)/2} dz_2 \right) dz_1 \\
&= \Phi(1) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z_1^2/2} \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{1 - t^2 e^{h(z_1)}} e^{-z_2^2/2} dz_2 \right) dz_1 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z_1^2/2} \left(\Phi(1) - \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{1 - t^2 e^{h(z_1)}} e^{-z_2^2/2} dz_2 \right) \right) dz_1 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z_1^2/2} \left(\Phi(1) - \Phi\left(1 - t^2 e^{h(z_1)}\right) \right) dz_1.
\end{aligned}$$

Here we used that $e^{-u^2/2}/\sqrt{2\pi}$ is the density of both ξ_1 and ξ_2, and, hence,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du = 1, \quad \Phi(\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tau} e^{-u^2/2} du. \tag{2.2}$$

Now fix an arbitrary $t \in (0,1)$. Then the quantile $\Phi^{-1}(1-t)$ is well defined, and since h is an increasing function (because Φ is so), we may infer that for all $z_1 \geq \Phi^{-1}(1-t)$ the following holds true:

$$h(z_1) \geq h\left(\Phi^{-1}(1-t)\right) = -1 - 2\log(1 - \Phi(\Phi^{-1}(1-t))) = -1 - 2\log t.$$

From here we derive that $t^2 e^{h(z_1)} \geq e^{-1}$ for all $z_1 \geq \Phi^{-1}(1-t)$, which leads, along with Φ being increasing, to

$$\Phi(1) - \Phi\left(1 - t^2 e^{h(z_1)}\right) \geq \Phi(1) - \Phi\left(1 - e^{-1}\right) =: \varepsilon > 0 \quad \forall z_1 \geq \Phi^{-1}(1-t).$$

Given that $\Phi(1) \geq \Phi(1 - t^2 e^{h(z_1)})$ for any z_1, we arrive at the following estimation:

$$\begin{aligned}
\varphi(0,1) - \varphi(t,1) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z_1^2/2} \left(\Phi(1) - \Phi\left(1 - t^2 e^{h(z_1)}\right)\right) dz_1 \\
&\geq \frac{1}{\sqrt{2\pi}} \int_{\Phi^{-1}(1-t)}^{\infty} e^{-z_1^2/2} \left(\Phi(1) - \Phi\left(1 - t^2 e^{h(z_1)}\right)\right) dz_1 \\
&\geq \varepsilon \frac{1}{\sqrt{2\pi}} \int_{\Phi^{-1}(1-t)}^{\infty} e^{-z_1^2/2} dz_1 \\
&= \varepsilon \left(1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\Phi^{-1}(1-t)} e^{-z_1^2/2} dz_1\right) \\
&= \varepsilon \left(1 - \Phi\left(\Phi^{-1}(1-t)\right)\right) \\
&= \varepsilon t.
\end{aligned}$$

Since $t \in (0,1)$ was arbitrary, we have proved (2.1) and thus the nondifferentiability of φ at $\bar{x}$. ∎

Figure 1 illustrates the graph of the nondifferentiable probability function φ constructed in Proposition 2.2.

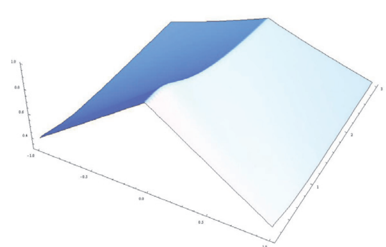

FIG. 1. *Graph of a nondifferentiable probability function.*

3. A gradient formula for parameter-dependent Gaussian probabilities in the convex case. In the following, we assume that $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a continuously differentiable function which is convex with respect to the second argument. We define

$$\varphi(x) := \mathbb{P}(g(x,\xi) \leq 0), \tag{3.1}$$

where $\xi \sim \mathcal{N}(0, R)$.

Remark 3.1. We recall that convex sets are Lebesgue measurable so that the probabilities in (3.1) are well defined by virtue of ξ having a density.

Remark 3.2. If ξ has a general nondegenerate Gaussian distribution, i.e., $\xi \sim \mathcal{N}(\mu, \Sigma)$ for some mean vector $\mu \in \mathbb{R}^m$ and some positive definite covariance matrix Σ of order (m, m), then one may define $\tilde{\xi} := D(\xi - \mu)$, where D is the diagonal matrix with elements $\Sigma_{ii}^{-1/2}$. Then, clearly, $\tilde{\xi} \sim \mathcal{N}(0, R)$, where R is the correlation matrix associated with Σ. Defining $\tilde{g} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ as

$$\tilde{g}(x, z) := g\left(x, D^{-1}z + \mu\right),$$

(3.1) can be rewritten as

$$\varphi(x) = \mathbb{P}\left(\tilde{g}(x, \tilde{\xi}) \leq 0\right),$$

where $\tilde{g}$ has the same properties as g (it is continuously differentiable and convex with respect to the second argument). Therefore, in (3.1), we may indeed assume without loss of generality that $\xi \sim \mathcal{N}(0, R)$.

By (1.2) and (3.1), we have, for all $x \in \mathbb{R}^n$, that

$$\varphi(x) = \int_{v \in \mathbb{S}^{m-1}} \mu_\eta\left(\{r \geq 0 : g(x, rLv) \leq 0\}\right) d\mu_\zeta = \int_{v \in \mathbb{S}^{m-1}} e(x, v) d\mu_\zeta \tag{3.2}$$

for

$$e(x, v) := \mu_\eta\left(\{r \geq 0 : g(x, rLv) \leq 0\}\right) \quad \forall x \in \mathbb{R}^n \ \forall v \in \mathbb{S}^{m-1}. \tag{3.3}$$

According to the possibility of evaluating (1.3) by Deàk's method, for instance, we can obtain a value $\varphi(x)$ for each fixed x. We now address the computation of $\nabla\varphi$. It is convenient to introduce the following two mappings $F, I : \mathbb{R}^n \rightrightarrows \mathbb{S}^{m-1}$ of directions with **f**inite and **i**nfinite intersection length:

$$F(x) := \left\{v \in \mathbb{S}^{m-1} | \exists r > 0 : g(x, rLv) = 0\right\},$$
$$I(x) := \left\{v \in \mathbb{S}^{m-1} | \forall r > 0 : g(x, rLv) \neq 0\right\}.$$

The following lemma collects some elementary properties that will be needed later.

LEMMA 3.1. *Let $x \in \mathbb{R}^n$ be such that $g(x, 0) < 0$. Then the following hold:*

1. *$v \in I(x)$ if and only if $g(x, rLv) < 0$ for all $r > 0$.*
2. *$F(x) \cup I(x) = \mathbb{S}^{m-1}$.*
3. *For $v \in F(x)$, let $r > 0$ be such that $g(x, rLv) = 0$. Then,*

$$\langle \nabla_z g(x, rLv), Lv \rangle \geq -\frac{g(x, 0)}{r}.$$

4. *If $v \in I(x)$, then $e(x, v) = 1$, where e is defined in* (3.3).

Proof. Statement 1 follows from the continuity of g, and statement 2 is evident from the definitions. The convexity of g with respect to the second argument yields

$$-\frac{1}{2}r\langle \nabla_z g(x, rLv), Lv \rangle = \left\langle \nabla_z g(x, rLv), \frac{1}{2}rLv - rLv \right\rangle \leq g\left(x, \frac{1}{2}rLv\right) - g(x, rLv)$$
$$= g\left(x, \frac{1}{2}rLv\right) \leq \frac{1}{2}g(x, 0) + \frac{1}{2}g(x, rLv) = \frac{1}{2}g(x, 0).$$

This proves statement 3. If $v \in I(x)$, then $e(x,v) = \mu_\eta(\mathbb{R}_+) = 1$ because $\mathbb{R}_+$ is the support of the chi-distribution. Therefore, statement 4 holds true. ∎

Next, we provide a local representation of the factor r as a function of x and v.

LEMMA 3.2. *Let (x,v) be such that $g(x,0) < 0$ and $v \in F(x)$. Then, there exist neighborhoods U of x and V of v as well as a continuously differentiable function $\rho^{x,v} : U \times V \to \mathbb{R}_+$ with the following properties:*

1. *For all $(x',v',r') \in U \times V \times \mathbb{R}_+$, the equivalence $g(x',r'Lv') = 0 \Leftrightarrow r' = \rho^{x,v}(x',v')$ holds true.*
2. *For all $(x',v') \in U \times V$ one has the gradient formula*

$$\nabla_x \rho^{x,v}(x',v') = -\frac{1}{\langle \nabla_z g(x',\rho^{x,v}(x',v')Lv'), Lv' \rangle} \nabla_x g(x',\rho^{x,v}(x',v')Lv').$$

Proof. By definition of $F(x)$ we have that $g(x,rLv) = 0$ for some $r > 0$. According to statement 3 in Lemma 3.1, we have that

$$\langle \nabla_z g(x,rLv), Lv \rangle \geq -\frac{g(x,0)}{r} > 0.$$

This allows us to apply the implicit function theorem to the equation $g(x,rLv) = 0$ and to derive the existence of neighborhoods U of x, V of v, and W of r along with a continuously differentiable function $\rho^{x,v} : U \times V \to W$ such that the equivalence

$$g(x',r'Lv') = 0,\ (x',v',r') \in U \times V \times W \Leftrightarrow\ r' = \rho^{x,v}(x',v'),\ (x',v') \in U \times V \tag{3.4}$$

holds true. By continuity of $\rho^{x,v}$, we may shrink the neighborhoods U and V such that $\rho^{x,v}$ maps into $\mathbb{R}_+$, and we may further shrink U such that $g(x',0) < 0$ for all $x' \in U$. Now, assume that $g(x',r^*Lv') = 0$ holds true for some $(x',v',r^*) \in U \times V \times (\mathbb{R}_+ \backslash W)$. Then, by "$\Leftarrow$" in (3.4), $g(x',\rho^{x,v}(x',v')Lv') = 0$, where $\rho^{x,v}(x',v') \in W$. Consequently, $r^* \neq \rho^{x,v}(x',v')$. On the other hand, $r^*, \rho^{x,v}(x',v') \in \mathbb{R}_+$. This contradicts the convexity of g with respect to the second argument and the fact that $g(x',0) < 0$. It follows that in (3.4) W may be replaced by $\mathbb{R}_+$, which proves statement 1. In particular, we have that $g(x',\rho^{x,v}(x',v')Lv') = 0$ for all $(x',v') \in U \times V$, which after differentiation gives the formula in statement 2. ∎

The preceding lemma allows us to observe the following.

LEMMA 3.3. *Let $x \in \mathbb{R}^n$ be such that $g(x,0) < 0$. Then the following hold:*

1. *If $v \in F(x)$, then there exist neighborhoods U of x and V of v such that $e(x',v') = F_\eta(\rho^{x,v}(x',v'))$ for all $(x',v') \in U \times V$, where e is defined in* (3.3)*, F_η is the cumulative distribution function of the chi-distribution with m degrees of freedom, and $\rho^{x,v}$ refers to the resolving function introduced in Lemma* 3.2.
2. *If $v \in I(x)$, then $\rho^{x_k,v_k}(x_k,v_k) \to \infty$ for any sequence $(x_k,v_k) \to (x,v)$ with $v_k \in F(x_k)$.*

Proof. By Lemma 3.2(1), we have for all (x',v') that $g(x',\rho^{x,v}(x',v')Lv') = 0$ and $g(x',r'Lv') \neq 0$ for all $r' \in \mathbb{R}_+$ with $r' \neq \rho^{x,v}(x',v')$. Now, (3.3) implies that

$$e(x',v') = \mu_\eta([0,\rho^{x,v}(x',v')]) = F_\eta(\rho^{x,v}(x',v')) - F_\eta(0) \quad \forall (x',v') \in U \times V.$$

Now, statement 1 of Lemma 3.3 follows upon observing that the chi-density is zero for negative arguments, whence $F_\eta(0) = 0$. Next, let $v \in I(x)$ and $(x_k,v_k) \to (x,v)$ with $v_k \in F(x_k)$. If $\rho^{x_k,v_k}(x_k,v_k)$ does not tend to ∞, then there exists a converging subsequence $\rho^{x_{k_l},v_{k_l}}(x_{k_l},v_{k_l}) \to r$ for some $r \geq 0$. Since $g(x,0) < 0$, we have that $g(x_{k_l}, 0) < 0$ for l sufficiently large. This allows us to apply Lemma 3.2 to the points (x_{k_l}, v_{k_l}), and so we infer from statement 1 of Lemma 3.3 that $g(x_{k_l}, \rho^{x_{k_l}, v_{k_l}}(x_{k_l}, v_{k_l}) L v_{k_l}) = 0$ for all l sufficiently large. By continuity of g we derive the contradiction $g(x, rLv) = 0$ with our assumption $v \in I(x)$. This proves statement 2. ◻

COROLLARY 3.4. *The function $e : \mathbb{R}^n \times \mathbb{S}^{m-1} \to \mathbb{R}$ defined in (3.3) is continuous at any $(x, v) \in \mathbb{R}^n \times \mathbb{S}^{m-1}$ such that $g(x, 0) < 0$.*

Proof. Let $(x, v) \in \mathbb{R}^n \times \mathbb{S}^{m-1}$ with $g(x, 0) < 0$ be arbitrarily given. Referring to the sets $F(x)$ and $I(x)$ characterized in Lemma 3.1, there are two possibilities: if $v \in F(x)$, then the function $\rho^{x,v}$ is defined on a neighborhood of (x, v) and is continuous there by Lemma 3.2. Moreover, in this case, e has the representation given in statement 1 of Lemma 3.3. But with the cumulative distribution function F_η of the chi-distribution being continuous, e is continuous, too, at (x, v) as a composition of continuous mappings. If, in contrast, $v \notin F(x)$, then $v \in I(x)$ by statement 2 of Lemma 3.1. From statement 4 of the same lemma, we know that $e(x, v) = 1$. Consider an arbitrary sequence $(x_k, v_k) \to (x, v)$ with $v_k \in \mathbb{S}^{m-1}$. Since $g(x, 0) < 0$, we have that $g(x_k, 0) < 0$ for k sufficiently large. Assume that $e(x_k, v_k) \to 1$ does not hold. Then, there are a subsequence (x_{k_l}, v_{k_l}) and some $\varepsilon > 0$ such that for all l

$$|e(x_{k_l}, v_{k_l}) - 1| \geq \varepsilon. \tag{3.5}$$

By Lemma 3.1(4), $v_{k_l} \notin I(x_{k_l})$, whence $v_{k_l} \in F(x_{k_l})$ for all l due to $v_{k_l} \in \mathbb{S}^{m-1}$ and Lemma 3.1(2). Then $\rho^{x_{k_l}, v_{k_l}}(x_{k_l}, v_{k_l}) \to \infty$ by Lemma 3.3(2). Since F_η is the distribution function of a random variable, it satisfies the relation $\lim_{t\to\infty} F_\eta(t) = 1$. Consequently, we may invoke statement 1 of Lemma 3.3 to verify that

$$\lim_{l\to\infty} e(x_{k_l}, v_{k_l}) = \lim_{l\to\infty} F_\eta(\rho^{x_{k_l}, v_{k_l}}(x_{k_l}, v_{k_l})) = 1.$$

This contradicts (3.5), and, hence, again by Lemma 3.1(4),

$$\lim_{k\to\infty} e(x_k, v_k) = 1 = e(x, v).$$

This proves continuity of e at (x, v). ◻

COROLLARY 3.5. *For any $x \in \mathbb{R}^n$ with $g(x, 0) < 0$ and $v \in F(x)$, the partial derivative with respect to x of the function $e : \mathbb{R}^n \times \mathbb{S}^{m-1} \to \mathbb{R}$ defined in (3.3) exists and is given by*

$$\nabla_x e(x, v) = -\frac{\chi(\rho^{x,v}(x, v))}{\langle \nabla_z g(x, \rho^{x,v}(x, v) Lv), Lv\rangle} \nabla_x g(x, \rho^{x,v}(x, v) Lv),$$

where χ is the density of the chi-distribution with m degrees of freedom and $\rho^{x,v}$ refers to the function introduced in Lemma 3.2.

Proof. By statement 1 of Lemma 3.3 we have that $e(x', v') = F_\eta(\rho^{x,v}(x', v'))$ for all x' in a neighborhood of x and all v' in a neighborhood of v. Differentiation with respect to x yields

$$\nabla_x e(x', v') = \chi(\rho^{x,v}(x', v')) \nabla_x \rho^{x,v}(x', v') \tag{3.6}$$

due to $F_\eta'(\tau) = \chi(\tau)$ for $\tau > 0$. In particular, $\nabla_x e(x, v) = \chi(\rho^{x,v}(x, v)) \nabla_x \rho^{x,v}(x, v)$. Now the assertion follows from Lemma 3.2(2). ◻

Next we prove a relation which is the key to some desired continuity properties.

DEFINITION 3.6. *Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a differentiable function. We say that g satisfies the* polynomial growth condition *at x if there exist constants $C, \varkappa > 0$ and a neighborhood $U(x)$ such that*

$$\|\nabla_x g(x', z)\| \leq \|z\|^{\varkappa} \quad \forall x' \in U(x),\ \forall z : \|z\| \geq C.$$

LEMMA 3.7. *Let x be such that $g(x,0) < 0$ and such that g satisfies the polynomial growth condition at x. Consider any sequence $(x_k, v_k) \to (x, v)$ for some $v \in I(x)$ such that $v_k \in F(x_k)$. Then,*

$$\lim_{k\to\infty} \nabla_x e(x_k, v_k) = 0.$$

Proof. First observe that $\rho^{x_k,v_k}(x_k, v_k) \to \infty$ by Lemma 3.3(2). Referring to the neighborhood $U(x)$ from Definition 3.6, we verify that for k sufficiently large

$$\|\nabla_x g(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k)\| \leq [\rho^{x_k,v_k}(x_k,v_k)]^{\varkappa} \|Lv_k\|^{\varkappa} \leq \|L\|^{\varkappa} [\rho^{x_k,v_k}(x_k,v_k)]^{\varkappa} \tag{3.7}$$

(recall that $\|v_k\| = 1$ due to $v_k \in F(x_k)$). Moreover, by continuity of g, there exists some $\delta_1 > 0$ such that $g(x_k, 0) \leq -\delta_1 < 0$ for k sufficiently large. Since $g(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k) = 0$ (see Lemma 3.2(1)), Lemma 3.1(3) provides that

$$\langle \nabla_z g(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k), Lv_k\rangle \geq -\frac{g(x_k,0)}{\rho^{x_k,v_k}(x_k,v_k)}.$$

Therefore,

$$\langle \nabla_z g(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k), Lv_k\rangle \geq \delta_1 [\rho^{x_k,v_k}(x_k,v_k)]^{-1} > 0. \tag{3.8}$$

Using the definition $\chi(y) = \delta_2 y^{m-1} e^{-y^2/2}$ of the density of the chi-distribution with m degrees of freedom (where $\delta_2 > 0$ is an appropriate factor), we may combine Corollary 3.5 with (3.7) and (3.8) in order to derive that

$$\begin{aligned} \|\nabla_x e(x_k, v_k)\| &= \left\| \frac{\chi(\rho^{x_k,v_k}(x_k,v_k))}{\langle \nabla_z g(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k), Lv_k\rangle} \nabla_x g(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k) \right\| \\ &\leq \delta_1^{-1} \rho^{x_k,v_k}(x_k,v_k) \cdot \delta_2 [\rho^{x_k,v_k}(x_k,v_k)]^{m-1} e^{-[\rho^{x_k,v_k}(x_k,v_k)]^2/2} \cdot \|L\|^{\varkappa} [\rho^{x_k,v_k}(x_k,v_k)]^{\varkappa} \\ &= \delta_1^{-1}\delta_2 \|L\|^{\varkappa} [\rho^{x_k,v_k}(x_k,v_k)]^{\varkappa+m} e^{-[\rho^{x_k,v_k}(x_k,v_k)]^2/2} \to_k 0, \end{aligned} \tag{3.9}$$

where the last limit follows from $\rho^{x_k,v_k}(x_k,v_k) \to \infty$ and the fact that $y^{\alpha} e^{-y^2/2} \to 0$ for $y \to \infty$, where $\alpha > 0$ is an arbitrary constant. This proves our assertion. □

Remark 3.3. One may observe from the proof of Lemma 3.7 that a weaker growth condition than that in Definition 3.6 (involving an exponential term) would suffice for proving the same result. One could, for instance, use the following exponential growth condition.

Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a differentiable function. We say that g satisfies the *exponential growth condition* at x if there exist constants $\delta_0, C > 0$ and a neighborhood $U(x)$ such that

$$\|\nabla_x g(x', z)\| \leq \delta_0 \exp(\|z\|) \quad \forall x' \in U(x),\ \forall z : \|z\| \geq C,$$

and we observe that the key estimate (3.9) of Lemma 3.7 becomes

$$\|\nabla_x e(x_k, v_k)\| \leq \delta_0 \delta_1^{-1} \delta_2 \left[\rho^{x_k,v_k}(x_k, v_k)\right]^m e^{-[\rho^{x_k,v_k}(x_k,v_k)]^2/2} e^{\|L\| \rho^{x_k,v_k}(x_k,v_k)}.$$

The same conclusion then easily follows.

In this paper, we do not put the emphasis on the weakest possible form of the growth condition but rather on its simplicity. It should be noted, however, that each of the following results requiring the polynomial growth condition holds upon requiring the above exponential growth condition instead.

COROLLARY 3.8. *Let x be such that $g(x, 0) < 0$ and such that g satisfies the polynomial growth condition at x. Then, for any $v \in \mathbb{S}^{m-1}$, the partial derivative with respect to x of the function e exists at (x, v) and is given by*

$$\nabla_x e(x, v) = \begin{cases} -\frac{\chi(\rho^{x,v}(x,v))}{\langle \nabla_z g(x, \rho^{x,v}(x,v)Lv), Lv \rangle} \nabla_x g\left(x, \rho^{x,v}(x, v) Lv\right) & \text{if } v \in F(x), \\ 0 & \text{else,} \end{cases}$$

where χ is the density of the chi-distribution with m degrees of freedom and $\rho^{x,v}$ refers to the function introduced in Lemma 3.2.

Proof. Thanks to Corollary 3.5 and Lemma 3.1(2), it is sufficient to show that $\nabla_x e(x, v) = 0$ for $v \in I(x)$. We shall show that, for any $i \in \{1, \ldots, m\}$,

$$\lim_{t \uparrow 0} \frac{e(x + tu_i, v) - e(x, v)}{t} = 0, \tag{3.10}$$

where u_i is the ith canonical unit vector in $\mathbb{R}^n$. In exactly the same way one can show that the corresponding limit for $t \downarrow 0$ equals zero. Altogether, this will prove that $\nabla_x e(x, v) = 0$. Assume that (3.10) is wrong. Since $e(x, v) = 1$ (by Lemma 3.1(4)) and $e(x + tu_i, v) \leq 1$ for all t (by definition of e as a probability in (3.3)), it follows that the quotient in (3.10) is always nonpositive, and, thus, negation of (3.10) implies the existence of some $\varepsilon > 0$ and of a sequence $t_k \uparrow 0$ such that

$$\frac{e(x + t_k u_i, v) - e(x, v)}{t_k} \geq \varepsilon. \tag{3.11}$$

In particular, $v \in F(x + t_k u_i)$ for all k because otherwise $v \in I(x + t_k u_i)$ and so $e(x + t_k u_i, v) = 1$ (again by Lemma 3.1(4)), thus contradicting (3.11). We may also assume that $g(x + t_k u_i, 0) < 0$ for all k. Now, fix an arbitrary k and define (recall that $t_k < 0$)

$$\alpha := \inf \left\{\tau \in [t_k, 0] \,|\, e(x + \tau u_i, v) = 1\right\}.$$

Due to $e(x, v) = 1$ we have that $\alpha \leq 0$. On the other hand, $e(x + t_k u_i, v) < 1$ and the continuity of e (see Corollary 3.4) provide that $\alpha > t_k$. We infer that $e(x + \tau u_i, v) < 1$ for all $\tau \in [t_k, \alpha)$ and, hence,

$$v \in F(x + \tau u_i) \quad \forall \tau \in [t_k, \alpha) \tag{3.12}$$

(once more by statements 2 and 4 of Lemma 3.1). But then the function

$$\beta(\tau) := e(x + \tau u_i, v)$$

is differentiable for all $\tau \in (t_k, \alpha)$ by virtue of Corollary 3.5, and its derivative is given by

$$\beta'(\tau) = \langle \nabla_x e(x + \tau u_i, v), u_i \rangle.$$

Therefore, the mean value theorem guarantees the existence of some $\tau_k^* \in (t_k, \alpha)$ such that

$$\beta'(\tau_k^*) = \frac{\beta(\alpha) - \beta(t_k)}{\alpha - t_k}$$

or, equivalently,

$$\langle \nabla_x e(x + \tau_k^* u_i, v), u_i \rangle = \frac{e(x + \alpha u_i, v) - e(x + t_k u_i, v)}{\alpha - t_k}.$$

By continuity of e and by definition of α, we have that $e(x + \alpha u_i, v) = 1 = e(x, v)$, whence, by $t_k < \alpha \leq 0$,

$$\langle \nabla_x e(x + \tau_k^* u_i, v), u_i \rangle = \frac{e(x, v) - e(x + t_k u_i, v)}{\alpha - t_k} \geq \frac{e(x, v) - e(x + t_k u_i, v)}{-t_k} \geq \varepsilon,$$

where the last relation follows from (3.11). Now, since k was arbitrarily fixed, we have constructed a sequence τ_k^* such that $t_k < \tau_k^* \leq 0$ such that

$$\langle \nabla_x e(x + \tau_k^* u_i, v), u_i \rangle \geq \varepsilon \quad \forall k. \tag{3.13}$$

Since $t_k \uparrow 0$, we also have that $\tau_k^* \uparrow 0$. Moreover, $v \in F(x + \tau_k^* u_i)$ by (3.12). Due to our assumption that g satisfies the polynomial growth condition at x and due to $v \in I(x)$, Lemma 3.7 yields that $\lim_{k\to\infty} \nabla_x e(x_k, v) = 0$, which contradicts (3.13). This proves Corollary 3.8. ∎

COROLLARY 3.9. *Let x be such that $g(x, 0) < 0$ and such that g satisfies the polynomial growth condition at x. Then, for any $v \in \mathbb{S}^{m-1}$, the partial derivative $\nabla_x e$ is continuous at (x, v).*

Proof. Let $x \in \mathbb{R}^n$ with $g(x, 0) < 0$ and $v \in \mathbb{S}^{m-1}$ be arbitrarily given. Also let $(x_k, v_k) \to (x, v)$ be an arbitrary sequence with $v_k \in \mathbb{S}^{m-1}$. If $v \in F(x)$, then relation (3.6) holds true locally around (x, v). In particular, for k large enough,

$$\begin{aligned} \nabla_x e(x_k, v_k) &= \chi\left(\rho^{x,v}(x_k, v_k)\right) \nabla_x \rho^{x,v}(x_k, v_k) \to \chi\left(\rho^{x,v}(x, v)\right) \nabla_x \rho^{x,v}(x, v) \\ &= \nabla_x e(x, v), \end{aligned}$$

where the convergence follows from the continuity of the chi-density and of $\nabla_x \rho^{x,v}$ as a result of Lemma 3.2. Hence, in the case of $v \in F(x)$, $\nabla_x e$ is continuous at (x, v). Now assume in contrast that $v \in I(x)$. Then $\nabla_x e(x, v) = 0$ by Corollary 3.8. Now assume that $\nabla_x e(x_k, v_k)$ does not converge to zero. Then $\|\nabla_x e(x_{k_l}, v_{k_l})\| \geq \varepsilon$ for some subsequence and some $\varepsilon > 0$. Then $v_{k_l} \in F(x_{k_l})$ for all l because otherwise $v_{k_l} \in I(x_{k_l})$ and, thus, $\nabla_x e(x_{k_l}, v_{k_l}) = 0$ due to Corollary 3.8 (applied to x_{k_l} rather than x; observe that the condition $g(x, 0) < 0$ and the polynomial growth condition at x are open conditions and hence continue to hold true for the x_{k_l}). Now Lemma 3.7 yields the contradiction

$$\lim_{l\to\infty} \nabla_x e(x_{k_l}, v_{k_l}) = 0$$

with $\|\nabla_x e(x_{k_l}, v_{k_l})\| \geq \varepsilon$. This proves Corollary 3.9. ∎

Now we are in a position to state our main result.

THEOREM 3.10. *Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable function which is convex with respect to the second argument. Consider the probability function φ defined in (3.1), where $\xi \sim \mathcal{N}(0, R)$ has a standard Gaussian distribution with correlation matrix R. Let the following assumptions be satisfied at some $\bar{x}$:*

1. $g(\bar{x},0)<0$.
2. *g satisfies the polynomial growth condition at $\bar{x}$ (Definition* 3.6*).*

Then φ is continuously differentiable on a neighborhood U of $\bar{x}$ and it holds that

$$\nabla\varphi(x) = -\int_{v\in F(x)} \frac{\chi\left(\rho^{x,v}(x,v)\right)}{\langle\nabla_z g\left(x,\rho^{x,v}(x,v)Lv\right),Lv\rangle}\nabla_x g\left(x,\rho^{x,v}(x,v)Lv\right)d\mu_\zeta(v) \quad \forall x\in U. \tag{3.14}$$

Here, μ_ζ is the law of the uniform distribution over $\mathbb{S}^{m-1}$, χ is the density of the chi-distribution with m degrees of freedom, L is a factor of the Cholesky decomposition $R=LL^T$, and $\rho^{x,v}$ is as introduced in Lemma 3.2*.*

Proof. Since $\xi\sim\mathcal{N}(0,R)$, the probability function φ has the representation (3.2). With $g(\bar{x},0)<0$, let U be a sufficiently small neighborhood of $\bar{x}$ such that for all $x\in U$ we still have that $g(x,0)<0$ and that the polynomial growth condition is satisfied at x. Then the partial derivative $\nabla_x e$ of the function e defined in (3.3) exists on $U\times\mathbb{S}^{m-1}$ by Corollary 3.8 and is continuous there by Corollary 3.9. By compactness of $\mathbb{S}^{m-1}$, there exists some $K>0$ such that

$$\|\nabla_x e(\bar{x},v)\|\le K \quad \forall v\in\mathbb{S}^{m-1}.$$

Again, continuity of $\nabla_x e$ on $U\times\mathbb{S}^{m-1}$ and compactness of $\mathbb{S}^{m-1}$ guarantee that the function $\alpha:U\to\mathbb{R}$ defined by

$$\alpha(x):=\max_{v\in\mathbb{S}^{m-1}}\|\nabla_x e(x,v)\|$$

is continuous. Since $\alpha(\bar{x})\le K$, we may assume, after possibly shrinking U, that $\alpha(x)\le 2K$ for all $x\in U$, whence

$$\|\nabla_x e(x,v)\|\le 2K \quad \forall x\in U,\ \forall v\in\mathbb{S}^{m-1}. \tag{3.15}$$

From $\mu_\zeta(\mathbb{S}^{m-1})=1$ for the law μ_ζ of the uniform distribution on $\mathbb{S}^{m-1}$ we infer that the constant $2K$ is an integrable function on $\mathbb{S}^{m-1}$ uniformly dominating $\|\nabla_x e(x,v)\|$ on $\mathbb{S}^{m-1}$ for all $x\in U$. Now Lebesgue's dominated convergence theorem allows us to differentiate (3.2) under the integral sign:

$$\nabla\varphi(\bar{x})=\int_{v\in\mathbb{S}^{m-1}}\nabla_x e(\bar{x},v)d\mu_\zeta.$$

As stated in the beginning of this proof, assumptions 1 and 2 imposed in Theorem 3.10 for the fixed point $\bar{x}$ continue to hold for all x in the neighborhood U. Therefore, we may derive that

$$\nabla\varphi(x)=\int_{v\in\mathbb{S}^{m-1}}\nabla_x e(x,v)d\mu_\zeta \quad \forall x\in U. \tag{3.16}$$

Exploiting once more the dominance argument from (3.15), the continuity of $\nabla_x e$ on $U\times\mathbb{S}^{m-1}$ and the compactness of $\mathbb{S}^{m-1}$ ensure by virtue of Lebesgue's dominated convergence theorem that $\nabla\varphi$ is continuous. Finally, formula (3.14) follows directly from Corollary 3.8. ∎

Remark 3.4. Evidently, formula (3.14) is explicit and can be used inside Deák's method in order to calculate $\nabla\varphi$ in parallel with φ by efficient sampling on $\mathbb{S}^{m-1}$. For each sampled point $v\in\mathbb{S}^{m-1}$ one first has to check whether the equation $g(x,rLv)=0$ has a solution $r\ge 0$ at all. If not, i.e., $v\in I(x)$, then such v does not contribute to the (approximated) integral in (3.14). Otherwise, i.e., $v \in F(x)$, one has to evaluate the integrand in (3.14), which amounts to finding the unique solution $r \geq 0$ of the equation $g(x, rLv) = 0$. In general, a few Newton–Raphson iterations should do the job.

We now want to focus our attention on the assumptions of Theorem 3.10. First, recall that assuming a standard Gaussian distribution $\xi \sim \mathcal{N}(0, R)$ does not mean any loss of generality by virtue of Remark 3.2. Also assumption 1 of the theorem is not restrictive. This will come as a consequence of the following proposition.

PROPOSITION 3.11. *With g and φ as in Theorem 3.10, let the following assumptions be satisfied at some $\bar{x}$:*

1. *There exists some $\bar{z}$ such that $g(\bar{x}, \bar{z}) < 0$.*
2. *$\varphi(\bar{x}) > 1/2$.*

Then $g(\bar{x}, 0) < 0$.

Proof. As in the proof of Theorem 3.10 we may assume that $\xi \sim \mathcal{N}(0, R)$ so that φ has the representation (3.2). Define the set $M := \{z \in \mathbb{R}^m | g(\bar{x}, z) \leq 0\}$. Clearly, M is convex and nonempty by our assumption 1. This same assumption (Slater point) guarantees that

$$\operatorname{int} M = \{z \in \mathbb{R}^m | g(\bar{x}, z) < 0\}.$$

Assume that $g(\bar{x}, 0) \geq 0$. Then $0 \notin \operatorname{int} M$, and, hence, one could separate 0 from M, which would mean that there exists some $c \in \mathbb{R}^m \backslash \{0\}$ such that

$$M \subseteq \left\{z \in \mathbb{R}^m | c^T z \leq 0\right\} =: \tilde{M}.$$

With ξ having a centered Gaussian distribution, the one-dimensional random variable $c^T \xi$ has a centered Gaussian distribution too, and, hence, we arrive with our assumption 3 at the contradiction

$$1/2 = \mathbb{P}\left(c^T \xi \leq 0\right) = \mathbb{P}\left(\xi \in \tilde{M}\right) \geq \mathbb{P}\left(\xi \in M\right) = \varphi\left(\bar{x}\right) > 1/2.$$

The proposition has been proved. ∎

Proposition 3.11 means that violation of assumption 1 in Theorem 3.10 implies that $g(\bar{x}, z) \geq 0$ for all z or that $\varphi(\bar{x}) \leq 1/2$. A typical application of Theorem 3.10 is probabilistic programming, where one is imposing the chance constraint $\varphi(x) \geq p$ with some probability level p close to one. Since gradients of φ are usually calculated at or close to feasible points (e.g., by cutting planes), the case $\varphi(\bar{x}) \leq 1/2$ is very unlikely to occur. On the other hand, $g(\bar{x}, z) \geq 0$ for all z is a degenerate situation meaning that there exists no Slater point for the convex function $g(\bar{x}, \cdot)$. In such a situation it typically happens that the set $\{z | g(x, z) \leq 0\}$ becomes empty for x arbitrarily close to $\bar{x}$, which would entail a discontinuity of φ at $\bar{x}$. Then, of course, there is no hope of calculating a gradient at all.

Finally, turning to condition 2 of Theorem 3.10 (growth condition), it may require some technical effort to check it in concrete applications (see, e.g., the examples discussed in the following section). On the other hand, we shall see in a moment that we may do without this condition in the case that the set $\{z | g(\bar{x}, z) \leq 0\}$ is bounded. To formulate a corresponding statement we need the following two auxiliary results.

LEMMA 3.12. *Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be continuous. Moreover, let g be convex in the second argument. Then, for any $x \in \mathbb{R}^n$ with $g\,(x, 0) < 0$ one has that $I(x) = \emptyset$ if and only if $M(x) := \{z \in \mathbb{R}^m | g(x, z) \leq 0\}$ is bounded.*

Proof. Let x be arbitrary such that $g(x,0) < 0$. Obviously boundedness of $M(x)$ implies that $I(x) = \emptyset$, so let us assume that $I(x) = \emptyset$ and that $M(x)$ is unbounded. Then there is a sequence z_n with $g(x,z_n) \leq 0$ and $\|z_n\| \to \infty$. Without loss of generality, we may assume that $\|z_n\|^{-1} z_n \to z$ for some $z \in \mathbb{R}^m \backslash \{0\}$. Let $t \geq 0$ be arbitrary. Then $\|z_n\|^{-1} t \leq 1$ for n sufficiently large. From convexity of $g(x,\cdot)$, $g(x,0) < 0$, and $g(x,z_n) \leq 0$ we infer that $g(x, \|z_n\|^{-1} t z_n) \leq 0$ for n sufficiently large. Passing to the limit, we get that $g(x,tz) \leq 0$. Thus, as $t \geq 0$ was arbitrary,

$$g(x,tz) \leq 0 \quad \forall t \geq 0. \tag{3.17}$$

Assume that there was some $\tau \geq 0$ with $g(x,\tau z) = 0$. Then, again by convexity of $g(x,\cdot)$ and by $g(x,0) < 0$, one would arrive at the following contradiction with (3.17):

$$g(x,tz) > g(x,\tau z) = 0 \quad \forall t > \tau.$$

Hence, actually $g(x,tz) < 0$ for all $t \geq 0$. Putting $v := L^{-1}z / \left\|L^{-1}z\right\|$, where L is the (invertible) matrix appearing in the definition of $I(x)$, and observing that this definition is correct due to $z \neq 0$, we derive that $g\left(x, t\left\|L^{-1}z\right\| Lv\right) < 0$ for all $t \geq 0$. Since $\left\|L^{-1}z\right\| > 0$, this implies that $g(x, rLv) < 0$ for all $r \geq 0$. Hence we have the contradiction $v \in I(x)$ with our assumption $I(x) = \emptyset$. It follows that $M(x)$ is bounded as was to be shown. ∎

PROPOSITION 3.13. *Let g be as in Lemma 3.12, and let $\bar{x} \in \mathbb{R}^n$ with $g(\bar{x},0) < 0$. If $M(\bar{x})$ is bounded, then there is a neighborhood U of $\bar{x}$ such that $M(x)$ remains bounded for all $x \in U$.*

Proof. By continuity of g, we may choose U small enough that $g(x,0) < 0$ for all $x \in U$. If the assertion is not true, then by virtue of Lemma 3.12 there exists a sequence $x_n \to \bar{x}$ such that $I(x_n) \neq \emptyset$ for all $n \in \mathbb{N}$. By statement 1 in Lemma 3.1 this implies the existence of another sequence $v_n \in \mathbb{S}^{m-1}$ such that

$$g(x_n, rLv_n) < 0 \quad \forall r \geq 0,\ \forall n \in \mathbb{N}.$$

Without loss of generality, we may assume that $v_n \to \bar{v}$ for some $\bar{v} \in \mathbb{S}^{m-1}$. For each $r \geq 0$ we may pass to the limit in the relation above, in order to derive that $g(\bar{x}, rL\bar{v}) \leq 0$ for all $r \geq 0$. With the same reasoning as below (3.17), we may conclude that indeed $g(\bar{x}, rL\bar{v}) < 0$ for all $r \geq 0$. This means that $\bar{v} \in I(\bar{x})$, whence $M(\bar{x})$ is unbounded by Lemma 3.12. This is a contradiction with our assumption. ∎

Now we are in a position to state an alternate variant of Theorem 3.10 which does not require the verification of the growth condition.

THEOREM 3.14. *Theorem 3.10 remains true if the second condition (growth condition) is replaced by the condition that the set $\{z | g(\bar{x},z) \leq 0\}$ is bounded. Then (3.14) becomes*

$$\nabla\varphi(x) = -\int_{v \in \mathbb{S}^{m-1}} \frac{\chi\left(\rho^{x,v}(x,v)\right)}{\left\langle \nabla_z g\left(x, \rho^{x,v}(x,v) Lv\right), Lv \right\rangle} \nabla_x g\left(x, \rho^{x,v}(x,v) Lv\right) d\mu_\zeta(v) \quad \forall x \in U. \tag{3.18}$$

Proof. As in the proof of Theorem 3.10, the function e is continuous on $U \times \mathbb{S}^{m-1}$ by Corollary 3.4 because this result does not require the growth condition to hold. Moreover, $\nabla_x e$ exists on $U \times \mathbb{S}^{m-1}$. Indeed, our boundedness assumption ensures via Proposition 3.13 that—after possibly shrinking the neighborhood U of $\bar{x}$—the set

$\{z|g(x,z) \le 0\}$ remains bounded for all $x \in U$. Lemma 3.12 implies that $I(x) = \emptyset$ or, equivalently, according to Lemma 3.1(2), that $F(x) = \mathbb{S}^{m-1}$ for all $x \in U$. Then Corollary 3.5 yields that $\nabla_x e$ exists on $U \times \mathbb{S}^{m-1}$ and is given by

$$\nabla_x e(x,v) = -\frac{\chi\left(\rho^{x,v}\left(x,v\right)\right)}{\left\langle \nabla_z g\left(x,\rho^{x,v}\left(x,v\right)Lv\right),Lv\right\rangle} \nabla_x g\left(x,\rho^{x,v}\left(x,v\right)Lv\right).$$

Since all occurring functions are continuous, the same holds true for $\nabla_x e$. Now the same argument as in the proof of Theorem 3.10 allows us to derive (3.16) which along with the formula for $\nabla_x e$ above yields (3.18). □

The result of Theorem 3.14 which we have derived here as a special case of Theorem 3.10 also can be derived from [30, Proposition 2.2] under an implicit boundedness assumption (Assumption 2.2(i)) corresponding to the one imposed here. Note that in Theorem 3.10 we do not require boundedness. Not surprisingly, gradient formulae for probability functions may come in different guises. In [34, Theorem 2.1], $\nabla\varphi$ was represented—again under the restrictive boundedness assumption—in the form of a volume or surface integral over the (x-dependent) domain of integration, which is not directly related to the integral over the unit sphere considered here.

4. Selected examples. In this section we are going to discuss some instances of the probabilistic constraint (1.1) to which our gradient formulae obtained in Theorems 3.10 and 3.14 apply and thus could be used in the numerical solution of corresponding optimization problems.

4.1. Gaussian distributions. We assume first, as before, that the random vector has a Gaussian distribution. We shall focus on the particular model

$$\mathbb{P}\left(\langle f(\xi), h_1(x)\rangle \le h_2(x)\right) \ge p \tag{4.1}$$

with nonlinear mappings $f : \mathbb{R}^m \to \mathbb{R}^l$, $h_1 : \mathbb{R}^n \to \mathbb{R}^l$, and $h_2 : \mathbb{R}^n \to \mathbb{R}$ involving a coupling of the random and decision vectors.

PROPOSITION 4.1. *In the probabilistic constraint* (4.1)*, let f, h_1, h_2 be continuously differentiable, and let the components f_i of f be convex and the components $h_{1,i}$ of h_1 be nonnegative. Furthermore, let $\xi \sim \mathcal{N}(0,R)$ have a standard Gaussian distribution with correlation matrix R and associated Cholesky decomposition $R = LL^T$. Consider any $\bar{x}$ with $\langle f(0), h_1(\bar{x})\rangle < h_2(\bar{x})$. Finally, let f satisfy the following polynomial growth condition:*

$$\|f(z)\| \le \|z\|^{\varkappa} \quad \forall z : \|z\| \ge C$$

for certain $\varkappa, C > 0$. Then the probability function $\varphi(x) := \mathbb{P}\left(\langle f(\xi), h_1(x)\rangle \le h_2(x)\right)$ defining the constraint (4.1) *is continuously differentiable on a neighborhood U of $\bar{x}$, and its gradient is given by*

$$\begin{aligned}\nabla\varphi\left(x\right) = &\int_{v\in F(x)} \frac{\chi\left(\rho^{x,v}\left(x,v\right)\right)}{\left\langle h_1^T(x)\nabla f\left(\rho^{x,v}\left(x,v\right)Lv\right),Lv\right\rangle} \\ &\cdot\left(\nabla h_2(x) - \left[f\left(\rho^{x,v}\left(x,v\right)Lv\right)\right]^T\nabla h_1(x)\right)d\mu_\zeta(v) \quad \forall x \in U.\end{aligned} \tag{4.2}$$

Proof. In our setting the general function g in (3.1) becomes $g\left(x,z\right) = \langle f(z), h_1(x)\rangle - h_2(x)$. The continuous differentiability and convexity with respect to the second argument of g are evident from our assumptions. Moreover, $g\left(\bar{x},0\right) < 0$. As for the growth condition, let U be a neighborhood of $\bar{x}$ on which $\max\{\|\nabla h_1\|, \|\nabla h_2\|\} \le K$

for some $K > 0$. Then, taking—without loss of generality—the maximum norm, we have that

$$\begin{aligned}\|\nabla_x g(x,z)\| &= \left\|\nabla h_2(x) - [f(z)]^T \nabla h_1(x)\right\| \leq K(\|f(z)\|+1)\\ &\leq \|z\|^{2+\varkappa} \quad \forall x \in U,\ z : \|z\| \geq \max\{C,K,2\}.\end{aligned}$$

Consequently, we may apply Theorem 3.10. Equation (4.2) follows immediately from (3.14) for the given form of the function g. □

4.2. Gaussian-like distributions. We are now going to apply Theorem 3.10 to probabilistic constraints with random vectors having non-Gaussian distributions. In the first case, we consider a linear probabilistic constraint

$$\mathbb{P}\left(\langle \eta, x\rangle \leq b\right) \geq p \tag{4.3}$$

with a random vector η whose components η_i $(i = 1,\ldots,l)$ are independent and have a χ^2-distribution with n_i degrees of freedom. By definition, $\eta_i = \sum_{k=1}^{n_i} \xi_{i,k}^2$, where the $\xi_{i,k} \sim \mathcal{N}(0,1)$ are independent for $k = 1,\ldots,n_i$. We are interested in the gradient of the probability function $\varphi(x) := \mathbb{P}\left(\langle \eta, x\rangle \leq b\right)$. Define a Gaussian random vector with independent components

$$\xi := \left(\xi_{1,1},\ldots,\xi_{1,n_1},\ldots,\xi_{l,1},\ldots,\xi_{l,n_l}\right) \sim \mathcal{N}(0,I).$$

Clearly, $\eta \sim f(\xi)$, where $f_i(z) := \sum_{k=1}^{n_i} z_{i,k}^2$ for $i = 1,\ldots,l$ and z is partitioned in the same way as ξ above. Then, the probability function defining (4.3) becomes

$$\varphi(x) = \mathbb{P}\left(\langle \eta, x\rangle \leq b\right) = \mathbb{P}\left(\langle f(\xi), x\rangle \leq b\right).$$

We derive the following gradient formula which does not need the verification of a polynomial growth condition and which is even fully explicit with respect to the resolving function $\rho^{x,v}$.

PROPOSITION 4.2. *In* (4.3)*, let $b > 0$. Consider any feasible point $\bar{x}$ of* (4.3) *satisfying $\bar{x}_i > 0$ for $i = 1,\ldots,n$. Then the probability function φ is continuously differentiable on a neighborhood U of $\bar{x}$, and its gradient is given by*

$$\nabla\varphi(x) = -\frac{\sqrt{b}}{2}\int_{v\in\mathbb{S}^{m-1}} \frac{\chi\left(\sqrt{b/\langle f(v), x\rangle}\right)}{\langle f(v), x\rangle^{3/2}} [f(v)]^T \, d\mu_\zeta(v) \quad \forall x \in U. \tag{4.4}$$

Proof. In our setting the general function g in (3.1) becomes $g(x,z) = \langle f(z), x\rangle - b$, which is continuously differentiable. Since the components f_i are convex, $g(x,\cdot)$ is convex whenever $x \geq 0$, which by our assumption holds true in a neighborhood of $\bar{x}$. Evidently, the results of Theorems 3.10 and 3.14 are of local nature (differentiability around $\bar{x}$), so they actually do not need convexity of $g(x,\cdot)$ for all $x \in \mathbb{R}^n$ but only for x in a neighborhood of $\bar{x}$ which is satisfied here. Next observe that $g(\bar{x},0) = -b < 0$. Finally, recalling that $\bar{x}_i > 0$ for $i = 1,\ldots,n$, we obtain the estimate

$$\begin{aligned}\{z | g(\bar{x},z) \leq 0\} = \{z | \langle f(z), \bar{x}\rangle \leq b\} &\subseteq \left\{z \left| \left(\min_{i=1,\ldots,n} \bar{x}_i\right) \sum_{i=1}^n f_i(z) \leq b\right.\right\}\\ &= \left\{z \left| \|z\|^2 \leq b\left(\min_{i=1,\ldots,n} \bar{x}_i\right)^{-1}\right.\right\},\end{aligned}$$

whence the set on the left-hand side is bounded. Altogether, this allows us to invoke Theorem 3.14 and to derive the validity of formula (3.18). We now specify this formula in our setting. First observe that, given $\xi \sim \mathcal{N}(0, I)$, we have that $R = I$; hence we have $L = I$ for the Cholesky decomposition $R = LL^T$. Next we calculate explicitly the function $\rho^{x,v}(x, v)$ which is the unique solution in $r \geq 0$ of the equation $\langle f(rLv), x\rangle = b$. Now, by definition of f,

$$\langle f(rLv), x\rangle = r^2 \langle f(v), x\rangle = b,$$

whence

$$r = \sqrt{b / \langle f(v), x\rangle}. \tag{4.5}$$

Next, we calculate

$$\begin{aligned}
\nabla_x g(x, \rho^{x,v}(x,v) Lv) &= [f(\rho^{x,v}(x,v) v)]^T = [\rho^{x,v}(x,v)]^2 [f(v)]^T \\
&= (b / \langle f(v), x\rangle) [f(v)]^T, \\
\langle \nabla_z g(x, \rho^{x,v}(x,v) Lv), Lv\rangle &= \langle \nabla_z g(x, \rho^{x,v}(x,v) v), v\rangle \\
&= \left\langle \sum\nolimits_{i=1}^n x_i \nabla f_i(\rho^{x,v}(x,v) v), v \right\rangle \\
&= \sum\nolimits_{i=1}^n x_i \langle \nabla f_i(\rho^{x,v}(x,v) v), v\rangle \\
&= 2\rho^{x,v}(x,v) \sum\nolimits_{i=1}^n x_i \sum\nolimits_{k=1}^{n_i} v_{i,k}^2 \\
&= 2\rho^{x,v}(x,v) \langle f(v), x\rangle = 2\sqrt{b \langle f(v), x\rangle}.
\end{aligned}$$

Combining these last relations with (4.5) provides formula (4.4). □

As a second instance for a non-Gaussian but Gaussian-like distribution, we consider the multivariate log-normal distribution. Recall that a random vector η follows a multivariate log-normal distribution if the vector $\xi := \log \eta$ (componentwise logarithm) has a Gaussian distribution. We now consider a probabilistic constraint of type

$$\mathbb{P}(\langle \eta, x\rangle \leq h(x)) \geq p, \tag{4.6}$$

where η is an m-dimensional random vector with log-normal distribution and $h : \mathbb{R}^m \to \mathbb{R}$ is some function. We are interested in the gradient of the associated probability function $\varphi(x) := \mathbb{P}(\langle \eta, x\rangle \leq h(x))$. We denote by $\xi := \log \eta$ the Gaussian random vector associated with η. Without loss of generality (see Remark 3.2) we may assume that $\xi \sim \mathcal{N}(0, R)$ for some correlation matrix R. We denote by L the associated factor in the Cholesky decomposition $R = LL^T$.

PROPOSITION 4.3. *In the setting above, assume that $\bar{x}$ satisfies $\bar{x}_i > 0$ for $i = 1, \ldots, m$. Assume, moreover, that h is continuously differentiable and that $h(\bar{x}) > \sum_{i=1}^m \bar{x}_i$. Then*

$$\nabla \varphi(x) = -\int_{\{v \in \mathbb{S}^{m-1} \mid \exists i: L_i v > 0\}} \frac{\chi(\rho^{x,v}(x,v))}{\sum_{i=1}^m x_i e_i^{\rho^{x,v}(x,v) L_i v} L_i v} \left[e^{\rho^{x,v}(x,v) Lv} - \nabla h(x) \right] d\mu_\zeta(v) \qquad \forall x \in U.$$

Here, L_i refers to the ith row of L, and the expression e^z has to be understood componentwise.

Proof. In our setting the general function g in (3.1) becomes $g(x,z) = \langle e^z, x\rangle - h(x)$. Clearly, g is continuously differentiable and convex with respect to z for all x close to $\bar{x}$ (as mentioned in the proof of Proposition 4.3, this weakened condition is enough in the context of Theorem 3.10). Moreover, $g(\bar{x}, 0) = \sum_{i=1}^m \bar{x}_i - h(\bar{x}) < 0$. In order to apply Theorem 3.10, it is sufficient to verify the exponential growth condition of Remark 3.3 (note that the originally imposed polynomial growth condition would not hold true here). To this aim, let U be a neighborhood of $\bar{x}$ on which $\|\nabla h\| \leq K$ for some $K > 0$. Then, with respect to the maximum norm, we get that

$$\|\nabla_x g(x', z)\| \leq \|e^z\| + \|\nabla h(x')\| \leq e^{\|z\|} + K \leq 2e^{\|z\|} \quad \forall x' \in U(x),\ \forall z : \|z\| \geq \log K.$$

Hence, the exponential growth condition of Remark 3.3 is satisfied. This allows us to apply Theorem 3.10. Inserting the corresponding derivative formulae for g, we derive that φ is continuously differentiable on a neighborhood U of $\bar{x}$ and its gradient is given by

$$(4.7)\quad \nabla\varphi(x) = -\int_{v\in F(x)} \frac{\chi(\rho^{x,v}(x,v))}{\sum_{i=1}^m x_i e^{\rho^{x,v}(x,v)\langle L_i, v\rangle}\langle L_i, v\rangle}\left[e^{\rho^{x,v}(x,v)Lv} - \nabla h(x)\right] d\mu_\zeta(v)$$

for all $x \in U$. Here, L_i denotes the ith row of the Cholesky factor L. To complete the proof, we have to verify the representation of the integration domain $F(x)$ asserted in the statement of this proposition. Without loss of generality, we assume the neighborhood U of $\bar{x}$ in the formula above to be small enough that $g(x,0) < 0$ and $x_i > 0$ for $i = 1, \ldots, m$ and for all $x \in U$ (recall that $g(\bar{x}, 0) < 0$ and $\bar{x}_i > 0$ for $i = 1, \ldots, m$). We claim that for all $x \in U$ the set $I(x)$ introduced below (3.3) can be written as

$$(4.8)\qquad I(x) = \left\{v \in \mathbb{S}^{m-1} | Lv \leq 0\right\}.$$

Indeed, let $x \in U$ and $v \in \mathbb{S}^{m-1}$ with $Lv \leq 0$ be arbitrary. Then, for all $r > 0$,

$$g(x, rLv) = \left\langle e^{rLv}, x\right\rangle - h(x) \leq \left\langle e^0, x\right\rangle - h(x) = g(x,0) < 0,$$

whence $v \in I(x)$ by Lemma 3.1(1). Conversely, let $x \in U$ and $v \in I(x)$ be arbitrary. Then $\langle e^{rLv}, x\rangle < h(x)$ for all $r > 0$. Define $J := \{i | L_i v > 0\}$. It follows from $x_i > 0$ for $i = 1, \ldots, m$ that

$$h(x) > \sum_{i\in J} x_i e^{r\langle L_i, v\rangle}.$$

If $J \neq \emptyset$, then the sum on the right-hand side tends to ∞ for $r \to \infty$, which is a contradiction to this sum being bounded from above by $h(x)$ for all $r > 0$. Consequently, $J = \emptyset$, proving $Lv \leq 0$ and, thus, the reverse inclusion of (4.8). Since, by definition, $F(x) = \mathbb{S}^{m-1} \backslash I(x)$, we may plug the information from (4.8) into (4.7) in order to derive the asserted formula. ∎

4.3. Student's (or *t*-) distribution. As a last application, we are going to consider probabilistic constraints of type (1.1), where the random vector ξ follows a so-called multivariate Student's or t-distribution. This is an important type of distribution, in particular, due to its application in the context of copulas. We recall that $\xi \sim \mathcal{T}(\mu, \Sigma, \nu)$, i.e., ξ obeys a multivariate t-distribution with parameters μ, Σ, ν, if $\xi = \mu + \vartheta\sqrt{\frac{\nu}{u}}$, where $\vartheta \sim \mathcal{N}(0, \Sigma)$ has a multivariate Gaussian distribution with mean 0 and covariance matrix Σ, $u \sim \chi^2(\nu)$ has a chi-squared distribution with

ν degrees of freedom, and ϑ and u are independent [22]. We are interested in the probability function (3.1) but this time for a t-variable rather than for a Gaussian variable.

Remark 4.1. Using the definition of a t-distribution, we may duplicate the arguments of Remark 3.2 in order to convince ourselves that in the consideration of (3.1) we may assume without loss of generality that $\xi \sim \mathcal{T}(0, R, \nu)$, where R is a correlation matrix. In particular, this can be arranged without disturbing the assumption of g in (3.1) being continuously differentiable and convex with respect to the second argument.

In a first step, we provide an expression for the probability function (3.1) in the case of a t-distribution.

THEOREM 4.4. *Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable function which is convex with respect to the second argument. Moreover, let $\xi \sim \mathcal{T}(0, R, \nu)$ for some correlation matrix R. Consider a point $\bar{x}$ such that $g(\bar{x}, 0) < 0$. Then there exists a neighborhood U of $\bar{x}$ such that the probability function* (3.1) *admits the representation*

$$\varphi(x) = \int_{v \in \mathbb{S}^{m-1}} \tilde{e}(x, v)\, d\mu_\zeta \quad \forall x \in U,$$

where for all $x \in U$ and $v \in \mathbb{S}^{m-1}$

$$\tilde{e}(x, v) := \begin{cases} F_{m,\nu}(m^{-1}[\rho^{x,v}(x, v)]^2), & v \in F(x), \\ 1, & v \in I(x), \end{cases}$$

and $F_{m,\nu}$ refers to the distribution function of the Fisher–Snedecor distribution with m and ν degrees of freedom. Moreover, $\rho^{x,v}$ is as introduced in Lemma 3.2, *and $F(x)$ and $I(x)$ are defined as in Lemma* 3.1.

The proof of this theorem is left for the appendix. Now we may duplicate the proof of Corollary 3.4 but with the function e there replaced by the function $\tilde{e}$ introduced above and with the expression $F_\eta(\rho^{x,v}(x', v'))$ in statement 1 of Lemma 3.3 replaced by the expression $F_{m,\nu}(m^{-1}[\rho^{x,v}(x', v')]^2)$ in order to derive the continuity of $\tilde{e}$ at any $x \in U$, where U is defined as in Theorem 4.4. Next we may copy the proof of Corollary 3.5 (again with the appropriate replacements) and get the following.

COROLLARY 4.5. *For any $x \in \mathbb{R}^n$ with $g(x, 0) < 0$ and $v \in F(x)$, the partial derivative with respect to x of the function $\tilde{e} : \mathbb{R}^n \times \mathbb{S}^{m-1} \to \mathbb{R}$ defined in Theorem* 4.4 *exists and is given by*

(4.9)
$$\nabla_x \tilde{e}(x, v) = -2\rho^{x,v}(x, v) \frac{f_{m,\nu}\left(m^{-1}[\rho^{x,v}(x, v)]^2\right)}{m \left\langle \nabla_z g(x, \rho^{x,v}(x, v) Lv), Lv \right\rangle} \nabla_x g(x, \rho^{x,v}(x, v) Lv),$$

where

$$f_{m,\nu}(t) = \begin{cases} \frac{\Gamma(m/2+\nu/2)}{\Gamma(m/2)\Gamma(\nu/2)} m^{m/2} \nu^{\nu/2} t^{m/2-1} (mt + \nu)^{-(m+\nu)/2}, & t \geq 0, \\ 0, & t < 0, \end{cases} \tag{4.10}$$

is the density of the Fisher–Snedecor distribution with m and ν degrees of freedom, $\rho^{x,v}$ refers to the function introduced in Lemma 3.2, *and L is a factor of the Cholesky decomposition $R = LL^T$.*

The following "equivalent" of Lemma 3.7 requires some additional conditions and work.

LEMMA 4.6. *Let x be such that $g(x,0) < 0$ and such that g satisfies the polynomial growth condition at x with coefficient $\varkappa < \nu$ (Definition* 3.6*). Consider any sequence $(x_k, v_k) \to (x, v)$ for some $v \in I(x)$ such that $v_k \in F(x_k)$. Then*

$$\lim_{k\to\infty} \nabla_x \tilde{e}(x_k, v_k) = 0.$$

Proof. First observe that $\rho^{x_k,v_k}(x_k, v_k) \to \infty$ by Lemma 3.3(2). The arguments of Lemma 3.7 allow us to deduce that for k sufficiently large the estimates (3.7) and (3.8) still hold. Using (4.10), we may combine Corollary 4.5 with (3.7) and (3.8) in order to derive that

$$\begin{aligned}
\|\nabla_x \tilde{e}(x_k, v_k)\| &= \left\| \frac{2\rho^{x_k,v_k}(x_k, v_k) f_{m,\nu}(m^{-1}\left[\rho^{x,v}(x,v)\right]^2)}{m\left\langle \nabla_z g\left(x_k, \rho^{x_k,v_k}(x_k,v_k) L v_k\right), L v_k\right\rangle} \nabla_x g\left(x_k, \rho^{x_k,v_k}(x_k, v_k) L v_k\right) \right\| \\
&\leq 2\nu^{\nu/2} \frac{\Gamma(m/2+\nu/2)}{\Gamma(m/2)\Gamma(\nu/2)} \|L\|^{\varkappa} \delta_1^{-1} \rho^{x_k,v_k}(x_k, v_k)^{m+\varkappa} \\
&\quad \cdot \left(1 + \frac{\rho^{x_k,v_k}(x_k, v_k)^2}{\nu}\right)^{-\frac{m+\nu}{2}} \to_k 0,
\end{aligned}$$

where the last limit follows from $\rho^{x_k,v_k}(x_k, v_k) \to \infty$ and $\varkappa < \nu$. ◻

Having established Lemma 4.6, the same arguments as in Corollary 3.8 can be used to show that $\tilde{e}$ is differentiable with respect to x and to derive a similar formula. This can be done since the proof of Corollary 3.8 uses only the properties of e and we have established the same properties for $\tilde{e}$. Accordingly, $\nabla_x \tilde{e}(x, v)$ is given by formula (4.9) if $v \in F(x)$, and $\nabla_x \tilde{e}(x, v) = 0$ if $v \in I(x)$. In the same way as in Corollary 3.9 one establishes the continuity of $\nabla_x \tilde{e}$ upon noting that $f_{m,\nu}(t)$ defined as in (4.10) is also continuous. We thus arrive at the following key result, the proof of which is a verbatim copy of that of Theorem 3.10 (again, e and $\tilde{e}$ have the same properties).

THEOREM 4.7. *Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable function which is convex with respect to the second argument. Consider the probability function φ defined in* (3.1)*, where $\xi \sim \mathcal{T}(0, R, \nu)$. Let the following assumptions be satisfied at some $\bar{x}$:*

1. $g(\bar{x}, 0) < 0$.
2. *g satisfies the polynomial growth condition at $\bar{x}$ (Definition* 3.6*) with coefficient $\varkappa < \nu$.*

Then φ is continuously differentiable on a neighborhood U of $\bar{x}$, and it holds that

(4.11)
$$\nabla\varphi(x) = \int_{v\in F(x)} -\frac{2\rho^{x,v}(x,v) f_{m,\nu}(m^{-1}\left[\rho^{x,v}(x,v)\right]^2)}{m\left\langle \nabla_z g\left(x, \rho^{x,v}(x,v) Lv\right), Lv\right\rangle} \nabla_x g\left(x, \rho^{x,v}(x,v) Lv\right) d\mu_\zeta(v)$$

for all $x \in U$. Here, μ_ζ is the law of the uniform distribution over $\mathbb{S}^{m-1}$, $f_{m,\nu}$ is the density of the Fisher–Snedecor distribution with m and ν degrees of freedom, and $\rho^{x,v}$ is as introduced in Lemma 3.2.

In the above result, the degrees of freedom ν of $\xi \sim \mathcal{T}(0, R, \nu)$ impose an important restriction on the growth condition and, hence, on the mappings g for which the result holds. In Theorem 3.14 we were able to replace the growth condition by a boundedness assumption. This can also be done now. Again the proof of the following result is a verbatim copy of that of Theorem 3.14.

THEOREM 4.8. *Theorem* 4.7 *remains true if the second condition (growth condition) is replaced by the condition that the set* $\{z|g(\bar{x},z) \leq 0\}$ *is bounded. Then,* (4.11) *becomes*

(4.12)

$$\nabla\varphi(x) = \int_{v\in\mathbb{S}^{m-1}} -\frac{2\rho^{x,v}(x,v)f_{m,\nu}(m^{-1}[\rho^{x,v}(x,v)]^2)}{m\langle\nabla_z g(x,\rho^{x,v}(x,v)Lv),Lv\rangle}\nabla_x g(x,\rho^{x,v}(x,v)Lv)\,d\mu_\zeta(v)$$

for all $x \in U$. *Moreover, this result holds for all* $\nu \geq 1$.

Remark 4.2. Theorem 4.8 in particular covers the case when ξ follows a multivariate Cauchy distribution, i.e., $\xi \sim \mathcal{T}(0,R,1)$. This case was excluded in Theorem 4.7.

5. Concluding remarks. In this paper we have provided representations of the gradients to convex probability functions as integrals with respect to uniform distribution over the unit sphere. This was possible in the case of Gaussian or alternative distributions (like log-normal or Student's). Having such representations, one may hope to solve corresponding probabilistically constrained optimization problems by applying nonlinear programming methods and exploiting Deák's sampling scheme of the sphere to simultaneously approximate values and gradients of the given probability functions. Proving the usefulness of this approach for numerical purposes will be the object of future research. A generalization from single random inequalities toward random inequality systems seems to be possible with an appropriate adaptation of the ideas developed here.

Appendix.

Proof of Theorem 4.4. Let U be a neighborhood of $\bar{x}$ small enough such that $g(x,0) < 0$ for all $x \in U$. Fix an arbitrary $x \in U$. According to the definition of ξ, there exist $\vartheta \sim \mathcal{N}(0,R)$ and $u \sim \chi^2(\nu)$ such that ϑ and u are independent and

$$\varphi(x) = \mathbb{P}\left(g\left(x,\vartheta\sqrt{\frac{\nu}{u}}\right) \leq 0\right) = \int_{\left\{(y,t)|t>0,g(x,y\sqrt{\frac{\nu}{t}})\leq 0\right\}} f_{\vartheta,u}(y,t)\,dydt,$$

where $f_{\vartheta,u}$ denotes the joint density of the vector (ϑ,u). By independence, $f_{\vartheta,u}(y,t) = f_\vartheta(y)\,f_u(t)$, where f_ϑ and f_u are the densities of ϑ and u, respectively. In particular, with Γ referring to the Gamma function, it holds that

$$f_u(t) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)}t^{\nu/2-1}e^{-t/2}, & t \geq 0, \\ 0, & t < 0. \end{cases} \tag{A.1}$$

Therefore,

$$\begin{aligned}\varphi(x) &= \int_0^\infty \left(\int_{\left\{y|g(x,y\sqrt{\frac{\nu}{t}})\leq 0\right\}} f_\vartheta(y)\,dy\right) f_u(t)\,dt \\ &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)}\int_0^\infty \mathbb{P}\left(g\left(x,\vartheta\sqrt{\frac{\nu}{t}}\right) \leq 0\right) t^{\nu/2-1}e^{-t/2}dt.\end{aligned} \tag{A.2}$$

With $M := \{z \in \mathbb{R}^m | g(x,z) \leq 0\}$ one has that, for $t > 0$,

$$\mathbb{P}\left(g\left(x,\vartheta\frac{\sqrt{\nu}}{t}\right) \leq 0\right) = \mathbb{P}\left(\vartheta \in \frac{t}{\sqrt{\nu}}M\right).$$

Since $\vartheta \sim \mathcal{N}(0, R)$, (1.2) yields that, for all $t > 0$,

$$\mathbb{P}\left(g\left(x, \vartheta\frac{\sqrt{\nu}}{t}\right) \le 0\right) = \int_{v\in\mathbb{S}^{m-1}} \mu_\eta\left(\left\{r \ge 0 | \frac{\sqrt{\nu}}{t} rLv \in M\right\}\right) d\mu_\zeta$$
$$= \int_{v\in\mathbb{S}^{m-1}} \mu_\eta\left(\left\{r \ge 0 | g\left(x, \frac{\sqrt{\nu}}{t} rLv\right) \le 0\right\}\right) d\mu_\zeta,$$

where η has a chi-distribution with m degrees of freedom and ζ has a uniform distribution over $\mathbb{S}^{m-1}$. Moreover, L is a factor of the Cholesky decomposition $R = LL^T$. Let $t > 0$ be arbitrary. Assume first that $v \in F(x)$. With $g(x, 0) < 0$, let $\rho^{x,v} : \tilde{U} \times \tilde{V} \to \mathbb{R}_+$ be the function defined on certain neighborhoods $\tilde{U}, \tilde{V}$ of x and v, respectively. It follows from statement 1 in Lemma 3.2 that

$$\left\{r \ge 0 | g\left(x, \frac{\sqrt{\nu}}{t} rLv\right) \le 0\right\} = \left[0, \frac{t}{\sqrt{\nu}}\rho^{x,v}(x, v)\right].$$

If in contrast $v \in I(x)$, then $g(x, rLv) < 0$ for all $r \ge 0$, whence

$$\left\{r \ge 0 | g\left(x, \frac{\sqrt{\nu}}{t} rLv\right) \le 0\right\} = \mathbb{R}_+.$$

Combining this with (A.2), we conclude that

(A.3)

$$\varphi(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty \left(\int_{v\in F(x)} \mu_\eta\left(\left[0, \frac{t}{\sqrt{\nu}}\rho^{x,v}(x, v)\right]\right) d\mu_\zeta \right.$$
$$\left. + \int_{v\in I(x)} \mu_\eta(\mathbb{R}_+) d\mu_\zeta\right) t^{\nu/2-1} e^{-t/2} dt$$
$$= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty \left(\int_{v\in F(x)} F_\eta\left(\frac{t}{\sqrt{\nu}}\rho^{x,v}(x, v)\right) d\mu_\zeta + \mu_\zeta(I(x))\right) t^{\nu/2-1} e^{-t/2} dt$$
$$= \mu_\zeta(I(x)) + \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty t^{\nu/2-1} e^{-t/2} \int_{v\in F(x)} F_\eta\left(\frac{t}{\sqrt{\nu}}\rho^{x,v}(x, v)\right) d\mu_\zeta dt,$$

where F_η denotes the distribution function of η and we exploited that $F_\eta(0) = 0$, $\mu_\eta(\mathbb{R}_+) = 1$, and

$$\frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty t^{\nu/2-1} e^{-t/2} dt = \int_{\mathbb{R}} f_u(t) dt = 1.$$

Now, let $r \ge 0$ be arbitrary and let ζ have a Fisher–Snedecor distribution with m and ν degrees of freedom. Then $\zeta = (\nu U_m)/(m U_\nu)$, where U_m and U_ν are independent and follow χ^2-distributions with m and ν degrees of freedom, respectively. Denoting by $F_{m,\nu}$ the distribution function of ζ, we derive that

$$F_{m,\nu}(m^{-1}r^2) = \mathbb{P}\left(U_\nu^{-1} U_m \le \nu^{-1} r^2\right) = \int_{\{(\tau,t) | \nu\tau \le tr^2\}} f_{U_m,U_\nu}(\tau, t) d\tau dt,$$

where f_{U_m,U_ν} denotes the joint density of the vector (U_m, U_ν). By independence, $f_{U_m,U_\nu}(\tau, t) = f_{U_m}(\tau) f_{U_\nu}(t)$, where the single χ^2-densities are defined with appropriate degrees of freedom in (A.1). It follows that

$$\begin{aligned}F_{m,\nu}(m^{-1}r^2) &= \int_0^\infty \int_0^{tr^2/\nu} f_{U_\nu}(t) f_{U_m}(\tau)\, d\tau dt \\ &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty t^{\nu/2-1} e^{-t/2} \frac{1}{2^{m/2}\Gamma(m/2)} \int_0^{tr^2/\nu} \tau^{m/2-1} e^{-\tau/2} d\tau dt \\ &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty t^{\nu/2-1} e^{-t/2} \frac{1}{2^{m/2-1}\Gamma(m/2)} \int_0^{r\sqrt{t/\nu}} s^{m-1} e^{-s^2/2} ds dt \\ &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty t^{\nu/2-1} e^{-t/2} \int_0^{r\sqrt{t/\nu}} f_\eta(s) ds dt.\end{aligned}$$

Here, we used that the variable η introduced above has a chi-distribution with m degrees of freedom and so its density is given by

$$f_\eta(s) = \begin{cases} \frac{1}{2^{m/2-1}\Gamma(m/2)} s^{m-1} e^{-s^2/2}, & s \geq 0, \\ 0, & s < 0. \end{cases}$$

Consequently, with F_η denoting the distribution function of η,

$$\begin{aligned}F_{m,\nu}(m^{-1}r^2) &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty t^{\nu/2-1} e^{-t/2} F_\eta\left(r\sqrt{t/\nu}\right) dt \\ &= \frac{1}{2^{\nu/2-1}\Gamma(\nu/2)} \int_0^\infty s^{\nu-1} e^{-s^2/2} F_\eta\left(sr/\sqrt{\nu}\right) ds. \qquad \text{(A.4)}\end{aligned}$$

Consequently, exploiting the definition $\tilde{e}$ in the statement of Theorem 4.4, putting $r := \rho^{x,v}(x,v)$ in (A.4), and applying Fubini's theorem, we end up via (A.3) at

$$\begin{aligned}&\int_{v\in\mathbb{S}^{m-1}} \tilde{e}(x,v)\, d\mu_\zeta \\ &= \mu_\zeta(I(x)) + \int_{v\in F(x)} F_{m,\nu}(m^{-1}\left[\rho^{x,v}(x,v)\right]^2) d\mu_\zeta \\ &= \mu_\zeta(I(x)) + \frac{1}{2^{\nu/2-1}\Gamma(\nu/2)} \int_0^\infty \int_{v\in F(x)} s^{\nu-1} e^{-s^2/2} F_\eta\left(s\rho^{x,v}(x,v)/\sqrt{\nu}\right) d\mu_\zeta ds \\ &= \varphi(x). \qquad \square\end{aligned}$$

Proof of (1.5). With the notation introduced in and below (1.4), we have that

$$\begin{aligned}\mathrm{Var}_\zeta\left(\mu_\eta([0,\rho(\zeta)])\right) &= \mathbb{E}_\zeta(\mu_\eta^2([0,\rho(\zeta)])) - p^2 \\ &\leq \mathbb{E}_\zeta(\mu_\eta([0,\rho(\zeta)])) - p^2 = \int_{\mathbb{S}^{m-1}\times\mathbb{R}_+} \mathbb{I}_{[0,\rho(v)]}(r)\, d\mu_\eta(r) \times d\mu_\zeta(v) - p^2 \\ &= \mathbb{E}_\xi(\mathbb{I}_M(\xi)) - p^2 = \mathbb{E}_\xi(\mathbb{I}_M^2(\xi)) - p^2 = \mathrm{Var}_\xi\left(\mathbb{I}_M(\xi)\right),\end{aligned}$$

where we used that $\mu_\eta \leq 1$ (as a probability), that $\mathbb{I}^2 = \mathbb{I}$, and that $\xi \in M$ if and only if $\eta \leq \rho(\zeta)$ (by (1.4)). $\square$

Acknowledgment. We would like to express our gratitude to Prof. J. O. Royset for stimulating discussions and for pointing us to references [11] and [30].

REFERENCES

[1] T. Arnold, R. Henrion, A. Möller, and S. Vigerske, *A mixed-integer stochastic nonlinear optimization problem with joint probabilistic constraints*, Pac. J. Optim., 10 (2014), pp. 5–20.

[2] A. Ben-Tal and A. Nemirovski, *Robust solutions of linear programming problems contaminated with uncertain data*, Math. Program., 88 (2000), pp. 411–424.

[3] J. S. Brauchart, E. B. Saff, I. H. Sloan, and R. S. Womersley, *QMC Designs: Optimal Order Quasi Monte Carlo Integration Schemes on the Sphere*, preprint, arXiv:1208.3267v1 [math.NA], 2012.

[4] J. Bukszár, *Hypermultitrees and sharp Bonferroni inequalities*, Math. Inequal. Appl., 6 (2003), pp. 727–743.

[5] G. C. Calafiore and M. C. Campi, *The scenario approach to robust control design*, IEEE Trans. Automat. Control, 51 (2006), pp. 742–753.

[6] A. Charnes and W. W. Cooper, *Chance-constrained programming*, Management Sci., 6 (1959/1960), pp. 73–79.

[7] I. Deák, *Computing probabilities of rectangles in case of multinormal distribution*, J. Stat. Comput. Simul., 26 (1986), pp. 101–114.

[8] I. Deák, *Subroutines for computing normal probabilities of sets—Computer experiences*, Ann. Oper. Res., 100 (2000), pp. 103–122.

[9] D. Dentcheva and G. Martinez, *Regularization methods for optimization problems with probabilistic constraints*, Math. Program., 138 (2013), pp. 223–251.

[10] D. Dentcheva, A. Prékopa, and A. Ruszczyński, *Concavity and efficient points for discrete distributions in stochastic programming*, Math. Program., 89 (2000), pp. 55–77.

[11] O. Ditlevsen and H. O. Madsen, *Structural Reliability Methods*, John Wiley & Sons, Chichester, UK, 1996.

[12] Y. M. Ermoliev, T. Y. Ermolieva, G. J. Macdonald, and V. I. Norkin, *Stochastic optimization of insurance portfolios for managing exposure to catastrophic risks*, Ann. Oper. Res., 99 (2000), pp. 207–225.

[13] J. Garnier, A. Omrane, and Y. Rouchdy, *Asymptotic formulas for the derivatives of probability functions and their Monte Carlo estimations*, European J. Oper. Res., 198 (2009), pp. 848–858.

[14] A. Geletu, A. Hoffmann, M. Klöppel, and P. Li, *Monotony analysis and sparse-grid integration for nonlinear chance constrained process optimization*, Eng. Optim., 43 (2011), pp. 1019–1041.

[15] A. Genz, *Numerical computation of multivariate normal probabilities*, J. Comput. Graph Statist., 1 (1992), pp. 141–149.

[16] A. Genz and F. Bretz, *Computation of Multivariate Normal and t Probabilities*, Lecture Notes in Statist. 195, Springer, Dordrecht, The Netherlands, 2009.

[17] R. Henrion, *Gradient estimates for Gaussian distribution functions: Application to probabilistically constrained optimization problems*, Numer. Algebra Control Optim., 2 (2012), pp. 655–668.

[18] R. Henrion and A. Möller, *A gradient formula for linear chance constraints under Gaussian distribution*, Math. Oper. Res., 37 (2012), pp. 475–488.

[19] A. I. Kibzun and S. Uryas'ev, *Differentiability of probability function*, Stoch. Anal. Appl., 16 (1998), pp. 1101–1128.

[20] F. Y. Kuo and I. H. Sloan, *Quasi-Monte Carlo methods can be efficient for integration over products of spheres*, J. Complexity, 21 (2005), pp. 196–210.

[21] K. Marti, *Differentiation of probability functions: The transformation method*, Comput. Math. Appl., 30 (1995), pp. 361–382.

[22] S. Nadarajah and S. Kotz, *Mathematical properties of the multivariate t distribution*, Acta Appl. Math., 89 (2005), pp. 53–84.

[23] A. Nemirovski and A. Shapiro, *Convex approximations of chance constrained programs*, SIAM J. Optim., 17 (2006), pp. 969–996.

[24] N. J. Olieman and B. van Putten, *Estimation method of multivariate exponential probabilities based on a simplex coordinates transform*, J. Stat. Comput. Simul., 80 (2010), pp. 355–361.

[25] B. Pagnoncelli, S. Ahmed, and A. Shapiro, *Sample average approximation method for chance constrained programming: Theory and applications*, J. Optim. Theory Appl., 142 (2009), pp. 399–416.

[26] G. Ch. Pflug and H. Weisshaupt, *Probability gradient estimation by set-valued calculus and applications in network design*, SIAM J. Optim., 15 (2005), pp. 898–914.

[27] A. Prékopa, *On logarithmic concave measures and functions*, Acta Sci. Math. (Szeged), 34 (1973), pp. 335–343.
[28] A. Prékopa, *Stochastic Programming*, Kluwer, Dordrecht, The Netherlands, 1995.
[29] A. Prékopa, *Probabilistic programming*, in Stochastic Programming, Handbooks Oper. Res. Management Sci. 10, A. Ruszczyński and A. Shapiro, eds., Elsevier, Amsterdam, 2003, pp. 267–351.
[30] J. O. Royset and E. Polak, *Extensions of stochastic optimization results to problems with system failure probability functions*, J. Optim. Theory Appl., 133 (2007), pp. 1–18.
[31] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, MPS-SIAM Ser. Optim. 9, SIAM, MPS, Philadelphia, 2009.
[32] T. Szántai, *Evaluation of a special multivariate gamma distribution function*, Math. Programming Stud., no. 27 (1986), pp. 1–16.
[33] T. Szántai, *Improved bounds and simulation procedures on the value of the multivariate normal probability distribution function*, Ann. Oper. Res., 100 (2000), pp. 85–101.
[34] S. Uryas'ev, *Derivatives of probability functions and integrals over sets given by inequalities*, J. Comput. Appl. Math., 56 (1994), pp. 197–223.
[35] S. Uryas'ev, *Derivatives of probability functions and some applications*, Ann. Oper. Res., 56 (1995), pp. 287–311.
[36] S. Uryas'ev, *Derivatives of probability and integral functions: General theory and examples*, in Encyclopedia of Optimization, 2nd ed., C. A. Floudas and P. M. Pardalos, eds., Springer, New York, 2009, pp. 658–663.
[37] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati, *On probabilistic constraints induced by rectangular sets and multivariate normal distributions*, Math. Methods Oper. Res., 71 (2010), pp. 535–549.
[38] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati, *On joint probabilistic constraints with Gaussian coefficient matrix*, Oper. Res. Lett., 39 (2011), pp. 99–102.
[39] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati, *Joint chance constrained programming for hydro reservoir management*, Optim. Eng., 15 (2014), pp. 509–531.