Weierstraß-Institut

für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions

Wim van Ackooij[1], René Henrion[2]

submitted: June 17, 2013

[1] Electricité de France R&D
OSIRIS
1 avenue du Général de Gaulle
92141 Clamart
France
E-Mail: wim.van-ackooij@edf.fr

[2] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: rene.henrion@wias-berlin.de

No. 1799
Berlin 2013

2010 *Mathematics Subject Classification.* 90C15.

Key words and phrases. Stochastic optimization, probabilistic constraints, chance constraints, gradients of probability functions.

This work was supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

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. In order 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 successfully done by analytically reducing these values and gradients to values of Gaussian distribution functions and computing the latter, for instance, by Genz' 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 in order 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. Later, the result is extended to alternative distributions with an emphasis on the multivariate Student (or T-) distribution.

1 Introduction

A probabilistic constraint is an inequality of the type

$$\mathbb{P}\left(g(x,\xi) \leq 0\right) \geq p, \tag{1}$$

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) expresses the requirement that a decision vector x is feasible if and only if the random inequality system $g(x,\xi) \leq 0$ is satisfied at least with probability $p \in [0,1]$. 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 on the theory, numerics and applications of probabilistic constraints, we refer to, e.g., [27], [28], [30].

Initiated by Charnes and Cooper [8] and pioneered by Prékopa (e.g., by his celebrated log-concavity-Theorem [29]) the analysis of probabilistic constraints has attracted much attention in recent years with a focus on algorithmic approaches. Without providing an exhaustive list, we refer here to models like robust optimization [5], penalty approach [13], p- efficient points [11, 12], scenario approximation [7], sample average approximation [25] or convex approximation [23].

The present paper is motivated by the traditional nonlinear programming approach to the solution of probabilistically constrained optimization problems: from a formal viewpoint, (1) is a conventional inequality constraint $\alpha(x) \geq p$ with $\alpha(x) := \mathbb{P}\left(g(x,\xi) \leq 0\right)$. On the other hand, a major difficulty arises from the fact that typically no analytical expression is available for α. All one can hope for, in general, are tools for numerically approximating α. Beyond crude Monte Carlo estimation of the probability defining α, there exist a lot of more efficient approaches based, for instance, on graph-theoretical arguments [6], variance reduction [32], Quasi-Monte-Carlo (QMC) techniques or sparse grid numerical integration [15]. 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 α reduces to a multivariate Gaussian distribution function), Genz [16, 17] developed a numerical integration scheme combining separation and reordering of variables with randomized QMC. Using this method, one may compute values of the Gaussian distribution function at fairly good precision in reasonable time even for a few hundred random variables.

A similar technique has been proposed for the multivariate Student (or T-) distribution [17]. The numerical evaluation of other multivariate distribution functions such as Gamma or exponential distribution has been discussed, e.g., in [31, 24].

For an efficient solution of probabilistically constrained problems via numerical nonlinear optimization it is evidently not sufficient to calculate just functional values of α, one also has to have access to gradients of α. The need to calculate gradients of probability functions has been recognized a long time ago and has given rise to many papers on representing such gradients (e.g., [21], [33], [20], [26], [14]). In the separable case with Gaussian distribution mentioned above, it is well-known [27, p. 203], that partial derivatives of α can be reduced analytically to function values $\tilde{\alpha}$ of a Gaussian distribution with different parameters. This fact has three important consequences: first it allows one to employ the same efficient method by Genz available for values of Gaussian distribution functions in order to compute gradients simultaneously; second, doing so, the error in calculating ∇a can be controlled by that in calculating α [18]; third, the mentioned analytic relation can be applied inductively, in order to get similar analytic relations between function values and higher-order derivatives. Fortunately, this very special circumstance can be extended to more general models: it has been demonstrated in [1, 2, 19] how for general linear probabilistic constraints $\alpha(x) := \mathbb{P}\left(T(x)\xi \leq a(x)\right) \geq p$ under Gaussian distribution and with possibly nonregular, nonquadratic matrix $T(x)$ not only the computation of α (which is evident) but also of ∇a can be analytically reduced to the computation of Gaussian distribution functions. Combining appropriately these ideas with Genz' 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, 2, 3, 4, 19]. It seems that the same approach can be elaborated also for the multivariate Student distribution, whereas it would work for the Log-normal distribution only in the special case of $\alpha(x) = \mathbb{P}\left(b(x) \leq \xi \leq a(x)\right)$.

When considering models which are nonlinear in ξ, a reduction to distribution functions seems not to be possible any more. In that case, another approach, the so-called spherical-radial decomposition of Gaussian random vectors (see, e.g., [17]) seems to be promising both for calculating function values and gradients of α. 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

$$\mathbb{P}\left(\xi \in M\right) = \int\limits_{v \in \mathbb{S}^{m-1}} \mu_\eta\left(\{r \geq 0 : rLv \cap M \neq \emptyset\}\right) d\mu_\zeta, \tag{2}$$

where μ_η and μ_ζ are the laws of η and ζ, respectively. This probability can be numerically computed by employing an efficient sampling scheme on $\mathbb{S}^{m-1}$ proposed by Deák [9, 10]. More generally, one may approximate the integral

$$\int\limits_{v \in \mathbb{S}^{m-1}} h(v) d\mu_\zeta \tag{3}$$

for any Lebesgue measurable function $h : \mathbb{S}^{m-1} \to \mathbb{R}$. In particular, for

$$h(v) := \mu_\eta\left(\{r \geq 0 : rLv \cap M \neq \emptyset\}\right),$$

we obtain the probability (2). In this paper, we will show how - with different functions $h(v)$ - the same efficient sampling scheme can be employed in order to simultaneously compute derivatives of this probability with respect to an exterior parameter. The results may serve as a basis for a numerical treatment of nonlinear convex probabilistic constraints with Gaussian and alternative distributions via nonlinear optimization. In Section 2, a rigorous justification for differentiating under the integral sign will be given. Doig so, we arrive at sufficient conditions for continuous differentiability of probability functions in the convex and Gaussian case as well as at an explicit integrand in (3). In Section 3, the obtained results are applied to various examples involving Gaussian and alternative distributions. Particular attention is paid to the multivariate Student distribution.

2 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

$$\varphi(x) := \mathbb{P}(g(x,\xi) \leq 0), \tag{4}$$

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

Remark 2.1 *We recall that convex sets are Lebesgue measurable so that the probabilities in (4) are well-defined by virtue of ξ having a density.*

Remark 2.2 *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*

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

(4) 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 (4), we may indeed assume without loss of generality, that $\xi \sim \mathcal{N}(0,R)$.

By (2) and (4), we have, for all $x \in \mathbb{R}^n$, that

$$\varphi(x) = \int\limits_{v \in \mathbb{S}^{m-1}} \mu_\eta(\{r \geq 0 : g(x, rLv) \leq 0\})\, d\mu_\zeta = \int\limits_{v \in \mathbb{S}^{m-1}} e(x,v) d\mu_\zeta \tag{5}$$

for

$$e(x,v) := \mu_\eta(\{r \geq 0 : g(x,rLv) \leq 0\}) \quad \forall x \in \mathbb{R}^n \ \forall v \in \mathbb{S}^{m-1}. \tag{6}$$

According to the possibility of evaluating (3) for instance by Deàk's method, 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:

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

The following Lemma collects some elementary properties needed later:

Lemma 2.1 *Let $x \in \mathbb{R}^n$ be such that $g(x,0) < 0$. Then,*

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 (6).*

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

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

This proves 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, 4. holds true.

$\square$

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

Lemma 2.2 *Let (x, v) be such that $g(x, 0) < 0$ and $v \in F(x)$. Then, there exist neighbourhoods 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 3. in Lemma 2.1, we have that

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

This allows to apply the Implicit Function Theorem to the equation $g(x, rLv) = 0$ and to derive the existence of neighbourhoods 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

$$g(x', r'Lv') = 0, (x', v', r') \in U \times V \times W \Leftrightarrow r' = \rho^{x,v}(x', v'), \ (x', v') \in U \times V \tag{7}$$

holds true. By continuity of $\rho^{x,v}$, we may shrink the neighbourhoods 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 (7), $g(x', \rho^{x,v}(x', v')Lv') = 0$, 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 (7) W may be replaced by $\mathbb{R}_+$ which proves 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 2.

$\square$

The preceding Lemma allows us to observe the following:

Lemma 2.3 *Let $x \in \mathbb{R}^n$ be such that $g(x,0) < 0$. Then,*

1 *If $v \in F(x)$ then there exist neighbourhoods 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 (6), 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 2.2.*

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 1. in Lemma 2.2, 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, (6) implies that

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

Now, 1. 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 not $\rho^{x_k,v_k}(x_k,v_k) \to \infty$, 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 2.2 to the points (x_{k_l},v_{k_l}), and so we infer from 1. in this Lemma that $g\left(x_{k_l},\rho^{x_{k_l},v_{k_l}}(x_{k_l},v_{k_l})\,Lv_{k_l}\right) = 0$ 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 2.

□

Corollary 2.1 *The function $e : \mathbb{R}^n \times \mathbb{S}^{m-1} \to \mathbb{R}$ defined in (6) 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 2.1, there are two possibilities: if $v \in F(x)$, then the function $\rho^{x,v}$ is defined on a neighbourhood of (x,v) and is continuous there by Lemma 2.2. Moreover, in this case, e has the representation given in 1. of Lemma 2.3. 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 2. of Lemma 2.1. From 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 not $e(x_k,v_k) \to 1$. Then, there is a subsequence (x_{k_l},v_{k_l}) and some $\varepsilon > 0$ such that for all l

$$|e(x_{k_l},v_{k_l}) - 1| \geq \varepsilon. \tag{8}$$

By 4. in Lemma 2.1, $v_{k_l} \notin I(x_{k_l})$, whence $v_{k_l} \in F(x_{k_l})$ for all l due to $v_{k_l} \in \mathbb{S}^{m-1}$ and 2. in Lemma 2.1. Then, $\rho^{x_{k_l},v_{k_l}}(x_{k_l},v_{k_l}) \to \infty$ by 2. of Lemma 2.3. 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 1. of Lemma 2.3 in order 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 (8) and, hence, again by 4. in Lemma 2.1,

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

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

□

Corollary 2.2 *For any $x \in \mathbb{R}^n$ with $g(x,0) < 0$ and $v \in F(x)$ the partial derivative w.r.t x of the function $e : \mathbb{R}^n \times \mathbb{S}^{m-1} \to \mathbb{R}$ defined in (6) exists and is given by*

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

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 2.2.

Proof. By 1. in Lemma 2.3 we have that $e(x',v') = F_\eta(\rho^{x,v}(x',v'))$ for all x' in a neighbourhood of x and v' in a neighbourhood of v. Differentiation with respect to x yields

$$\nabla_x e(x',v') = \chi\left(\rho^{x,v}\left(x',v'\right)\right)\nabla_x \rho^{x,v}\left(x',v'\right) \tag{9}$$

due to $F'_\eta(\tau) = \chi(\tau)$ for $\tau > 0$. In particular, $\nabla_x e(x,v) = \chi\left(\rho^{x,v}(x,v)\right)\nabla_x \rho^{x,v}(x,v)$. Now, the assertion follows from 2. in Lemma 2.2.

□

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

Definition 2.1 *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 neighbourhood $U(x)$ such that*

$$\left\|\nabla_x g\left(x',z\right)\right\| \le \|z\|^{\varkappa} \quad \forall x' \in U(x)\ \forall z : \|z\| \ge C.$$

Lemma 2.4 *Let x be such that $g(x,0) < 0$ and 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 2. in Lemma 2.3. Referring to the neighbourhood $U(x)$ from Definition 2.1, we verify that for k sufficiently large

$$\left\|\nabla_x g\left(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k\right)\right\| \le \left[\rho^{x_k,v_k}(x_k,v_k)\right]^{\varkappa} \|Lv_k\|^{\varkappa} \le \|L\|^{\varkappa}\left[\rho^{x_k,v_k}(x_k,v_k)\right]^{\varkappa} \tag{10}$$

(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 $g(x_k,0) \le -\delta_1 < 0$ for k sufficiently large. Since $g\left(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k\right) = 0$ (see 1. in Lemma 2.2), 3. in Lemma 2.1 provides that

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

Therefore,

$$\left\langle \nabla_z g\left(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k\right), Lv_k\right\rangle \ge \delta_1\left[\rho^{x_k,v_k}(x_k,v_k)\right]^{-1} > 0. \tag{11}$$

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 2.2 with (10) and (11) in order to derive that

$$\begin{aligned} \|\nabla_x e(x_k,v_k)\| = \left\|\frac{\chi\left(\rho^{x_k,v_k}(x_k,v_k)\right)}{\left\langle \nabla_z g\left(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k\right), Lv_k\right\rangle}\nabla_x g\left(x_k, \rho^{x_k,v_k}(x_k,v_k) Lv_k\right)\right\| &\le \\ \delta_1^{-1}\rho^{x_k,v_k}(x_k,v_k)\cdot\delta_2\left[\rho^{x_k,v_k}(x_k,v_k)\right]^{m-1} e^{-\left[\rho^{x_k,v_k}(x_k,v_k)\right]^2/2}\cdot\|L\|^{\varkappa}\left[\rho^{x_k,v_k}(x_k,v_k)\right]^{\varkappa} &= \\ \delta_1^{-1}\delta_2\|L\|^{\varkappa}\left[\rho^{x_k,v_k}(x_k,v_k)\right]^{\varkappa+m} e^{-\left[\rho^{x_k,v_k}(x_k,v_k)\right]^2/2} \to_k \quad &0, \end{aligned} \tag{12}$$

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 2.3 *One may observe from the proof of Lemma 2.4 that a weaker growth condition than that in Definition 2.1 (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 neighbourhood $U(x)$ such that*

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

and observe that the key estimate (12) of Lemma 2.4 becomes

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

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 hold upon requiring the above exponential growth condition instead.

Corollary 2.3 *Let x be such that $g\left(x, 0\right) < 0$ and that g satisfies the polynomial growth condition at x. Then, for any $v \in \mathbb{S}^{m-1}$ the partial derivative w.r.t x of the function e exists at (x, v) and is given by*

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

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 2.2.

Proof. Thanks to Corollary 2.2 and to 2. in Lemma 2.1 it is sufficient to show that $\nabla_x e(x, v) = 0$ for $v \in I(x)$. We shall show that, for any $i \in \{1, \ldots, m\}$

$$\lim_{t \uparrow 0} \frac{e(x + tu_i, v) - e(x, v)}{t} = 0, \tag{13}$$

where u_i is the i-th 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 (13) is wrong. Since $e(x, v) = 1$ (by 4. in Lemma 2.1) and $e(x + tu_i, v) \leq 1$ for all t (by definition of e as a probability in (6)), it follows that the quotient in (13) is always non-positive and, thus, negation of (13) implies the existence of some $\varepsilon > 0$ and of a sequence $t_k \uparrow 0$ such that

$$\frac{e(x + t_k u_i, v) - e(x, v)}{t_k} \geq \varepsilon. \tag{14}$$

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 4. in Lemma 2.1), thus contradicting (14). 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 2.1) provide that $\alpha > t_k$. We infer that $e(x + \tau u_i, v) < 1$ for all $\tau \in [t_k, \alpha)$ and, hence,

$$v \in F(x + \tau u_i) \quad \forall \tau \in [t_k, \alpha) \tag{15}$$

(once more by 2. and 4. in Lemma 2.1). 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 2.2 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 (14). Now, since k was arbitrarily fixed, we have constructed a sequence τ_k^* such that $t_k < \tau_k^* \leq 0$ such that

$$\langle \nabla_x e(x + \tau_k^* u_i, v), u_i \rangle \geq \varepsilon \quad \forall k. \tag{16}$$

Since $t_k \uparrow 0$, we also have that $\tau_k^* \uparrow 0$. Moreover, $v \in F(x + \tau_k^* u_i)$ by (15). Due to our assumption that g satisfies the polynomial growth condition at x and due to $v \in I(x)$, Lemma 2.4 yields that $\lim_{k \to \infty} \nabla_x e(x_k, v) = 0$ which contradicts (16). This proves our Corollary.

□

Corollary 2.4 *Let x be such that $g(x, 0) < 0$ and 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. Let also $(x_k, v_k) \to (x, v)$ be an arbitrary sequence with $v_k \in \mathbb{S}^{m-1}$. If $v \in F(x)$, then relation (9) holds true locally around (x, v). In particular, for k large enough,

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

where the convergence follows from the continuity of the chi-density and of $\nabla_x \rho^{x,v}$ as a result of Lemma 2.2. Hence, in 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 2.3. 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 2.3 (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, hence continue to hold true for the x_{k_l}). Now, Lemma 2.4 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 our Corollary.

□

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

Theorem 2.2 *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 (4), 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 $\bar{x}$ (Def. 2.1).

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

$$\nabla\varphi\left(x\right)=-\int\limits_{v\in F(x)}\frac{\chi\left(\rho^{x,v}\left(x,v\right)\right)}{\left\langle\nabla_{z}g\left(x,\rho^{x,v}\left(x,v\right)Lv\right),Lv\right\rangle}\nabla_{x}g\left(x,\rho^{x,v}\left(x,v\right)Lv\right)d\mu_{\zeta}(v)\quad\forall x\in U.\tag{17}$$

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 2.2.

Proof. Since $\xi\sim\mathcal{N}\left(0,R\right)$, the probability function φ gets the representation (5). With $g\left(\bar{x},0\right)<0$, let U be a sufficiently small neighbourhood 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 (6) exists on $U\times\mathbb{S}^{m-1}$ by Corollary 2.3 and is continuous there by Corollary 2.4. By compactness of $\mathbb{S}^{m-1}$, there exists some $K>0$ such that

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

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

$$\|\nabla_x e(x,v)\|\leq 2K\quad\forall x\in U\;\forall v\in\mathbb{S}^{m-1}.\tag{18}$$

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 (5) under the integral sign:

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

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

$$\nabla\varphi\left(x\right)=\int\limits_{v\in\mathbb{S}^{m-1}}\nabla_x e(x,v)d\mu_\zeta\quad\forall x\in U.\tag{19}$$

Exploiting once more the dominance argument from (18), 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 (17) follows directly from Corollary 2.3.

□

Remark 2.4 *Evidently, formula (17) 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 $r\geq 0$ at all. If not so ($v\in I(x)$), then such v does not contribute to the (approximated) integral in (17). Otherwise ($v\in F(x)$), one has to evaluate the integrand in (17) 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 2.2. 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 2.2. Also assumption 1. of the Theorem is not restrictive. This will come as a consequence of the following proposition:

Proposition 2.1 *With g and φ as in Theorem 2.2, 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 2.2 we may assume that $\xi \sim \mathcal{N}(0, R)$ so that φ gets the representation (5). 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 means that violation of the first assumption in Theorem 2.2 implies that $g(\bar{x}, z) \geq 0$ for all z or that $\varphi(\bar{x}) \leq 1/2$. A typical application of Theorem 2.2 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 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 to calculate a gradient at all.

Finally, turning to condition 2. of Theorem 2.2 (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 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 2.5 *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 $g\left(x, \|z_n\|^{-1} t z_n\right) \leq 0$ for n sufficiently large. Passing to the limit, we get that $g(x, tz) \leq 0$. Thus, as $t \geq 0$ was arbitrary,

$$g(x, tz) \leq 0 \quad \forall t \geq 0. \tag{20}$$

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 (20):

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

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 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 2.2 *Let g be as in Lemma 2.5 and $\bar{x} \in \mathbb{R}^n$ with $g(\bar{x}, 0) < 0$. If $M(\bar{x})$ is bounded, then there is a neighbourhood 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 was not true, then by virtue of Lemma 2.5 there exists a sequence $x_n \to \bar{x}$ such that $I(x_n) \neq \emptyset$ for all $n \in \mathbb{N}$. By 1. in Lemma 2.1 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 (20) 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 2.5. This is a contradiction with our assumption.

□

Now we are in a position to state an alternative variant of Theorem 2.2 which does not require the verification of the growth condition:

Theorem 2.3 *Theorem 2.2 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, (17) becomes*

$$\nabla\varphi(x) = -\int\limits_{v\in\mathbb{S}^{m-1}} \frac{\chi\left(\rho^{x,v}(x,v)\right)}{\left\langle \nabla_z g\left(x, \rho^{x,v}(x,v)Lv\right), Lv\right\rangle} \nabla_x g\left(x, \rho^{x,v}(x,v)Lv\right) d\mu_\zeta(v) \quad \forall x \in U. \tag{21}$$

Proof. As in the proof of Theorem 2.2, the function e is continuous on $U \times \mathbb{S}^{m-1}$ by Corollary 2.1 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 2.2 that - after possibly shrinking the neighbourhood U of $\bar{x}$ - the set $\{z|g(x,z) \leq 0\}$ remains bounded for all $x \in U$. Lemma 2.5 implies that $I(x) = \emptyset$ or, equivalently according to 2. in Lemma 2.1 - that $F(x) = \mathbb{S}^{m-1}$ for all $x \in U$. Then, Corollary 2.2 yields that $\nabla_x e$ exists on $U \times \mathbb{S}^{m-1}$ and is given by

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

Since all occurring functions are continuous, the same holds true for $\nabla_x e$. Now the same argument as in the proof of Theorem 2.2 allows us to derive (19) which along with the formula for $\nabla_x e$ above yields (21).

□

3 Selected Examples

In this section we are going to discuss some instances of the probabilistic constraint (1) to which our gradient formulae obtained in Theorems 2.2 and 2.3 apply and thus could be used in the numerical solution of corresponding optimization problems.

3.1 Gaussian distributions

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

$$\mathbb{P}[\langle f(\xi), h_1(x)\rangle \leq h_2(x)] \geq p \tag{22}$$

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

Proposition 3.1 *In the probabilistic constraint (22), let f, h_1, h_2 be continuously differentiable, 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:*

$$\|f(z)\| \leq \|z\|^{\varkappa} \quad \forall z : \|z\| \geq C$$

for certain $\varkappa, C > 0$. Then the probability function $\varphi(x) := \mathbb{P}\left[\langle f(\xi), h_1(x)\rangle \leq h_2(x)\right]$ defining the constraint (22) is continuously differentiable on a neighbourhood U of $\bar{x}$ and its gradient is given by

$$\nabla\varphi(x) = \int\limits_{v\in F(x)} \frac{\chi\left(\rho^{x,v}(x,v)\right)}{\left\langle h_1^T(x)\nabla f\left(\rho^{x,v}(x,v)Lv\right), Lv\right\rangle}\left(\nabla h_2(x) - \left[f(\rho^{x,v}(x,v)Lv)\right]^T\nabla h_1(x)\right) d\mu_\zeta(v) \quad \forall x \in U. \tag{23}$$

Proof. In our setting the general function g in (4) becomes $g(x,z) = \langle f(z), h_1(x)\rangle - h_2(x)$. The continuous differentiability and convexity with respect to the second argument of g are evident from our assumptions. Moreover, $g(\bar{x}, 0) < 0$. As for the growth condition, let U be a neighbourhood of $\bar{x}$ on which $\max\{\|\nabla h_1\|, \|\nabla h_2\|\} \leq K$ for some $K > 0$. Then, taking without loss of generality, the maximum norm, we have that

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

Consequently, we may apply Theorem 2.2. (23) follows immediately from (17) for the given form of the function g.

□

3.2 Gaussian-like distributions

We are now going to apply Theorem 2.2 to probabilistic constraints with random vectors having non-Gaussian distributions. In a first case, we consider a linear probabilistic constraint

$$\mathbb{P}[\langle \eta, x\rangle \leq b] \geq p \tag{24}$$

with a random vector η whose components η_i $(i = 1,\dots,l)$ 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 $k = 1,\dots,n_i$. We are interested in the gradient of the probability function $\varphi(x) := \mathbb{P}[\langle\eta, x\rangle \leq b]$. Define a Gaussian random vector with independent components

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

Clearly, $\eta \sim f(\xi)$, where $f_i(z) := \sum_{k=1}^{n_i} z_{i,k}^2$ for $i = 1,\dots,l$ and z is partitioned in the same way as ξ above. Then, the probability function defining (24) becomes

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

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 3.2 *In (24), let $b > 0$. Consider any feasible point $\bar{x}$ of (24) satisfying $\bar{x}_i > 0$ for $i = 1, \ldots, n$. Then the probability function φ is continuously differentiable on a neighbourhood U of $\bar{x}$ and its gradient is given by*

$$\nabla\varphi(x) = -\frac{\sqrt{b}}{2} \int\limits_{v\in\mathbb{S}^{m-1}} \frac{\chi\left(\sqrt{b/\langle f(v), x\rangle}\right)}{\langle f(v), x\rangle^{3/2}} [f(v)]^T \, d\mu_\zeta(v) \quad \forall x \in U. \tag{25}$$

Proof. In our setting the general function g in (4) 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 neighbourhood of $\bar{x}$. Evidently, the result of Theorems 2.2 and 2.3 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 neighbourhood 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

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

whence the set on the left-hand side is bounded. Altogether, this allows us to invoke Theorem 2.3 and to derive the validity of formula (21). 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

$$r = \sqrt{b/\langle f(v), x\rangle}. \tag{26}$$

Next, we calculate

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

Combination of these last relations with (26) provides formula (25).

□

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 lognormal distribution if the vector $\xi := \log\eta$ (componentwise logarithm) has a Gaussian distribution. We consider now a probabilistic constraint of type

$$\mathbb{P}[\langle\eta, x\rangle \leq h(x)] \geq p \tag{27}$$

where η is an m-dimensional random vector with lognormal 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 2.2) 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 3.3 *In the setting above, assume that $\bar{x}$ satisfies $\bar{x}_i > 0$ for $i = 1, \dots, m$. Assume moreover that h is continuously differentiable and that $h(\bar{x}) > \sum_{i=1}^m \bar{x}_i$. Then,*

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

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 (4) 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 (24) this weakened condition is enough in the context of Theorem 2.2). Moreover, $g(\bar{x},0) = \sum_{i=1}^m \bar{x}_i - h(\bar{x}) < 0$. In order to apply Theorem 2.2, it is sufficient to verify the exponential growth condition of Remark 2.3 (note that the originally imposed polynomial growth condition would not hold true here). To this aim, let U be a neighbourhood of $\bar{x}$ on which $\|\nabla h\| \leq K$ for some $K > 0$. Then, with respect to the maximum norm, we get that

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

Hence, the exponential growth condition of Remark 2.3 is satisfied. This allows us to apply Theorem 2.2. Inserting the corresponding derivative formulae for g, we derive that φ is continuously differentiable on a neighbourhood U of $\bar{x}$ and its gradient is given by

$$\nabla\varphi(x) = -\int\limits_{v\in F(x)} \frac{\chi\left(\rho^{x,v}(x,v)\right)}{\sum_{i=1}^m x_i e^{\rho^{x,v}(x,v)\langle L_i, v\rangle} \langle L_i, v\rangle} \left[e^{\rho^{x,v}(x,v)Lv} - \nabla h(x)\right] d\mu_\zeta(v) \quad \forall x \in U. \tag{28}$$

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 neighbourhood U of $\bar{x}$ in the formula above to be small enough that $g(x,0) < 0$ and $x_i > 0$ for $i = 1, \dots, m$ and for all $x \in U$ (recall that $g(\bar{x},0) < 0$ and $\bar{x}_i > 0$ for $i = 1, \dots, m$). We claim that for all $x \in U$ the set $I(x)$ introduced below (6) can be written as

$$I(x) = \left\{v \in \mathbb{S}^{m-1} | Lv \leq 0\right\}. \tag{29}$$

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 1. of Lemma 2.1. Conversely, let $x \in U$ and $v \in I(x)$ be arbitrary. Then, $\left\langle e^{rLv}, x\right\rangle < h(x)$ for all $r > 0$. Define $J := \{i | L_i v > 0\}$. It follows from $x_i > 0$ for $i = 1, \dots, m$ 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 (29). Since, by definition, $F(x) = \mathbb{S}^{m-1} \backslash I(x)$, we may plug the information from (29) into (28) in order to derive our asserted formula.

□

3.3 Student (or T-) distribution

As a last application, we are going to consider probabilistic constraints of type (1), where the random vector ξ follows a so-called multivariate Student 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 μ 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 (4) but this time for a T-variable rather than for a Gaussian one.

Remark 3.1 *Using the definition of a T- distribution, we may copy the arguments of Remark 2.2 in order to convince ourselves that in the consideration of (4) 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 (4) being continuously differentiable and convex with respect to the second argument.*

In a first step, we provide an expression for the probability function (4) in case of a T-distribution:

Theorem 3.1 *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 neighbourhood U of $\bar{x}$ such that the probability function (4) admits the representation*

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

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 2.2 and $F(x)$ and $I(x)$ are defined in Lemma 2.1.

The proof of this Theorem is left for the Appendix. Now, we may copy the proof of Corollary 2.1 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 2.3 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 in the Theorem above. Next, we may copy the proof of Corollary 2.2 (again with the appropriate replacements) and get the following:

Corollary 3.1 *For any $x \in \mathbb{R}^n$ with $g(x, 0) < 0$ and $v \in F(x)$ the partial derivative w.r.t x of the function $\tilde{e} : \mathbb{R}^n \times \mathbb{S}^{m-1} \to \mathbb{R}$ defined in Theorem 3.1 exists and is given by*

$$\nabla_x \tilde{e}(x, v) = -2\rho^{x,v}(x, v)\frac{f_{m,\nu}\left(m^{-1}[\rho^{x,v}(x, v)]^2\right)}{m\langle\nabla_z g(x, \rho^{x,v}(x, v)Lv), Lv\rangle}\nabla_x g(x, \rho^{x,v}(x, v)Lv) \tag{30}$$

where

$$f_{m,\nu}(t) = \begin{cases} \frac{\Gamma(m/2+\nu/2)}{\Gamma(m/2)\Gamma(\nu/2)} m^{m/2}\nu^{\nu/2}t^{m/2-1}(mt+\nu)^{-(m+\nu)/2} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{31}$$

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 2.2 and L is a factor of the Cholesky decomposition $R = LL^T$.

It is the equivalent of Lemma 2.4 that requires some additional conditions and work:

Lemma 3.1 *Let x be such that $g(x,0) < 0$ and that g satisfies the polynomial growth condition at x with coefficient $\varkappa < \nu$ (Def. 2.1). 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 2. in Lemma 2.3. The arguments of Lemma 2.4 allow us to deduce that for k sufficiently large the estimates (10) and (11) still hold. Using (31), we may combine Corollary 3.1 with (10) and (11) in order to derive that

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

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

□

Upon having established Lemma 3.1 the same arguments of Corollary 2.3 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 2.3 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 (30) 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 2.4 one establishes the continuity of $\nabla_x \tilde{e}$ upon noting that $f_{m,\nu}(t)$ defined in (31) is also continuous. We thus arrive at the following key result, of which the proof is a verbatim copy of that of Theorem 2.2 (Again e and $\tilde{e}$ have the same properties).

Theorem 3.2 *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 (4), 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 $\bar{x}$ (Def. 2.1) with coefficient $\varkappa < \nu$.*

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

$$\nabla\varphi(x) = \int\limits_{v\in F(x)} -\frac{2\rho^{x,v}(x,v) f_{m,\nu}(m^{-1}[\rho^{x,v}(x,v)]^2)}{m\langle \nabla_z g(x, \rho^{x,v}(x,v) Lv), Lv\rangle} \nabla_x g(x, \rho^{x,v}(x,v) Lv)\, d\mu_\zeta(v) \quad \forall x \in U. \tag{32}$$

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 2.2.

4 Concluding Remarks

We have provided in this paper 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). Having such representation, one may hope for solving corresponding probabilistically constrained optimization problems by applying nonlinear programming methods and exploit Deák's sampling scheme of the sphere in order to simultaneously approximate values and gradients of the given probability functions. To prove the usefulness of this approach for numerical purposes will be the object of future research. A generalization from single random inequalities towards random inequality systems seems to be possible with an appropriate adaptation of the ideas developed here.

5 Appendix

Proof of Theorem 3.1. Let U be a neighbourhood 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

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

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

$$f_u(t) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} t^{\nu/2-1} e^{-t/2} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{33}$$

Therefore,

$$\varphi(x) = \int_0^\infty \left(\int\limits_{\left\{y|g(x,y\sqrt{\frac{\nu}{t}})\leq 0\right\}} f_\vartheta(y)\, dy \right) f_u(t)\, dt = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty \mathbb{P}\left(g(x, \vartheta\sqrt{\frac{\nu}{t}}) \leq 0\right) t^{\nu/2-1} e^{-t/2} dt. \tag{34}$$

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

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

Since $\vartheta \sim \mathcal{N}(0,R)$, (2) yields that for all $t > 0$

$$\begin{aligned} \mathbb{P}\left(g(x, \vartheta\frac{\sqrt{\nu}}{t}) \leq 0\right) &= \int\limits_{v \in \mathbb{S}^{m-1}} \mu_\eta\left(\{r \geq 0 | \frac{\sqrt{\nu}}{t} rLv \in M\}\right) d\mu_\zeta \\ &= \int\limits_{v \in \mathbb{S}^{m-1}} \mu_\eta\left(\{r \geq 0 | g(x, \frac{\sqrt{\nu}}{t} rLv) \leq 0\}\right) d\mu_\zeta, \end{aligned}$$

where η has a χ-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 neighbourhoods $\tilde{U}, \tilde{V}$ of x and v, respectively. It follows from 1. in Lemma 2.2 that

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

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

$$\{r \geq 0 | g(x, \frac{\sqrt{\nu}}{t} rLv) \leq 0\} = \mathbb{R}_+.$$

Combining this with (34), we conclude that

$$\begin{aligned}
\varphi(x) &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty \left(\int_{v\in F(x)} \mu_\eta \left(\left[0, \frac{t}{\sqrt{\nu}} \rho^{x,v}(x,v) \right] \right) d\mu_\zeta + \int_{v\in I(x)} \mu_\eta(\mathbb{R}_+)\, d\mu_\zeta \right) t^{\nu/2-1} e^{-t/2} dt \\
&= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty \left(\int_{v\in F(x)} F_\eta \left(\frac{t}{\sqrt{\nu}} \rho^{x,v}(x,v) \right) d\mu_\zeta + \mu_\zeta(I(x)) \right) t^{\nu/2-1} e^{-t/2} dt \\
&= \mu_\zeta(I(x)) + \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty t^{\nu/2-1} e^{-t/2} \int_{v\in F(x)} F_\eta \left(\frac{t}{\sqrt{\nu}} \rho^{x,v}(x,v) \right) d\mu_\zeta dt, \qquad (35)
\end{aligned}$$

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 $r \geq 0$ be arbitrary and let ζ have a Fisher-Snedecor (F-) 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- squared distributions with m and ν degrees of freedom, respectively. Denoting by $F_{m,\nu}$ the distribution function of ζ, we derive that

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

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 (33). 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 χ-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,

$$\begin{aligned}
F_{m,\nu}(m^{-1}r^2) &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty t^{\nu/2-1} e^{-t/2} F_\eta\left(r\sqrt{t/\nu} \right) dt \\
&= \frac{1}{2^{\nu/2-1}\Gamma(\nu/2)} \int_0^\infty s^{\nu-1} e^{-s^2/2} F_\eta\left(sr/\sqrt{\nu} \right) ds. \qquad (36)
\end{aligned}$$

Consequently, exploiting the definition $\tilde{e}$ in the statement of Theorem 3.1, putting $r := \rho^{x,v}(x,v)$ in (36) and applying Fubini's theorem, we end up via (35) at

$$\begin{aligned}
\int\limits_{v\in\mathbb{S}^{m-1}} \tilde{e}(x,v)\, d\mu_\zeta &= \\
\mu_\zeta(I(x)) + \int\limits_{v\in F(x)} F_{m,\nu}(m^{-1}[\rho^{x,v}(x,v)]^2) d\mu_\zeta &= \\
\mu_\zeta(I(x)) + \frac{1}{2^{\nu/2-1}\Gamma(\nu/2)} \int\limits_0^\infty \int\limits_{v\in F(x)} s^{\nu-1} e^{-s^2/2} F_\eta\left(s\rho^{x,v}(x,v)/\sqrt{\nu}\right) d\mu_\zeta ds &= \varphi(x).
\end{aligned}$$

$\square$

References

[1] W. Van Ackooij, R. Henrion, A. Möller and R. Zorgati, On probabilistic constraints induced by rectangular sets and multivariate normal distributions, Mathematical Methods of Operations Research 71 (2010), 535–549.

[2] W. Van Ackooij, R. Henrion, A. Möller and R. Zorgati, On joint probabilistic constraints with Gaussian coefficient matrix, Operations Research Letters 39 (2011), 99–102.

[3] W. Van Ackooij, R. Zorgati, R. Henrion and A. Möller, Joint Chance Constrained Programming for Hydro Reservoir Management, submitted.

[4] T. Arnold, R. Henrion, A. Möller and S. Vigerske, A mixed-integer stochastic nonlinear optimization problem with joint probabilistic constraints, MATHEON-Preprint 1009 (2013), submitted.

[5] A. Ben-Tal and A. Nemirovski, Robust solutions of Linear Programming problems contaminated with uncertain data, Mathematical Programming 88 (2002), 411–424.

[6] J. Bukszár, Hypermultitrees and sharp Bonferroni inequalities, Mathematical Inequalities and Applications 6 (2003), 727–743.

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

[8] A. Charnes and W. Cooper, Chance-Constrained Programming, Management Science 6 (1959), 73–89.

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

[10] I. Deák, Subroutines for Computing Normal Probabilities of Sets – Computer Experiences, Annals of Operations Research, 100 (2000), 103–122.

[11] D. Dentcheva, A. Prékopa and A. Ruszczyński, Concavity and efficient points of discrete distributions in probabilistic programming, Mathematical Programming 89 (2000), 55–77.

[12] D. Dentcheva, G. Martinez, Regularization methods for optimization problems with probabilistic constraints, Mathematical Programming 138 (2013), 223–251.

[13] Y.M. Ermoliev, T.Y. Ermolieva, G.J. Macdonald and V.I. Norkin, Stochastic Optimization of Insurance Portfolios for Managing Exposure to Catastrophic Risk, Annals of Operations Research 99 (2000), 207–225.

[14] 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), 848-858.

[15] A. Geletu, A. Hoffmann, M. Klöppel and P. Li, Monotony analysis and sparse-grid integration for nonlinear chance constrained process optimization, Engineering Optimization 43 (2011), 1019–1041.

[16] A. Genz, Numerical Computation of Multivariate Normal Probabilities, Journal of Computational and Graphical Statistics 1 (1992), 141–149.

[17] A. Genz and F. Bretz, *Computation of Multivariate Normal and t Probabilities*, Lecture Notes in Statistics, vol. 195, Springer, Heidelberg, 2009.

[18] R. Henrion, Gradient estimates for Gaussian distribution functions: Application to probabilistically constrained optimization problems, Numerical Algebra, Control and Optimization 2 (2012), 655–668.

[19] R. Henrion and A. Möller, A gradient formula for linear chance constraints under Gaussian distribution, Mathematics of Operations Research 37 (2012), 475–488.

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

[21] K. Marti, Differentiation of probability functions: The transformation method, Computers Math. Applic. 30 (1995), 361–382.

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

[23] A. Nemirovski and A. Shapiro, Convex approximations of chance constrained programs, SIAM J. Optim. 17 (2006), 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), 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), 399–416.

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

[27] A. Prékopa, *Stochastic Programming*, Kluwer, Dordrecht, 1995.

[28] A. Prékopa, *Probabilistic Programming*, Stochastic Programming (A. Ruszczyński and A. Shapiro, eds.), Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.

[29] A. Prékopa, On logarithmic concave measures and functions, Acta Scientiarium Mathematicarum (Szeged) 34 (1973), 335–343.

[30] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming*, MPS-SIAM series on optimization 9, 2009.

[31] T. Szántai, Evaluation of a special multivariate Gamma distribution, Mathematical Programming Study 27 (1996), 1–16.

[32] T. Szántai, Improved Bounds and Simulation Procedures on the Value of the Multivariate Normal Probability Distribution Function, Annals of Operations Research 100 (2000), 85–101.

[33] S. Uryas'ev, *Derivatives of probability functions and integrals over sets given by inequalities*, J. Comp. Appl. Math. 56 (1994), 197–223.