Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

(Sub-) Gradient formulae for probability functions of random inequality systems under Gaussian distribution

Wim van Ackooij[1], René Henrion[2]

submitted: February 22, 2016

[1] EDF 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. 2230
Berlin 2016

2010 *Mathematics Subject Classification.* 90C15.

Key words and phrases. Stochastic optimization, gradients of probability functions, spheric radial decomposition, multivariate Gaussian distribution, Clarke subdifferential, Mordukhovich subdifferential, probabilistic constraint.

The second author gratefully acknowledges support by the *FMJH Program Gaspard Monge in optimization and operations research* including support to this program by EDF as well as support by the *Deutsche Forschungsgemeinschaft* within Project B04 in CRC TRR 154.

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

We consider probability functions of parameter-dependent random inequality systems under Gaussian distribution. As a main result, we provide an upper estimate for the Clarke subdifferential of such probability functions without imposing compactness conditions. A constraint qualification ensuring continuous differentiability is formulated. Explicit formulae are derived from the general result in case of linear random inequality systems. In the case of a constant coefficient matrix an upper estimate for even the smaller Mordukhovich subdifferential is proven.

1 Introduction

A probability function has the form

$$\varphi(x) := \mathbb{P}\left(g(x,\xi) \leq 0\right), \tag{1}$$

where $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ is a mapping defining a (random) inequality system, $x \in \mathbb{R}^n$ is a decision vector and ξ is an m- dimensional random vector defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The inequality sign in (1) is to be understood componentwise. Throughout the paper we shall make the following basic assumptions:

$$\begin{array}{l} g \text{ is continuously differentiable} \\ \text{the mappings } g_j(x,\cdot) \text{ are convex for all } x \in \mathbb{R}^n \text{ and all } j = 1,\dots,p \\ \xi \sim \mathcal{N}(0,R) \text{ is nondegenerate Gaussian with } R_{ii} = 1 \ (i = 1,\dots,m). \end{array} \tag{2}$$

Here, we refer to the commonly used notation $\mathcal{N}(\mu,\Sigma)$ for a Gaussian distribution with expectation μ and covariance matrix Σ. Our assumption implies that ξ is standard Gaussian with components that are centered and have unit variances. In other words, the (nondegenerate) covariance matrix is actually a correlation matrix. This assumption is no restriction because it can always be achieved under an affine linear transformation of ξ whose action on the mapping g would not affect the properties imposed in (2).

Probability functions (1) play a fundamental role in stochastic optimization problems either as an objective (reliability) to be maximized or when defining a constraint ensuring the robustness of decisions (probabilistic or chance constraint). Applications can be found in water management, telecommunications, electricity network expansion, mineral blending, chemical engineering etc. (see, e.g., [18,19,24]). Treating probability functions in the framework of optimization problems (with respect to the decision variable x) requires not only to calculate – or better: to approximate – the probability $\varphi(x)$ itself but also its gradient $\nabla\varphi$. This is why derivatives of probability functions have attracted much attention in the past (see, e.g., [7, 11, 12, 14, 17, 20–22, 25, 26, 28]). Many of these papers provide gradient formulae for fairly general classes of distributions for instance in the form of surface and/or volume integrals associated with the feasible set $K := \{z \in \mathbb{R}^m : g(\bar{x}, z) \leq 0\}$ where $\bar{x}$ is the point at which the derivative $\nabla\varphi$ is supposed to be computed. This generality comes with two drawbacks: first, the mentioned surface/volume integrals may be difficult to deal with numerically, at least for nonlinear g (see, e.g., [18, p. 207], [22, p.3]). Second, a principal assumption made in order to derive differentiability of

φ at all is the compactness of the set K (e.g., [25, p. 200, Assumptions (A2)], [22, Assumption 2.2 (i)], [17, p. 902]). Indeed, without compactness, one cannot expect differentiability of φ even with the nicest data. In [27, Prop. 2.2] an example of even a single inequality $g(x,\xi) \leq 0$ (i.e., $p=1$) is provided, where the basic assumptions (2) are fulfilled and where the set K satisfies Slater's constraint qualification, yet φ fails to be differentiable. On the other hand, compactness of K is a quite restrictive assumption in probabilistic programming and one would be interested in identifying situations, where differentiability of φ holds true even in the unbounded case. There seems to be a good chance to do so in case of Gaussian or Gaussian-like (e.g., Student- or log-normally distributed) random vectors.

Indeed, the compactness issue disappears in the case of mappings g which are linear in ξ, when ξ has a multivariate Gaussian distribution. Extending a classical differentiability result for the Gaussian distribution function (e.g., [18, p. 204]), corresponding gradient formulae could be found for mappings $g(x,\xi) = A(x)\xi \leq b(x)$ in (1) with surjective $A(x)$ [28] or with possibly nonsurjective $A(x) \equiv A$ under the Linear Independence Constraint Qualification for the set K [11]. The important fact about all these gradient formulae is that they provide a fully explicit reduction of partial derivatives of φ to values of Gaussian distribution functions again. In this way, efficient tools for computing the latter, such as Genz' code [8] can be employed not only to calculate values of φ but also gradients $\nabla\varphi$ at the same time. Moreover, using induction on the obtained formulae, explicit reductions to Gaussian distribution functions are easily found for any higher order derivative of φ. Finally, the precision for calculating $\nabla\varphi$ can be controlled by that for calculating Gaussian distributions functions, for instance, in Genz' code [10, p. 662].

This methodology fails, however, when g is nonlinear in ξ. In such a case, while keeping the Gaussian character of the random vector, one may resort to the so-called spheric-radial decomposition of Gaussian distributions [3,4,8] (see Section 2.1). Now, unlike the linear situation, differentiability of φ can no longer be taken for granted (not even under a constraint qualification and if g has just one component, see the counterexample mentioned above). Gradient formulae based on spheric-radial decomposition can be found in [6] (without rigorous proof) or in [21,22] albeit under the restrictive compactness assumption on the set K. In order to overcome this assumption, the main intention of [27] consisted in identifying an easy to check growth condition on the partial derivatives of g guaranteeing differentiability of φ without compactness of K. A corresponding result was derived for the setting of (2) with a single component of g (i.e., $p=1$) upon imposing Slater's condition on K. When considering systems of random inequalities rather than a single one (as it is typical in most applications), Slater's condition is no longer sufficient to guarantee differentiability of φ even if K were compact and g a linear mapping:

Example 1.1 *Let ξ have a one-dimensional standard Gaussian distribution and define*

$$g(x_1,x_2,x_3,\xi) := (\xi - x_1, \xi - x_2, -\xi - x_3).$$

Then, with Φ referring to the one-dimensional standard Gaussian distribution function, one has that

$$\varphi(x_1,x_2) = \max\{\min\{\Phi(x_1),\Phi(x_2)\} - \Phi(x_3), 0\}.$$

Clearly φ fails to be differentiable at $\bar{x} := (0,0,-1)$, while $K = [-1,0]$ is compact and satisfies Slater's condition in the description via g.

This inherent nondifferentiability motivates us in the present paper not only to look for conditions allowing us to generalize the differentiability result in [27] from a single inequality to inequality systems but even to take a more general, namely nonsmooth analysis perspective for viewing at probability functions. We will show that the already mentioned growth condition (but now imposed on each component of g) implies the local Lipschitz continuity of φ. This motivates the computation of subdifferentials $\partial\varphi$ in the sense of Clarke or Mordukhovich (see Section 2.3). For related work on the use of subdifferentials in settings similar to, but different from ours, we refer, for instance, to [5, 29]. As a main result, we will derive in Section 3 an upper estimate for the Clarke subdifferential of φ under the assumption that g is continuously differentiable and component-wise convex in ξ (no further assumption w.r.t. x). This result allows us in Section 4 to identify constraint qualifications – similar to those considered in a more general framework (but with compactness assumed for K) in [22, Assumption 2.2 (iv)] and [12, Theorem 2.4 and 3.1] – ensuring the (continuous) differentiability of φ. The obtained gradient formula is specialized then in Section 5 to linear random inequality systems, thus providing new representations in different disguise of the gradient formulae from [11, 28] mentioned above, which were formulated in terms of Gaussian distribution functions. Finally, in Section 6 the paper addresses the issue of refining the nonsmooth formula towards the use of Mordukhovich's rather than the bigger Clarke's subdifferential. This will be possible in the case of linear mappings g and thus improves the results on Clarke subdifferentials of singular Gaussian distribution functions in [29].

We note that the (sub-) differentiability results in this paper are not only of theoretical but also of practical interest in that they provide easy to implement gradient formulae. This relies on the fact that both, values and partial derivatives of φ, are represented as surface integrals with respect to the uniform distribution on the unit sphere. In contrast, surface integrals in the general derivative formulae mentioned above are typically taken over the boundary of the set K which may be difficult to compute. For the sphere, efficient sampling schemes are reported, for instance, in [1, 4]. Those schemes can be employed in order to simultaneously update approximations of φ and $\nabla\varphi$ with the same sample generated on the sphere. Finally, we emphasize, that the methodology described here for Gaussian distributions can be easily adapted to Gaussian-like distributions (like Student, log-normal etc.) by reducing them to Gaussian ones after an appropriate transformation of the mapping g. We do not discuss this issue here in detail because it is exactly the same methodology as the one presented in the case of a single inequality in [27].

2 Preliminaries

2.1 Spheric-radial decomposition of a Gaussian distribution

Let ξ be an m-dimensional Gaussian random vector normally distributed according to $\xi \sim \mathcal{N}(0, R)$ for some positive definite correlation matrix R. Then, $\xi = \eta L\zeta$, where $R = LL^T$ is some factorization (e.g., Cholesky decomposition) of R, η has a Chi-distribution with m degrees of freedom and ζ has a uniform distribution on the Euclidean unit sphere

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

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

$$\mathbb{P}(\xi \in M) = \int\limits_{v \in \mathbb{S}^{m-1}} \mu_\eta \left(\{ r \geq 0 : rLv \cap M \neq \emptyset \} \right) d\mu_\zeta, \tag{3}$$

where μ_η and μ_ζ are the laws of η and ζ, respectively. The consideration of distributions $\mathcal{N}(0, R)$ is no loss of generality because this standardized form is well-known to be achieved under a linear transformation of a given general Gaussian random vector. Then, (3) keeps holding true upon transforming accordingly the set M.

2.2 Probability function in spheric-radial form and preliminary results

Given the constraint mapping g in (1), we pass to the maximum function $g^m : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ over its components by defining

$$g^m(x,z) = \max_{j=1,\dots,p} g_j(x,z), \tag{4}$$

Evidently, the probability function (1) can be written as $\varphi(x) = \mathbb{P}(g^m(x,\xi) \le 0)$. By (3) we have that

$$\varphi(x) = \int_{v\in\mathbb{S}^{m-1}} \mu_\eta\left(\{r \ge 0 : g^m(x,rLv) \le 0\}\right) d\mu_\zeta = \int_{v\in\mathbb{S}^{m-1}} e(x,v) d\mu_\zeta \tag{5}$$

where

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

As a consequence of (2), g^m is convex in the second argument. In [27], probability functions of a single continously differentiable inequality, convex in the Gaussian random vector ξ, have been investigated. Because our inequality $g^m(x,\xi) \le 0$ fails to be differentiable as a maximum function, we cannot directly apply those results. Nonetheless, several of them are useful for the generalization to our setting.

Throughout the paper we will consider arguments x for which $g^m(x,0) < 0$, i.e., for which 0 is a Slater point of the inequality system $g(x,z) \le 0$ in z. This is no severe restriction because in case that $g^m(x,0) \ge 0$, the feasible set $\{z|g(x,z) \le 0\}$ would be a subset of some halfspace containing zero by convexity of $g^m(x,\cdot)$. As a consequence of ξ having a symmetric and centered distribution (see (2)), the probability of this halfspace would be 0.5 implying that $\varphi(x) \le 0.5$. In many practical applications, however, values of probability functions close to one are considered.

The assumption $g^m(x,0) < 0$ along with the convexity of $g^m(x,\cdot)$ implies that for each $x \in \mathbb{R}^n$ and each $v \in \mathbb{S}^{m-1}$, (6) can be simplified as

$$e(x,v) = \mu_\eta([0,r^*]),$$

where $r^* = \infty$ in case that $g^m(x,rLv) < 0$ for all $r > 0$ or r^* is the unique solution of $g^m(x,rLv) = 0$ in $r \ge 0$. Since this case distinction is essential when dealing with potentially unbounded sets, we are led to the definition of the following setvalued mappings $F_j, I_j, F, I : \mathbb{R}^n \rightrightarrows \mathbb{S}^{m-1}$ for $j = 1,\dots,p$:

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

The following Lemma collects some elementary properties needed later:

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

1 $F_j(x) \cup I_j(x) = F(x) \cup I(x) = \mathbb{S}^{m-1}$ *for all* $j = 1,\dots,p$.

2 *For* $j \in \{1,\dots,p\}$ *and* $v \in F_j(x)$ *let* $r > 0$ *be such that* $g_j(x,rLv) = 0$. *Then,*

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

3 $F(x) = \cup_{j=1}^{p} F_j(x)$, $I(x) = \cap_{j=1}^{p} I_j(x)$.

4 $e(x,v) = 1$ *if* $v \in I(x)$ *and* $e(x,v) < 1$ *if* $v \in F(x)$.

Proof. 1. and 3. are obvious. 2. follows easily from the convexity of $g(x,\cdot)$ (see [27, Lemma 3.1]). As for 4., $v \in I(x)$ entails that

$$\{r \geq 0 : g(x, rLv) \leq 0\} = \mathbb{R}_+$$

and, hence, by (6), $e(x,v) = \mu_\eta(\mathbb{R}_+) = 1$ because the support of the Chi-distribution is $\mathbb{R}_+$. Otherwise, if $v \in F(x)$, then again via 1. and by convexity of $g(x,\cdot)$, we see that

$$\{r \geq 0 : g(x, rLv) \leq 0\} = [0, R]$$

for some $R > 0$, whence $e(x,v) = \mu_\eta([0,R]) = 1 - \mu_\eta([R,\infty))$. With the Chi-density being strictly positive for all arguments, we conclude that $\mu_\eta([R,\infty)) > 0$, such that $e(x,v) < 1$. $\square$

Lemma 2.2 (Lemma 3.2 in [27]) *Let $j = 1,...,p$ be arbitrary and let (x,v) be such that $g_j(x,0) < 0$ and $v \in F_j(x)$. Then, there exist neighbourhoods U_j of x and V_j of v as well as a continuously differentiable function $\rho_j^{x,v} : U_j \times V_j \to \mathbb{R}_+$ with the following properties:*

1 *For all $(x',v',r') \in U_j \times V_j \times \mathbb{R}_+$ the equivalence $g_j(x', r'Lv') = 0 \Leftrightarrow r' = \rho_j^{x,v}(x',v')$ holds true.*

2 *For all $(x',v') \in U_j \times V_j$ one has the gradient formula*

$$\nabla_x \rho_j^{x,v}(x',v') = -\frac{1}{\left\langle \nabla_z g_j(x', \rho_j^{x,v}(x',v')Lv'), Lv' \right\rangle} \nabla_x g_j(x', \rho_j^{x,v}(x',v')Lv').$$

Definition 2.1 *Let $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a differentiable function. We say that h satisfies the exponential growth condition at x if there exist constants C, δ_0 and a neighbourhood $U(x)$ such that*

$$\|\nabla_x h(x',z)\| \leq \delta_0 e^{\|z\|} \quad \forall x' \in U(x)\ \forall z : \|z\| \geq C.$$

Lemma 2.3 (Lemma 3.3 and Lemma 3.7 in [27]) *Let $j = 1,...,p$ be arbitrary and let $x \in \mathbb{R}^n$ be such that $g_j(x,0) < 0$. Moreover, let $v \in I_j(x)$ and consider any sequence $(x_k, v_k) \to (x,v)$ with $v_k \in F_j(x_k)$. Then $\rho_j^{x_k,v_k}(x_k,v_k) \to_k \infty$. If, in addition, g_j satisfies the exponential growth condition at x, then also*

$$\chi\left(\rho_j^{x_k,v_k}(x_k,v_k)\right) \nabla_x \rho_j^{x_k,v_k}(x_k,v_k) \to_k 0.$$

Here, χ is the density of the chi-distribution with m degrees of freedom and $\rho_j^{x_k,v_k}$ is the resolving function defined in a neighbourhood of (x_k,v_k) as in Lemma 2.2.

2.3 Clarke and Mordukhovich subdifferential

In this section, we recall the definitions of some well-known subdifferentials of nonsmooth functions (see [2, 15]).

Definition 2.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary function and fix any $\bar{x} \in \mathbb{R}^n$. Then,*

- *the Fréchet subdifferenital of f at $\bar{x}$ is the set*

$$\hat{\partial} f(\bar{x}) = \left\{ x^* \in \mathbb{R}^n \mid \liminf_{x \to \bar{x}} \frac{f(x) - f(\bar{x}) - \langle x^*, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0 \right\}$$

- *the Mordukhovich or limiting subdifferential of f at $\bar{x}$ is the set*

$$\partial^M f(\bar{x}) = \left\{ x^* \in \mathbb{R}^n \mid \exists x_n \to \bar{x}, x_n^* \to x^* : f(x_n) \to f(\bar{x}), x_n^* \in \hat{\partial} f(x_n) \right\}$$

- *if f is locally Lipschitz continuous around $\bar{x}$, then the Clarke subdifferential of f at $\bar{x}$ is the set*

$$\partial^c f(\bar{x}) = \mathrm{Co} \left\{ x^* \in \mathbb{R}^n \mid \exists x_n \to \bar{x}, \nabla f(x_n) \to x^* \right\},$$

  where 'Co' refers to the convex hull.

Note that, thanks to Rademacher's Theorem, a locally Lipschitz continuous function is differentiable almost everywhere and, hence, its Clarke subdifferential is nonempty. Moreover, for such functions, the Clarke subdifferential is always the closed convex hull of the Mordukhovich subdifferential, the latter being a nonconvex set and, thus, strictly smaller than the former, in general. The partial subdifferential of a function depending on two variables is defined as the subdifferential of the partial function, similar to the definition of partial derivatives.

3 Clarke subdifferential of φ

The aim of this section is to provide an upper estimate for the Clarke subdifferential of the probability function (1). The main result of this section is formulated in Theorem 3.1. It will be based on interchanging subdifferentiation and integration in (5). This requires to calculate the Clarke subdifferential of the function e in (6) first. To start with, we prove the following auxiliary result:

Lemma 3.1 *Let $x \in \mathbb{R}^n$ be such that $g^m(x, 0) < 0$ and let $v \in F(x)$. Then, introducing the index set $J_F^{x,v} := \{j \in \{1, ..., p\} | v \in F_j(x)\}$, the functions $\rho_j^{x,v}$ from Lemma 2.2 are well-defined for $j \in J_F^{x,v}$ on the neighbourhood $\tilde{U} \times \tilde{V}$ of (x, v), where, with U_j, V_j from Lemma 2.2,*

$$\tilde{U} := \cap_{j \in J_F} U_j, \quad \tilde{V} := \cap_{j \in J_F} V_j.$$

Moreover, there exist neighbourhoods $U \subseteq \tilde{U}$ of x and $V \subseteq \tilde{V}$ of v with the following properties:

1 *For all $(x', v', r') \in U \times V \times \mathbb{R}_+$ the equivalence $g^m(x', r'Lv') = 0 \Leftrightarrow r' = \rho^{x,v}(x', v')$ holds true, where $\rho^{x,v} : \tilde{U} \times \tilde{V} \to \mathbb{R}_+$ is defined as*

$$\rho^{x,v}(x', v') := \min_{j \in J_F^{x,v}} \rho_j^{x,v}(x', v') \quad \forall (x', v') \in \tilde{U} \times \tilde{V}. \tag{7}$$

2 *For all $(x', v') \in U \times V$, the partial Clarke-sub-differential of $\rho^{x,v}$ (w.r.t. x) is given by*

$$\partial_x^c \rho^{x,v}(x', v') = \mathrm{Co} \left\{ \nabla_x \rho_j^{x,v}(x', v') \ : j \in \mathcal{J}^{x,v}(x', v') \right\}, \tag{8}$$

where "Co"denotes the convex hull and $\mathcal{J}^{x,v}(x', v') := \{j \in J_F^{x,v} | \rho_j^{x,v}(x', v') = \rho^{x,v}(x', v')\}$.

Proof. Our assumptions and Lemma 2.1 (3.) imply that $g_j(x,0) < 0$ for all $j \in \{1,...,p\}$ and $J_F^{x,v} \neq \emptyset$. Hence, the set $\tilde{U} \times \tilde{V}$ defined in the statement of this lemma is indeed a neighbourhood of (x,v) and Lemma 2.2 (1.) yields the equivalence

$$g_j(x', r'Lv') = 0 \Leftrightarrow r' = \rho_j^{x,v}(x', v') \quad \forall\, (x', v', r') \in \tilde{U} \times \tilde{V} \times \mathbb{R}_+ \; \forall j \in J_F^{x,v}, \tag{9}$$

In particular, the min-function $\rho^{x,v}$ in (7) is well-defined and continuous on $\tilde{U} \times \tilde{V}$. We may clearly shrink $\tilde{U} \times \tilde{V}$ to a neighbourhood $U \times V$ of (x,v) which is bounded and – by continuity of g^m – satisfies that $g^m(x', 0) < 0$ for all $x' \in U$. Boundedness of $U \times V$ and continuity of $\rho^{x,v}$ imply the existence of some $R > 0$ with

$$\rho^{x,v}(x', v') \leq R \quad \forall\, (x', v') \in U \times V. \tag{10}$$

Moreover, since $j \in (J_F^{x,v})^c$ entails $v \in I_j(x)$ (by Lemma 2.1 (1.) and (3.)), Lemma 2.3 allows us to shrink $U \times V$ once more such that

$$\rho_j^{x',v'}(x', v') \geq R + 1 \quad \forall\, (x', v') \in U \times V : v' \in F_j\,(x') \;\; \forall j \in (J_F^{x,v})^c. \tag{11}$$

Here, $\rho_j^{x',v'}$ refers to the resolving function in Lemma 2.2 whose existence around (x', v') is guaranteed by $v' \in F_j\,(x')$.

Now, in order to prove statement 1. of this Lemma, let $(x', v', r') \in U \times V \times \mathbb{R}_+$ be such that $g^m(x', r'Lv') = 0$. Assuming that $r' > \rho^{x,v}(x', v')$, there would exist some $j \in J_F^{x,v}$ with $r' > \rho_j^{x,v}(x', v')$. From (9), we then derive the contradiction

$$0 = g_j(x', \rho_j^{x,v}(x', v')Lv') \leq g^m(x', \rho_j^{x,v}(x', v')Lv') < g^m(x', r'Lv') = 0,$$

where the strict inequality follows from $g^m(x', 0) < 0$ and from the convexity of $g^m(x', \cdot)$. Hence, $r' \leq \rho^{x,v}(x', v')$. If, in contrast, $r' < \rho^{x,v}(x', v')$, then with the same arguments as before, we arrive at

$$g_j(x', \rho_j^{x,v}(x', v')Lv') = 0 = g^m(x', r'Lv') < g^m(x', \rho^{x,v}(x', v')Lv') \quad \forall j \in J_F^{x,v}. \tag{12}$$

Hence, for any $j \in J_F^{x,v}$, we have the relations

$$g_j(x', 0) \leq g^m(x', 0) < 0, \quad g_j(x', \rho_j^{x,v}(x', v')Lv') = 0, \quad \rho^{x,v}(x', v') \leq \rho_j^{x,v}(x', v').$$

Now, convexity of $g_j(x', \cdot)$ provides that $g_j(x', \rho^{x,v}(x', v')Lv') \leq g_j(x', \rho_j^{x,v}(x', v')Lv')$. This allows us to conclude from (12) that

$$g_j(x', \rho^{x,v}(x', v')Lv') < g^m(x', \rho^{x,v}(x', v')Lv') \quad \forall j \in J_F^{x,v}.$$

Consider now an arbitrary $j \in (J_F)^c$. In the case of $v' \in I_j\,(x')$ one has that

$$g_j(x', \rho^{x,v}(x', v')Lv') < 0 < g^m(x', \rho^{x,v}(x', v')Lv') \tag{13}$$

with the first inequality following from the definition of $I_j\,(x')$ and the second one following from (12). In the opposite case, one has that $v' \in F_j\,(x')$ by Lemma 2.1 (1.). Then, exploiting (10) and (11), we end up with $\rho_j^{x',v'}(x', v') > \rho^{x,v}(x', v')$. Hence, with the same convexity argument as before,

$$0 = g_j(x', \rho_j^{x',v'}(x', v')Lv') > g_j(x', \rho^{x,v}(x', v')Lv'). \tag{14}$$

Combining this with (13), we have shown that

$$g_j(x', \rho^{x,v}(x', v')Lv') < g^m(x', \rho^{x,v}(x', v')Lv') \quad \forall j \in (J_F^{x,v})^c.$$

Together with (12), one arrives at the contradiction

$$g_j(x', \rho^{x,v}(x', v')Lv') < g^m(x', \rho^{x,v}(x', v')Lv') \quad \forall j \in J_F^{x,v} \cup (J_F^{x,v})^c = \{1, \dots, p\}$$

with the definition of g^m. Summarizing we have proven that $r' = \rho^{x,v}(x', v')$ which shows the part '⇒' in the equivalence claimed in statement 1. of this Lemma.

Conversely, assume that $r' = \rho^{x,v}(x', v')$ for some $(x', v', r') \in U \times V \times \mathbb{R}_+$. Select any $j^* \in J_F^{x,v}$ with $\rho^{x,v}(x', v') = \rho_{j^*}^{x,v}(x', v')$. Then, by (9),

$$g_{j^*}(x', r'Lv') = g_{j^*}(x', \rho_{j^*}^{x,v}(x', v')Lv') = 0. \tag{15}$$

On the other hand, if $j \in J_F^{x,v}$ is arbitrary, then $r' = \rho^{x,v}(x', v') \leq \rho_j^{x,v}(x', v')$ and

$$0 = g_j(x', \rho_j^{x,v}(x', v')Lv') \geq g_j(x', r'Lv')$$

by $g_j(x', 0) < 0$ and convexity of $g_j(x', \cdot)$. Finally, for $j \in (J_F^{x,v})^c$ one has that $v \in I_j(x)$. In the case where also $v' \in I_j(x')$, we have that $g_j(x', r'Lv') < 0$. In the opposite case of $v' \in F_j(x')$ (10) and (11) yield that $\rho_j^{x',v'}(x', v') > \rho^{x,v}(x', v')$. Then, by Lemma 2.2 (1.) and applying the same convexity argument as before, we get

$$0 = g_j(x', \rho_j^{x',v'}(x', v')Lv') > g_j(x', \rho^{x,v}(x', v')Lv') = g_j(x', r'Lv').$$

Summarizing, we have shown that $g_j(x', r'Lv') \leq 0$ for all $j = 1, \dots, p$, which together with (15) leads to the desired relation $g^m(x', r'Lv') = 0$. This proves statement 1. of our Lemma.

As for statement 2., we may apply [2, Proposition 2.3.12] to the relation $-\rho^{x,v} = \max_{j \in J_F^{x,v}} -\rho_j^{x,v}$ in order to derive the equality

$$\partial_x^c(-\rho^{x,v}(x', v')) = \mathrm{Co}\,\{-\nabla_x \rho_j^{x,v}(x', v') \mid j \in \mathcal{J}^{x,v}(x', v')\}.$$

On the other hand $\partial_x^c(-\rho^{x,v}(x', v') = -\partial_x^c \rho^{x,v}(x', v')$ by [2, Proposition 2.3.1], which allows us to prove (8) since $\mathrm{Co}\,(-A) = -\mathrm{Co}\,A$ for any set A. □

If one dealt with a single component of g only (i.e., $p = 1$), then trivially the functions g^m in (4) and $\rho^{x,v}$ in (7) would be continuously differentiable and, hence, Lemma 3.1 (1.) would allow us to invoke two results [27, Lemma 3.3 and Corollary 3.4] derived in this restricted setting. Of course, for $p > 1$, g^m and $\rho^{x,v}$ are just locally Lipschitz continuous and in particular continuous. Continuity is indeed immediate from the given max- and min- operations in (4) and (7) applied to the (differentiable) components g_j and $\rho_j^{x,v}$, respectively. Since, none of the two above mentioned results exploits differentiability arguments and only continuity is needed there, we do not provide a proof of the following Lemma which is literally a copy of the proofs of those results:

Lemma 3.2 *Let $x \in \mathbb{R}^n$ be such that $g^m(x, 0) < 0$. The following holds:*

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 and $\rho^{x,v}$ are defined in (6) and (7), respectively, and F_η is the cumulative distribution function of the Chi-distribution with m degrees of freedom.*

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 $v_k \in F(x_k)$.

3 The function e is continuous at (x,v) for any $v \in \mathbb{S}^{m-1}$.

Corollary 3.1 *Let $x \in \mathbb{R}^n$ be such that $g^m(x,0) < 0$ and $v \in F(x)$. Then, there exists a neighbourhood $U \times V$ of (x,v) such that e is Lipschitz on $U \times V$ and*

$$\partial_x^c e(x',v') = \mathrm{Co}\,\left\{\chi\left(\rho^{x,v}(x',v')\right)\nabla_x\rho_j^{x,v}(x',v')\ : j \in \mathcal{J}^{x,v}(x',v')\right\} \quad \forall (x',v') \in U \times V.$$

Here χ is the density of the Chi-distribution with m degrees of freedom, and $\mathcal{J}^{x,v}$ as introduced in Lemma 3.1.

Proof. From Lemma 3.2 (1.), we know that $e = F_\eta \circ \rho^{x,v}$ in a neighbourhood $U \times V$ of (x,v). We may assume this neighbourhood small enough so that $\rho^{x,v}$ is Lipschitz there as a minimum of smooth functions by (7). Since the mapping F_η is continuously differentiable with $F'_\eta = \chi$, Clarke's chain rule ([2, Theorem 2.3.9 (ii)]) yields that

$$\partial_x^c e(x',v') = \chi\left(\rho^{x,v}(x',v')\right)\partial_x^c \rho^{x,v}(x',v') \quad \forall (x',v') \in U \times V.$$

The assertion now follows from (8). □

In the following we want to to generalize Corollary 3.1 and to establish the local Lipschitz continuity of the partial mapping $e(\cdot,v)$ around any $x \in \mathbb{R}^n$ with $g^m(x,0) < 0$ and any $v \in \mathbb{S}^{m-1}$ and to provide a formula for its Clarke subdifferential. To this aim, we need the following auxiliary results:

Lemma 3.3 *Let $x \in \mathbb{R}^n$ be such that $g^m(x,0) < 0$ and assume that all components g_j of g satisfy the exponential 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,*

$$\lim_{k\to\infty} \partial_x^c e(x_k,v_k) = \{0\},$$

where the latter means that for each $\varepsilon > 0$ there exists an index $K > 0$ such that $\partial_x^c e(x_k,v_k) \subseteq B(0,\varepsilon)$ for all $k \geq K$, where $B(0,\varepsilon)$ is the ball of radius ε centered at $0 \in \mathbb{R}^n$.

Proof. By Corollary 3.1 it follows that any $s_k \in \partial_x^c e(x_k,v_k)$ can be written as

$$s_k = \chi(\rho^{x_k,v_k}(x_k,v_k)) \cdot \sum_{j \in \mathcal{J}^{x_k,v_k}(x_k,v_k)} \lambda_j^{(k)} \nabla_x \rho_j^{x_k,v_k}(x_k,v_k),$$

where $\lambda_j^{(k)} \geq 0$ for all $j \in \mathcal{J}^{x_k,v_k}(x_k,v_k)$ and $\sum_{j\in\mathcal{J}^{x_k,v_k}(x_k,v_k)} \lambda_j^{(k)} = 1$. Since according to Lemma 3.1 (2.), $\rho^{x_k,v_k}(x_k,v_k) = \rho_j^{x_k,v_k}(x_k,v_k)$ for $j \in \mathcal{J}^{x_k,v_k}(x_k,v_k)$, one may characterize s_k alternatively by

$$s_k = \sum_{j \in \mathcal{J}^{x_k,v_k}(x_k,v_k)} \lambda_j^{(k)} \chi(\rho_j^{x_k,v_k}(x_k,v_k)) \nabla_x \rho_j^{x_k,v_k}(x_k,v_k) = \sum_{j=1}^{p} \mu_j^{(k)},$$

where we have put

$$\mu_j^{(k)} := \begin{cases} \lambda_j^{(k)} \chi(\rho_j^{x_k,v_k}(x_k,v_k)) \nabla_x \rho_j^{x_k,v_k}(x_k,v_k) & (j \in \mathcal{J}^{x_k,v_k}(x_k,v_k)) \\ 0 & (j \in \{1,\ldots,p\} \backslash \mathcal{J}^{x_k,v_k}(x_k,v_k)) \end{cases}.$$

The assertion of our Lemma will follow if we can show that $\mu_j^{(k)} \to_k 0$ for all $j \in \{1,\ldots,p\}$. In order to do so fix any $j \in \{1,\ldots,p\}$. If there is only a finite number of indices k with $j \in \mathcal{J}^{x_k,v_k}(x_k,v_k)$, then $\mu_j^{(k)} = 0$ for all k large enough, whence the claimed convergence holds true. Otherwise, consider the subsequence k_l consisting of all indices k with $j \in \mathcal{J}^{x_k,v_k}(x_k, v_k)$. Then, $(x_{k_l}, v_{k_l}) \to_l (x, v)$ and $v_{k_l} \in F(x_{k_l})$ for all l. Moreover, our assumption $v \in I(x)$ implies that $v \in I_j(x)$ by Lemma 2.1 (3.). Therefore, Lemma 2.3 allows us to conclude that

$$\chi(\rho_j^{x_{k_l},v_{k_l}}(x_{k_l}, v_{k_l}))\nabla_x \rho_j^{x_{k_l},v_{k_l}}(x_{k_l}, v_{k_l}) \to_l 0,$$

whence $\mu_j^{(k_l)} \to_l 0$ due to $\lambda_j^{(k_l)} \in [0, 1]$. Consequently, if $\varepsilon > 0$ is arbitrarily given, then there exists some l' such that

$$\left\|\mu_j^{(k_l)}\right\| \leq \varepsilon \quad \forall l \geq l'. \tag{16}$$

Put $k' := k_{l'}$. Then, for any $k \geq k'$ one either has that $j \in \mathcal{J}^{x_k,v_k}(x_k, v_k)$ in which case $k = k_l$ for some $l \geq l'$ and, hence, (16) holds true. Otherwise, $j \notin \mathcal{J}^{x_k,v_k}(x_k, v_k)$ in which case $\mu_j^{(k)} = 0$. It follows again the claimed convergence $\mu_j^{(k)} \to_k 0$, i.e., $\|\mu_j^{(k)}\| \leq \varepsilon$ for all $k \geq k'$. □

Corollary 3.2 *Let x be such that $g^m(x, 0) < 0$ and that g_j satisfies the exponential growth condition at x for all $j = 1, ..., p$. Then, for any $v \in \mathbb{S}^{m-1}$, the function $e(\cdot, v)$ is Lipschitz continuous in a neighbourhood of x and its Clarke subdifferential is given by*

$$\partial_x^c e(x, v) = \begin{cases} \operatorname{Co}\left\{-\frac{\chi(\rho^{x,v}(x,v))}{\langle \nabla_z g_j(x,\rho^{x,v}(x,v)Lv), Lv\rangle}\nabla_x g_j\left(x, \rho^{x,v}\left(x, v\right) Lv\right) \ j \in \mathcal{J}^{x,v}(x, v)\right\} & \text{if } v \in F(x) \\ \{0\} & \text{if } v \in I(x) \end{cases}.$$

Here χ is the density of the chi-distribution with m degrees of freedom, $\rho^{x,v}$ refers to the resolving function and $\mathcal{J}^{x,v}(x, v)$ is the active index set both introduced in Lemma 3.1.

Proof. Fix arbitrary x and v as indicated above. If $v \in F(x)$, then $e(\cdot, v) = F_\eta(\rho^{x,v}(\cdot, v))$ in a neighbourhood of x by Lemma 3.2 (1.), hence $e(\cdot, v)$ is Lipschitz continuous on this neighbourhood and the asserted formula for $\partial_x^c e(x, v)$ follows from Corollary 3.1 and Lemma 2.2 (2.). Therefore, we may assume $v \in I(x)$ now. We start by verifying local Lipschitz continuity of $e(\cdot, v)$ around x. If this were not true, then there would exist sequences $x_k \to_k x$ and $y_k \to_k x$ with

$$|e(x_k, v) - e(y_k, v)| > k \|x_k - y_k\| \quad \forall k \in \mathbb{N}. \tag{17}$$

By Lemma 3.2 (3.), we may assume that all x_k, y_k are contained in a ball around x such that $e(\cdot, v)$ is continuous in this ball. Moreover, we may assume that this ball is small enough to guarantee that

$$g^m(x', 0) < 0 \quad \forall x' \in [x_k, y_k] \ \forall k \in \mathbb{N}. \tag{18}$$

We will show that for all $k \in \mathbb{N}$ there exist $z_k \in [x_k, y_k]$ and $x_k^* \in \partial_x^c e(z_k, v)$ such that

$$v \in F(z_k) \text{ and } |e(x_k, v) - e(y_k, v)| \leq \left(\|x_k^*\| + k^{-1}\right) \|x_k - y_k\|. \tag{19}$$

To show this claim, let us fix an arbitrary k now. If $v \in I(x_k) \cap I(y_k)$, then Lemma 2.1 (4.) leads to a contradiction with (17). Hence, without loss of generality, $v \in F(x_k)$. Define $x^t := (1-t)\,x_k + ty_k$ for all $t \in [0, 1]$ and

$$\tau := \sup\left\{t \in [0, 1] \,|\, e(x^{t'}, v) < 1 \ \forall t' \in [0, t]\right\}.$$

Since $e(x^0, v) = e(x_k, v) < 1$ by Lemma 2.1 (4.), the continuity of $e(\cdot, v)$ on the line segment $[x_k, y_k]$ provides that $\tau \in (0, 1]$. Moreover, we may find an $\alpha \in (0, \tau)$ with

$$|e(x^\alpha, v) - e(x^\tau, v)| \leq k^{-1} \|x_k - y_k\|.$$

Since $\alpha \in (0,\tau)$ this implies that $e(x^{t'},v) < 1$ for all $t' \in [0,\alpha]$, Lemma 2.1 (5.) yields that

$$v \in F(x^{t'}) \quad \forall t' \in [0,\alpha]\,. \tag{20}$$

Taking into account (20) and that, by (18), $g^m(x^{t'},0) < 0$ for all $t' \in [0,\alpha]$, Corollary 3.1 yields that $e(\cdot,v)$ is locally Lipschitz continuous on an open neighbourhood of the line segment $[x^0,x^\alpha]$. This allows us to invoke Lebourg's mean value theorem [13, Theorem 1.7], in order to derive the existence of some $t^* \in [0,\alpha]$ and some $x^* \in \partial_x^c e\left(x^{t^*},v\right)$ such that

$$\left|e\left(x^0,v\right) - e\left(x^\alpha,v\right)\right| \le \|x^*\| \left\|x^0 - x^\alpha\right\|.$$

Therefore, recalling that $x_k = x^0$ and that $x^\alpha \in [x_k,y_k]$, we arrive at

$$|e\left(x_k,v\right) - e\left(x^\tau,v\right)| \le \|x^*\| \left\|x_k - x^\alpha\right\| + k^{-1}\left\|x_k - y_k\right\| \le \left(\|x^*\| + k^{-1}\right)\left\|x_k - y_k\right\|. \tag{21}$$

Clearly, $v \in F(x^{t^*})$ by (20). If $\tau = 1$, then $x^\tau = y_k$ and (19) follows upon putting $z_k := x^{t^*}$ and $x_k^* := x^*$. Otherwise, $\tau < 1$ and then $e\left(x^\tau,v\right) = 1$ by continuity of $e(\cdot,v)$ on the line segment $[x_k,y_k]$. We have to distinguish two cases: first, if $v \in I(y_k)$, then $e\left(y_k,v\right) = 1$ by Lemma 2.1 (4.) and so (19) follows from (21) and $e\left(y_k,v\right) = e\left(x^\tau,v\right)$ with the same z_k, x_k^* as before. In the second case, $v \in F(y_k)$, so the roles of x_k and y_k can be interchanged in deriving (21). Therefore, we may assume without loss of generality that $e(y_k,v) \ge e(x_k,v)$. Then, with e being bounded from above by 1,

$$|e\left(x_k,v\right) - e\left(x^\tau,v\right)| = 1 - e\left(x_k,v\right) \ge e(y_k,v) - e\left(x_k,v\right) = |e(x_k,v) - e(y_k,v)|\,.$$

Now, (19) follows once more from (21) with the same z_k, x_k^* as before. Since $k \in \mathbb{N}$ was chosen arbitrary, we have altogether verified (19). Clearly, $z_k \in [x_k,y_k]$ implies $z_k \to_k x$. Since also $v \in F(z_k)$ and $v \in I(x)$, Lemma 3.3 yields that $\|x_k^*\| + k^{-1}$ is a bounded sequence contradicting (17). Summarizing, we have proven Lipschitz continuity of $e(\cdot,v)$ around x.

It remains to calculate the Clarke subdifferential of $e(\cdot,v)$. By [2, Theorem 2.5.1] we have that

$$\partial_x^c e(x,v) = \mathrm{Co}\left\{x^* | \exists x_l \to x :\ e(\cdot,v) \text{ differentiable at } x_l \text{ and } \nabla_x e(x_l,v) \to_l x^*\right\}.$$

Therefore, in order to prove the remaining assertion $\partial_x^c e(x,v) = 0$ of our Corollary, we have to show that $\nabla_x e(x_l,v) \to_l 0$ holds true for any sequence $x_l \to x$ with $e(\cdot,v)$ differentiable at all x_l. Let us fix any such sequence and assume that the asserted convergence would not hold true. Then,

$$\|\nabla_x e(x_{l_k},v)\| \ge \varepsilon \quad \forall k \tag{22}$$

for some subsequence and some $\varepsilon > 0$. If $v \in I(x_{l_k})$, for some k, then $e(\cdot,v)$ reaches its maximum possible value at x_{l_k} (see Lemma 2.1 (4.)). Since $e(\cdot,v)$ is differentiable at x_{l_k}, it follows the contradiction $\nabla_x e(x_{l_k},v) = 0$ with (22). Hence, $v \in F(x_{l_k})$ for all k and so by Lemma 3.3 we have that

$$\{0\} = \lim_{k\to\infty} \partial_x^c e(x_{l_k},v).$$

On the other hand, $\nabla_x e(x_{l_k},v) \in \partial_x^c e(x_{l_k})$ by [2, Proposition 2.2.2], whence $\nabla_x e(x_{l_k},v) \to_k 0$, which is a contradiction with (22) again. Summarizing, we have shown that $\nabla_x e(x_l,v) \to 0$ along any sequence x_l at which $e(\cdot,v)$ is differentiable. This finishes the proof. □

Now, we are in a position to prove the main result of this paper. The set-valued integral appearing in (23) has to be interpreted as explained in Remark 3.1 below.

Theorem 3.1 *In addition to our basic assumptions* (2)*, let the following conditions be satisfied at some fixed* $\bar{x} \in \mathbb{R}^n$*:*

1 $g^m(\bar{x}, 0) < 0$.

2 g_j *satisfies the exponential growth condition at* $\bar{x}$ *(Def. 2.1) for all* $j = 1, ..., p$.

Then, φ *in* (1) *is locally Lipschitz continuous on a neighbourhood* U *of* $\bar{x}$ *and it holds that*

$$\partial^c\varphi(x) \subseteq \int\limits_{v\in F(x)} \mathrm{Co}\left\{-\frac{\chi(\hat{\rho}(x,v))}{\langle\nabla_z g_j(x,\hat{\rho}(x,v)Lv), Lv\rangle}\nabla_x g_j(x,\hat{\rho}(x,v)Lv)\middle| j\in\hat{\mathcal{J}}(x,v)\right\} d\mu_\zeta(v) \tag{23}$$

for all $x \in U$*. Here,* L *is a factor in some decomposition* $R = LL^T$ *(e.g., Cholesky decomposition),* $\hat{\rho}(x, v)$ *refers to the unique solution in* $r \geq 0$ *of the equation* $g^m(x, rLv) = 0$ *and*

$$\hat{\mathcal{J}}(x,v) := \{j \in \{1,\dots,p\} | g_j(x,\hat{\rho}(x,v)Lv) = 0\} \quad (v \in F(x))$$

Proof. Since assumptions 1. and 2. are open, there exists an open neighbourhood $\tilde{U}$ of $\bar{x}$ such that

$$g^m(x,0) < 0 \quad \text{and the } g_j \text{ satisfy the exponential growth condition} \quad \forall j = 1,\dots,p \ \forall x \in \tilde{U}. \tag{24}$$

According to Lemma 3.2 (3.) e is continuous on $\tilde{U} \times \mathbb{S}^{m-1}$. Consequently, for each $x \in \tilde{U}$ the mapping $v \in \mathbb{S}^{m-1} \mapsto e(x, v)$ is measurable. Next, we show that the function

$$\alpha(x,v) := \max\{\|s\| \ : \ s \in \partial_x^c e(x,v)\}.$$

is upper semi-continuous on $\tilde{U} \times \mathbb{S}^{m-1}$. In order to do so, fix an arbitrary $(x, v) \in \tilde{U} \times \mathbb{S}^{m-1}$ and an arbitrary sequence $(x_k, v_k) \to_k (x, v)$ with $(x_k, v_k) \in \tilde{U} \times \mathbb{S}^{m-1}$ for all k. Assume first that $v \in F(x)$. By a continuity argument, there exists k_0 such that

$$\mathcal{J}^{x,v}(x_k,v_k) \subseteq \mathcal{J}^{x,v}(x,v) \quad \forall k \geq k_0$$

holds true for the index set mapping $\mathcal{J}^{x,v}$ introduced in Lemma 3.1. Then, by Corollary 3.1,

$$\begin{aligned}
\alpha(x_k,v_k) &= \max\left\{\|s\| \,|s \in \mathrm{Co}\,\left\{\chi(\rho^{x,v}(x_k,v_k))\nabla_x\rho_j^{x,v}(x_k,v_k) \ : j\in\mathcal{J}^{x,v}(x_k,v_k)\right\}\right\}\\
&\leq \max\left\{\|s\| \,|s \in \mathrm{Co}\,\left\{\chi(\rho^{x,v}(x_k,v_k))\nabla_x\rho_j^{x,v}(x_k,v_k) \ : j\in\mathcal{J}^{x,v}(x,v)\right\}\right\}\\
\to_k \quad & \max\left\{\|s\| \,|s \in \mathrm{Co}\,\left\{\chi(\rho^{x,v}(x,v))\nabla_x\rho_j^{x,v}(x,v) \ : j\in\mathcal{J}^{x,v}(x,v)\right\}\right\} = \alpha(x,v).
\end{aligned}$$

Since the sequence $(x_k, v_k) \to_k (x, v)$ was arbitrarily chosen, it follows that

$$\limsup_{(x',v')\to(x,v),\,(x',v')\in\tilde{U}\times\mathbb{S}^{m-1}} \alpha(x',v') \leq \alpha(x,v)$$

which is the upper semi-continuity of α at (x, v). Now assume that $v \in I(x)$, whence $\alpha(x, v) = 0$. We claim that $\alpha(x_k, v_k) \to_k 0$. If this was not the case, then $\alpha(x_{k_l}, v_{k_l}) \geq \varepsilon$ for some subsequence $(x_{k_l}, v_{k_l}) \to_l (x, v)$ and some $\varepsilon > 0$. Assume that $v_{k_l} \in I(x_{k_l})$ for some l. Since the two assumptions of our Theorem are open with respect to x they may be assumed to continue to hold at x_{k_l}. Then, $\partial_x^c e(x_{k_l}, v_{k_l}) = \{0\}$ by Corollary 3.2, whence the contradiction $\alpha(x_{k_l}, v_{k_l}) = 0$. Therefore, $v_{k_l} \in F(x_{k_l})$ for all l and, hence, by Lemma 3.3,

$$\lim_{l\to\infty} \partial_x^c e(x_{k_l}, v_{k_l}) = \{0\}.$$

This yields once more a contradiction $\alpha\left(x_{k_l}, v_{k_l}\right) \to_l 0$ with $\alpha\left(x_{k_l}, v_{k_l}\right) \geq \varepsilon$. Consequently, $\alpha\left(x_k, v_k\right) \to_k 0$ as claimed so that α is continuous at (x, v). Summarizing, we have shown that α is upper semi-continuous on $\tilde{U} \times \mathbb{S}^{m-1}$.

Let $\mathbb{B}\left(\bar{x}; r\right)$ be a closed ball centered at $\bar{x}$ and with radius $r > 0$ such that $\mathbb{B}\left(\bar{x}; r\right) \subseteq \tilde{U}$. By the Weierstrass Theorem, the upper semi-continuous function α realizes its maximum on the compact set $\mathbb{B}\left(\bar{x}; r\right) \times \mathbb{S}^{m-1}$, hence α is bounded on this set by some constant $M > 0$. Define the open neighbourhood $U := \operatorname{int} \mathbb{B}\left(\bar{x}; r\right)$ and choose arbitrary $x, y \in U$ and $v \in \mathbb{S}^{m-1}$. Observe that $e(\cdot, v)$ is locally Lipschitz continuous on $U \subseteq \tilde{U}$ by Corollary 3.2 and as a consequence of (24). Lebourg's mean value theorem [13, Theorem 1.7] then implies the existence of some $\tilde{x}$ in the line segment $[x, y]$ and of some $s^* \in \partial_x^c e(\tilde{x}, v)$ such that

$$e(x, v) - e(y, v) = \langle s^*, x - y \rangle .$$

Since $\tilde{x} \in \mathbb{B}\left(\bar{x}; r\right)$, we conclude that $\|s^*\| \leq \alpha(\tilde{x}, v) \leq M$. Summarizing, we have shown that

$$|e(x, v) - e(y, v)| \leq M \|x - y\| \quad \forall x, y \in U \ \forall v \in \mathbb{S}^{m-1}.$$

This property allows us to invoke Clarke's Theorem on the interchange of integral and subdifferential [2, Theorem 2.7.2] in order first to conclude that φ is locally Lipschitz continuous on U and second to derive from (5) and from Corollary 3.2 the formula

$$\partial^c \varphi\left(x\right) = \partial^c \int\limits_{v \in \mathbb{S}^{m-1}} e(x, v) d\mu_\zeta \subseteq \int\limits_{v \in \mathbb{S}^{m-1}} \partial^c e(x, v) d\mu_\zeta = \int\limits_{v \in F(x)} \partial^c e(x, v) d\mu_\zeta$$

$$= \int\limits_{v \in F(x)} \operatorname{Co} \left\{ -\frac{\chi\left(\rho^{x,v}\left(x, v\right)\right)}{\left\langle \nabla_z g_j\left(x, \rho^{x,v}\left(x, v\right) L v\right), L v \right\rangle} \nabla_x g_j\left(x, \rho^{x,v}\left(x, v\right) L v\right) \middle| \, j \in \mathcal{J}^{x,v}(x, v) \right\} d\mu_\zeta(v).$$

By Lemma 3.1 (1.), $\rho^{x,v}\left(x, v\right)$ is the unique solution in r of the equation $g^m(x, rLv) = 0$, hence $\hat{\rho}(x, v) = \rho^{x,v}\left(x, v\right)$ with $\hat{\rho}$ as introduced in the statement of this Theorem. It remains to show that

$$\mathcal{J}^{x,v}(x, v) = \hat{\mathcal{J}}(x, v) \quad \forall x \in U \quad \forall v \in F(x)$$

for $\hat{\mathcal{J}}$ as introduced in the statement of this Theorem. To this aim, fix arbitrary $x \in U$ and $v \in F(x)$. Let also $j \in \mathcal{J}^{x,v}(x, v)$ be arbitrarily given. By definition, $\rho_j^{x,v}(x, v) = \rho^{x,v}(x, v) = \hat{\rho}(x, v)$, whence

$$g_j(x, \hat{\rho}(x, v) L v) = g_j(x, \rho_j^{x,v}(x, v) L v) = 0$$

and so $j \in \hat{\mathcal{J}}(x, v)$. Conversely, let $j \in \hat{\mathcal{J}}(x, v)$ be arbitrary. Then, $g_j(x, \hat{\rho}(x, v) L v) = 0$ which entails that $v \in F_j(x)$ and that $j \in J_F^{x,v}$ with the latter set as introduced in Lemma 3.1. By Lemma 2.2 (1.), $\rho_j^{x,v}(x, v)$ is the unique solution in $r \geq 0$ of the equation $g_j(x, rLv) = 0$. Consequently, by (7)

$$\rho_j^{x,v}(x, v) = \hat{\rho}(x, v) = \rho^{x,v}(x, v) = \min_{j' \in J_F^{x,v}} \rho_{j'}^{x,v}(x, v).$$

This shows that $j \in \mathcal{J}^{x,v}(x, v)$ and finishes the proof of the Theorem. □

Remark 3.1 *The integral in* (23) *is to be understood as the set of integrals over all measurable selections of the set-valued integrand. More precisely,* (23) *means that for any* $x^* \in \partial^c \varphi(x)$ *there exists a measurable function* β *such that for* μ_ζ *– almost every* $v \in F(x)$

$$\beta(v) \in \operatorname{Co} \left\{ -\frac{\chi\left(\hat{\rho}\left(x, v\right)\right)}{\left\langle \nabla_z g_j\left(x, \hat{\rho}\left(x, v\right) L v\right), L v \right\rangle} \nabla_x g_j\left(x, \hat{\rho}\left(x, v\right) L v\right) \middle| \, j \in \hat{\mathcal{J}}(x, v) \right\}$$

and $x^* = \int_{v \in F(x)} \beta(v) d\mu_\zeta(v)$.

4 Differentiability of φ and a gradient formula

Theorem 3.1 provides an immediate characterization for the differentiability of the probability function φ:

Theorem 4.1 *In addition to the assumptions of Theorem 3.1 suppose that*

$$\mu_\zeta(\{v \in F(\bar{x}) | \#\hat{\mathcal{J}}(\bar{x}, v) \geq 2\}) = 0. \tag{25}$$

Then, φ is Fréchet differentiable at $\bar{x}$ and

$$\nabla\varphi(\bar{x}) = - \int\limits_{v \in F(\bar{x}), \#\hat{\mathcal{J}}(\bar{x},v)=1} \frac{\chi\left(\hat{\rho}\left(\bar{x}, v\right)\right)}{\left\langle \nabla_z g_{j(v)}\left(\bar{x}, \hat{\rho}\left(\bar{x}, v\right) Lv\right), Lv \right\rangle} \nabla_x g_{j(v)}\left(\bar{x}, \hat{\rho}\left(\bar{x}, v\right) Lv\right) d\mu_\zeta(v), \tag{26}$$

where $\hat{\rho}\left(\bar{x}, v\right)$ is the unique solution in $r \geq 0$ of the equation $g^m(\bar{x}, rLv) = 0$ and $j(v)$ is the unique index $j \in \{1, \ldots, k\}$ satisfying $g_j(\bar{x}, \hat{\rho}\left(\bar{x}, v\right) Lv) = 0$. If (25) holds locally around $\bar{x}$, i.e., if there is a neighbourhood U of $\bar{x}$ such that

$$\mu_\zeta(\{v \in F(x) | \#\hat{\mathcal{J}}(x, v) \geq 2\}) = 0 \quad \forall x \in U, \tag{27}$$

then φ is continuously differentiable in U.

Proof. Under (25), the integrand in (23) (for $x := \bar{x}$) is single-valued μ_ζ – almost everywhere on $F(\bar{x})$, hence the integral is single-valued. Since $\partial^c\varphi(\bar{x})$ is nonempty by local Lipschitz continuity of φ on the one hand [2, Proposition 2.1.2] and is contained in the single-valued integral by (23) on the other hand, it follows that $\partial^c\varphi(\bar{x})$ coincides with the integral. In particular, $\partial^c\varphi(\bar{x})$ is single-valued and, hence φ is Fréchet differentiable [2, Proposition 2.2.4]. Moreover, $\partial^c\varphi(\bar{x}) = \{\nabla\varphi(\bar{x})\}$ and (26) follows from (23) upon observing that the integration domain can be reduced to those $v \in F(\bar{x})$ for which $\hat{\mathcal{J}}(\bar{x}, v)$ is a singleton (by (25)) and recalling the definition of $\hat{\mathcal{J}}(\bar{x}, v)$. The second assertion of the Theorem follows from [2, Corollary to Proposition 2.2.4]. □

Condition (27) may be difficult to verify in a concrete context as it refers to the uniform measure on the sphere of the radial projection of some set. In the following, we want to identify an explicit constraint qualification for the inequality system $g(x, z) \leq 0$ under which φ is (continuously) differentiable. In order to do so, we need the following characterization of the uniform measure over $\mathbb{S}^{m-1}$ as a so-called cone measure (see also [16]):

Lemma 4.1 *Let $A \subseteq \mathbb{S}^{m-1}$ be a Borel measurable subset. Then, the uniform measure μ_ζ on $\mathbb{S}^{m-1}$ can be represented as*

$$\mu_\zeta(A) = \frac{1}{\lambda(\mathbb{B})} \lambda(\mathrm{cone}(A) \cap \mathbb{B}), \tag{28}$$

where $\mathbb{B}$ is the closed unit ball, λ is the Lebesgue measure in $\mathbb{R}^m$ and $\mathrm{cone}(A)$ is the cone generated by the set A.

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

$$\mathcal{I}\left(x, z\right) := \{j \in \{1, \ldots, p\} \,|g_j\left(x, z\right) = 0\} \tag{29}$$

the active index set of g at (x, z). We say that the inequality system $g(x, z) \leq 0$ satisfies the *Rank-2-Constraint Qualification* $(R2CQ)$ at $x \in \mathbb{R}^n$ if

$$\text{rank} \ \ \{\nabla_z g_j(x, z), \nabla_z g_i(x, z)\} = 2 \quad \forall i, j \in \mathcal{I}(x, z), i \neq j \quad \forall z \in \mathbb{R}^m : g(x, z) \leq 0. \tag{R2CQ}$$

Note that $(R2CQ)$ is substantially weaker than the usual *Linear Independence Constraint Qualification (LICQ)* common in nonlinear optimization and requiring the linear independence of all gradients to active constraints.

Lemma 4.2 *Let g be as in Theorem 3.1. Moreover, let $\bar{x} \in \mathbb{R}^n$ be given such that*

1 $g^m(\bar{x}, 0) < 0.$

2 g *satisfies* $(R2CQ)$ *at* $\bar{x}$.

Then, $\mu_\zeta(M') = 0$ *for* $M' := \{v \in \mathbb{S}^{m-1} | \exists r > 0 : g(\bar{x}, rLv) \leq 0, \ \#\mathcal{I}(\bar{x}, rLv) \geq 2\}$*, where* L *is the regular matrix in the decomposition* $R = LL^T$.

Proof. For $i, j \in \{1, \ldots, k\}$, let

$$M_{i,j} := \left\{v \in \mathbb{S}^{m-1} | \exists r > 0 : g(\bar{x}, rLv) \leq 0, \ g_i(\bar{x}, rLv) = g_j(\bar{x}, rLv) = 0\right\}.$$

Since the union

$$M' = \bigcup_{i,j \in \{1,\ldots,k\}, i<j} M_{i,j}.$$

is finite, it is evidently sufficient to show that $\mu_\zeta(M_{i,j}) = 0$ for any $i, j \in \{1, \ldots, k\}$ with $i < j$. Without loss of generality, it is enough to verify that $\mu_\zeta(M_{1,2}) = 0$. Define

$$M^*_{1,2} := \{z \in \mathbb{R}^m | g(\bar{x}, z) \leq 0, \ g_1(\bar{x}, z) = g_2(\bar{x}, z) = 0\}.$$

and observe that $\mathbb{R}_+ M_{1,2} = L^{-1}\left(\mathbb{R}_+ M^*_{1,2}\right)$. We note first that $M_{1,2}$ is a Borel measurable subset of $\mathbb{S}^{m-1}$. Indeed, for any $l \in \mathbb{N}$, the set $[0, l] \cdot \left(M^*_{1,2} \cap \mathbb{B}(0, l)\right)$ is closed by closedness of $M^*_{1,2}$. Consequently

$$\mathbb{R}_+ M^*_{1,2} = \bigcup\nolimits_{l \in \mathbb{N}} [0, l] \cdot \left(M^*_{1,2} \cap \mathbb{B}(0, l)\right)$$

is Borel measurable in $\mathbb{R}^m$ and so is $\mathbb{R}_+ M_{1,2} = L^{-1}\left(\mathbb{R}_+ M^*_{1,2}\right)$. Since trivially $M_{1,2} = \mathbb{R}_+ M_{1,2} \cap \mathbb{S}^{m-1}$, it follows that $M_{1,2}$ is a Borel measurable subset of $\mathbb{S}^{m-1}$. This allows us to apply Lemma 4.2, in order to derive that

$$\mu_\zeta(M_{1,2}) = \frac{\lambda(\mathbb{R}_+ M_{1,2} \cap \mathbb{B})}{\lambda(\mathbb{B})} = \frac{\lambda\left(L^{-1}\left(\mathbb{R}_+ M^*_{1,2}\right) \cap \mathbb{B}\right)}{\lambda(\mathbb{B})}.$$

Hence, in order to prove the Lemma, it will be sufficient to show that

$$\lambda\left(L^{-1}\left(\mathbb{R}_+ M^*_{1,2}\right) \cap \mathbb{B}\right) = 0. \tag{30}$$

In order to do so, notice that $\text{rank} \ \ \{\nabla_z g_j(x, z)\}_{j=1,2} = 2$ for all $z \in M^*_{1,2}$ as a consequence of assumption 2. One may define for each $z \in M^*_{1,2}$ an open neighbourhood $W(z)$ such that the rank condition above extends to the whole neighbourhood. Then, $W := \bigcup_{z \in M^*} W(z)$ is an open set containing $M^*_{1,2}$ such that

$$\text{rank} \ \ \{\nabla_z g_j(x, z)\}_{j=1,2} = 2 \quad \forall z \in W. \tag{31}$$

Defining $\hat{M} := \{z \in W | g_j(\bar{x}, z) = 0 \quad (j = 1, 2)\}$, the respective definitions yield that $M_{1,2}^* \subseteq \hat{M}$. We show next that the set $L^{-1}\left(\mathbb{R}_+\hat{M}\right)\backslash\{0\}$ is a differentiable manifold of dimension $m-1$. Observe first, that by assumption 1., we have the following equivalence:

$$w \in L^{-1}\left(\mathbb{R}_+\hat{M}\right)\backslash\{0\} \Leftrightarrow \exists t > 0 : g_j(\bar{x}, tLw) = 0 \quad (j = 1, 2) \quad \text{and} \quad tLw \in W. \tag{32}$$

Let $\bar{t} > 0$ and $\bar{w}$ be arbitrarily chosen such that $\bar{t}L\bar{w} \in W$ and $g_j(\bar{x}, \bar{t}L\bar{w}) = 0$ for $j = 1, 2$. In particular, $\bar{w} \in F_j(\bar{x})$ for $j = 1, 2$. Define a mapping β by $\beta_j(w, t) := g_j(\bar{x}, tLw)$ for $j = 1, 2$. Then,

$$\nabla\beta(\bar{w}, \bar{t}) = \nabla_z g(\bar{x}, \bar{t}L\bar{w})(\bar{t}L|L\bar{w}) =: (A|b). \tag{33}$$

Thanks to (31), the matrix A is surjective, hence it contains a quadratic submatrix $\tilde{A}$ of order $(2, 2)$ which is regular. Without loss of generality, we may assume that $\tilde{A}$ consists of the first 2 columns of A. On the other hand, Lemma 2.1 (2.) that $\langle\nabla_z g_j(\bar{x}, \bar{t}L\bar{w}), L\bar{w}\rangle > 0$ for $j = 1, 2$. As a consequence, $\nabla_t\beta(\bar{w}, \bar{t}) = b \neq 0$ in (33). Therefore, we can exchange a suitable column in the regular matrix $\tilde{A}$ with the vector b without destroying its regularity. Assume, without loss of generality, that the last column of $\tilde{A}$ can be replaced by b such that the resulting matrix A' remains regular. Then, by the Implicit Function Theorem, the equations $\beta_j(w, t) = 0 \quad (j = 1, 2)$ can be resolved in a neighbourhood $U_{\bar{w}} \times U_{\bar{t}}$ of $(\bar{w}, \bar{t})$ as

$$w_1 = \tilde{\varphi}_1(w_2, \ldots, w_m) \tag{34}$$
$$t = \tilde{\varphi}_2(w_2, \ldots, w_m). \tag{35}$$

with certain continuously differentiable functions $\tilde{\varphi}_j \quad (j = 1, 2)$. Since $\bar{t} > 0$ and $\bar{t}L\bar{w} \in W$, we may further assume $U_{\bar{w}} \times U_{\bar{t}}$ to be small enough such that

$$tLw \in W, t > 0 \quad \forall (w, t) \in U_{\bar{w}} \times U_{\bar{t}}. \tag{36}$$

Now, $g_1(\bar{x}, \bar{t}L\bar{w}) = 0$ and (32) imply that $\bar{w} \neq 0$ and $\|\bar{w}\|^{-1}\bar{w} \in F_1(\bar{x})$. Hence, Lemma 2.2 guarantees the existence of a neighbourhood V of $\|\bar{w}\|^{-1}\bar{w}$ and a continuously differentiable function $\alpha : V \to \mathbb{R}_+$ such that for all $(v, r) \in V \times \mathbb{R}_+$ the equivalence

$$g_1(\bar{x}, rLv) = 0 \Leftrightarrow r = \alpha(v) \tag{37}$$

holds true. In particular, $\bar{t} = \|\bar{w}\|^{-1}\alpha\left(\|\bar{w}\|^{-1}\bar{w}\right)$. This allows us to define a neighbourhood $\tilde{U} \subseteq U_{\bar{w}}$ of $\bar{w}$ such that for all $w \in \tilde{U}$ one has that $\|w\|^{-1}w \in V$ and $\|w\|^{-1}\alpha\left(\|w\|^{-1}w\right) \in U_{\bar{t}}$. We claim that

$$w \in \tilde{U} \cap L^{-1}\left(\left[\mathbb{R}_+\hat{M}\right]\backslash\{0\}\right) \Leftrightarrow w \in \tilde{U} \quad \text{and} \quad w_1 = \tilde{\varphi}_j(w_2, \ldots, w_m). \tag{38}$$

Indeed, if $w \in \tilde{U} \cap L^{-1}\left(\left[\mathbb{R}_+\hat{M}\right]\backslash\{0\}\right)$, then by (32), there is some $t > 0$ such that $g_j(\bar{x}, tLw) = 0$ for all $j = 1, \ldots, l$. Since $\|w\|^{-1}w \in V$, we infer from (37) that $t = \|w\|^{-1}\alpha\left(\|w\|^{-1}w\right) \in U_{\bar{t}}$. Hence, $(w, t) \in U_{\bar{w}} \times U_{\bar{t}}$ and the direction '$\Rightarrow$' of our asserted equivalence follows from (34). Conversely, let $w \in \tilde{U}$ satisfy (34). Then, with t defined by (35), one has that $g_j(\bar{x}, tLw) = 0$ for all $j = 1, 2$. Taking into account (36), the direction '$\Leftarrow$' of our asserted equivalence then follows from (32).

In conclusion, as $\bar{w} \in L^{-1}\left(\left[\mathbb{R}_+\hat{M}\right]\backslash\{0\}\right)$ was arbitrary, the equivalence (38) shows that $L^{-1}\left(\left[\mathbb{R}_+\hat{M}\right]\backslash\{0\}\right)$ is a differentiable manifold of dimension $m - 1 < m$. As a consequence,

$\lambda\left(L^{-1}\left(\left[\mathbb{R}_+\hat{M}\right]\setminus\{0\}\right)\right) = 0$ (e.g., [9, Lemma 1.5]). Since $M^*_{1,2} \subseteq \hat{M}$, we infer that $\lambda\left(L^{-1}\left(\left[\mathbb{R}_+M^*_{1,2}\right]\setminus\{0\}\right)\right) = \lambda\left(L^{-1}\left[\mathbb{R}_+M^*_{1,2}\right]\right)\setminus\{0\}\right) = 0$, whence $\lambda\left(L^{-1}\left[\mathbb{R}_+M^*_{1,2}\right]\right)\right) = 0$ implying (30) as desired. This completes the proof. □

By combination of Lemma 4.2 and Theorem 4.1, we arrive at the main result of this section:

Corollary 4.1 *In addition to the assumptions of Theorem 3.1, suppose that $(R2CQ)$ is satisfied at $\bar{x}$. Then, φ is Fréchet differentiable at $\bar{x}$ and the gradient formula (26) holds true. If $(R2CQ)$ is satisfied locally around $\bar{x}$, then, φ is continuously differentiable at $\bar{x}$.*

Remark 4.1 *We note that neither* (27) *nor (R2CQ) are new conditions for ensuring differentiability of probability functions. They can be found in [22, Assumption 2.2 (iv)] in the context of spheric radial decomposition and in [12, Theorem 3.1, Assumption (vi)] in a general setting, respectively. However, in both references, compactness of the set $\{z|g(\bar{x},z)\leq 0\}$ is needed, which we do not impose here.*

5 Probability functions for linear random inequality systems

In this section, we are going to apply the previously obtained results to probability functions for linear random inequality systems:

$$\varphi(x) := \mathbb{P}\left(A(x)\xi \leq b(x)\right), \tag{39}$$

i.e., in (1) we have $g\left(x,\xi\right) = A(x)\xi - b(x)$ for matrix and vector functions $A : \mathbb{R}^n \to \mathbb{R}^{p\times m}, b : \mathbb{R}^n \to \mathbb{R}^p$. In this special case not only the resulting gradient formula becomes more explicit but, more importantly, several assumptions made before (exponential growth condition, local validity of (R2CQ)) can be omitted. The subsequently derived gradient formulae are fully explicit and 'ready-to-use' similar to those obtained in [11,18,28] for the same probability function but in a different disguise. The different representations of the same gradient may turn out to be advantageous depending on the concrete problem considered. In the following, for a matrix P we denote by P_j its jth row and by $P_{j,i}$ its entry in row j and column i.

Theorem 5.1 *In (39) let A, b be continuously differentiable and let $\xi \sim \mathcal{N}\left(0,R\right)$ for some positive definite correlation matrix R admitting a decomposition $R = LL^T$. Fix any $\bar{x}\in\mathbb{R}^n$ such that $b_j(\bar{x}) > 0$ for all $j\in\{1,\ldots,p\}$. Finally assume that any two rows of the matrix $A(\bar{x})$ are linearly independent. Then, φ in* (1) *is continuously differentiable at $\bar{x}$ and it holds that*

$$\nabla\varphi(\bar{x}) = -\int\limits_{\{v\in\mathbb{S}^{m-1}|J^*(v)\neq\emptyset,\#J^{**}(v)=1\}} \frac{\chi\left(\hat{\rho}\left(v\right)\right)}{A_{j(v)}(\bar{x})Lv}\left(\hat{\rho}\left(v\right)\sum_{i=1}^m \nabla A_{j(v),i}(\bar{x})L_i v - \nabla b_{j(v)}(\bar{x})\right) d\mu_\zeta(v), \tag{40}$$

where

$$\begin{aligned} J^*(v) &:= \left\{j\in\{1,\ldots,p\}\,|A_j(\bar{x})Lv > 0\right\}, \\ \hat{\rho}\left(v\right) &:= \min_{j\in J^*(v)}\left\{b_j(\bar{x})/(A_j(\bar{x})Lv)\right\}, \\ J^{**}(v) &:= \left\{j\in J^*(v)|\hat{\rho}\left(v\right) = b_j(\bar{x})/(A_j(\bar{x})Lv\right\}. \end{aligned}$$

and $j(v)$ is the unique element of the index set $J^{**}(v)$, i.e., $j(v)$ is the unique index $j\in\{1,\ldots,p\}$ satisfying $A_j(\bar{x})Lv > 0$ and $b_j(\bar{x}) = \hat{\rho}\left(v\right)A_j(\bar{x})Lv$.

Proof. In order to prove the result, we want to apply Corollary 4.1. To do so, we have first to check the assumptions of Theorem 3.1. The general assumptions of this Theorem as well as assumption 1. are clearly satisfied by the hypotheses we made. Concerning assumption 2. of Theorem 3.1, we claim that the exponential growth condition (Def. 2.1) is satisfied for all $j = 1, ..., p$. Indeed, by A being continuously differentiable, there exists a neighbourhood U of $\bar{x}$ and a constant K such that $\|\nabla A_{j,i}(x)\| \leq K$ for all $x \in U$ and all $i, j \in \{1, \ldots, p\}$. Then, Def. 2.1 holds true because of

$$\|\nabla_x g_j(x, z)\| = \left\|\sum_{i=1}^{m} z_i \nabla A_{j,i}(x)\right\| \leq K \|z\|_1 \leq K e^{\|z\|_1} \quad \forall z \in \mathbb{R}^m.$$

In order to verify the asserted continuous differentiability of φ via Corollary 4.1, it remains to check that the constraint qualification $(R2CQ)$ is satisfied on a neighbourhood of $\bar{x}$. Clearly, our assumption on pairwise linear independence of the rows of $A(\bar{x})$ implies $(R2CQ)$ to hold at $\bar{x}$ itself. If it didn't hold locally around $\bar{x}$, then there would be sequences $x_k \in \mathbb{R}^n$ $z_k \in \mathbb{R}^m$, $\lambda_k \in \mathbb{R}$ and $i_k, j_k \in \{1, \ldots, p\}$ such that

$$x_k \to \bar{x},\ A_{i_k}(x_k) z_k = b_{i_k}(x_k),\ A_{j_k}(x_k) z_k = b_{j_k}(x_k),\ A_{i_k}(x_k) = \lambda_k A_{j_k}(x_k),\ i_k \neq j_k.$$

By passing to a subsequence which we do not relabel, we may assume the existence of $i, j \in \{1, \ldots, p\}$ such that

$$A_i(x_k) z_k = b_i(x_k),\ A_j(x_k) z_k = b_j(x_k),\ A_i(x_k) = \lambda_k A_j(x_k),\ i \neq j.$$

We infer that $\lambda_k b_j(x_k) = b_i(x_k)$ for all k. From $b_i(x_k) \to b_i(\bar{x}) > 0$ and $b_j(x_k) \to b_j(\bar{x}) > 0$ we conclude that

$$\lambda_k \to \lambda := \frac{b_i(\bar{x})}{b_j(\bar{x})} \neq 0,$$

whence the contradiction $A_i(\bar{x}) = \lambda A_j(\bar{x})$ with our assumption on pairwise linear independence of the rows of $A(\bar{x})$. Consequently, we have shown that φ is continuously differentiable at $\bar{x}$. It remains to prove that the general gradient formula (26) ensured by Corollary 4.1 reduces in the special case of (39) to the asserted formula (40). This follows easily upon specifying the partial derivatives of g, the concrete shape of $\hat{\rho}(v)$ and upon observing the relations

$$F(\bar{x}) = \left\{v \in \mathbb{S}^{m-1} | J^*(v) \neq \emptyset\right\}, \quad \hat{\mathcal{J}}(\bar{x}, v) = J^{**}(v).$$

□

Next, we specialize the previous result to linear inequality systems $Az \leq b(x)$ with constant coefficient matrix. Without loss of generality, we may assume that $b(x) = x$ because the difference in the resulting gradient formulae consists just in a post-multiplication by the explicit derivative Db according to the chain rule. Hence, consider now the probability function

$$\varphi(x) := \mathbb{P}(A\xi \leq x). \tag{41}$$

This specialization of (39) not only leads to a substantially simpler gradient formula but also to a weakened constraint qualification, where only active rows of the matrix A come into play now in order to guarantee continuous differentiability of φ:

Corollary 5.1 *In (41) let $\xi \sim \mathcal{N}(0, R)$ for some positive definite correlation matrix R admitting a decomposition $R = LL^T$. Fix any $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}_j > 0$ for all $j \in \{1, \ldots, p\}$. Finally assume that any two active rows of the matrix A are linearly independent:*

$$Az \leq \bar{x},\ A_i z = \bar{x}_i,\ A_j z = \bar{x}_j,\ i \neq j \Longrightarrow \text{rank } \{A_i, A_j\} = 2. \tag{42}$$

Then, φ in (1) *is continuously differentiable at $\bar{x}$ and it holds that*

$$\frac{\partial \varphi}{\partial x_j}(\bar{x}) = \int\limits_{\{v \in \mathbb{S}^{m-1} | A_j L v > 0, \bar{x}_j = \hat{\rho}(v) A_j L v\}} \frac{\chi\left(\hat{\rho}\left(v\right)\right)}{A_j L v} \, d\mu_\zeta(v) \quad (j = 1, \dots, p) \,. \tag{43}$$

Proof. Clearly, the gradient formula (43) follows from (40) in the special setting of (41). Evidently, (42) corresponds to the general constraint qualification $(R2CQ)$. In order to derive continuous differentiability of φ via Corollary (4.1), it is sufficient to verify that it automatically holds locally round $\bar{x}$. If this was not the case, we could repeat the argument form the proof of Theorem 5.1 in order to derive the existence of sequences x_k, z_k, λ_k and of indices $i \neq j$ such that $x_k \to \bar{x}$, $A_i = \lambda_k A_j$ and $i, j \in \mathcal{I}\left(x_k, z_k\right)$. As in that proof it follows that $\lambda_k \to \lambda := \bar{x}_i / \bar{x}_j > 0$ and $A_i = \lambda A_j$. As a consequence of the Hausdorff continuity of the mapping $x \mapsto \{z | Az \leq x\}$, the relation $i \in \mathcal{I}\left(x_k, z_k\right)$ implies the existence of some $\bar{z}$ such that $A\bar{z} \leq \bar{x}$ and $A_i \bar{z} = \bar{x}_i$. Then, also $A_j z = \lambda^{-1} \bar{x}_i = \bar{x}_j$ but $\operatorname{rank} \ \{A_i, A_j\} = 1$, a contradiction with (42). □

Finally, we consider a gradient formula for the cumulative distribution function

$$F_\xi(x) := \mathbb{P}\left(\xi \leq x\right) \tag{44}$$

associated with a Gaussian random vector:

Corollary 5.2 *Let $\xi \sim \mathcal{N}\left(0, R\right)$ for some positive definite correlation matrix R admitting a decomposition $R = LL^T$. Fix any $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}_j > 0$ for all $j \in \{1, \dots, p\}$. Then, φ in* (1) *is continuously differentiable at $\bar{x}$ and it holds that*

$$\frac{\partial \varphi}{\partial x_j}(\bar{x}) = \int\limits_{\{v \in \mathbb{S}^{m-1} | L_j v > 0, \bar{x}_j = \hat{\rho}(v) L_j v\}} \frac{\chi\left(\hat{\rho}\left(v\right)\right)}{L_j v} \, d\mu_\zeta(v) \quad (j = 1, \dots, p) \,.$$

Proof. (44) follows from (41) by putting $A := I$. Clearly, any two rows of I are linearly independent. The gradient formula follows from $A_j L v = L_j v$ in this special case. □

6 Mordukhovich subdifferential of probability functions for linear random inequality systems

We reconsider the probability function φ in (41) under the assumptions of Corollary 5.1 except the constraint qualification (42). Without this constraint qualification, we cannot hope for differentiability of φ (see Example 1.1). Nevertheless, it is still locally Lipschitzian and admits an upper estimate for its Mordukhovich subdifferential which is more precise than its Clarke subdifferential. This allows us to sharpen the upper estimate in the general result (23) for this special class of problems. In order to prepare a corresponding result, we introduce the following equivalence class within the index set $\{1, \dots, p\}$ of rows of the matrix A in (41):

$$i \sim j \Longleftrightarrow \exists \lambda \in \mathbb{R} : A_i = \lambda A_j, \; \bar{x}_i = \lambda \bar{x}_j.$$

By the assumption $\bar{x}_j > 0$ for all $j \in \{1, \dots, p\}$ made in Corollary 5.1, $i \sim j$ implies that $\lambda > 0$ in the defining relation. Similarly, $i \nsim j$ implies that (42) is satisfied. Denote by $\tilde{p} \leq p$ the number of different equivalence classes $[i]$. Without loss of generality, we may assume that the first $\tilde{p}$ rows of A belong to different equivalence classes. Then, it obviously holds for any $i = 1, \ldots, \tilde{p}$ that

$$A_j z \leq x_j \quad \forall j \in [i] \Longleftrightarrow A_i z \leq h_i(x) := \min_{j \in [i]} \lambda_j^{-1} x_j. \tag{45}$$

We denote by $\tilde{A}$ the submatrix of first $\tilde{p}$ rows of A.

Theorem 6.1 *In (41) let $\xi \sim \mathcal{N}(0, R)$ for some positive definite correlation matrix R admitting a decomposition $R = LL^T$. Fix any $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}_j > 0$ for all $j \in \{1, \ldots, p\}$. Then, φ is locally Lipschitz continuous and its Mordukhovich subdifferential can be estimated from above by*

$$\partial^M \varphi(\bar{x}) \subseteq \sum_{i=1}^{\tilde{p}} \int\limits_{\left\{v \in \mathbb{S}^{m-1} | \tilde{A}_i L v > 0, \bar{y}_i = \hat{\rho}(v) \tilde{A}_i L v\right\}} \frac{\chi(\hat{\rho}(v))}{\tilde{A}_i L v} d\mu_\zeta(v) \cdot \bigcup \left\{\lambda_j^{-1} e_j | j \in [i] : \lambda_j^{-1} \bar{x}_j = h_i(\bar{x})\right\},$$

where

$$\hat{\rho}(v) := \min \left\{\bar{y}_j / (A_j L v) | j \in \{1, \ldots, \tilde{p}\} : \tilde{A}_j L v > 0\right\}.$$

Proof. We introduce the modified probability function $\tilde{\varphi}(y) := \mathbb{P}\left(\tilde{A}\xi \leq y\right)$ (for $y \in \mathbb{R}^{\tilde{p}}$) and observe that thanks to (45) the original probability function φ in (41) can be written as the composition $\varphi = \tilde{\varphi} \circ h$ with a Lipschitz continuous mapping $h = (h_1, \ldots, h_{\tilde{p}})$. Since the rows of $\tilde{A}$ refer to rows belonging to different equivalence classes in A, they satisfy (42) (see remarks preceding the statement of this theorem). Furthermore, $\bar{y} := h(\bar{x})$ satisfies $\bar{y}_i > 0$ for $i = 1, \ldots, \tilde{p}$. This allows us to derive from Corollary 5.1 that $\tilde{\varphi}$ is continuously differentiable in a neighbourhood of $\bar{y}$ and that

$$\frac{\partial \tilde{\varphi}}{\partial y_i}(\bar{y}) = \int\limits_{\left\{v \in \mathbb{S}^{m-1} | \tilde{A}_i L v > 0, \bar{y}_i = \hat{\rho}(v) \tilde{A}_i L v\right\}} \frac{\chi(\hat{\rho}(v))}{\tilde{A}_i L v} d\mu_\zeta(v) \quad (i = 1, \ldots, \tilde{p}),$$

where $\hat{\rho}(v)$ is defined in the statement of this theorem. The chain rule for the Mordukhovich subdifferential [15, Theorem 1.110 (ii)] now yields that

$$\begin{aligned} \partial^M \varphi(\bar{x}) &= \partial^M \langle \nabla \tilde{\varphi}(h(\bar{x})), h \rangle (\bar{x}) = \partial^M \left(\sum_{i=1}^{\tilde{p}} \frac{\partial \tilde{\varphi}}{\partial y_i}(h(\bar{x})) \cdot h_i\right)(\bar{x}) \\ &\subseteq \sum_{i=1}^{\tilde{p}} \partial^M \left(\frac{\partial \tilde{\varphi}}{\partial y_i}(h(\bar{x})) \cdot h_i(\bar{x})\right), \end{aligned}$$

where the last inclusion follows from the sum rule in [15, Theorem 2.33 (c)]. Next, we observe that $\frac{\partial \tilde{\varphi}}{\partial y_i}(h(\bar{x})) \geq 0$ for all $i = 1, \ldots, \tilde{p}$ because $\tilde{\varphi}$ is evidently nondecreasing with respect to the partial order of $\mathbb{R}^{\tilde{p}}$. This allows us by [15, p. 112] to continue the previous relation as

$$\partial^M \varphi(\bar{x}) \subseteq \sum_{i=1}^{\tilde{p}} \frac{\partial \tilde{\varphi}}{\partial y_i}(h(\bar{x})) \cdot \partial^M h_i(\bar{x}). \tag{46}$$

Given the definition of components h_i in (45), we conclude from [15, Theorem 1.113] that

$$\partial^M h_i(\bar{x}) \subseteq \bigcup \left\{\lambda_j^{-1} e_j | j \in [i] : \lambda_j^{-1} \bar{x}_j = h_i(\bar{x})\right\} \quad (i = 1, \ldots, \tilde{p}),$$

where e_j refers to the jth canonical unit vector in $\mathbb{R}^n$. (Actually, as the components h_i are minima over linear functions, it is easy to show that even equality holds in the previous relation; we cannot benefit, however, from this improvement because (46) already involves an inclusion anyway). ☐

References

[1] 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. *Mathematics of Computation*, 83:2821–2851, 2014.

[2] F.H. Clarke. *Optimisation and Nonsmooth Analysis*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1987.

[3] I. Deák. Computing probabilities of rectangles in case of multinormal distribution. *Journal of Statistical Computation and Simulation*, 26(1-2):101–114, 1986.

[4] I. Deák. Subroutines for computing normal probabilities of sets - computer experiences. *Annals of Operations Research*, 100:103–122, 2000.

[5] D. Dentcheva. *Optimisation Models with Probabilistic Constraints. Chapter 4 in [24]*. MPS-SIAM series on optimization. SIAM and MPS, Philadelphia, 2009.

[6] O. Ditlevsen and H.O. Madsen. *Structural Reliability Methods*. Wiley, 1st edition, 1996.

[7] J. Garnier, A. Omrane, and Y. Rouchdy. Asymptotic formulas for the derivatives of probability functions and their Monte Carlo estimations. *European Journal of Operations Research*, 198:848–858, 2009.

[8] A. Genz and F. Bretz. *Computation of multivariate normal and t probabilities.* Number 195 in Lecture Notes in Statistics. Springer, Dordrecht, 2009.

[9] M. Golubitsky and V. Guillemin. *Stable Mappings and Their Singularities*, volume 14 of *Graduate Texts in Mathematics*. Springer, New York, 1973.

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

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

[12] A.I. Kibzun and S. Uryas'ev. Differentiability of probability function. *Stochastic Analysis and Applications*, 16:1101–1128, 1998.

[13] G. Lebourg. Generic differentiability of lipschitzian functions. *Transactions of the American Mathematical Society*, 256:125–144, 1979.

[14] K. Marti. Differentiation of probability functions: The transformation method. *Computers and Mathematics with Applications*, 30:361–382, 1995.

[15] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation I. Basic Theory.* Springer, Heidelberg, 2006.

[16] A. Naor and D. Romik. Projecting the surface measure of the sphere of ℓ_p^n. *Ann. I.H. Poincaré*, 39(2):241–261, 2003.

[17] G. Pflug and H. Weisshaupt. Probability gradient estimation by set-valued calculus and applications in network design. *SIAM Journal on Optimization*, 15:898–914, 2005.

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

[19] A. Prékopa. *Probabilistic programming. In [23] (Chapter 5)*, pages 267–352. Elsevier, Amsterdam, 2003.

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

[21] J.O. Royset and E. Polak. Implementable algorithm for stochastic optimization using sample average approximations. *Journal of Optimization Theory and Applications*, 122(1):157–184, 2004.

[22] J.O. Royset and E. Polak. Extensions of stochastic optimization results to problems with system failure probability functions. *Journal of Optimization Theory and Applications*, 133(1):1–18, 2007.

[23] A. Ruszczyński and A. Shapiro. *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 2003.

[24] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming. Modeling and Theory*, volume 9 of *MPS-SIAM series on optimization*. SIAM and MPS, Philadelphia, 2009.

[25] S. Uryas'ev. Derivatives of probability functions and integrals over sets given by inequalities. *Journal of Computational and Applied Mathematics*, 56(1-2):197–223, 1994.

[26] S. Uryas'ev. Derivatives of probability functions and some applications. *Annals of Operations Research*, 56:287–311, 1995.

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

[28] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati. On joint probabilistic constriants with Gaussian Coefficient Matrix. *Operations Research Letters*, 39:99–102, 2011.

[29] W. van Ackooij and M. Minoux. A characterization of the subdifferential of singular Gaussian distribution functions. *Set-Valued and Variational Analysis*, 23:465–483, 2015.