SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 5, pp. 63–87

(Sub-)Gradient Formulae for Probability Functions of Random Inequality Systems under Gaussian Distribution*

Wim van Ackooij† and R. Henrion‡

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 the 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.

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

AMS subject classification. 90C15

DOI. 10.1137/16M1061308

1. Introduction.

A probability function has the form

$$\varphi(x) := \mathbb{P}(g(x,\xi) \le 0), \tag{1.1}$$

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 (1.1) is to be understood componentwise. Throughout the paper we shall make the following basic assumptions:

$$\begin{array}{l} g \text{ is continuously differentiable,} \\ \text{the mappings } g_j(x,\cdot) \text{ are convex for all } x \in \mathbb{R}^n \text{ and all } j = 1,\dots,p, \\ \xi \sim \mathcal{N}(0,R) \text{ is nondegenerate Gaussian with } R_{ii} = 1\ (i = 1,\dots,m). \end{array} \tag{1.2}$$

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

*Received by the editors February 12, 2016; accepted for publication (in revised form) September 27, 2016; published electronically January 10, 2017.

http://www.siam.org/journals/juq/5/M106130.html

Funding: The work of the second author was supported by the FMJH Program Gaspard Monge in optimization and operations research including support to this program by the EDF as well as support by the Deutsche Forschungsgemeinschaft within Project B04 in CRC TRR 154.

†OSIRIS, EDF R&D, 91120 Palaiseau, France (wim.van-ackooij@edf.fr).

‡Nonlinear Optimization and Inverse Problems, Weierstrass Institute Berlin, Germany 10117 (henrion@wias-berlin.de).

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 (1.2).

Probability functions (1.1) 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., [20, 21, 26]). Treating probability functions in the framework of optimization problems (w.r.t. the decision variable x) requires calculating—or, better, approximating—not only the probability $\varphi(x)$ itself but also its gradient $\nabla\varphi$. This is why derivatives of probability functions have attracted so much attention in the past (see, e.g., [8, 12, 14, 16, 19, 22, 24, 25, 28, 29, 31, 32]). Indeed, historically a first formula for the gradient of a probability function in the form of an involved surface integral was derived in [22]. This formula was later generalized into a combination of a surface and volume integral in [28, 29]. Here is it also worthwhile to mention that [16] first suggested transforming the probability function into a volume integral in order to ensure that the integration domain does not depend on the decision vector. 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., [20, p. 207], [25, 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., [28, p. 200, Assumptions (A2)], [25, Assumption 2.2(i)], [19, p. 902]). Indeed, without compactness, one cannot expect differentiability of φ even with the nicest data. In [30, Proposition 2.2] an example of only a single inequality $g(x, \xi) \leq 0$ (i.e., $p = 1$) is provided, where the basic assumptions (1.2) 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., [20, p. 204]), corresponding gradient formulae could be found for mappings $g(x, \xi) = A(x)\xi \leq b(x)$ in (1.1) with surjective $A(x)$ [31] or with possibly nonsurjective $A(x) \equiv A$ under the linear independence constraint qualification (LICQ) for the set K [12]. 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's code [9], can be employed to calculate not only 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 (see also [32]). Finally, the precision for calculating $\nabla\varphi$ can be controlled by that for calculating Gaussian distributions functions, for instance in Genz's code [11, 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 [4, 5, 9] (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 [7] (without rigorous proof) or in [24, 25], albeit under the restrictive compactness assumption on the set K. In order to overcome this assumption, the work [30] identified 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 (1.2) 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 is typical in most applications), Slater's condition is no longer sufficient to guarantee differentiability of φ even if K is compact and g a linear mapping, as shown next.

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 [30] from a single inequality to inequality systems, but even to take a more general, namely nonsmooth, analysis perspective for viewing 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 the reader to, for instance, [6, 33]. 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 componentwise 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 [25, Assumption 2.2(iv)] and [14, Theorems 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 [31, 12] 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 toward the use of the Mordukhovich rather than the bigger Clarke 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 [33].

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 w.r.t. 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 [2, 5]. 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 was presented in the case of a single inequality in [30].

Notation. Throughout this paper R will be a positive definite $m \times m$ correlation matrix with associated matrix L resulting from the decomposition (e.g., Cholesky) $R = LL^\mathsf{T}$. For a given set $A \subseteq \mathbb{R}^s$ and $s \geq 1$, $\mathrm{Co}(A)$ will denote the convex hull of A, and the notation $\mathbb{R}_+A$ is to be interpreted as $\mathbb{R}_+A := \{ra \ : r \geq 0, a \in A\}$. For a given finite subset $J \subseteq \mathbb{N}$, $\#J$ will denote its cardinal. Finally, for a $k \times p$ matrix B and vector $b \in \mathbb{R}^k$, $(B|b)$ denotes the $k \times (p+1)$ matrix resulting from B by appending b as the $(p+1)$th column.

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 $R = LL^\mathsf{T}$ 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

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

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

$$\mathbb{P}(\xi \in M) = \int_{v \in \mathbb{S}^{m-1}} \mu_\eta\left(\{r \geq 0 : rLv \cap M \neq \emptyset\}\right) d\mu_\zeta(v), \tag{2.1}$$

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, (2.1) keeps holding true upon transforming the set M accordingly.

2.2. Probability function in spheric-radial form and preliminary results. Given the constraint mapping g in (1.1), we pass to the maximum function $g^m : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ over its components by defining

$$g^m(x, z) = \max_{j=1,\dots,p} g_j(x, z). \tag{2.2}$$

Evidently, the probability function (1.1) can be written as $\varphi(x) = \mathbb{P}(g^m(x,\xi) \leq 0)$. By (2.1) we have that

$$\varphi(x) = \int_{v\in\mathbb{S}^{m-1}} \mu_\eta\left(\{r \geq 0 : g^m(x, rLv) \leq 0\}\right) d\mu_\zeta(v) = \int_{v\in\mathbb{S}^{m-1}} e(x,v) d\mu_\zeta(v), \tag{2.3}$$

where

$$e(x,v) := \mu_\eta\left(\{r \geq 0 : g^m(x, rLv) \leq 0\}\right) \quad \forall x \in \mathbb{R}^n,\ \forall v \in \mathbb{S}^{m-1}. \tag{2.4}$$

As a consequence of (1.2), g^m is convex in the second argument. In [30], probability functions of a single continuously differentiable inequality, convex in the Gaussian random vector ξ, have been investigated. Because our inequality $g^m(x,\xi) \leq 0$ 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 $g(x,z) \leq 0$ in z. This is no severe restriction because in the case that $g^m(x,0) \geq 0$, the feasible set $\{z | g(x,z) \leq 0\}$ would be a subset of some half-space containing zero by convexity of $g^m(x,\cdot)$. As a consequence of ξ having a symmetric and centered distribution (see (1.2)), the probability of this half-space would be 0.5, implying that $\varphi(x) \leq 0.5$. In many practical applications, however, values of probability functions close to 1 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 $v \in \mathbb{S}^{m-1}$, (2.4) can be simplified as

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

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

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

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 the following hold:*

1. *$F_j(x) \cup I_j(x) = F(x) \cup I(x) = \mathbb{S}^{m-1}$ for all $j = 1, \ldots, p$.*
2. *For $j \in \{1, \ldots, p\}$ and $v \in F_j(x)$ let $r > 0$ be such that $g_j(x, rLv) = 0$. Then,*

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

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. Statements 1 and 3 are obvious, whereas 2 follows easily from the convexity of $g(x,\cdot)$ (see [30, 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 (2.4), $e(x,v) = \mu_\eta(\mathbb{R}_+) = 1$ because the support of the Chi-distribution is $\mathbb{R}_+$. Otherwise, if $v \in F(x)$, then again via point 1 and by the 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 $e(x,v) < 1$. ■

Lemma 2.2 (Lemma 3.2 in [30]). *Let $j = 1, \ldots, p$ be arbitrary, and let (x,v) be such that $g_j(x,0) < 0$ and $v \in F_j(x)$. Then there exist neighborhoods 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 2.3. *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 neighborhood $U(x)$ such that*

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

Lemma 2.4 (Lemmas 3.3 and 3.7 in [30]). *Let $j = 1, \ldots, p$ 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 neighborhood 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 [3, 17]).

Definition 2.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary function, and fix any $\bar{x} \in \mathbb{R}^n$. Then,*

- *the Fréchet subdifferential 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*

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

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 (1.1). The main result of this section is formulated in Theorem 3.6. It will be based on interchanging subdifferentiation and integration in (2.3). This requires calculating the Clarke subdifferential of the function e in (2.4) first. To start, 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 $J_F^{x,v} := \{j \in \{1,\dots,p\} | v \in F_j(x)\}$, the functions $\rho_j^{x,v}$ from Lemma 2.2 are well defined for $j \in J_F^{x,v}$ on the neighborhood $\tilde{U} \times \tilde{V}$ of (x,v), where, with U_j, V_j from Lemma 2.2,*

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

Moreover, there exist neighborhoods $U \subseteq \tilde{U}$ 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 the resolving function $\rho^{x,v} : \tilde{U} \times \tilde{V} \to \mathbb{R}_+$ is defined as*

$$\rho^{x,v}(x',v') := \min_{j \in J_F^{x,v}} \rho_j^{x,v}(x',v') \quad \forall \left(x',v'\right) \in \tilde{U} \times \tilde{V}. \tag{3.1}$$

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

$$\partial_x^c \rho^{x,v}(x',v') = \mathrm{Co}\left\{ \nabla_x \rho_j^{x,v}(x',v') \;: j \in \mathcal{J}^{x,v}(x',v') \right\}, \tag{3.2}$$

where $\mathcal{J}^{x,v}(x',v') := \{j \in J_F^{x,v} | \rho_j^{x,v}(x',v') = \rho^{x,v}(x',v')\}$ is the active index set.

Proof. Our assumptions and Lemma 2.1.3 imply that $g_j(x,0) < 0$ for all $j \in \{1,\dots,p\}$ 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 neighborhood of (x,v), and Lemma 2.2.1 yields the equivalence

$$g_j(x', r'Lv') = 0 \Leftrightarrow r' = \rho_j^{x,v}(x',v') \quad \forall \left(x',v',r'\right) \in \tilde{U} \times \tilde{V} \times \mathbb{R}_+, \forall j \in J_F^{x,v}. \tag{3.3}$$

In particular, the min-function $\rho^{x,v}$ in (3.1) is well defined and continuous on $\tilde{U} \times \tilde{V}$. We may clearly shrink $\tilde{U} \times \tilde{V}$ to a neighborhood $U \times V$ 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

$$\rho^{x,v}(x',v') \le R \quad \forall \left(x',v'\right) \in U \times V. \tag{3.4}$$

Moreover, since $j \in \left(J_F^{x,v}\right)^c$ entails $v \in I_j(x)$ (by Lemma 2.1 parts 1 and 3), Lemma 2.4 allows us to shrink $U \times V$ once more such that

$$\rho_j^{x',v'}(x',v') \ge R+1 \quad \forall \left(x',v'\right) \in U \times V : v' \in F_j\left(x'\right), \forall j \in \left(J_F^{x,v}\right)^c. \tag{3.5}$$

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 (3.3), we then derive the contradiction

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

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

$$g_j(x',\rho_j^{x,v}(x',v')Lv') = 0 = g^m(x',r'Lv') < g^m(x',\rho^{x,v}(x',v')Lv') \quad \forall j \in J_F^{x,v}. \tag{3.6}$$

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

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

Now, convexity of $g_j(x',\cdot)$ provides that $g_j(x',\rho^{x,v}(x',v')Lv') \le g_j(x',\rho_j^{x,v}(x',v')Lv')$. This allows us to conclude from (3.6) 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

$$g_j(x',\rho^{x,v}(x',v')Lv') < 0 < g^m(x',\rho^{x,v}(x',v')Lv'), \tag{3.7}$$

with the first inequality following from the definition of $I_j(x')$ and the second one following from (3.6). In the opposite case, one has that $v' \in F_j(x')$ by Lemma 2.1.1. Then, exploiting (3.4) and (3.5), we end up with $\rho_j^{x',v'}(x',v') > \rho^{x,v}(x',v')$. Hence, with the same convexity argument as before,

$$0 = g_j(x',\rho_j^{x',v'}(x',v')Lv') > g_j(x',\rho^{x,v}(x',v')Lv'). \tag{3.8}$$

Combining this with (3.7), we have shown that

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

Together with (3.6), this brings us to the contradiction

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

with the definition of g^m. Summarizing, we have proven that $r' = \rho^{x,v}(x',v')$, which shows the part "$\Rightarrow$" 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 (3.3),

$$g_{j^*}(x',r'Lv') = g_{j^*}(x',\rho_{j^*}^{x,v}(x',v')Lv') = 0. \tag{3.9}$$

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

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

by $g_j(x',0) < 0$ and the convexity of $g_j(x',\cdot)$. Finally, for $j \in \left(J_F^{x,v}\right)^c$ 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 $v' \in F_j(x')$, (3.4) and (3.5) 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 $g_j(x',r'Lv') \le 0$ for all $j = 1,\dots,p$, which, together with (3.9), 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 [3, Proposition 2.3.12] to $-\rho^{x,v} = \max_{j\in J_F^{x,v}} -\rho_j^{x,v}$ 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 [3, Proposition 2.3.1], which allows us to prove (3.2) 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 (2.2) and $\rho^{x,v}$ in (3.1) would be continuously differentiable, and hence Lemma 3.1.1 would allow us to invoke two results [30, 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 (2.2) and (3.1) applied to the (differentiable) components g_j and $\rho_j^{x,v}$, respectively. Since neither of the two above-mentioned results exploits differentiability arguments and since 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 hold:*

1. *If $v \in F(x)$, then there exist neighborhoods U of x and V of v such that $e(x',v') = F_\eta(\rho^{x,v}(x',v'))$ for all $(x',v') \in U \times V$, where e and $\rho^{x,v}$ are defined in (2.4) and (3.1), respectively, and F_η is the 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 3.3. *Let* $x \in \mathbb{R}^n$ *be such that* $g^m(x, 0) < 0$ *and* $v \in F(x)$*. Then there exists a neighborhood* $U \times V$ *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* $\mathcal{J}^{x,v}$ *is as introduced in Lemma* 3.1*.*

Proof. From Lemma 3.2.1, we know that $e = F_\eta \circ \rho^{x,v}$ in a neighborhood $U \times V$ of (x, v). We may assume this neighborhood small enough so that $\rho^{x,v}$ is Lipschitz there as a minimum of smooth functions, by (3.1). Since the mapping F_η is continuously differentiable with $F'_\eta = \chi$, Clarke's chain rule [3, 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 (3.2). ■

In the following we want to generalize Corollary 3.3 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 3.4. *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 is to be understood as the Painlevé–Kuratowski limit.

Proof. By Corollary 3.3 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

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

The assertion of our lemma will follow if we can show that $\mu_j^{(k)} \to_k 0$ for all $j \in \{1,\dots,p\}$. In order to do so, fix any $j \in \{1,\dots,p\}$. 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 2.4 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

$$\left\|\mu_j^{(k_l)}\right\| \le \varepsilon \quad \forall l \ge l'. \tag{3.10}$$

Set $k' := k_{l'}$. Then, for any $k \ge k'$ one either has that $j \in \mathcal{J}^{x_k,v_k}(x_k,v_k)$, in which case $k = k_l$ for some $l \ge l'$ and hence (3.10) holds true, or $j \notin \mathcal{J}^{x_k,v_k}(x_k,v_k)$, in which case $\mu_j^{(k)} = 0$. The claimed convergence again follows, $\mu_j^{(k)} \to_k 0$, i.e., $\|\mu_j^{(k)}\| \le \varepsilon$ for all $k \ge k'$. ∎

Corollary 3.5. *Let x be such that $g^m(x,0) < 0$ and that g_j satisfies the exponential growth condition at x for all $j = 1,\dots,p$. Then, for any $v \in \mathbb{S}^{m-1}$, the function $e(\cdot,v)$ is Lipschitz continuous in a neighborhood of x, and its Clarke subdifferential is given by*

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

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 (the latter two were 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 neighborhood of x by Lemma 3.2.1; hence $e(\cdot,v)$ is Lipschitz continuous on this neighborhood, and the asserted formula for $\partial_x^c e(x,v)$ follows from Corollary 3.3 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

$$|e(x_k,v) - e(y_k,v)| > k\,\|x_k - y_k\| \quad \forall k \in \mathbb{N}. \tag{3.11}$$

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

$$g^m(x',0) < 0 \quad \forall x' \in [x_k,y_k], \forall k \in \mathbb{N}. \tag{3.12}$$

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

$$v \in F(z_k) \quad \text{and} \quad |e(x_k, v) - e(y_k, v)| \le \left(\|x_k^*\| + k^{-1}\right)\|x_k - y_k\|. \tag{3.13}$$

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 (3.11). Hence, without loss of generality, $v \in F(x_k)$. Define $x^t := (1-t)\, x_k + t y_k$ 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

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

Since $\alpha \in (0,\tau)$, this implies that $e(x^{t'}, v) < 1$ for all $t' \in [0, \alpha]$, and Lemma 2.1.5 yields that

$$v \in F(x^{t'}) \quad \forall t' \in [0, \alpha]. \tag{3.14}$$

Taking into account (3.14) and that, by (3.12), $g^m(x^{t'}, 0) < 0$ for all $t' \in [0, \alpha]$, Corollary 3.3 yields that $e(\cdot, v)$ is locally Lipschitz continuous on an open neighborhood of the line segment $[x^0, x^\alpha]$. This allows us to invoke Lebourg's mean value theorem [15, Theorem 1.7] in order to derive the existence of some $t^* \in [0, \alpha]$ and some $x^* \in \partial_x^c e(x^{t^*}, v)$ 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

$$|e(x_k, v) - e(x^\tau, v)| \le \|x^*\|\|x_k - x^\alpha\| + k^{-1}\|x_k - y_k\| \le \left(\|x^*\| + k^{-1}\right)\|x_k - y_k\|. \tag{3.15}$$

Clearly, $v \in F(x^{t^*})$ by (3.14). If $\tau = 1$, then $x^\tau = y_k$, and (3.13) follows upon putting $z_k := x^{t^*}$ and $x_k^* := x^*$. Otherwise, $\tau < 1$, and then $e(x^\tau, v) = 1$ 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 $e(y_k, v) = 1$ by Lemma 2.1.4, and so (3.13) follows from (3.15) and $e(y_k, v) = e(x^\tau, v)$ 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 (3.15). 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,

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

Now, (3.13) follows once more from (3.15) with the same z_k, x_k^* as before. Since $k \in \mathbb{N}$ was chosen arbitrarily, we have altogether verified (3.13). 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 3.4 yields that $\|x_k^*\| + k^{-1}$ is a bounded sequence contradicting (3.11). Summarizing, we have proven Lipschitz continuity of $e(\cdot, v)$ around x.

It remains to calculate the Clarke subdifferential of $e(\cdot, v)$. By [3, 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,

$$\|\nabla_x e(x_{l_k}, v)\| \geq \varepsilon \quad \forall k \tag{3.16}$$

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}, the contradiction $\nabla_x e(x_{l_k}, v) = 0$ follows with (3.16). Hence, $v \in F(x_{l_k})$ for all k, and so by Lemma 3.4 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 [3, Proposition 2.2.2], whence $\nabla_x e(x_{l_k}, v) \to_k 0$, which is a contradiction with (3.16) 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 (3.17) has to be interpreted as explained in Remark 3.1 below.

Theorem 3.6. *In addition to our basic assumptions* (1.2)*, 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* $\bar{x}$ *(Definition* 2.3*) for all* $j = 1, \ldots, p$.

Then, φ *in* (1.1) *is locally Lipschitz continuous on a neighborhood* U *of* $\bar{x}$*, and it holds that*

$$\partial^c \varphi(x) \subseteq \int_{v\in F(x)} \mathrm{Co}\left\{ -\frac{\chi(\hat{\rho}(x,v))}{\langle \nabla_z g_j(x, \hat{\rho}(x,v) Lv), Lv\rangle} \nabla_x g_j(x, \hat{\rho}(x,v) Lv) \,\middle|\, j \in \hat{\mathcal{J}}(x,v) \right\} d\mu_\zeta(v) \tag{3.17}$$

for all $x \in U$*. Here,* $\hat{\rho}(x,v)$ *refers to the unique solution in* $r \geq 0$ *of the equation* $g^m(x, rLv) = 0$ *and*

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

Proof. Assumptions 1 and 2 continue to hold on an appropriate neighborhood of $\bar{x}$. Hence, there exists an open neighborhood $\tilde{U}$ of $\bar{x}$ such that

$$g^m(x,0) < 0, \quad g_j \text{ satisfy the exponential growth condition} \quad \forall j = 1,\ldots,p,\ \forall x \in \tilde{U}. \tag{3.18}$$

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 semicontinuous 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 3.3,

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

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

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

which is the upper semicontinuity 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 assumptions of the theorem hold on a neighborhood of 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 3.5, 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 3.4,

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

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

Let $\mathbb{B}(\bar{x}; r)$ be a closed ball centered at $\bar{x}$ and with radius $r > 0$ such that $\mathbb{B}(\bar{x}; r) \subseteq \tilde{U}$. By the Weierstrass theorem, the upper semicontinuous function α realizes its maximum on the compact set $\mathbb{B}(\bar{x}; r) \times \mathbb{S}^{m-1}$; hence α is bounded on this set by some constant $M > 0$. Define the open neighborhood $U := \mathrm{int}\,\mathbb{B}(\bar{x}; r)$, 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 3.5 and as a consequence of (3.18). Lebourg's mean value theorem [15, 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 $\tilde{x} \in \mathbb{B}(\bar{x}; r)$, 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 [3, Theorem 2.7.2] in order first to conclude that φ is locally Lipschitz continuous on U and second to derive from (2.3) and from Corollary 3.5 the formula

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

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

By Lemma 3.1.1, $\rho^{x,v}(x,v)$ is the unique solution in r of the equation $g^m(x,rLv)=0$, and hence $\hat{\rho}(x,v)=\rho^{x,v}(x,v)$ with $\hat{\rho}$ as introduced in the statement of this theorem. It remains to show that

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

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

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

and so $j\in\hat{\mathcal{J}}(x,v)$. Conversely, let $j\in\hat{\mathcal{J}}(x,v)$ be arbitrary. Then, $g_j(x,\hat{\rho}(x,v)Lv)=0$, 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 (3.1),

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

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

Remark 3.1. The integral in (3.17) is to be understood as the set of integrals over all measurable selections (see, e.g., [23, Chapter 14]) of the set-valued integrand. More precisely, (3.17) means that for any $x^*\in\partial^c\varphi(x)$ there exists a measurable function β such that for μ_ζ-almost every $v\in F(x)$

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

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

Remark 3.2. Let us end this section with a remark concerning the interest of disposing of outer characterizations of the subdifferential. Frequently, in optimization problems, wherein a constraint or objective function involving the mapping φ of the type (1.1) appears, first order optimality conditions involve a condition of the type $0\in\partial^c\varphi(x)$. Disposing of an outer-estimate of the subdifferential, as in (3.17), then allows us to formulate a relaxed condition which always admits a feasible solution whenever the original equation admitted one. This is not true for an inner-characterization, which would lead to a more restrictive condition, potentially not admitting any solution. Outer-approximations are also frequently used in mathematical programs with equilibrium constraints (see, e.g., [27, Theorem 4.1] and the subsequent discussion).

4. Differentiability of φ and a gradient formula. Theorem 3.6 provides an immediate characterization for the differentiability of the probability function φ, given next.

Theorem 4.1. *In addition to the assumptions of Theorem 3.6 suppose that*

$$\mu_\zeta(\{v \in F(\bar{x}) | \#\hat{\mathcal{J}}(\bar{x}, v) \geq 2\}) = 0. \tag{4.1}$$

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

$$\nabla\varphi(\bar{x}) = -\int_{v \in F(\bar{x}), \#\hat{\mathcal{J}}(\bar{x},v)=1} \frac{\chi\left(\hat{\rho}\left(\bar{x}, v\right)\right)}{\left\langle \nabla_z g_{j(v)}\left(\bar{x}, \hat{\rho}\left(\bar{x}, v\right) Lv\right), Lv\right\rangle} \nabla_x g_{j(v)}\left(\bar{x}, \hat{\rho}\left(\bar{x}, v\right) Lv\right) d\mu_\zeta(v), \tag{4.2}$$

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 $j \in \{1, \dots, k\}$ satisfying $g_j(\bar{x}, \hat{\rho}\left(\bar{x}, v\right) Lv) = 0$. If (4.1) holds locally around $\bar{x}$, i.e., if there is a neighborhood U of $\bar{x}$ such that

$$\mu_\zeta(\{v \in F(x) | \#\hat{\mathcal{J}}(x, v) \geq 2\}) = 0 \quad \forall x \in U, \tag{4.3}$$

then φ is continuously differentiable in U.

Proof. Under (4.1), the integrand in (3.17) (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 [3, Proposition 2.1.2], and is contained in the single-valued integral by (3.17), 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 [3, Proposition 2.2.4]. Moreover, $\partial^c\varphi(\bar{x}) = \{\nabla\varphi(\bar{x})\}$ and (4.2) follows from (3.17) 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 (4.1)) and recalling the definition of $\hat{\mathcal{J}}(\bar{x}, v)$. The second assertion of the theorem follows from [3, corollary to Proposition 2.2.4]. ■

Condition (4.3) 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 [18]).

Lemma 4.2. *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*

$$\mu_\zeta(A) = \frac{1}{\lambda(\mathbb{B})}\lambda(\mathrm{cone}(A) \cap \mathbb{B}), \tag{4.4}$$

where $\mathbb{B}$ is the closed unit ball, λ is the Lebesgue measure in $\mathbb{R}^m$, and 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

$$\mathcal{I}\left(x, z\right) := \{j \in \{1, \dots, p\} \,|\, g_j\left(x, z\right) = 0\} \tag{4.5}$$

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

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

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 4.3. *Let g be as in Theorem 3.6. 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 $M':=\left\{v\in\mathbb{S}^{m-1}|\exists r>0: g(\bar{x},rLv)\le 0,\ \#\mathcal{I}(\bar{x},rLv)\ge 2\right\}$.

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

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

Since the union

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

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

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

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 $[0,l]\cdot(M_{1,2}^*\cap\mathbb{B}(0,l))$ is closed by the closedness of $M_{1,2}^*$. Consequently

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

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 4.3, in order to derive that

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

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

$$\lambda\left(L^{-1}\left(\mathbb{R}_+M_{1,2}^*\right)\cap\mathbb{B}\right)=0. \tag{4.6}$$

In order to do this, notice that $\operatorname{rank}\{\nabla_z g_j(x,z)\}_{j=1,2}=2$ 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 neighborhood $W(z)$ such that the rank condition above extends to the whole neighborhood. Then, $W := \bigcup_{z\in M^*} W(z)$ is an open set containing $M^*_{1,2}$ such that

$$\text{rank}\,\{\nabla_z g_j\,(x,z)\}_{j=1,2} = 2 \quad \forall z \in W. \tag{4.7}$$

Defining $\hat{M} := \{z \in W | g_j\,(\bar{x},z) = 0 \;\; (j=1,2)\}$, the respective definitions yield that $M^*_{1,2} \subseteq \hat{M}$.

We show next that the set $L^{-1}\big(\mathbb{R}_+\hat{M}\big)\backslash\{0\}$ is a differentiable manifold of dimension $m-1$. Observe first that, by assumption 1, we have the following equivalence:

$$w \in L^{-1}\left(\mathbb{R}_+\hat{M}\right)\backslash\{0\} \Leftrightarrow \exists t>0 : g_j\,(\bar{x},tLw) = 0 \quad (j=1,2) \quad \text{and} \quad tLw \in W. \tag{4.8}$$

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,

$$\nabla\beta\,(\bar{w},\bar{t}) = \nabla_z g\,(\bar{x},\bar{t}L\bar{w})\,(\bar{t}L|L\bar{w}) =: (A|b)\,. \tag{4.9}$$

Thanks to (4.7), 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 two columns of A. On the other hand, we know from 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 (4.9). 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 $\beta_j\,(w,t) = 0 \;\; (j=1,2)$ can be resolved in a neighborhood $U_{\bar{w}} \times U_{\bar{t}}$ of $(\bar{w},\bar{t})$ as

$$w_1 = \tilde{\varphi}_1\,(w_2,\ldots,w_m)\,, \tag{4.10}$$

$$t = \tilde{\varphi}_2\,(w_2,\ldots,w_m)\,, \tag{4.11}$$

with certain continuously differentiable functions $\tilde{\varphi}_j \;\; (j=1,2)$. 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

$$tLw \in W,\;\; t>0, \quad \forall\,(w,t) \in U_{\bar{w}} \times U_{\bar{t}}. \tag{4.12}$$

Now, $g_1\,(\bar{x},\bar{t}L\bar{w}) = 0$ and (4.8) 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 neighborhood 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

$$g_1\,(\bar{x},rLv) = 0 \Leftrightarrow r = \alpha\,(v) \tag{4.13}$$

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

$$w \in \tilde{U} \cap L^{-1}\left(\left[\mathbb{R}_+\hat{M}\right]\backslash\{0\}\right) \Leftrightarrow w \in \tilde{U} \quad \text{and} \quad w_1 = \tilde{\varphi}_j\,(w_2,\ldots,w_m)\,. \tag{4.14}$$

Indeed, if $w \in \tilde{U} \cap L^{-1}\big([\mathbb{R}_+\hat{M}]\backslash\{0\}\big)$, then by (4.8) there is some $t > 0$ such that $g_j\,(\bar{x}, tLw) = 0$ for all $j = 1, \dots, l$. Since $\|w\|^{-1} w \in V$, we infer from (4.13) that $t = \|w\|^{-1} \alpha\big(\|w\|^{-1} w\big) \in U_{\bar{t}}$. Hence, $(w, t) \in U_{\bar{w}} \times U_{\bar{t}}$, and the direction "$\Rightarrow$" of our asserted equivalence follows from (4.10). Conversely, let $w \in \tilde{U}$ satisfy (4.10). Then, with t defined by (4.11), one has that $g_j\,(\bar{x}, tLw) = 0$ for all $j = 1, 2$. Taking into account (4.12), the direction "$\Leftarrow$" of our asserted equivalence then follows from (4.8).

In conclusion, as $\bar{w} \in L^{-1}\big([\mathbb{R}_+\hat{M}]\backslash\{0\}\big)$ was arbitrary, the equivalence (4.14) shows that $L^{-1}\big([\mathbb{R}_+\hat{M}]\backslash\{0\}\big)$ is a differentiable manifold of dimension $m - 1 < m$. As a consequence, $\lambda\big(L^{-1}\big([\mathbb{R}_+\hat{M}]\backslash\{0\}\big)\big) = 0$ (e.g., [10, Lemma 1.5]). Since $M^*_{1,2} \subseteq \hat{M}$, we infer that $\lambda\left(L^{-1}\left(\left[\mathbb{R}_+M^*_{1,2}\right]\backslash\{0\}\right)\right) = \lambda\left(L^{-1}\left[\mathbb{R}_+M^*_{1,2}\right]\backslash\{0\}\right) = 0$, whence $\lambda\big(L^{-1}\left[\mathbb{R}_+M^*_{1,2}\right]\big) = 0$, implying (4.6) as desired. This completes the proof. ∎

By combination of Lemma 4.3 and Theorem 4.1, we arrive at the main result of this section.

Corollary 4.4. *In addition to the assumptions of Theorem 3.6, suppose that* (R2CQ) *is satisfied at $\bar{x}$. Then φ is Fréchet differentiable at $\bar{x}$, and the gradient formula (4.2) holds true. If* (R2CQ) *is satisfied locally around $\bar{x}$, then φ is continuously differentiable on an appropriate neighborhood of $\bar{x}$.*

Remark 4.1. We note that neither (4.3) nor (R2CQ) is a new condition for ensuring differentiability of probability functions. They can be found in [25, Assumption 2.2(iv)] in the context of spheric-radial decomposition, and in [14, Theorem 3.1, Assumption(vi)] in a general setting. However, in both references, compactness of the set $\{z | g(\bar{x}, z) \le 0\}$ 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:

$$\varphi(x) := \mathbb{P}\left(A(x)\xi \le b(x)\right); \tag{5.1}$$

i.e., in (1.1) we have $g\,(x, \xi) = A(x)\xi - b(x)$ 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 does 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 [31, 12, 20] for the same probability function but in a different form. 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 (5.1) let A, b be continuously differentiable, and let $\xi \sim \mathcal{N}\,(0, R)$ for some positive definite correlation matrix R admitting a decomposition $R = LL^\mathsf{T}$. Fix any $\bar{x} \in \mathbb{R}^n$ such that $b_j(\bar{x}) > 0$ for all $j \in \{1, \dots, p\}$. Finally, assume that any two rows of the matrix $A(\bar{x})$ are linearly independent. Then, φ in (1.1) is continuously differentiable at $\bar{x}$, and it* *holds that*

(5.2)

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

where

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

and $j(v)$ is the unique element of the index set $J^{**}(v)$, i.e., $j(v)$ is the unique index $j\in\{1,\dots,p\}$ 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 4.4. To do so, we have first to check the assumptions of Theorem 3.6. The general assumptions of this theorem as well as assumption 1 are clearly satisfied by the hypotheses we made. Concerning assumption 2 of Theorem 3.6, we claim that the exponential growth condition (Definition 2.3) is satisfied for all $j=1,\dots,p$. Indeed, by A being continuously differentiable, there exists a neighborhood U of $\bar{x}$ and a constant K such that $\|\nabla A_{j,i}(x)\|\leq K$ for all $x\in U$ and all $i,j\in\{1,\dots,p\}$. Then, Definition 2.3 holds true because of

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

In order to verify the asserted continuous differentiability of φ via Corollary 4.4, it remains to check that the constraint qualification (R2CQ) is satisfied on a neighborhood of $\bar{x}$. Clearly, our assumption on pairwise linear independence of the rows of $A(\bar{x})$ implies that (R2CQ) holds 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 $i_k,j_k\in\{1,\dots,p\}$ such that

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

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

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

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 (4.2) ensured by Corollary 4.4 reduces in the special case of (5.1) to the asserted formula (5.2). This follows easily upon specifying the partial derivatives of g and the concrete shape of $\hat{\rho}(v)$, and upon observing the relations

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

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 postmultiplication by the explicit derivative Db according to the chain rule. Hence, consider now the probability function

$$\varphi(x) := \mathbb{P}\left(A\xi \leq x\right). \tag{5.3}$$

This specialization of (5.1) leads not only 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 5.2. *In (5.3) let $\xi \sim \mathcal{N}(0, R)$ for some positive definite correlation matrix R admitting a decomposition $R = LL^{\mathsf{T}}$. 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:*

$$Az \leq \bar{x}, \quad A_i z = \bar{x}_i, \quad A_j z = \bar{x}_j, \qquad i \neq j \Longrightarrow \text{rank } \{A_i, A_j\} = 2. \tag{5.4}$$

Then, φ in (1.1) is continuously differentiable at $\bar{x}$, and it holds that

$$\frac{\partial \varphi}{\partial x_j}(\bar{x}) = \int_{\{v \in \mathbb{S}^{m-1} | A_j L v > 0, \bar{x}_j = \hat{\rho}(v) A_j L v\}} \frac{\chi\left(\hat{\rho}\left(v\right)\right)}{A_j L v} \, d\mu_\zeta(v) \quad (j = 1, \ldots, p). \tag{5.5}$$

Proof. Clearly, the gradient formula (5.5) follows from (5.2) in the special setting of (5.3). Evidently, (5.4) corresponds to the general constraint qualification (R2CQ). In order to derive continuous differentiability of φ via Corollary (4.4), it is sufficient to verify that it automatically holds locally around $\bar{x}$. If this were not the case, we could repeat the argument from 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 $i, j \in \mathcal{I}(x_k, z_k)$. 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 set-valued mapping $x \mapsto \{z | Az \leq x\}$ (see [1, Theorem 3.4.1], [13, p. 121]), the relation $i \in \mathcal{I}(x_k, z_k)$ 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 rank $\{A_i, A_j\} = 1$, a contradiction with (5.4). $\blacksquare$

Finally, we consider a gradient formula for the distribution function

$$F_\xi(x) := \mathbb{P}\left(\xi \leq x\right) \tag{5.6}$$

associated with a Gaussian random vector.

Corollary 5.3. *Let $\xi \sim \mathcal{N}(0,R)$ for some positive definite correlation matrix R admitting a decomposition $R = LL^{\mathsf{T}}$. Fix any $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}_j > 0$ for all $j \in \{1,\dots,p\}$. Then, φ in (1.1) is continuously differentiable at $\bar{x}$, and it holds that*

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

Proof. Equation (5.6) follows from (5.3) by setting $A := I$. 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 (5.3) under the assumptions of Corollary 5.2 except the constraint qualification (5.4). 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 (3.17) 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 (5.3):

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

By the assumption $\bar{x}_j > 0$ for all $j \in \{1,\dots,p\}$ made in Corollary 5.2, $i \sim j$ implies that $\lambda > 0$ in the defining relation. Similarly, $i \nsim j$ implies that (5.4) is satisfied. Denote by $\tilde{p} \le p$ 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 $i = 1,\dots,\tilde{p}$ that

$$A_j z \le x_j \quad \forall j\in[i] \iff A_i z \le h_i(x) := \min_{j\in[i]} \lambda_j^{-1} x_j. \tag{6.1}$$

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

Theorem 6.1. *In (5.3) let $\xi \sim \mathcal{N}(0,R)$ for some positive definite correlation matrix R admitting a decomposition $R = LL^{\mathsf{T}}$. Fix any $\bar{x} \in \mathbb{R}^n$ such that $\bar{x}_j > 0$ for all $j \in \{1,\dots,p\}$. Then, φ is locally Lipschitz continuous, and its Mordukhovich subdifferential can be estimated from above by*

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

where

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

Proof. We introduce the modified probability function $\tilde{\varphi}(y) := \mathbb{P}(\tilde{A}\xi \le y)$ (for $y \in \mathbb{R}^{\tilde{p}}$) and observe that, thanks to (6.1), the original probability function φ in (5.3) 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 (5.4) (see the 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 5.2 that $\tilde{\varphi}$ is continuously differentiable in a neighborhood of $\bar{y}$ and that

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

where $\hat{\rho}\left(v\right)$ is defined in the statement of this theorem. The chain rule for the Mordukhovich subdifferential [17, Theorem 1.110(ii)] now yields that

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

where the last inclusion follows from the sum rule in [17, 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 w.r.t. the partial order of $\mathbb{R}^{\tilde{p}}$. This allows us by [17, p. 112] to continue the previous relation as

$$\partial^M \varphi(\bar{x}) \subseteq \sum_{i=1}^{\tilde{p}} \frac{\partial \tilde{\varphi}}{\partial y_i}(h(\bar{x})) \cdot \partial^M h_i(\bar{x}). \tag{6.2}$$

Given the definition of components h_i in (6.1), we conclude from [17, 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 from this improvement, however, because (6.2) already involves an inclusion anyway.) ∎

7. Conclusions and perspectives.

In this paper we have investigated differentiability of probability functions acting on nonlinear systems of inequalities. The underlying random vectors have been assumed to be multivariate Gaussian or Gaussian-like. The underlying system of inequalities need not define a compact set, and under mild conditions an outer-approximation of the Clarke subdifferential of the probability function is obtained. In the case that this system of inequalities is linear with constant coefficient matrix, we also obtain an upper estimate of the smaller Mordukhovich subdifferential. The presented formulae for the subdifferentials and/or gradients are explicit in the nominal entry data and can therefore be readily implemented by a practitioner familiar with Monte Carlo sampling. We are keen to investigate extensions of these results by relaxing the convexity assumption of the system in the argument represented by the random vector.

Acknowledgment. This paper is dedicated to the memory of András Prékopa.

REFERENCES

[1] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer, *Non-Linear Parametric Optimization*, Birkhäuser, Basel, 1982.

[2] 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*, Math. Comput., 83 (2014), pp. 2821–2851, https://doi.org/10.1090/S0025-5718-2014-02839-1.

[3] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics in Appl. Math. 5, SIAM, Philadelphia, 1990, https://doi.org/10.1137/1.9781611971309.

[4] I. Deák, *Computing probabilities of rectangles in case of multinormal distribution*, J. Statist. Comput. Simul., 26 (1986), pp. 101–114, https://doi.org/10.1080/00949658608810951.

[5] I. Deák, *Subroutines for computing normal probabilities of sets—Computer experiences*, Ann. Oper. Res., 100 (2000), pp. 103–122, https://doi.org/10.1023/A:1019215116991.

[6] D. Dentcheva, *Optimization models with probabilistic constraints*, in Lectures on Stochastic Programming. Modeling and Theory, A. Shapiro, D. Dentcheva, and A. Ruszczyński, eds., MPS-SIAM Ser. Optim. 9, SIAM and MPS, Philadelphia, 2009, pp. 87–153.

[7] O. Ditlevsen and H. O. Madsen, *Structural Reliability Methods*, Wiley, New York, 1996.

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

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

[10] M. Golubitsky and V. Guillemin, *Stable Mappings and Their Singularities*, Grad. Texts Math. 14, Springer-Verlag, New York, 1973.

[11] R. Henrion, *Gradient estimates for Gaussian distribution functions: Application to probabilistically constrained optimization problems*, Numer. Algebra Control Optim., 2 (2012), pp. 655–668, https://doi.org/10.3934/naco.2012.2.655.

[12] R. Henrion and A. Möller, *A gradient formula for linear chance constraints under Gaussian distribution*, Math. Oper. Res., 37 (2012), pp. 475–488, https://doi.org/10.1287/moor.1120.0544.

[13] R. Henrion and W. Römisch, *Lipschitz and differentiability properties of quasi-concave and singular normal distribution functions*, Ann. Oper. Res., 177 (2010), pp. 115–125.

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

[15] G. Lebourg, *Generic differentiability of Lipschitzian functions*, Trans. Amer. Math. Soc., 256 (1979), pp. 125–144, https://doi.org/10.2307/1998104.

[16] K. Marti, *Differentiation of probability functions: The transformation method*, Comput. Math. Appl., 30 (1995), pp. 361–382, https://doi.org/10.1016/0898-1221(95)00113-1.

[17] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation* I. *Basic Theory*, Grundl. Math. Wiss. 330, Springer, Heidelberg, 2006.

[18] A. Naor and D. Romik, *Projecting the surface measure of the sphere of* ℓ_p^n, Ann. Inst. H. Poincaré Probab. Statist., 39 (2003), pp. 241–261, https://doi.org/10.1016/S0246-0203(02)00008-0.

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

[20] A. Prékopa, *Stochastic Programming*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995.

[21] A. Prékopa, *Probabilistic programming*, in Stochastic Programming, Handbooks Oper. Res. Management Sci. 10, A. Ruszczyński and A. Shapiro, eds., Elsevier, Amsterdam, 2003, pp. 267–351.

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

[23] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, 3rd ed., Grundl. Math. Wiss. 317, Springer-Verlag, Berlin, 2009.

[24] J. O. Royset and E. Polak, *Implementable algorithm for stochastic optimization using sample average approximations*, J. Optim. Theory Appl., 122 (2004), pp. 157–184.

[25] J. O. Royset and E. Polak, *Extensions of stochastic optimization results to problems with system failure probability functions*, J. Optim. Theory Appl., 133 (2007), pp. 1–18.

[26] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming. Modeling and Theory*, MPS-SIAM Ser. Optim. 9, SIAM and MPS, Philadelphia, 2009.

[27] T. M. Surowiec, *Explicit Stationarity Conditions and Solution Characterization for Equilibrium Problems with Equilibrium Constraints*, Ph.D. thesis, Humboldt-Universität zu Berlin, Berlin, 2010.

[28] S. Uryas'ev, *Derivatives of probability functions and integrals over sets given by inequalities*, J. Comput. Appl. Math., 56 (1994), pp. 197–223, https://doi.org/10.1016/0377-0427(94)90388-3.

[29] S. Uryas'ev, *Derivatives of probability functions and some applications*, Ann. Oper. Res., 56 (1995), pp. 287–311.

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

[31] W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati, *On joint probabilistic constraints with Gaussian coefficient matrix*, Oper. Res. Lett., 39 (2011), pp. 99–102, https://doi.org/10.1016/j.orl.2011.01.005.

[32] W. van Ackooij and J. Malick, *Second-order differentiability of probability functions*, Optim. Lett., (2016), https://doi.org/10.1007/s11590-016-1015-7.

[33] W. van Ackooij and M. Minoux, *A characterization of the subdifferential of singular Gaussian distribution functions*, Set-Valued Var. Anal., 23 (2015), pp. 465–483, https://doi.org/10.1007/s11228-015-0317-8.