OPTIMIZATION
https://doi.org/10.1080/02331934.2019.1576670

Generalized gradients for probabilistic/robust (probust) constraints

Wim van Ackooij [a], René Henrion [b] and Pedro Pérez-Aros [c]

[a] R&D. Department OSIRIS, Électricité de France, Palaiseau, France; [b] Weierstrass Institute, Berlin, Germany; [c] Instituto de Ciencias de la Ingeniera, Universidad de O'Higgins, Rancagua, Chile

ABSTRACT
Probability functions are a powerful modelling tool when seeking to account for uncertainty in optimization problems. In practice, such uncertainty may result from different sources for which unequal information is available. A convenient combination with ideas from robust optimization then leads to probust functions, i.e. probability functions acting on generalized semi-infinite inequality systems. In this paper we employ the powerful variational tools developed by Boris Mordukhovich to study generalized differentiation of such probust functions. We also provide explicit outer estimates of the generalized subdifferentials in terms of nominal data.

ARTICLE HISTORY
Received 11 December 2018
Accepted 26 January 2019

KEYWORDS
Stochastic optimization; probabilistic constraints; chance constraints; gradients of probability functions; probust constraints

AMS CLASSIFICATION
90C15

1. Introduction

The theory of optimization and optimal control has enormously benefitted from the pioneering work of Boris Mordukhovich in the areas of variational analysis and generalized differentiation. Starting with his introduction of the fundamental limiting normal cone [1] – being small (non-convex) and robust at the same time -, new tools of generalized differentiation (subdifferential, coderivative etc.) grew out of this initial concept in a natural way and opened a powerful perspective for more efficient characterizations of stability or necessary conditions in set-valued analysis and optimization. A striking account of this progress has been given in Mordukhovich's celebrated two-volume monograph [2] on the Theory and Applications of Variational Analysis and Generalized Differentiation. An update of recent developments, for instance in hierarchical or semi-infinite optimization problems, can be found in [3]. Not to the least, the theory developed by Boris Mordukhovich has also found fruitful applications in probabilistic programming, be it for the stability theory of probabilistic constraints [4] or be it for the derivation of 'small' subdifferential formulae for probability functions [5,6].

CONTACT René Henrion henrion@wias-berlin.de
This paper is dedicated to the 70th birthday of Prof. Boris Mordukhovich

The current work is devoted to the characterization of the Mordukhovich subdifferential of probability functions as they arise in optimization problems with probabilistic constraints or in problems of reliability maximization. The classical form of a probability function considered in operations research is

$$\varphi(x) = \mathbb{P}(g_i(x,\xi) \leq 0, \quad (i = 1, \dots, m)), \tag{1}$$

where x is a finite dimensional decision vector, ξ is a finite dimensional random vector, and g represents a constraint mapping defining some finite random inequality system. The probability function assigns to each decision x the probability of satisfying the given random inequality system. Typically, they are embedded into optimization problems in one of the two ways

$$\begin{aligned} &\min\{f(x) \mid \varphi(x) \geq p\} \quad \text{(probabilistic constraint);} \\ &\max\{\varphi(x) \mid x \in X\} \quad \text{(reliability maximization).} \end{aligned}$$

Applications are abundant in engineering and particularly in power management (see, e.g. [7]). It is well recognized that such probability function are inherently non-smooth even if all input data (mapping g and distribution of ξ) are smooth (see, e.g. [5, Ex. 1], [8, Prop. 2.2], [6, Ex. 1.1]). Without further conditions, just continuity can be expected to hold true by imposing some standard constraint qualification. There are basically two reasons for inherent non-smoothness: first, even for a single stochastic inequality (i.e. $m = 1$ in (1)), the effect of an unbounded support for the distribution may cause a failure of differentiability for the parameter dependent improper integral defining φ; second, even in the case of bounded support but in the presence of several inequalities, some constraint qualification (additional to that ensuring continuity) has to be imposed (e.g. the so-called 'rank-2-constraint qualification' [9, Th. 3.1], [6, Lemma 4.3]).

In order to figure out those additional conditions finally guaranteeing differentiability, it turned out to be useful to investigate probability functions by means of tools from variational analysis and generalized differentiation [5,6,10,11]. A powerful tool to carry out this investigation in the context of Gaussian or Gaussian like distributions of ξ is the so-called *spheric-radial decomposition* of an m-dimensional Gaussian random vector $\xi \sim (\mu, \Sigma)$, which allows to represent the Gaussian probability of a Borel measurable set M as

$$\mathbb{P}(\xi \in M) = \int_{\mathbb{S}^{m-1}} \mu_\eta(\{r \geq 0 \mid \mu + rLv \in M\})\, \mathrm{d}\mu_\zeta(v), \tag{2}$$

where μ_η is the one-dimensional Chi-distribution with m degrees of freedom, μ_ζ is the uniform distribution on the unit sphere $\mathbb{S}^{m-1}$ and L originates from a factorization $\Sigma = LL^T$ of the covariance matrix of ξ. This decomposition has been used in the computation of Gaussian probabilities in early papers by Deák [12,13]. However, in the context of optimization problems, the set M would depend on the decision vector x and so, in addition to the computation of probabilities, one would also be interested in the sensitivity of this probability with respect to the decision vector. In [14] it was shown that the gradient with respect to x of the probability (2) (with M replaced by some explicitly described moving set $M(x)$) can be represented – just by differentiation under the integral sign – as a spheric integral much like (2) but with a modified integrand. This allows for a simultaneous sampling scheme of the uniform distribution on the sphere in order to determine probabilities and sensitivities at a time in the framework of some nonliear optimization solver.

While in [14] more general constraints (finite unions of finite intersections of smooth inequalities) were admitted than in (1), the authors make a boundedness assumption which may be quite restrictive for applications. In [8] it was shown for a single inequality constraint, how to circumvent this boundedness assumption by verifying some growth condition which turns out to apply in almost all practically relevant situations. In [6] the analogous result was proven for several inequaities and a general (Clarke-) subgradient formula provided for the probability function in absence of further constraint qualifications. If, however, the rank-2-constraint qualification mentioned above, applies to the stochastic inequality system, then φ could be shown to be differentiable and its gradient represented in the form of a spheric integral again (without relying on a boundedness assumption). Further improvements concerning the application of spheric radial decomposition to probabilistic programming were obtained by extending the setting to infinite-dimensional decisions [5], to the class of elliptical distributions [11] and to the derivation of second-order derivative formulae for φ [15].

The aim of this work is to characterize the Mordukhovich subdifferential of probability functions of the type

$$\varphi(x) = \mathbb{P}(g(x, y, \xi) \leq 0 \quad \forall\, y \in T(x)), \tag{3}$$

where $g : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}$ is a constraint function, $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ is a multifunction representing some moving index set for the inequality system $g(x, y, \xi) \leq 0$, and ξ is an m-dimensional random vector. We note first, that φ is correctly defined by justifying that the set, of which the probability is taken, is Borel measurable. Indeed, it is closed as an intersection of sets which are closed by continuity of g. Observe that the stochastic inequality system, over which the probability is taken in (3) has the typical form of a constraint set in generalized semi-infinite programming [16]. Such probability functions have recently attracted much attention in optimization problems with so-called *probust* (= **pro**babilistic/ro**bust**) constraints (see [17–21]. These arise in a quite natural way, when uncertainty parameters in the inequalities of some constraint system have a mixed nature with some parameters being endowed with stochastic information and others not. Such is the case for instance in gas transport optimization, where the loads at the exits of the gas networks are random (and modelled by multivariate statistical distributions based on historical data) while friction coefficients of the pipes underground are just uncertain within some given range [18]. Then it makes sense, to model both types of uncertainty in a single joint model which is a probabilistic constraint with respect to the stochastic parameter ξ over an infinite inequality system reflecting the robust part of uncertainty.

A recent investigation of model (3) specialized to the setting of a constant (yet continuous) index set $T(x) \equiv T$ can be found in [21]. The consideration of decision-dependent index sets as in (3) adds another twist to the non-smooth character of the probability function φ. Our (first) practical motivation for analyzing this model comes again from optimization problems in gas transportation as, for instance, the maximization of free capacities by the owner of a gas network (see [19] for details). As in previous work, our intention is – starting with tools from generalized differentiation – to provide explicit conditions under which φ is differentiable along with a gradient formula as a spheric integral in the vein of the discussion above.

This paper has the following organization. In Section 2 we provide a brief overview of the employed concepts and background information. Our derivation involves an auxiliary mapping, which can be represented as a 'marginal function'. The careful study of its continuity, representation as a marginal function and derivation of subdifferential estimates are the topic of Section 3. The final Section 4 is devoted to the study of subdifferential estimates for the probability function itself, carefully discusses the employed assumptions and provides an application.

2. Basic assumptions and concepts

We start with the following definitions of well-known properties of multifunctions:

Definition 2.1: Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ be a set-valued mapping. Then,

(1) S is closed if its graph $\operatorname{gph} S := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^p \mid y \in S(x)\}$ is a closed set.
(2) S is locally bounded if for every $x \in \mathbb{R}^n$ there exists a neighbourhood U of x such that $S(U) := \cup_{x' \in U} S(x')$ is bounded.
(3) S is inner (or: lower) semicontinuous at $(\bar{x}, \bar{y}) \in \operatorname{gph} S$ if for every sequence $x_k \to \bar{x}$ there exists a sequence $y_k \in S(x_k)$ with $y_k \to \bar{y}$. Moreover, we say that S is inner semicontinuous at $\bar{x}$ if it is inner semicontinuous at every $(\bar{x}, \bar{y}) \in \operatorname{gph} S$.

Throughout this paper, we make the following standing assumptions on (3):

$$g \text{ is continuously differentiable and it is convex with respect to the third variable.} \tag{4}$$

$$T \text{ is a closed and locally bounded multifunction.} \tag{5}$$

$$\xi \text{ has an elliptical distribution according to } \xi \sim \mathcal{E}(\mu, \Sigma, \theta)$$
$$\text{(see Definition 2.8) with continuous generator } \theta. \tag{6}$$

We note that (5) implies T to have compact values, i.e. $T(x)$ is compact for all $x \in \mathbb{R}^n$. Moreover, we have the following immediate observation:

Proposition 2.2: *From (5) it follows that, if $T(\bar{x}) = \emptyset$ for some $\bar{x} \in \mathbb{R}^n$, then $T(x) = \emptyset$ and $\varphi = 1$ locally around $\bar{x}$ for φ in (3).*

Proof: Assume that there exists some sequence $x_n \to \bar{x}$ with $T(x_n) \neq \emptyset$ and choose $y_n \in T(x_n)$. Then, by local boundedness of T, one has that $y_{n_k} \to_k \bar{y}$ for some subsequence and for some $\bar{y}$. The closedness of T now implies the contradiction $\bar{y} \in T(\bar{x})$ with $T(\bar{x}) = \emptyset$. Consequently, there is some neighbourhood U of $\bar{x}$ such that $T(x) = \emptyset$ for all $x \in U$. This entails that for every $x \in U$ the inequality system

$$g(x, y, z) \leq 0 \quad \forall\, y \in T(x) = \emptyset$$

is trivially satisfied for all $z \in \mathbb{R}^m$. Therefore, $\varphi(x) = 1$. ■

Clearly, the last proposition shows, that our probability function φ in (3) behaves trivially around arguments at which T is empty. That is why we will exclude this case from the corresponding results below.

In the following, we collect some basic concepts of variational analysis and generalized differentiation used in the paper (see, e.g. [3,22]).

Definition 2.3: Let $C \subseteq \mathbb{R}^n$ be a closed set and $\bar{x} \in C$. The contingent cone $T_C(\bar{x})$, the Fréchet normal cone $\hat{N}_C(\bar{x})$ and the Mordukhovich normal cone $N_C(\bar{x})$ to C at $\bar{x}$ are respectively defined as:

$$T_C(\bar{x}) := \left\{ d \in \mathbb{R}^n \mid \exists t_n \downarrow 0,\ \exists\, x_n \in C, t_n^{-1}(x_n - \bar{x}) = d \right\}$$
$$\hat{N}_C(\bar{x}) := \left\{ x^* \in \mathbb{R}^n \mid \langle x^*, d \rangle \leq 0\, \forall\, d \in T_C(\bar{x}) \right\}$$
$$N_C(\bar{x}) := \{ x^* \in \mathbb{R}^n \mid \exists (x_n, x_n^*) \to (\bar{x}, x^*), x_n \in C, x_n^* \in \hat{N}_C(x_n) \}.$$

We recall that the Mordukhovich normal cone (unlike the Fréchet normal cone) has closed graph [22, Prop. 6.6]. The normal cones above induce the following subdifferentials for lower semicontinuous functions $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$:

Definition 2.4: The Fréchet subdifferential $\hat{\partial} f(\bar{x})$, the Mordukhovich subdifferential $\partial f(\bar{x})$ and the singular Mordukhovich subdifferential $\partial^\infty f(\bar{x})$ of f at $\bar{x}$ are respectively defined as:

$$\hat{\partial} f(\bar{x}) := \{ x^* \in \mathbb{R}^n \mid (x^*, -1) \in \hat{N}_{\mathrm{epi} f}(\bar{x}, f(\bar{x})) \}$$
$$\partial f(\bar{x}) := \{ x^* \in \mathbb{R}^n \mid (x^*, -1) \in N_{\mathrm{epi} f}(\bar{x}, f(\bar{x})) \}$$

$$\partial^\infty f(\bar{x}) := \{x^* \in \mathbb{R}^n \mid (x^*, 0) \in N_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))\},$$

where $\operatorname{epi} f := \{(x,t) \in \mathbb{R}^{n+1} \mid f(x) \le t\}$ refers to the epigraph of f.

The following representation of ∂f in terms of $\hat{\partial} f$ is well known [3, Proposition 1.20]:

$$\partial f(\bar{x}) = \{x^* \in \mathbb{R}^n \mid \exists (x_n, x_n^*) \to (\bar{x}, x^*), f(x_n) \to f(\bar{x}), x_n^* \in \hat{\partial} f(x_n)\}. \qquad (7)$$

The Mordukhovich normal cone provides the possibility to introduce a derivative concept for general set-valued mappings:

Definition 2.5: Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ be a set-valued mapping with closed graph. The coderivative of S at $(\bar{x}, \bar{y}) \in \operatorname{gph} S$ is defined as

$$D^* S(\bar{x}, \bar{y})(y^*) := \{x^* \in \mathbb{R}^n \mid (x^*, -y^*) \in N_{\operatorname{gph} S}(\bar{x}, \bar{y})\}.$$

The fact that N is a cone implies immediately the following simple rule:

$$D^* S(\bar{x}, \bar{y})(\lambda y^*) = \lambda D^* S(\bar{x}, \bar{y})(y^*) \quad \forall\, \lambda \ge 0\, \forall\, y^*. \qquad (8)$$

The following concept is a Lipschitz like property of set-valued mappings:

Definition 2.6: The set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ has the Aubin property at $(\bar{x}, \bar{y}) \in \operatorname{gph} S$ if there exists $K > 0$ together with neighbourhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$d\left(y, S(x_2)\right) \le K d(x_1, x_2) \quad \forall\, y \in S(x_1) \cap V \quad \forall\, x_1, x_2 \in U,$$

where, on the left-hand side, 'd' refers to the usual point-to-set distance.

In finite dimensions, the Aubin property can be equivalently characterized by means of the co-derivative. This is the object of the so-called *Mordukhovich criterium* (see, e.g. [22, Theorem 9.40]):

Theorem 2.7: *Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ be a set-valued mapping with closed graph and $(\bar{x}, \bar{y}) \in \operatorname{gph} S$. Then S has the Aubin property at $(\bar{x}, \bar{y})$ if and only if $D^* S(\bar{x}, \bar{y})(0) = \{0\}$.*

In this work we will consider elliptically distributed random vectors [23]:

Definition 2.8: We say that the m-dimensional random vector ξ is elliptically distributed with mean μ, positive definite covariance matrix Σ and generator $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ ($\xi \sim \mathcal{E}(\mu, \Sigma, \theta)$ for short) if and only if its density $f_\xi : \mathbb{R}^m \to \mathbb{R}_+$ is given by

$$f_\xi(z) = (\det \Sigma)^{-1/2}\, \theta\left((z-\mu)^\top \Sigma^{-1} (z-\mu)\right),$$

where the generator must satisfy $\int_0^\infty t^{m/2} \theta(t)\, \mathrm{d}t < \infty$.

The class of elliptical distributions contains, for instance, the multivariate Gaussian, Student (t-), symmetric multivariate stable, symmetric multivariate Laplace and multivariate logistic distributions. Now, consider a decomposition

$$\Sigma = LL^T \tag{9}$$

of Σ (e.g. Cholesky decomposition). Then, it can be shown that ξ admits a representation as

$$\xi = \mu + \mathcal{R}L\zeta,$$

which we will refer to as the *spherical radial* decomposition. Herein ζ has a uniform distribution over the m-dimensional Euclidean unit sphere $\mathbb{S}^{m-1} := \{z \in \mathbb{R}^m : \sum_{i=1}^m z_i^2 = 1\}$ and $\mathcal{R}$ – being stochastically independent of ζ – possesses a density, which is given by

$$f_{\mathcal{R}}(r) := \begin{cases} \dfrac{2\pi^{m/2}}{\Gamma\left(\frac{m}{2}\right)} r^{m-1}\theta(r^2) & \text{if } r \geq 0, \\ 0 & \text{if } r < 0. \end{cases} \tag{10}$$

Let us consider the probability function (3) with $\xi \sim \mathcal{E}(\mu, \Sigma, \theta)$. Then, similar to the previously discussed representation (2) in the special case of Gaussian distributions, one has that

$$\mathbb{P}(\xi \in M) = \int_{\mathbb{S}^{m-1}} \mu_{\mathcal{R}}(\{r \geq 0 \mid \mu + rLv \in M\})\, \mathrm{d}\mu_\zeta(v)$$

for any Borel measurable subset $M \subseteq \mathbb{R}^m$, where μ_ζ refers to the uniform measure on $\mathbb{S}^{m-1}$ and $\mu_{\mathcal{R}}$ to the one-dimensional probability measure induced by the density (10) (which would reduce to the Chi-distribution in (2)). Applying this representation to the set

$$M := \{z \in \mathbb{R}^m \,|\, g(x, y, z) \leq 0 \;\forall\, y \in T(x)\},$$

one derives the corresponding form of the probability function (3):

$$\begin{aligned} \varphi(x) &= \int_{v \in \mathbb{S}^{m-1}} \mu_{\mathcal{R}}(\{r \geq 0 : g(x, y, \mu + rLv) \leq 0 \quad \forall\, y \in T(x)\})\, \mathrm{d}\mu_\zeta(v) \\ &= \int_{v \in \mathbb{S}^{m-1}} e(x, v)\, \mathrm{d}\mu_\zeta(v) \quad \left(x \in \mathbb{R}^n\right). \end{aligned} \tag{11}$$

The study of the integrand $e : \mathbb{R}^n \times \mathbb{S}^{m-1} \to [0, 1]$ defined by

$$e(x, v) := \mu_{\mathcal{R}}(\{r \geq 0 \,|\, g(x, y, \mu + rLv) \leq 0,\; \forall\, y \in T(x)\}). \tag{12}$$

will be of great importance in this paper. We will to refer to it as the *radial probability function.*

3. Continuity properties

3.1. (Semi-) continuity of the probability function

When studying the infinite inequality system $[g(x,y,z) \leq 0 \ \forall\, y \in T(x)]$, it will be useful to consider the following *maximum* function $g_T : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ given by:

$$g_T(x,z) := \begin{cases} \max\limits_{y \in T(x)} g(x,y,z) & \text{if } T(x) \neq \emptyset, \\ -\infty & \text{if } T(x) = \emptyset. \end{cases} \tag{13}$$

Note that writing 'max' in this definition is justified from assumptions (4)–(5) by g being continuous and T having compact values. Clearly, g_T is convex in the second argument as a consequence of our convexity assumption in (4). The following equivalence is immediate:

$$g(x,y,z) \leq 0 \quad \forall\, y \in T(x) \Longleftrightarrow g_T(x,z) \leq 0. \tag{14}$$

In particular, our probability function φ in (3) may be equivalently written as

$$\varphi(x) = \mathbb{P}\left(g_T(x,\xi) \leq 0\right). \tag{15}$$

The results of the following Lemma are well known (see, e.g. [16, p.401]):

Lemma 3.1: *Let $(\bar{x},\bar{z}) \in \mathbb{R}^n \times \mathbb{R}^m$ be arbitrary such that $T(\bar{x}) \neq \emptyset$. Then, under our assumptions (4)–(5) it holds that*

(1) *g_T is upper semicontinuous at $(\bar{x},\bar{z})$.*
(2) *If, in addition, T is inner semicontinuous at $\bar{x}$ (i.e. inner semicontinuous at every $(\bar{x},y) \in \operatorname{gph} T$), then g_T is also lower semicontinuous, hence continuous at $(\bar{x},\bar{z})$.*

From here, we derive the following properties for the probability function φ in (3):

Proposition 3.2: *Let $\bar{x} \in \mathbb{R}^n$ be arbitrary such that $T(\bar{x}) \neq \emptyset$. Then, under our assumptions (4)–(6) it holds that*

(1) *If there is some $\bar{z} \in \mathbb{R}^m$ and some $\varepsilon > 0$ such that*

$$g(\bar{x},y,\bar{z}) \leq -\varepsilon \quad \forall\, y \in T(\bar{x}), \tag{16}$$

then φ is lower semicontinuous at $\bar{x}$.
(2) *If T is inner semicontinuous at $\bar{x}$, then φ is upper semicontinuous at $\bar{x}$.*

Proof: Observe first, that (16) implies $\bar{z}$ to be a strong Slater point for the convex inequality system $g(\bar{x},y,z) \leq 0$ $(y \in T(x))$ in the variable z. It follows that

$g_T(\bar{x}, \bar{z}) \leq -\varepsilon < 0$ by compactness of $T(\bar{x})$, whence $\bar{z}$ is a Slater point of the (single) convex inequality $g_T(\bar{x}, z) \leq 0$ in the variable z. This provides that (with 'bd' = boundary)

$$\left\{z \in \mathbb{R}^m \mid g_T(\bar{x}, z) = 0\right\} \subseteq \text{bd}\left\{z \in \mathbb{R}^m \mid g_T(\bar{x}, z) \leq 0\right\}.$$

Since the boundary of a convex set has Lebesgue measure zero, the set on the left-hand side itself has Lebesgue measure zero. Since our random vector ξ is absolutely continuous with respect to the Lebesgue measure (by having a density according to (6)), one infers that

$$\mathbb{P}\left(g_T(\bar{x}, \xi) = 0\right).$$

Moreover, we know from (1) in Lemma 3.1 that g_T is upper semicontinuous at $(\bar{x}, z)$ for any $z \in \mathbb{R}^m$. Then, taking into account (15), it follows from [17, Lemma 2] (with the inequality system and thus lower and upper semicontinuity reversed there), that φ is lower semicontinuous at $\bar{x}$ as claimed in (1). As for (2), (2) in Lemma 3.1 yields that g_T is lower semicontinuous at $(\bar{x}, z)$ for any $z \in \mathbb{R}^m$. Then, again from [17, Lemma 2], one derives that φ is upper semicontinuous at $\bar{x}$. ∎

Of course, joining all assumptions in the previous Proposition would ensure the continuity of the probability function φ in (3).

We complete this section by an openness result for the Aubin property needed later:

Proposition 3.3: *Fix $\bar{x} \in \mathbb{R}^n$. Assume that our set-valued index set mapping $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ satisfies the Aubin property at all $(\bar{x}, y) \in \text{gph}\, T$. Then, under assumption (5) there is a neighbourhood U of $\bar{x}$, such that T has the Aubin property at all $(x, y) \in \text{gph}\, T$ with $x \in U$ and with some common (independent of x and y) modulus $K \geq 0$. Moreover,*

$$\left\|x^*\right\| \leq K \left\|y^*\right\| \quad \forall\, y^*\, \forall\, x^* \in D^*T(x, y)\left(y^*\right)\, \forall\, y \in T(x)\, \forall\, x \in U. \tag{17}$$

Proof: Since T satisfies the Aubin property at all $(\bar{x}, y)$ with $y \in T(\bar{x})$, a standard compactness argument with respect to the set $T(\bar{x})$ (which is compact by (5)) yields the existence of a neighbourhood U of $\bar{x}$ of a neighbourhood V of the compact set $T(\bar{x})$ and of a constant $K \geq 0$ such that

$$d\left(y, T\left(x_2\right)\right) \leq K d\left(x_1, x_2\right) \quad \forall\, y \in T\left(x_1\right) \cap V\, \forall\, x_1, x_2 \in U. \tag{18}$$

We claim that T has the Aubin property at all $(x, y) \in \text{gph}\, T$ with $x \in U$ with the common modulus K. Assuming the contrary would provide us with a sequence $(x_k, y_k) \in \text{gph}\, T$ such that $x_k \to_k \bar{x}$ and T fails to have the Aubin property at (x_k, y_k) with modulus K. The local boundedness of T at $\bar{x}$ (see (5)) implies that for $y_{k_l} \to_l \bar{y}$ for some subsequence and some $\bar{y}$. By gph T being closed (see (5)), we have that $(\bar{x}, \bar{y}) \in \text{gph } T$. Since T fails to have the Aubin property at (x_{k_l}, y_{k_l}) with modulus K, we may find a sequence $(\tilde{x}_l, \hat{x}_l, \tilde{y}_l)$ satisfying

$$\begin{aligned} & d\left(\tilde{x}_l, x_{k_l}\right), d\left(\hat{x}_l, x_{k_l}\right), d\left(\tilde{y}_l, y_{k_l}\right) \leq l^{-1}, \quad \tilde{y}_l \in T\left(\tilde{x}_l\right), \\ & d\left(\tilde{y}_l, T\left(\hat{x}_l\right)\right) > Kd\left(\tilde{x}_l, \hat{x}_l\right) \ \forall\, l. \end{aligned} \tag{19}$$

Clearly, $\tilde{x}_l, \hat{x}_l \to_l \bar{x}$ and $\tilde{y}_l \to_l \bar{y} \in T(\bar{x}) \subseteq V$. In particular, for l large enough, we have that $\tilde{y}_l \in T(\tilde{x}_l) \cap V$ and $\tilde{x}_l, \hat{x}_l \in U$. Hence, $d(\tilde{y}_l, T(\hat{x}_l)) \leq Kd(\tilde{x}_l, \hat{x}_l)$ for l large enough by (18) which contradicts (19).

It remains to prove (17). From [3, (3.10)] we infer with the result proven so far, that

$$\left\|x^*\right\| \leq K \quad \forall\, x^* \in D^*T(x,y)(y^*) : \left\|y^*\right\| \leq 1 \,\forall\, y \in T(x) \,\forall\, x \in U.$$

Combining this with (8) yields

$$\frac{\|x^*\|}{\|y^*\|} \leq K \quad \forall\, y^* \neq 0 \,\forall\, x^* \in D^*T(x,y)(y^*) \,\forall\, y \in T(x) \,\forall\, x \in U.$$

On the other hand, $D^*T(x,y)(0) = \{0\}$ by Theorem 2.7. Altogether, this proves (17). ■

3.2. (Lipschitz-) continuity and subdifferential of the radius function and of the radial probability function

We define the following function $\rho : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}_+ \cup \{\infty\}$:

$$\rho(x, y, v) := \sup\{r \geq 0 \,|\, g(x, y, \mu + rLv) \leq 0\}. \tag{20}$$

(compare (12)). If the set of $r \geq 0$ with $g(x, y, \mu + rLv) \leq 0$ is empty, then we define $\rho(x, y, v) := 0$. It follows immediately from the definition that

$$\rho(x, y, tv) = t^{-1}\rho(x, y, v) \quad \forall\, t > 0. \tag{21}$$

Let $\bar{x} \in \mathbb{R}^n$, $\bar{y} \in \mathbb{R}^p$ and $\bar{v} \in \mathbb{R}^m$ be given. Suppose that $g(\bar{x}, \bar{y}, \mu) < 0$. Then, by continuity and convexity of $g(\bar{x}, \bar{y}, \cdot)$ as assumed in (4), either $\rho(\bar{x}, \bar{y}, \bar{v}) = \infty$ – in which case $g(\bar{x}, \bar{y}, \mu + rL\bar{v}) < 0$ for all $r \geq 0$ – or $\rho(\bar{x}, \bar{y}, \bar{v}) < \infty$ is the unique solution of $g(\bar{x}, \bar{y}, \mu + rL\bar{v}) = 0$. The following lemma follows from Lemmas 3.1, 3.2 and 3.3 in [8]. The latter ones were proven just in case of a single constraint $g(x, z) \leq 0$, i.e. with missing variable y for indexing the inequality in our possibly infinite system. It turns out, however, that by treating the couple (x, y) exactly as the single variable x has been treated in [8], one may copy the original proofs to get the following results:

Lemma 3.4: *Assume that $\bar{x} \in \mathbb{R}^n, \bar{y} \in \mathbb{R}^p$ and $\bar{v} \in \mathbb{S}^{m-1}$ are such that $g(\bar{x}, \bar{y}, \mu) < 0$. Then, the extended-valued function ρ is continuous at $(\bar{x}, \bar{y}, \bar{v})$ with respect to the topology of $\mathbb{R}_+ \cup \{\infty\}$. Moreover, if $\rho(\bar{x}, \bar{y}, \bar{v}) < \infty$, then, ρ is continuously differentiable on a neighbourhood W of $(\bar{x}, \bar{y}, \bar{v})$ and*

$$\nabla_{x/y}\rho(x, y, v) = -\frac{\nabla_{x/y}g(x, y, \mu + \rho(x, y, v)Lv)}{\langle \nabla_z g(x, y, \mu + \rho(x, y, v)Lv), Lv\rangle} \tag{22}$$

$$\langle \nabla_z g(x, y, \mu + \rho(x, y, v)Lv), Lv\rangle \geq -\frac{g(x, y, \mu)}{\rho(x, y, v)} > 0 \tag{23}$$

holds true for all $(x, y, v) \in W$.

As an immediate consequence of Lemma 3.4 and of (21), we have

Corollary 3.5: *Assume that $\bar{x} \in \mathbb{R}^n$, $\bar{y} \in \mathbb{R}^p$ are such that $g(\bar{x}, \bar{y}, \mu) < 0$. Then, the extended-valued function ρ is continuous at $(\bar{x}, \bar{y}, \bar{v})$ for every $\bar{v} \in \mathbb{R}^m \backslash \{0\}$ with respect to the topology of $\mathbb{R}_+ \cup \{\infty\}$.*

Next, we define the radius function $\rho_T : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+ \cup \{\infty\}$ by

$$\rho_T(x, v) := \inf_{y \in T(x)} \rho(x, y, v). \tag{24}$$

Proposition 3.6: *Under (4)–(6), for every $v \in \mathbb{S}^{m-1}$ and every x with $g_T(x, \mu) < 0$ it holds that*

$$\{r \geq 0 \mid g_T(x, \mu + rLv) \leq 0\} = [0, \rho_T(x, v)], \tag{25}$$

where $[0, \infty] = [0, \infty)$ is intended. Moreover, if $\rho_T(x, v)$ is finite, then it is the unique solution r of the equation $g_T(x, \mu + rLv) = 0$.

Proof: If $T(x) = \emptyset$, then, by definitions in (13) and (24) we have that $\rho_T(x, v) = \infty$ and $g_T(x, \mu + rLv) = -\infty$ for every $r \geq 0$, hence both sets in (25) coincide. Next, assume that $T(x) \neq \emptyset$. If $\rho_T(x, v) = \infty$, then, $\rho(x, y, v) = \infty$ for all $y \in T(x)$ which entails that $g(x, y, \mu + rLv) \leq 0$ for all $r \geq 0$ and all $y \in T(x)$. It follows that $g_T(x, \mu + rLv) \leq 0$ for all $r \geq 0$ and, hence, (25) holds true again. Therefore, we may assume now in addition that $\rho_T(x, v) < \infty$.

Let $r \in [0, \rho_T(x, v)]$ be given. Then, by (24) and (20)

$$\sup\{r' \geq 0 \mid g(x, y, \mu + r'Lv) \leq 0\} = \rho(x, y, v) \geq r \quad \forall\, y \in T(x).$$

Let $y \in T(x)$ be arbitrary. If there existed some $r^* \geq r$ with $g(x, y, \mu + r^*Lv) \leq 0$, then by our assumption $g_T(x, \mu) < 0$ and by convexity of $g(x, y, \cdot)$ it would follow that $g(x, y, \mu + rLv) \leq 0$. Otherwise, the relation above entails the existence of some sequence $r_k \uparrow r$ such that $g(x, y, \mu + r_kLv) \leq 0$. Then, again $g(x, y, \mu +$

$rLv) \leq 0$. Since $y \in T(x)$ was arbitrary, one infers that $g_T(x, \mu + rLv) \leq 0$ which establishes the inclusion $' \supseteq'$ of (25).

Conversely, let $g_T(x, \mu + rLv) \leq 0$ for some $r \geq 0$. Then, $g(x, y, \mu + rLv) \leq 0$ and, hence, $r \leq \rho(x, y, v)$ for all $y \in T(x)$ which entails that $\rho_T(x, v) \geq r$. This finally proves (25).

Concerning the final statement of the proposition, note first that the assumption $\rho_T(x, v) < \infty$ along with the already proven identity (25) yields that

$$g_T(x, \mu + \rho_T(x, v)Lv) \leq 0; \quad g_T(x, \mu + (\rho_T(x, v) + 1/k)Lv) > 0 \quad \forall\, k.$$

The assumption $\rho_T(x, v) < \infty$ implies that $T(x) \neq \emptyset$. Consequently, Lemma 3.1 ensures the upper semicontinuity of g_T. Thus,

$$g_T(x, \mu + \rho_T(x, v)Lv) \geq \limsup_k g_T(x, \mu + (\rho_T(x, v) + 1/k)Lv) \geq 0,$$

so that $\rho_T(x, v)$ is a solution r of the equation $g_T(x, \mu + rLv) = 0$. Uniqueness of this solution now follows from the convexity of $g_T(x, \cdot)$ along with our assumption $g_T(x, \mu) < 0$. ■

Corollary 3.7: *Assume that* (4)–(6) *holds true. If $g_T(x, \mu) < 0$, then the radial probability function in* (12) *can be represented as*

$$e(x, v) = \begin{cases} F_{\mathcal{R}}(\rho_T(x, v)) & \text{if } \rho_T(x, v) < \infty \\ 1 & \text{if } \rho_T(x, v) = \infty \end{cases} \qquad \forall\, x \in \mathbb{R}^n\ \forall\, v \in \mathbb{S}^{m-1}, \tag{26}$$

where $F_{\mathcal{R}}$ is the cumulative distribution function of the one-dimensional probability measure $\mu_{\mathcal{R}}$.

Proof: From (12), (14) and (25), we derive

$$e(x, v) = \mu_{\mathcal{R}}(\{r \geq 0 \,|\, g_T(x, \mu + rLv) \leq 0\}) = \mu_{\mathcal{R}}([0, \rho_T(x, v)]).$$

If $\rho_T(x, v) = \infty$, then $[0, \rho_T(x, v)] = \mathbb{R}_+$, and so

$$e(x, v) = \mu_{\mathcal{R}}(\mathbb{R}_+) = \int_0^\infty f_{\mathcal{R}}(t)\, \mathrm{d}t = 1,$$

where the last identity follows from the fact that the density $f_{\mathcal{R}}$ takes value zero for negative arguments (see (10)). If, in contrast, $\rho_T(x, v) < \infty$, then $\mu_{\mathcal{R}}([0, \rho_T(x, v)]) = F_{\mathcal{R}}(\rho_T(x, v))$ by definition of a distribution function and again by (see (10)). ■

The next result characterizes the continuity of the radial function ρ_T.

Theorem 3.8: *Assume that* (4)–(6) *hold true. Let x be such that $g_T(x, \mu) < 0$. Then, ρ_T (as a possibly extended-valued function) is lower semicontinous at (x, v)*

for every $v \in \mathbb{S}^{m-1}$ in the topology of $\mathbb{R}_+ \cup \{\infty\}$. If, moreover, T is inner semicontinuous at x, then ρ_T is also upper semicontinous, hence, continuous at (x, v) for every $v \in \mathbb{S}^{m-1}$ in the topology of $\mathbb{R}_+ \cup \{\infty\}$.

Proof: As for the verification of lower semicontinuity, consider a sequence $(x_k, v_k) \to (x, v)$ with

$$\rho_T(x_k, v_k) \to \alpha := \liminf_{(x',v') \to (x,v)} \rho_T(x', v').$$

We have to show that $\rho_T(x, v) \le \alpha$. This is trivial if $\alpha = \infty$, hence assume that $\alpha < \infty$. Consequently, $\rho_T(x_k, v_k) < \infty$ for k large enough. This entails, on the one hand, that $T(x_k) \neq \emptyset$ by (24), whence $T(x) \neq \emptyset$ according to Proposition 2.2. On the other hand, by the last statement of Proposition 3.6,

$$g_T(x_k, \mu + \rho_T(x_k, v_k)Lv_k) = 0$$

for k large enough. Now, we may exploit the upper semicontinuity of g_T at $(x, \alpha Lv)$, which is guaranteed by Lemma 3.1, in order to derive that

$$0 = \lim_k g_T(x_k, \mu + \rho_T(x_k, v_k)Lv_k) \le g_T(x, \mu + \alpha Lv). \tag{27}$$

If $\rho_T(x, v) = \infty$, then $g_T(x, \mu + rLv) \le 0$ for all $r \ge 0$ by (25). In particular $g_T(x, \mu + \alpha Lv) \le 0$ and also $g_T(x, \mu + (\alpha + 1)Lv) \le 0$, whence $g_T(x, \mu + \alpha Lv) = 0$ by (27). This provides a contradiction with $g_T(x, \mu) < 0$ and the convexity of $g_T(x, \cdot)$. Therefore, $\rho_T(x, v) < \infty$, and so, by the last statement of Proposition 3.6, $g_T(x, \mu + \rho_T(x, v)Lv) = 0$. The assumption $\rho_T(x, v) > \alpha$ would then imply with $g_T(x, \mu) < 0$ and the convexity of $g_T(x, \cdot)$ that $g_T(x, \mu + \alpha Lv) < 0$ contradicting (27). Hence, $\rho_T(x, v) \le \alpha$, as was to be shown.

The upper semicontinuity of ρ_T at (x, v) is trivial in case that $\rho_T(x, v) = \infty$, hence we assume that $\rho_T(x, v) < \infty$ which in particular implies that $T(x) \neq \emptyset$. Then, by (13), for an arbitrary $\varepsilon > 0$ there exists some $y_\varepsilon \in T(x)$ such that

$$g(x, \mu + y_\varepsilon, (\rho_T(x, v) + \varepsilon)Lv) = g_T(x, \mu + (\rho_T(x, v) + \varepsilon)Lv) > 0,$$

where the strict inequality follows from (25). Let $(x_k, v_k) \to (x, v)$ be a sequence with

$$\rho_T(x_k, v_k) \to \beta := \limsup_{(x',v') \to (x,v)} \rho_T(x', v').$$

By the inner semicontinuity of T at x, there exists a sequence $y_k \to y_\varepsilon$ with $y_k \in T(x_k)$. Then, by continuity of g,

$$\lim_k g(x_k, y_k, \mu + (\rho_T(x, v) + \varepsilon)Lv_k) = g(x, y_\varepsilon, \mu + (\rho_T(x, v) + \varepsilon)Lv) > 0.$$

It follows that

$$g_T(x_k, \mu + (\rho_T(x, v) + \varepsilon)Lv_k) \ge g(x_k, y_k, \mu + (\rho_T(x, v) + \varepsilon)Lv_k) > 0$$

for k large enough. Therefore, again by (25), one has that $\rho_T(x, v) + \varepsilon > \rho_T(x_k, v_k)$ for k sufficiently large. Hence, $\beta \leq \rho_T(x, v) + \varepsilon$. Since $\varepsilon > 0$ was chosen arbitrarily, the claimed upper semicontinuity of ρ_T at (x, v) follows. ■

Corollary 3.9: *Assume that* (4)–(6) *hold true. Let x be such that $g_T(x, \mu) < 0$. Then, for every $v \in \mathbb{S}^{m-1}$, e is lower semicontinuous at (x, v). If, moreover, T is inner semicontinuous at x, then for every $v \in \mathbb{S}^{m-1}$, e is also upper semicontinous, hence continuous at (x, v).*

Proof: Let $v \in \mathbb{S}^{m-1}$ be arbitrarily given. Observe first that $g_T(x', \mu) < 0$ holds locally around x by $g_T(x, \mu) < 0$ and by upper semicontinuity of g_T at (x, μ), (see Proposition 3.1). Suppose that e fails to be lower semicontinuous at (x, v). Then, there exist sequences $x_k \to_k x$ and $v_k \to_k v$ such that $e(x, v) > \lim_k e(x_k, v_k)$. Observe first that, for k large enough,

$$e\left(x_k, v_k\right) < e\left(x, v\right) \leq 1$$

because e, as a probability function, takes values not larger than one. In particular, for k large enough, Corollary 3.7 implies that

$$\rho_T\left(x_k, v_k\right) < \infty, \quad e\left(x_k, v_k\right) = F_{\mathcal{R}}\left(\rho_T\left(x_k, v_k\right)\right).$$

In the case of $\rho_T(x, v) = \infty$, the lower semicontinuity of ρ_T at (x, v) guaranteed by Theorem 3.8 entails that $\liminf_k \rho_T(x_k, v_k) = \infty$, whence $\rho_T(x_k, v_k) \to_k \infty$. $F_{\mathcal{R}}$, as a one-dimensional distribution function, satisfies $\lim_{t\to\infty} F_{\mathcal{R}}(t) = 1$. This allows us to derive the contradiction

$$e\left(x, v\right) > \lim_k e\left(x_k, v_k\right) = \lim_k F_{\mathcal{R}}\left(\rho_T\left(x_k, v_k\right)\right) = 1 \geq e\left(x, v\right).$$

Consequently, we may assume that $\rho_T(x, v) < \infty$. Then, once more, we establish a contradiction

$$\begin{aligned} e\left(x, v\right) &> \lim_k e\left(x_k, v_k\right) = \lim_k F_{\mathcal{R}}\left(\rho_T\left(x_k, v_k\right)\right) \geq F_{\mathcal{R}}\left(\liminf_k \rho_T\left(x_k, v_k\right)\right) \\ &\geq F_{\mathcal{R}}\left(\rho_T\left(x, v\right)\right) = e\left(x, v\right). \end{aligned}$$

Here we exploited the fact that $F_{\mathcal{R}}$ as a one-dimensional distribution function having a density, is non-decreasing and continuous, so that the liminf of the function can be estimated from below by the function value at the liminf. The remaining relations follow once more from the lower semicontinuity of ρ_T at (x, v) and from Corollary 3.7. Altogether, this shows the claimed lower semicontinuity of e at (x, v). The verification of continuity of e at (x, v) under the additional assumption of T being inner semicontinuous at x follows similar lines upon exploiting the continuity of ρ_T thanks to Theorem 3.8. ■

Our next step is to provide an upper estimate for the Mordukhovich subdifferential of our radius function ρ_T. To this aim, we will apply the following result:

Theorem 3.10 ([3], Theorem 4.1): *We consider the marginal function $\beta : \mathbb{R}^{n_1} \to \bar{\mathbb{R}}$ defined by*

$$\beta\left(z\right) := \inf\left\{\alpha\left(z,y\right) \mid y \in G(z)\right\},$$

where $\alpha : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \bar{\mathbb{R}}$ is a lower semicontinuous extendended-valued function and $G : \mathbb{R}^{n_1} \rightrightarrows \mathbb{R}^{n_2}$ is a multifunction with closed graph. Define the argmin-mapping $\Psi : \mathbb{R}^{n_1} \rightrightarrows \mathbb{R}^{n_2}$ by

$$\Psi\left(z\right) := \left\{y \in G(z) \mid \beta\left(z\right) = \alpha\left(z,y\right)\right\}.$$

Fix some $\bar{z} \in \mathbb{R}^{n_1}$ with $\beta(\bar{z}) < \infty$ and $\Psi(\bar{z}) \neq \emptyset$. Assume that Ψ is locally bounded around $\bar{z}$ and that the condition

$$\partial^{\infty}\alpha\left(\bar{z},y\right) \cap -N_{\operatorname{gph} G}\left(\bar{z},y\right) = \{0\} \quad \forall\, y \in \Psi\left(\bar{z}\right) \tag{28}$$

is satisfied (see Definition 2.4). Then, the following inclusion holds true:

$$\partial\beta\left(\bar{z}\right) \subseteq \bigcup \left\{z^* + D^*G\left(\bar{z},y\right)\left(y^*\right) \mid \left(z^*,y^*\right) \in \partial\alpha\left(\bar{z},y\right), y \in \Psi\left(\bar{z}\right)\right\}.$$

Before applying this result, we introduce the argmin mapping $M : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^p$ associated with our problem and defined by

$$M(x,v) := \{y \in T(x) \mid \rho_T\left(x,v\right) = \rho\left(x,y,v\right)\}. \tag{29}$$

Proposition 3.11: *Assume that (4)–(6) hold true. Let $(\bar{x},\bar{v}) \in \mathbb{R}^n \times \mathbb{S}^{m-1}$ be such that $g_T(\bar{x},\mu) < 0$, $T(\bar{x}) \neq \emptyset$ and $\rho_T(\bar{x},\bar{v}) < \infty$. Then,*

$$\begin{aligned}\partial\rho_T\left(\bar{x},\bar{v}\right) \subseteq \bigcup_{y\in M(\bar{x},\bar{v})} & \left[\left\{\nabla_x\rho\left(\bar{x},y,\bar{v}\right)\right\} + D^*T\left(\bar{x},y\right)\left(\nabla_y\rho\left(\bar{x},y,\bar{v}\right)\right)\right] \\ & \times \left\{\nabla_v\rho\left(\bar{x},y,\bar{v}\right)\right\}.\end{aligned}$$

Proof: By upper semicontinuity of g_T (see (1) in Lemma 3.1), there exists a compact neighbourhood $\bar{U}$ of $\bar{x}$ such that $g_T(x,\mu) < 0$ for all $x \in \bar{U}$. Then,

$$g(x,y,\mu) < 0 \quad \forall\, x \in \bar{U}\, \forall\, y \in T(x). \tag{30}$$

Define $B := (\bar{U} \times \mathbb{R}^p) \cap \operatorname{gph} T$ which is closed since $\operatorname{gph} T$ is so (see (5)). By local boundedness of T (see (5)), we may assume $\bar{U}$ to be small enough, such that $T(x) \subseteq W$ for all $x \in \bar{U}$ and some bounded set W. Consequently, B is a closed subset of the bounded set $\bar{U} \times W$ and, hence, is compact. Since (30) may be written as $g(x, y, \mu) < 0$ for all $(x, y) \in B$, the continuity of g and the compactness of B ensure the existence of some compact set $\bar{V}$ with $B \subseteq \operatorname{int} \bar{V}$ and

$$g(x, y, \mu) < 0 \quad \forall \left(x, y\right) \in \bar{V}.$$

Then, by Corollary 3.5, ρ is continuous on the compact set

$$A := \bar{V} \times \left\{ v \in \mathbb{R}^m \,\middle|\, \frac{1}{2} \leq \|v\| \leq 2 \right\}$$

with respect to the topology of $\mathbb{R}_+ \cup \{\infty\}$. Observe that $B \times \mathbb{S}^{m-1} \subseteq \operatorname{int} A$. We define the function $\tilde{\rho} : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}_+ \cup \{\infty\}$ as $\tilde{\rho} := \rho + i_A$ where i_A is the indicator function of A. Then, $\tilde{\rho}$ is lower semicontinuous. Define $\tilde{\rho}_T : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+ \cup \{\infty\}$ by

$$\tilde{\rho}_T(x, v) := \inf_{y \in T(x)} \tilde{\rho}\left(x, y, v\right). \tag{31}$$

For $x \in \bar{U}$, $y \in T(x)$ and v with $\frac{1}{2} \leq \|v\| \leq 2$, one has that $(x, y, v) \in A$, so that $\rho(x, y, v) = \tilde{\rho}(x, y, v)$. Consequently

$$\begin{aligned} \rho_T(x, v) &= \inf_{y \in T(x)} \rho(x, y, v) = \inf_{y \in T(x)} \tilde{\rho}(x, y, v) = \tilde{\rho}_T(x, v) \\ &\forall\, x \in \bar{U}\, \forall\, v : \frac{1}{2} \leq \|v\| \leq 2. \end{aligned} \tag{32}$$

Now, $\tilde{\rho}_T$ above can be formally written as

$$\tilde{\rho}_T(x, v) = \inf_{y \in K(x, v)} \tilde{\rho}(x, y, v), \tag{33}$$

where the multifunction $K : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^p$ is defined as $K(x, v) := T(x)$ for all $(x, v) \in \mathbb{R}^n \times \mathbb{R}^m$. In this form, $\tilde{\rho}_T$ is a marginal function as is β in Theorem 3.10. We check that the assumptions of that Theorem are fulfilled for the setting of $z := (x, v)$, $\beta(z) := \tilde{\rho}_T(x, v)$, $\alpha(z, y) := \tilde{\rho}(x, y, v)$, $G(z) := K(x, v)$ and $\Psi(z) := \tilde{M}(x, v)$, where $\tilde{M} : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^p$ is defined by

$$\tilde{M}(x, v) := \{y \in K(x, v) \mid \tilde{\rho}_T(x, v) = \tilde{\rho}\left(x, y, v\right)\}.$$

First, recall that $\tilde{\rho}$ is lower semicontinuous and that $\operatorname{gph} K$ is closed due to $\operatorname{gph} T$ being closed. Moreover, we easily observe from the respective definitions and (32) that

$$\tilde{M}(x, v) = M(x, v) \subseteq T(x) \quad \forall\, x \in \bar{U}\, \forall\, v : \frac{1}{2} \leq \|v\| \leq 2. \tag{34}$$

for M introduced in (29). As a consequence, $\tilde{M}$ is locally bounded at $(\bar{x}, \bar{v})$ because T is so by (5). Next, since $T(\bar{x}) \neq \emptyset$ by assumption, there exists a sequence $y_k \in T(\bar{x})$ with $\rho(\bar{x}, y_k, \bar{v}) \to_k \rho_T(\bar{x}, \bar{v})$. By compactness of $T(\bar{x})$, one has that $y_{k_l} \to_l \bar{y}$ for some subsequence and some $\bar{y} \in T(\bar{x})$. Now, the continuity of ρ on A with respect to the topology of $\mathbb{R}_+ \cup \{\infty\}$ ensures that

$$\rho\left(\bar{x}, y_{k_l}, \bar{v}\right) \to_l \rho\left(\bar{x}, \bar{y}, \bar{v}\right) = \rho_T\left(\bar{x}, \bar{v}\right).$$

Hence, $M(\bar{x}, \bar{v}) \neq \emptyset$ and $\tilde{M}(\bar{x}, \bar{v}) \neq \emptyset$ by (34). Finally, $\tilde{\rho}_T(\bar{x}, \bar{v}) = \rho_T(\bar{x}, \bar{v}) < \infty$ by assumption and by (32). Therefore, it remains to check condition (28). Clearly it would be sufficient to show that

$$\partial^\infty \tilde{\rho}\left(\bar{x}, y, \bar{v}\right) = \{0\} \quad \forall\, y \in \tilde{M}(\bar{x}, \bar{v}) = M(\bar{x}, \bar{v}).$$

Fix an arbitrary such $\bar{y} \in M(\bar{x}, \bar{v})$. By definition and assumption,

$$\rho\left(\bar{x}, \bar{y}, \bar{v}\right) = \rho_T\left(\bar{x}, \bar{v}\right) < \infty.$$

Moreover, our assumption $g_T(\bar{x}, \mu) < 0$ implies that $g(\bar{x}, \bar{y}, \mu) < 0$. This allows us to invoke Lemma 3.4, in order to derive that ρ is continuously differentiable in a neighbourhood of $(\bar{x}, \bar{y}, \bar{v})$. In particular, it is locally Lipschitz there, which implies that $\partial^\infty \rho(\bar{x}, \bar{y}, \bar{v}) = \{0\}$ (see [3, Theorem 1.22]). On the other hand (see above),

$$\left(\bar{x}, \bar{y}, \bar{v}\right) \in B \times \mathbb{S}^{m-1} \subseteq \operatorname{int} A.$$

Since $\tilde{\rho}$ and ρ agree on A, they agree on a neighbourhood of $(\bar{x}, \bar{y}, \bar{v})$. Hence,

$$\partial^\infty \tilde{\rho}(\bar{x}, \bar{y}, \bar{v}) = \partial^\infty \rho(\bar{x}, \bar{y}, \bar{v}) = \{0\}.$$

Summarizing, all assumptions of Theorem 3.10 applied to the marginal function (33) are satisfied and we derive the uper estimate

$$\begin{aligned} \partial \tilde{\rho}_T\left(\bar{x}, \bar{v}\right) \subseteq \bigcup \Big\{ \left(x^*, v^*\right) + D^*K\left(\bar{x}, \bar{v}, y\right)\left(y^*\right) \mid \left(x^*, y^*, v^*\right) \\ \in \partial \tilde{\rho}\left(\bar{x}, y, \bar{v}\right), y \in \tilde{M}(\bar{x}, \bar{v}) \Big\}. \end{aligned} \tag{35}$$

Once more we exploit that ρ is continuously differentiable and agrees with $\tilde{\rho}$ in a neighbourhood of $(\bar{x}, y, \bar{v})$ for an arbitrary $y \in \tilde{M}(\bar{x}, \bar{v}) = M(\bar{x}, \bar{v})$. Similarly, $\tilde{\rho}_T$ and ρ_T coincide on a neighbourhood of $(\bar{x}, \bar{v})$ by (32).Therefore,

$$\partial \tilde{\rho}\left(\bar{x}, y, \bar{v}\right) = \partial \rho\left(\bar{x}, y, \bar{v}\right) = \left\{\nabla \rho\left(\bar{x}, y, \bar{v}\right)\right\}; \quad \partial \tilde{\rho}_T\left(\bar{x}, \bar{v}\right) = \partial \rho_T\left(\bar{x}, \bar{v}\right).$$

Furthermore, the definition of the coderivative and $K(x, v) = T(x)$ immediately yield that

$$D^*K\left(\bar{x}, \bar{v}, y\right)\left(y^*\right) = D^*T\left(\bar{x}, y\right)\left(y^*\right) \times \{0\}.$$

Summarizing, (35) implies the inclusion claimed in this proposition. ■

Theorem 3.12: *If, in addition to the assumptions of Proposition 3.11, T has the Aubin property at every $(\bar{x}, y) \in \operatorname{gph} T$, then ρ_T is locally Lipschitz continuous around $(\bar{x}, \bar{v})$. Moreover, the partial subdifferential (subdifferential of partial function) of ρ satisfies*

$$\partial_x \rho_T(\bar{x}, \bar{v}) \subseteq \bigcup_{y \in M(\bar{x}, \bar{v})} \left\{\nabla_x \rho\left(\bar{x}, y, \bar{v}\right)\right\} + D^* T\left(\bar{x}, y\right)\left(\nabla_y \rho\left(\bar{x}, y, \bar{v}\right)\right).$$

Proof: Referring to the marginal function $\tilde{\rho}_T$ in (31) and to the derivations made in Proposition 3.11, it follows from Corollary 4.3. in [3] that $\tilde{\rho}_T$ is locally Lipschitzian at $(\bar{x}, \bar{v})$ if it is lower semicontinuous around this point. From (32) we know that $\tilde{\rho}_T$ and ρ_T coincide on a neighbourhood of $(\bar{x}, \bar{v})$. On the other hand, ρ_T is lower semicontinuous on a neighbourhood of $(\bar{x}, \bar{v})$. To see this, observe first that ρ_T is lower semicontinuous on $\bar{U} \times \mathbb{S}^{m-1}$ for $\bar{U}$ defined in the beginning of the proof of Proposition 3.11. Since, by (24) and (21), $\rho_T(x, rv) = r^{-1}\rho_T(x, v)$ for any $r > 0$, we conclude that ρ_T is lower semicontinuous on $\bar{U} \times (\mathbb{R}^m \backslash \{0\})$ with respect to the topology of $\mathbb{R}_+ \cup \{\infty\}$. As a consequence $\tilde{\rho}_T$ is lower semicontinuous itself around $(\bar{x}, \bar{v})$. This proves the local Lipschitz continuity of $\tilde{\rho}_T$ at $(\bar{x}, \bar{v})$. Conversely, as $\tilde{\rho}_T$ and ρ_T coincide locally around $(\bar{x}, \bar{v})$, ρ_T itself must be locally Lipschitzian around $(\bar{x}, \bar{v})$ which also implies that $\partial^\infty \rho_T(\bar{x}, \bar{v}) = \{(0, 0)\}$ (see [3, Theorem 1.22]). Now, [22, Cor. 10.11] yields the inclusion

$$\partial_x \rho_T(\bar{x}, \bar{v}) \subseteq \left\{x^* \in \mathbb{R}^n \mid \exists v^* \in \mathbb{R}^m : \left(x^*, v^*\right) \in \partial \rho_T(\bar{x}, \bar{v})\right\}.$$

Then, the desired formula follows from the upper estimate obtained in Proposition 3.11. ■

Corollary 3.13: *Under the assumptions of Theorem 3.12, the partial radial probability function $e(\cdot, v)$ is locally Lipschitz around $\bar{x}$ for every $v \in \mathbb{S}^{m-1}$ close to $\bar{v}$ with some common Lipschitz constant independent of v. Moreover,*

$$\partial_x e(x, v) \subseteq f_{\mathcal{R}}(\rho_T(x, v))\partial_x \rho_T(x, v)$$

for x and $v \in \mathbb{S}^{m-1}$ locally around $(\bar{x}, \bar{v})$, where $f_{\mathcal{R}}$ refers to the density in (10).

Proof: By assumption (3), the generator θ of the density $f_{\mathcal{R}}$ and, hence, $f_{\mathcal{R}}$ itself is continuous. Consequently, the associated one-dimensional cumulative distribution function $F_{\mathcal{R}}$ is continuously differentiable with $F'_{\mathcal{R}} = f_{\mathcal{R}}$. Since, moreover, ρ_T is locally Lipschitz around $(\bar{x}, \bar{v})$ by Theorem 3.12, it follows from $\rho_T(\bar{x}, \bar{v}) < \infty$ that $\rho_T(x, v) < \infty$ locally around $(\bar{x}, \bar{v})$. Hence, by Corollary 3.7,

$$e(x, v) = F_{\mathcal{R}}(\rho_T(x, v)) \tag{36}$$

for x and $v \in \mathbb{S}^{m-1}$ locally around $(\bar{x}, \bar{v})$. In particular, since $F_{\mathcal{R}}$ is locally Lipschitz and ρ_T is locally Lipschitz around $(\bar{x}, \bar{v})$, it follows that $e(\cdot, v)$ is locally

Lipschitz around $\bar{x}$ for every $v \in \mathbb{S}^{m-1}$ close to $\bar{v}$ with some common Lipschitz constant independent of v. Then, the chain rule for subdifferentials [2, Corollary 3.43] yields for x and $v \in \mathbb{S}^{m-1}$ locally around $(\bar{x}, \bar{v})$:

$$\begin{aligned}\partial_x e(x, v) &= \partial e(\cdot, v)(x) = F'_{\mathcal{R}}(\rho_T(x, v))\partial \rho_T(\cdot, v)(x) \\ &= f_{\mathcal{R}}(\rho_T(x, v))\partial_x \rho_T(x, v).\end{aligned}$$

■

4. Subdifferential of the probability function φ

In this section, we present the main result of this paper, namely a fully explicit (in terms of the problem data) subdifferential formula for the (total) probability function φ defined in (3). We start by collecting the necessary preparations.

4.1. Preparatory results

As observed in [5,8] the probability function φ can not be expected to be differentiable, actually not even locally Lipschitzian, even in the simplest possible settings. This is mainly due to the unboundedness of the support of the given random vector ξ. The missing data property can be formulated as the following growth condition:

Definition 4.1 (Growth condition): We say that our problem data (g, T) from (3) satisfies the ψ-growth condition at $\bar{x}$ if for some $R, \varepsilon > 0$ it holds that:

$$\|\nabla_{(x,y)} g(x, y, z)\| \leq \psi(\|z\|) \quad \forall\, y \in T(x)\, \forall\, x \in \mathbb{B}(\bar{x}, \varepsilon)\, \forall\, \|z\| \geq R.$$

Here $\psi : \mathbb{R}_+ \to \mathbb{R}$ is a non-decreasing function such that (with $f_{\mathcal{R}}$ from (10)):

$$\lim_{r\to\infty} r f_{\mathcal{R}}(r)\psi(\delta r) = 0 \quad \forall\, \delta > 0.$$

For the existence and verification of such growth condition in concrete settings, we refer to the discussion of Section 4.3. The key for proving our main result in the next section is the following statement on the partial Fréchet subdifferential of the radial probability function:

Lemma 4.2: *Assume that (4)–(6) hold true. Let $\bar{x} \in \mathbb{R}^n$ be such that $g_T(\bar{x}, \mu) < 0$ and $T(\bar{x}) \neq \emptyset$. In addition, assume that T has the Aubin property at $(\bar{x}, \bar{y})$ for every $\bar{y} \in T(\bar{x})$ and that the data couple (g, T) satisfies the ψ-growth condition at $\bar{x}$ according to Definition 4.1. Then, there exists a neighbourhood U of $\bar{x}$ and a constant $K \geq 0$ such that, with $\mathbb{B}$ denoting the unit ball in $\mathbb{R}^n$,*

$$\hat{\partial}_x e(x, v) \subseteq K\mathbb{B} \quad \forall\ (x, v) \in U \times \mathbb{S}^{m-1}. \tag{37}$$

Proof: We recall first, that the assumed Aubin property of T at every $(\bar{x}, \bar{y})$ with $\bar{y} \in T(\bar{x})$ entails the inner semicontinuity of T at $\bar{x}$. The assertion will follow from a standard compactness argument (w.r.t. $\mathbb{S}^{m-1}$) if we are able to show that for every $\bar{v} \in \mathbb{S}^{m-1}$ there exist neighbourhoods $U_{\bar{v}}$ of $\bar{x}$ and $W_{\bar{v}}$ of $\bar{v}$ as well as a constant $K_{\bar{v}} \geq 0$ such that

$$\hat{\partial}_x e\,(x, v) \subseteq K_{\bar{v}}\mathbb{B} \quad \forall\ (x, v) \in U_{\bar{v}} \times \left[W_{\bar{v}} \cap \mathbb{S}^{m-1}\right]. \tag{38}$$

In order to verify (38), fix an arbitrary $\bar{v} \in \mathbb{S}^{m-1}$. Consider first the case $\rho_T(\bar{x}, \bar{v}) < \infty$. Then, Corollary 3.13 implies that $e(\cdot, v)$ is locally Lipschitz in a neighbourhood $U_{\bar{v}}$ of $\bar{x}$ for every $v \in \mathbb{S}^{m-1}$ in a neighbourhood $W_{\bar{v}}$ of $\bar{v}$ and with Lipschitz constant L not depending on v. By [2, Theorem 3.52],

$$\hat{\partial}_x e\,(x, v) \subseteq L\mathbb{B} \quad \forall\ (x, v) \in U_{\bar{v}} \times \left[W_{\bar{v}} \cap \mathbb{S}^{m-1}\right].$$

This yields (38) with $K_{\bar{v}} := L$.

We now turn to the more involved case $\rho_T(\bar{x}, \bar{v}) = \infty$. We will show a slightly stronger property than needed in (38), namely that for each $\eta > 0$ there exist neighbourhoods $U_{\bar{v}}$ of $\bar{x}$ and $W_{\bar{v}}$ of $\bar{v}$ such that

$$\hat{\partial}_x e\,(x, v) \subseteq \eta\mathbb{B} \quad \forall\ (x, v) \in U_{\bar{v}} \times \left[W_{\bar{v}} \cap \mathbb{S}^{m-1}\right], \tag{39}$$

Of course, this will imply (38) and finally prove our Proposition. Now, let $\eta > 0$ be arbitrary. As a first part of (39), we show that

$$\hat{\partial}_x e\,(x, v) \subseteq \eta\mathbb{B} \quad \forall\ (x, v) \in U_{\bar{v}} \times \left[W_{\bar{v}} \cap \mathbb{S}^{m-1}\right] : \rho_T(x, v) = \infty. \tag{40}$$

Thanks to the upper semicontinuity at (x, μ) of g_T ((1) in Lemma 3.1) we may assume that $g_T(x, \mu) < 0$ for all $x \in U_{\bar{v}}$. Corollary 3.9 then guarantees that $e(\cdot, v)$ is lower semicontinuous on $U_{\bar{v}}$ for any $v \in \mathbb{S}^{m-1}$ and Corollary 3.7 yields that

$$e\,(x, v) = 1 \quad \forall\ (x, v) \in U_{\bar{v}} \times \mathbb{S}^{m-1} : \rho_T(x, v) = \infty.$$

Since e is a probability function and as such takes values not larger than one, x must be a (global) maximizer of the function $e(\cdot, v)$ for every $x \in U_{\bar{v}}$ and every $v \in \mathbb{S}^{m-1}$ such that $\rho_T(x, v) = \infty$. However, it is easily shown that the Fr échet subdifferential of a lower semicontinuous function at a local maximizer must be contained in zero (possibly empty), see, e.g. Proof of [5, Corollary 1(ii)]. In other words, we actually have that $\hat{\partial}_x e(x, v) \subseteq \{0\}$. This proves (40). It remains to verify the relation

$$\hat{\partial}_x e\,(x, v) \subseteq \eta\mathbb{B} \quad \forall\ (x, v) \in U_{\bar{v}} \times \left[W_{\bar{v}} \cap \mathbb{S}^{m-1}\right] : \rho_T(x, v) < \infty. \tag{41}$$

By virtue of the upper semicontinuity at (x, μ) of g_T ((1) in Lemma 3.1) we may define a neighbourhood $U_{\bar{v}}$ of $\bar{x}$ such that

$$g_T(x, \mu) \leq \frac{1}{2} g_T(\bar{x}, \mu) < 0, \quad \forall\, x \in U_{\bar{v}}. \tag{42}$$

We next address Definition 4.1: Thanks to the continuity of ρ_T at $(\bar{x}, \bar{v})$ with $\rho_T(\bar{x}, \bar{v}) = \infty$ in the topology of $\mathbb{R}_+ \cup \{\infty\}$ (see Theorem 3.8) we may further shrink $U_{\bar{v}}$ and also find a neighbourhood $W_{\bar{v}}$ of $\bar{v}$ such that, for R appearing in Definition 4.1, for L from (9) and for η fixed above,

$$\begin{aligned} \|\mu + \rho_T(x,v)Lv\| \geq R, \quad & \rho_T(x,v)\,\|L\| \geq \|\mu\| \\ \rho_T(x,v)f_{\mathcal{R}}(\rho_T(x,v))\psi(2\|L\|\rho_T(x,v)) &\leq \eta, \end{aligned} \tag{43}$$

for all $(x,v) \in U_{\bar{v}} \times [W_{\bar{v}} \cap \mathbb{S}^{m-1}]$. Here, in the first relation, we used the fact that the covariance matrix Σ (see Definition 2.8) and, hence, L from (9) are regular, so that

$$\|Lv\| \geq c := \min_{w \in \mathbb{S}^{m-1}} Lw > 0 \quad \forall\, v \in \mathbb{S}^{m-1}.$$

Next, we shrink $U_{\bar{v}}$ once more such that, thanks to Proposition 3.3, T has the Aubin property at all $(x,y) \in \operatorname{gph} T$ with $x \in U_{\bar{v}}$ and with some common (independent of x and y) modulus $K \geq 0$. In particular, (17) is satisfied then with $U := U_{\bar{v}}$.

Now, consider arbitrary $(x,v) \in U_{\bar{v}} \times [W_{\bar{v}} \cap \mathbb{S}^{m-1}]$ such that $\rho_T(x,v) < \infty$ and $x^* \in \hat{\partial}_x e(x,v)$. In particular, $T(x) \neq \emptyset$ by (24). Then, by our previous definitions of neighbourhoods, the assumptions of Theorem 3.12 are satisfied at (x,v), hence, along with Corollary 3.13, we derive the existence of some $y \in M(x,v)$ (with M defined in (29), hence $\rho(x,y,v) = \rho_T(x,v)$) and some

$$w^* \in D^*T\left(x,y\right)\left(\nabla_y \rho\left(x,y,v\right)\right)$$

such that

$$x^* = f_{\mathcal{R}}(\rho_T\left(x,v\right))\left(\nabla_x \rho\left(x,y,v\right) + w^*\right). \tag{44}$$

For the purpose of abbreviation, we put (see (23))

$$\lambda := \langle \nabla_z g(x,y,\mu + \rho_T(x,v)Lv), Lv \rangle > 0.$$

Then (8) yields $\lambda w^* \in D^*T(x,y)(\lambda \nabla_y \rho(x,y,v)$, which along with (17) and (22) gives

$$\left\|\lambda w^*\right\| \leq K\left\|\lambda \nabla_y \rho\left(x,y,v\right)\right\| = K\left\|\nabla_y g(x,y,\mu + \rho(x,y,v)Lv)\right\|. \tag{45}$$

Recall, that the constant K does not depend on x and y. Next, (42) and (23) lead to the estimate

$$\begin{aligned} 0 < \lambda^{-1} &= \frac{1}{\langle \nabla_z g(x,y,\mu + \rho(x,y,v)Lv), Lv \rangle} \leq -\frac{\rho(x,y,v)}{g(x,y,\mu)} \\ &\leq -\frac{\rho_T(x,v)}{g_T(x,\mu)} \leq \frac{2\rho_T(x,v)}{\left|g_T(\bar{x},\mu)\right|}. \end{aligned}$$

Now, (44), (22) and (45) provide

$$\begin{aligned}\left\|x^*\right\| &\leq \lambda^{-1} f_{\mathcal{R}}(\rho_T(x,v))\left(\left\|\nabla_x g(x,y,\mu+\rho(x,y,v)Lv)\right\| + \left\|\lambda w^*\right\|\right)\\ &\leq \frac{2\rho_T(x,v)}{\left|g_T(\bar{x},\mu)\right|} f_{\mathcal{R}}(\rho_T(x,v))\,(K+1)\left\|\nabla_{(x,y)} g(x,y,\mu+\rho(x,y,v)Lv)\right\|.\end{aligned}$$

From the first two relations of (43) and Definition 4.1 (recalling that ψ is non-decreasing) we infer that

$$\begin{aligned}\left\|\nabla_{(x,y)} g(x,y,\mu+\rho(x,y,v)Lv)\right\| &= \left\|\nabla_{(x,y)} g(x,y,\mu+\rho_T(x,v)Lv)\right\|\\ &\leq \psi\left(\|\mu+\rho_T(x,v)Lv)\|\right)\\ &\leq \psi\left(\|\mu\| + \rho_T(x,v)\,\|L\|\right)\\ &\leq \psi\left(2\,\|L\|\,\rho_T(x,v)\right).\end{aligned}$$

Consequently, there is a constant $C := 2(K+1)/|g_T(\bar{x},\mu)| > 0$ such that, along with the third relation of (43),

$$\left\|x^*\right\| \leq C\rho_T(x,v) f_{\mathcal{R}}(\rho_T(x,v))\psi\left(2\,\|L\|\,\rho_T(x,v)\right) \leq C\eta.$$

Since $\eta > 0$ was arbitrary and since C does not depend on η (because K doesn't), we may apply the result to $\tilde{\eta} := \eta/C$. Hence, we find neighbourhoods $U_{\bar{v}} \times W_{\bar{v}}$ such that $\|x^*\| \leq C\tilde{\eta} = \eta$ for every $x^* \in \hat{\partial}_x e(x,v)$ and every $(x,v) \in U_{\bar{v}} \times [W_{\bar{v}} \cap \mathbb{S}^{m-1}]$ such that $\rho_T(x,v) < \infty$. This proves (41) and the whole Lemma. ■

Corollary 4.3: *The assertion of Lemma 4.2 remains true if the ψ-growth condition is replaced by the following assumption: There exists some $y \in T(\bar{x})$ such that the set $\{z \in \mathbb{R}^m \mid g(\bar{x},y,z) \leq 0\}$ is bounded.*

Proof: The assumption implies by definition that $\rho(\bar{x},y,v) < \infty$ for every $v \in \mathbb{S}^{m-1}$. Hence, $\rho_T(\bar{x},v) < \infty$ for every $v \in \mathbb{S}^{m-1}$. In that case, the proof of Lemma 4.2 does not rely on the ψ-growth condition and is finished after the first paragraph upon proving (38). ■

Corollary 4.4: *Under the assumptions of Lemma 4.2 one has that $\partial_x e(\bar{x},\bar{v}) = \{0\}$ for every $\bar{v} \in \mathbb{S}^{m-1}$ such that $\rho_T(\bar{x},\bar{v}) = \infty$.*

Proof: The inclusion (37) yields, by virtue of [2, Theorem 3.52], that $e(\cdot,\bar{v})$ is locally Lipschitzian at $\bar{x}$. As a consequence, $\partial_x e(\bar{x},\bar{v}) \neq \emptyset$. From the proof of Lemma 4.2, we know the following result in the case of $\rho_T(\bar{x},\bar{v}) = \infty$ (see (39)): For each $\eta > 0$ there exist neighbourhoods $U_{\bar{v}}$ of $\bar{x}$ and $W_{\bar{v}}$ of $\bar{v}$ such that

$$\hat{\partial}_x e(x,v) \subseteq \eta\mathbb{B} \quad \forall\ (x,v) \in U_{\bar{v}} \times \left[W_{\bar{v}} \cap \mathbb{S}^{m-1}\right].$$

Then, it follows from (7) that $x^* \in \partial_x e(\bar{x},\bar{v})$ implies $x^* = 0$. ■

4.2. Main result

We are now in a position to formulate the main result of this paper, namely a subdifferential estimate for the (total) probability function φ in (3):

Theorem 4.5: *Assume that (4)–(6) hold true. Let $\bar{x} \in \mathbb{R}^n$ be such that $g_T(\bar{x}, \mu) < 0$ and $T(\bar{x}) \neq \emptyset$. In addition, assume that T has the Aubin property at $(\bar{x}, \bar{y})$ for every $\bar{y} \in T(\bar{x})$. Finally let one of the following conditions be satisfied:*

(1) *There exists some $y \in T(\bar{x})$ such that the set $\{z \in \mathbb{R}^m \mid g(\bar{x}, y, z) \leq 0\}$ is bounded.or:*
(2) *the data couple (g, T) satisfies the ψ-growth condition at $\bar{x}$ according to Definition 4.1.*

Then, φ in (3) is locally Lipschitzian at $\bar{x}$ and the following upper estimate for its subdifferential holds true:

$$\partial\varphi(\bar{x}) \subseteq \int_{F(\bar{x})} f_{\mathcal{R}}(\rho_T(\bar{x}, v)) \cdot \left[\bigcup_{y \in M(\bar{x}, v)} \left\{ \nabla_x \rho\left(\bar{x}, y, v\right) \right\} \right.$$
$$\left. + D^* T\left(\bar{x}, y\right)\left(\nabla_y \rho\left(\bar{x}, y, v\right)\right) \right] \mathrm{d}\mu_\zeta(v), \tag{46}$$

where $F(\bar{x}) := \{v \in \mathbb{S}^{m-1} \mid \rho_T(\bar{x}, v) < \infty\}$.

Proof: According to Lemma 4.2 and Corollary 4.3, either of the assumptions (1) or (2) imply the estimate (37) which in turn, by virtue of [2, Theorem 3.52], entails that there exists a neighbourhood U of $\bar{x}$ such that for every $v \in \mathbb{S}^{m-1}$ the partial function $e(\cdot, v)$ is locally Lipschitzian on U with a common modulus K (independent of v). Clearly, this modulus (considered as a constant function) is integrable with respect to the uniform measure on the sphere since the latter is compact. Moreover, for every $x \in U$, the function $e(x, \cdot)$ is lower semicontinuous, hence measurable on $\mathbb{S}^{m-1}$.

Altogether, this allows us to apply Clarke's Theorem for subdifferentiation of integral functionals [24, Theorem 2.7.2] to (11). First this Theorem guarantees that φ is locally Lipschitzian at $\bar{x}$. Second, it allows to interchange subdifferentiation and integration in the following way:

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

Here, the upper index 'C' is meant to indicate Clarke's subdifferential. In the following we make use of the well known relation $\partial^C f(\bar{x}) = \operatorname{co} \partial f(\bar{x})$ (with 'co' referring to the convex hull) for functions being locally Lipschitzian at $\bar{x}$ [3, (1.83)]. In particular, Corollary 4.4 yields that

$$\partial_x^C e(\bar{x}, v) = \mathrm{co}\partial_x e(\bar{x}, v) = \mathrm{co}\{0\} = \{0\}$$

for every $v \in \mathbb{S}^{m-1}$ such that $\rho_T(\bar{x}, v) = \infty$. Hence, the inclusion above simplifies to

$$\partial^C \varphi(\bar{x}) \subseteq \int_{\{v \in \mathbb{S}^{m-1} \mid \rho_T(\bar{x}, v) < \infty\}} \partial_x^C e(\bar{x}, v)\, \mathrm{d}\mu_\zeta(v).$$

On the other hand, the same inclusion is obtained for the Mordukhovich subdifferential due to

$$\begin{aligned}
\partial\varphi(\bar{x}) \subseteq \mathrm{co}\,\partial\varphi(\bar{x}) = \partial^C\varphi(\bar{x}) &\subseteq \int_{\{v \in \mathbb{S}^{m-1} \mid \rho_T(\bar{x}, v) < \infty\}} \partial_x^C e(\bar{x}, v)\, \mathrm{d}\mu_\zeta(v) \\
&= \int_{\{v \in \mathbb{S}^{m-1} \mid \rho_T(\bar{x}, v) < \infty\}} \mathrm{co}\partial_x e(\bar{x}, v)\, \mathrm{d}\mu_\zeta(v) \\
&= \int_{\{v \in \mathbb{S}^{m-1} \mid \rho_T(\bar{x}, v) < \infty\}} \partial_x e(\bar{x}, v)\, \mathrm{d}\mu_\zeta(v),
\end{aligned}$$

where the last equality is a consequence of Aumann's Theorem. Now, taking into account that for $(\bar{x}, v)$ with $\rho_T(\bar{x}, v) < \infty$ the assumptions of Theorem 3.12 and Corollary 3.13 are satisfied, we may derive our desired formula from the resulting inclusion

$$\partial_x e(\bar{x}, v) \subseteq f_{\mathcal{R}}(\rho_T(\bar{x}, v)) \cdot \bigcup_{y \in M(\bar{x}, v)} \left\{\nabla_x \rho\left(\bar{x}, y, v\right)\right\} + D^* T\left(\bar{x}, y\right)\left(\nabla_y \rho\left(\bar{x}, y, v\right)\right).$$

■

The interpretation of (46) is as follows: For every $x^* \in \partial\varphi(\bar{x})$, there exists a measurable function $\beta : \mathbb{S}^{m-1} \to \mathbb{R}$ such that

$$\beta(v) \in \partial_x e(\bar{x}, v)\, \mu_\zeta - \text{a.e.} \quad \text{and} \quad x^* = \int_{\{v \in \mathbb{S}^{m-1} \mid \rho_T(\bar{x}, v) < \infty\}} \beta(v)\, \mathrm{d}\mu_\zeta(v).$$

We conclude this section by formulating a condition for the differentiability of the probability function:

Corollary 4.6: *In addition to the assumption of Theorem 4.5, suppose that at $\bar{x}$ we have $\mu_\zeta(V_2 \backslash V_1) = 0$ for the two sets*

$$\begin{aligned}
V_2 &:= \left\{v \in \mathbb{S}^{m-1} \mid \rho_T(\bar{x}, v) < \infty\right\}, \\
V_1 &:= \Big\{v \in V_2 \mid \exists \left(\bar{y}, x^*\right) \in \mathbb{R}^p \times \mathbb{R}^n : M(\bar{x}, v) = \{\bar{y}\}, \\
&\qquad D^* T\left(\bar{x}, y\right)\left(\nabla_y \rho\left(\bar{x}, y, v\right)\right) = x^*\Big\}.
\end{aligned}$$

Then φ in (3) is (strictly) differentiable at $\bar{x}$ and the following (exact) gradient formula applies:

$$\nabla\varphi(\bar{x}) = \int_{\{v\in\mathbb{S}^{m-1} \mid \rho_T(\bar{x},v)<\infty\}} f_{\mathcal{R}}(\rho_T(\bar{x},v)) \cdot \left(\nabla_x\rho\left(\bar{x},\bar{y}(v),v\right)+x^*(v)\right) \mathrm{d}\mu_\zeta(v),$$

where $\bar{y}(v)$ is the unique element of $M(\bar{x},v)$ and $x^*(v) = D^*T(\bar{x},\bar{y}(v))(\nabla_y\rho(\bar{x},\bar{y}(v),v))$. Moreover, φ is continuously differentiable if the measure zero condition above holds true locally around $\bar{x}$.

Proof: The assumption implies that the integrand in (46) is single-valued $\mu_\zeta -$ *a.e.* and, hence, the whole integral on the right-hand side reduces to a singleton. On the other hand, $\partial\varphi(\bar{x})$ on the left-hand side is non-empty due to the already shown Lipschitz continuity of φ at $\bar{x}$. Hence, the only way for $\partial\varphi(\bar{x})$ being included in the right-hand side is to coincide with it. Consequently $\partial\varphi(\bar{x})$ is a singleton too. This implies first, that φ is strictly differentiable at $\bar{x}$ [3, Theorem 4.17] and second that the asserted gradient formula comes as consequence of (46). ■

4.3. Discussion of hypotheses

In this section we provide a short discussion of the assumptions we imposed in order to derive the results of the previous section:

- The assumption $g_T(\bar{x},\mu) < 0$ expresses the fact that the mean of the random parameter should be strictly feasible for the infinite inequality system $g(\bar{x},y,z) \le 0$ $(y \in T(\bar{x}))$. It can be easily seen (see, e.g. [8, Prop. 3.11] that, thanks to the convexity of g with respect to the third variable and thanks to the symmetry of elliptic distributions around their mean, this assumption will fulfilled be satisfied if $\varphi(\bar{x}) \ge 0.5$, i.e. if under the fixed decision $\bar{x}$ the probability (3) is not smaller than one half. This assumption is by no means restrictive when taking into account that in probabilistic programming probabilities close to one are required.
- The assumption $T(\bar{x}) \ne \emptyset$ is not restrictive either since the case $T(\bar{x}) = \emptyset$ entails the trivial situation $\nabla\varphi(\bar{x}) = 0$ as pointed out in Proposition 2.2.
- For T to have the Aubin property at $(\bar{x},\bar{y})$ for every $\bar{y} \in T(\bar{x})$ is a central stability requirement for the whole analysis. For the case of T describing a smooth finite parameter-dependent inequality system – as discussed in the following section – it is well-known to be equivalent with the classical Mangasarian–Fromovitz Constraint Qualification in nonlinear programming. Without this assumption we could not hope for the local Lipschitz continuity φ so that also the subdifferential formula (46) for a just continuous φ would become much more involved (see, e.g. [5] for fixed index mapping T).

- As for the ψ-growth condition, this is necessary to impose only if the set $\{z \in \mathbb{R}^m \mid g(\bar{x}, y, z) \leq 0\}$ is unbounded for all $y \in T(\bar{x})$, see Theorem 4.5. If so, then it turns out that the ψ-growth condition is satisfied in basically all practical applications. We refer to the discussion in [11, Section 4.2], which was done, however, in the context of a fixed and finite index mapping T. Nonetheless, the cases considered there, referring to a situation where the dependence on z of the mapping g is separable with respect to the second variable x can be carried over to the setting of this paper by keeping separability of g between z and the two remaining variables (x, y).
- The measure zero condition $\mu_\zeta(V_2 \backslash V_1) = 0$ in Corollary 4.6 is indispensable in order to derive (strict or continuous) differentiability of φ at $\bar{x}$. An similar condition can be already found in the early paper [14, Assumption 2.2(iv)]. Of course, such condition with respect to the measure of the uniform distribution on the sphere may be hard to verify from the originally given inequality system. In the context again of a fixed and finite index mapping T, this issue could be reduced to the verification of the so-called rank-2-constraint qualification for the original inequality system induced by g [6, Lemma 4.3]. This CQ is well-known and easy to check. It is in particular weaker than the standard Linear Independence Constraint Qualification considered in nonlinear programming. In the context of our infinite and moving index set $T(x)$ a corresponding result seems much harder to prove and will be the subject of further investigations.

4.4. Application

In many applications the moving index set $T(x)$ will have the concrete description as a finite parametric inequality system:

$$T(x) := \{y \in \mathbb{R}^p \mid h_j(x, y) \leq 0 \quad \big(j = 1, \ldots, q\big)\}. \tag{47}$$

Our aim is to ensure all assumptions of the main result related with T by means of concrete assumptions with respect to the description (47):

Theorem 4.7: *Consider the probability function* (3) *with a moving index set given by* (47). *Let $\bar{x} \in \mathbb{R}^n$ be such that $g_T(\bar{x}, \mu) < 0$. Apart from* (4) *and* (6), *we suppose that*

(1) *The h_j in* (47) *are continuously differentiable and they are convex with respect to y.*
(2) *There exists some $\bar{y} \in \mathbb{R}^p$ such that $h_j(\bar{x}, \bar{y}) < 0$ $(j = 1, \ldots, q)$ (Slater point).*
(3) *$T(\bar{x})$ is bounded.*
(4) *One of the assumptions* (1) *or* (2) *in Theorem* 4.5 *applies.*

Then, φ is locally Lipschitzian at $\bar{x}$ and the following upper estimate for its subdifferential holds true:

$$\partial\varphi(\bar{x}) \subseteq \int_{F(\bar{x})} f_{\mathcal{R}}(\rho_T(\bar{x},v)) \cdot \left[\bigcup_{y\in M(\bar{x},v)} \left\{\nabla_x \rho\left(\bar{x},y,v\right)\right\} + \bigcup_{\lambda\in\Lambda(y,v)} \sum_{j=1}^{q} \lambda_j \nabla_x h_j(\bar{x},y) \right] \mathrm{d}\mu_\zeta(v), \tag{48}$$

where, $F(\bar{x}) := \{v \in \mathbb{S}^{m-1} \mid \rho_T(\bar{x},v) < \infty\}$ and for $y \in T(\bar{x})$ and $v \in \mathbb{S}^{m-1}$

$$\Lambda\left(y,v\right) := \left\{\lambda \in \mathbb{R}^q_+ \middle| \nabla_y \rho\left(\bar{x},y,v\right) = -\nabla_y^T h(\bar{x},y)\lambda, \lambda^T h(\bar{x},y) = 0\right\}. \tag{49}$$

Proof: Our assumptions imply by well-known arguments that $T(\bar{x}) \neq \emptyset$ (due to (2)), that T has closed graph and is locally bounded (due to being bounded at $\bar{x}$ by (3) and convex-valued by (1)). Moreover, T has the Aubin property at $(\bar{x},\bar{y})$ for every $\bar{y} \in T(\bar{x})$ thanks to the existence of a Slater point in (2). Altogether, this allows us to derive the inclusion (46). The claimed formula then follows upon using the representation

$$D^*T\left(\bar{x},y\right)\left(\nabla_y \rho\left(\bar{x},y,v\right)\right) = \bigcup_{\lambda\in\Lambda(y,v)} \sum_{j=1}^{q} \lambda_j \nabla_x h_j(\bar{x},y) \tag{50}$$

of the coderivative of T with the set $\Lambda(y,v)$ introduced in the statement of this Theorem. This representation follows from [2, Corollary 4.35] upon noting that our Slater point assumption is equivalent with the so-called Mangasarian–Fromovitz Constraint Qualification in the setting of (47). ■

The previous result can be slightly improved if the Slater Condition is strengthened:

Corollary 4.8: *If, in Theorem 4.7, condition* (2) *is replaced by the Linear Independence Constraint Qualification* (*LICQ*)

$$\left\{\nabla_y h_j(\bar{x},y)\right\}_{\{j \mid h_j(\bar{x},y)=0\}} \text{ islinearlyindependentforall} y \in T\left(\bar{x}\right)$$

and, in addition, $\mu_\zeta(V_2 \backslash \tilde{V}_1) = 0$ for V_2 as defined in Corollary 4.6 and

$$\tilde{V}_1 := \{v \in V_2 \mid \#M(\bar{x},v) = 1\},$$

then φ is (*strictly*) *differentiable at $\bar{x}$ and the following* (*exact*) *gradient formula applies*:

$$\nabla\varphi(\bar{x}) = \int_{\{v\in\mathbb{S}^{m-1} \mid \rho_T(\bar{x},v)<\infty\}} f_{\mathcal{R}}(\rho_T(\bar{x},v))$$

$$\cdot \left(\nabla_x \rho\left(\bar{x}, \bar{y}(v), v\right) + \sum_{j=1}^{q} \lambda_j(y, v) \nabla_x h_j(\bar{x}, y) \right) \mathrm{d}\mu_\zeta(v),$$

where $\lambda(y, v)$ *is the unique element in* $\Lambda(y, v)$ *as defined in* (49).

Proof: The stronger (LICQ) yields that the Lagrange multiplier in Theorem 4.7 is uniquely defined, i.e. $\Lambda(y, v) = \{\lambda(y, v)\}$ for all $y \in T(\bar{x})$ and $v \in \mathbb{S}^{m-1}$. Then, by (50),

$$D^*T\left(\bar{x}, y\right)\left(\nabla_y \rho\left(\bar{x}, y, v\right)\right) = \sum_{j=1}^{q} \lambda_j(y, v) \nabla_x h_j(\bar{x}, y) =: x^*(v).$$

As a consequence, the sets $\tilde{V}_1$ and V_1 introduced above and in Corollary 4.6 coincide and, thus, this Corollary provides the claimed gradient formula. ■

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The first two authors author gratefully acknowledge support by the Fondation Mathématique Jacques Hadamard within the Program Gaspard Monge in optimization and operations research including support to this program by EDF.

ORCID

Wim van Ackooij http://orcid.org/0000-0002-9943-3572

References

[1] Mordukhovich BS. Maximum principle in problems of time optimal control with nonsmooth constraints. J Appl Math Mech. 1976;40:960–969.
[2] Mordukhovich BS. Variational analysis and generalized differentiation I. Basic theory. II. Applications. Heidelberg: Springer; 2006. (Grundlehren der mathematischen Wissenschaften; vol. 330).
[3] Mordukhovich BS. Variational analysis and applications. Cham: Springer; 2018. (Springer monographs in mathematics).
[4] Henrion R, Römisch W. Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. Math Program. 1999;84:55–88.
[5] Hantoute A, Henrion R, Pérez-Aros P. Subdifferential characterization of continuous probability functions under Gaussian distribution. Math Program. To appear (appeared online).
[6] van Ackooij W, Henrion R. (Sub-) Gradient formulae for probability functions of random inequality systems under Gaussian distribution. SIAM J Uncertain Quantification. 2017;5(1):63–87.
[7] Prékopa A. Stochastic programming. Dordrecht: Kluwer; 1995.

[8] van Ackooij W, Henrion R. Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions. SIAM J Optim. 2014;24(4):1864–1889.
[9] Kibzun A, Uryasev S. Differentiability of probability function. Stoch Anal Appl. 1998;16:1101–1128.
[10] Shapiro A, Dentcheva D, Ruszczyński A. Lectures on stochastic programming: modeling and theory. 2nd ed. Philadelphia: SIAM and MPS; 2014. (MPS-SIAM series on optimization; vol. 9).
[11] van Ackooij W, Aleksovska I, Zuniga MM. (sub-)differentiability of probability functions with elliptical distributions. Set-Val Var Anal. 2018;26(4):887–910.
[12] Deák I. Computing probabilities of rectangles in case of multinormal distribution. J Statist Comput Simul. 1986;26(1–2):101–114.
[13] Deák I. Subroutines for computing normal probabilities of sets – computer experiences. Ann Oper Res. 2000;100:103–122.
[14] Royset J, Polak E. Extensions of stochastic optimization results to problems with system failure probability functions. J Optim Theory Appl. 2007;133(1):1–18.
[15] van Ackooij W, Malick J. Second-order differentiability of probability functions. Optim Lett. 2017;11(1):179–194.
[16] Vázquez FG, Rückmann JJ, Stein O, et al. Generalized semi-infinite programming: a tutorial. J Comput Appl Math. 2008;217(2):394–419. Special issue: semi-infinite programming (SIP).
[17] Farshbaf-Shaker MH, Henrion R, Hömberg D. Properties of chance constraints in infinite dimensions with an application to pde constrained optimization. Set-Val Var Anal. 2018;26(4):821–841.
[18] González Gradón T, Heitsch H, Henrion R. A joint model of probabilistic/robust constraints for gas transport management in stationary networks. Comput Manage Sci. 2017;14:443–460.
[19] Heitsch H. On probabilistic capacity maximization in a stationary gas network. WIAS-Preprint 2540. Berlin: Weierstrass Institute; 2018.
[20] van Ackooij W, Frangioni A, de Oliveira W. Inexact stabilized Benders' decomposition approaches: with application to chance-constrained problems with finite support. Comput Optim Appl. 2016;65(3):637–669.
[21] van Ackooij W, Pérez-Aros P. Generalized differentiation of probability functions acting on an infinite system of constraints. Submitted draft – available soon on arxiv. 2018;1–24.
[22] Rockafellar R, Wets RJB. Variational analysis. 3rd ed. Berlin: Springer Verlag; 2009. (Grundlehren der mathematischen Wissenschaften; vol. 317).
[23] Fang K, Kotz S, Ng KW. Symmetric multivariate and related distributions. New York: Chapman and Hall; 1990. (Monographs on statistics and applied probability; vol. 36).
[24] Clarke F. Optimisation and nonsmooth analysis. New York: Wiley; 1983. (Classics in applied mathematics).