Open Journal of Mathematical Optimization

Wim van Ackooij & Pedro Pérez-Aros

Gradient formulae for probability functions depending on a heterogenous family of constraints

Volume 2 (2021), article no. 7 (29 pages)

https://doi.org/10.5802/ojmo.9

Article submitted on September 10, 2020, revised on August 9, 2021,
accepted on August 9, 2021.

Gradient formulae for probability functions depending on a heterogenous family of constraints

Wim van Ackooij
EDF R& D
7 Boulevard Gaspard Monge
91120 Palaiseau
France
wim.van-ackooij@edf.fr

Pedro Pérez-Aros
Instituto de Ciencias de la Ingenieria
Universidad de O'Higgins
Rancagua
Chile
pedro.perez@uoh.cl

Abstract

Probability functions measure the degree of satisfaction of certain constraints that are impacted by decisions and uncertainty. Such functions appear in probability or chance constraints ensuring that the degree of satisfaction is sufficiently high. These constraints have become a very popular modelling tool and are indeed intuitively easy to understand. Optimization problems involving probabilistic constraints have thus arisen in many sectors of the industry, such as in the energy sector. Finding an efficient solution methodology is important and first order information of probability functions play a key role therein. In this work we are motivated by probability functions measuring the degree of satisfaction of a potentially heterogenous family of constraints. We suggest a framework wherein each individual such constraint can be analyzed structurally. Our framework then allows us to establish formulae for the generalized subdifferential of the probability function itself. In particular we formally establish a (sub)-gradient formulæ for probability functions depending on a family of non-convex quadratic inequalities. The latter situation is relevant for gas-network applications.

Digital Object Identifier 10.5802/ojmo.9

2020 Mathematics Subject Classification 90C15.

Keywords Stochastic optimization, probabilistic constraints, chance constraints, generalized gradients.

Acknowledgments The second author partially supported by: CONICYT Grants: Fondecyt Regular 1190110 and Fondecyt Regular 1200283 and Programa Regional Mathamsud 20-Math-08 CODE: MATH190003.

1 Introduction

In many applications we are given a mapping $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ modelling constraints wherein the first argument represents a decision vector and the second argument a random vector. The use of probabilistic constraints to design safe decisions has become commonplace. We request that for a user given safety level $p \in (0, 1)$, the probability function $\varphi : \mathbb{R}^n \to [0, 1]$ defined as:

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

satisfies $\varphi(x) \geq p$. Here $\mathbb{P}$ is a Borel measure and $x \leq y$, $x, y \in \mathbb{R}^k$ is to be understood componentwise.

Although evidently, one can introduce the maximum mapping $g^m : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$:

$$g^m(x, z) = \max_{j=1,\ldots,k} g_j(x, z), \tag{2}$$

and observe that $\varphi(x) = \mathbb{P}[g^m(x, \xi) \leq 0]$, some analytical properties of g are lost in the process. Therefore the extension from $k = 1$ to $k > 1$ is, even when g is convex in the second argument, non-trivial, e.g., compare [32] to [31]. A property of the probability function φ that is particularly relevant is some degree of generalized differentiation.

It is the purpose of this work to present a general framework wherein the analysis of each component g_j of the function g separately allows us to derive information about the probability function φ as a whole. As a key finding we are able to combine various, but different, structure of component function g_j: non-convex quadratic, affine and convex in the second argument. Throughout this work we do assume however that the random vector $\xi \in \mathbb{R}^m$ is elliptically symmetric. This is not restrictive, or not as restrictive as it may first appear, when one realizes that g may have a non-linear dependency in the second argument. The analysis thus covers certain log-normal situations (g with an exponential convex dependency in the second argument), and much more. The use of Gaussian or Gaussian-like (elliptical distributions) is very frequent in applications, in particular in energy. The survey [30] on unit-commitment problems impacted by uncertainty makes this apparent. Indeed it is reasonable to some extent to model load following a Gaussian distribution, e.g., [4]. Likewise, energy generation of wind turbines can be linked to Gaussian distributions as done in [3]. Elliptical distributions are also common in gas-network applications, e.g., [14]. Furthermore whenever underlying uncertainty has a temporal aspect, time series models may be employed. The latter, can, if causal, be expressed in terms of an "innovation" process, typically taken to be Gaussian. Since our general framework allows for possibly non-linear transformations (at least with generic g) of elliptically symmetric random vectors, the framework is thus very rich to cover many forms of uncertainty vectors.

1.1 Relevance of first order information

First order information is an important ingredient to concretely solve probabilistically constrained optimization problems. Indeed most algorithms, except for "derivative free" ones, require first order information of some kind to compute successive iterates or approximate solutions. First order information is also of importance in expressing optimality conditions. Obtaining a workable formula for the gradient of the probability function has shown largely preferable to the use of finite differences (e.g., [1, 12, 14]). This results from the probability function typically being computed with a manageable, but present, approximation error. The availability of a formula for the gradient of the probability function, readily evaluated, allows one to obtain much better results, much faster. The formulæ given in this work too are such that they can be evaluated simultaneously with the probability function value and with the same cost. The provided formulæ naturally benefit from "variance reduction" but this can be further enhanced when making use of Quasi Monte-Carlo (QMC) type approaches (e.g., [12]).

The importance of first order information was recognized a long time ago and differentiability of probability functions studied under various assumptions. We refer the reader to e.g., [10, 19, 27] for a sample of these works. Although generally in these works the random vector ξ can be of arbitrary (continuous) distribution, several other assumptions limit the scope. A key assumption typically is that $\{z \in \mathbb{R}^m : g(x, z) \leq 0\}$ is bounded or compact near the point of interest $\bar{x}$. Such a condition is problematic in so much that it rules out several interesting structures. To give an example, an affine structure of the type $c_i(x)^\mathsf{T} z \leq d_i(x)$ is not compatible with this assumption. We do mention however that an abstract integrability condition is mentioned in [17, Remark 4.6] as a replacement for this compactness condition.

More recently, starting with [24, 25] a different way to investigate differentiability was initiated. This was achieved through the use of the spherical radial decomposition of elliptically symmetric random vectors. The latter family contains the multi-variate Gaussian, but also multi-variate Student random vectors. Although the restriction to a specific, yet broad, class of random vectors is made, as a result, other restrictions can be relaxed. For instance the above mentioned compactness assumption (that is still present in [25, Assumption 2.2 (i)], but relaxed in [31]). Furthermore abstract conditions can be naturally tied in with properties of the nominal underlying data. As an example, the abstract transversality condition [25, Assumption 2.2 (iii)] can be avoided when the mappings g_j are assumed to be convex in the second argument (in which case the transversality condition automatically holds). It can be shown to not be necessary, at the price of a non-trivial and novel analysis, when exploiting some specific further structure, such as the mapping g_j being non-convex quadratic (in which case the transversality condition can not hold!). In order to see why allowing for non-transversal directions is difficult, it is needed to recall that a "candidate" for the "gradient" of φ can be represented as an integral of a given expression. This last expression however showcases a vanishing term in the denominator of a certain fraction near non-transversal directions. It is thus not even clear, a priori, if the "candidate" is well defined to begin with. We refer to [37, Section 2.4] for further information. In this work we will not require the transversality condition and provide a general framework to extend the analysis of a situation involving a function g with a single component to a situation wherein several components are present.

For general introductions to the theory of probability constraints we refer to [7, 15, 21, 22]. We also refer to [28] for further context and references.

1.2 Discussion of bottlenecks and contributions

We briefly mentioned earlier the difficulty in extending results from the case $k = 1$ to the case $k > 1$, which will be the main focus of our current work. Let us intuitively try to explain the underlying reasons. First, from a general viewpoint, the mapping g^m given in (2), building the bridge between the case $k > 1$ and $k = 1$ is necessarily non-smooth, even if each component function g_j, $j = 1, \dots, k$ is smooth. Intuitively one may believe that such non-smoothness may occur only at a very select set of points and thus be "insufficient" to perturb properties of φ. This is however false, even in the most favourable cases as for instance [16, Example 1] shows (in this example $m = 1$, $k = 2$, $g_1(x, z) = z - x_1$, $g_2(x, z) = z - x_2$, $\xi \sim \mathcal{N}(0, 1)$). Non-smoothness as a result of moving from $k = 1$ to $k > 1$ thus needs to be carefully analysed. Making use of the spherical radial decomposition of ξ has shown to be very fruitful in understanding generalized differentiability of probability functions. Indeed, under this decomposition, one can represent the probability function as follows:

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

where μ_ζ is a probability measure on the (compact) Euclidian unit sphere $\mathbb{S}^{m-1}$ and $e(x, v)$ measures in itself the length of various intervals in $\mathbb{R}$, but not according to the Lebesgue measure, but according to a probability measure. The boundary points of each of these intervals depend in turn on the nominal data functions g_j, $j = 1, \dots, k$ and on the argument (x, v). A possible situation is illustrated in Figure 1.

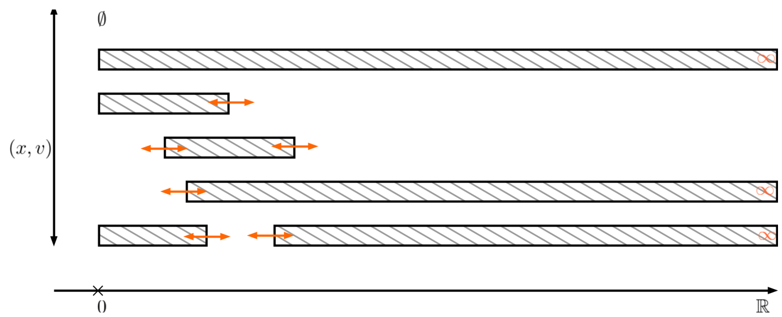

Figure 1 The mapping e is the probability measure of certain intervalles. Their number and shape depend on the arguments (x, v). This figure illustrates the possible situations when $k = 1$ and g_1 is (non-convex) quadratic in the second argument with non-vanishing quadratic term.

We can now highlight that several possible difficulties could appear:

- The number of possible intervals could be infinite.
- The number of possible intervals can depend on (x, v) and vary. Intervals can locally merge, split, or vanish - creating a specific form of non-smoothness.
- The boundary points of a given interval depend in a non-trivial way on the boundary points of "single component functions". This can be understood as follows: the union of intervals that needs to be measured is the intersection over $j = 1, \dots, k$ of similar such unions but based on each component function g_j only. As a result a complex mixture of a dependency on each components may appear.

We are now in measure to discuss how previous works have addressed some of these possible issues. The works [24, 25] assumed that g is "star-shaped" in the second argument so that e consists of measuring a single interval only. In [13, 31, 32, 33], the mapping g is assumed to be convex in the second argument. Under this assumption and a minor technical condition, that automatically holds at any x satisfying $\varphi(x) > \frac{1}{2}$, there is also only a single interval to analyse of the form $[0, \rho(x, v)]$, where ρ is an extended valued function. The technicality of moving from $k = 1$ to $k > 1$ resides in moving from classic analysis to non-smooth analysis and thus the need to adapt the set of tools. When g is not assumed to be convex in the second argument, the analysis is rendered more complex because non-transversal directions appear: the integrability of certain terms needs to be carefully studied. This was successfully done when $k = 1$ and g is quadratic in the second argument in [37], but the study was possible because an analytic form of the various involved terms was available. We note moreover in this case that we have to deal with 1 or possibly two intervals depending on (x, v). Now, already in this case the analysis can not be carried over to several components, because the terms that need to be studied are now a non-smooth combination of earlier studied terms. It does not appear viable (consider in [37, Definition 3.1]) to provide an exhaustive and comprehensive analytical study, of all the possible cases even when g is restricted to be quadratic in the second argument.

As it thus appears, there is substantial difficulty in levering properties, especially in the non-convex $k > 1$ case from each component to the probability function as a whole. This work suggests a high-level analysis making this possible. It thus allows us to consider the extension of [37] from $k = 1$ to $k > 1$, which is practically relevant for the study of gas-networks, e.g., [14]. It also makes possible the consideration of various mappings g_j with a different type of dependency in the second argument: “affine”, convex or quadratic. It may also provide a path for similar extensions from $k = 1$ to $k > 1$ for still to be explored dependencies in the second argument such as polynomial ones. Then, the study of differentiability of φ may remain concerned with the study of the case $k = 1$, which should considerably simplify the analysis.

We also care to mention two further works in progress, [34, 38], that are concerned with other extensions. The first work provides an extension while assuming a convex like dependency in the second argument, but allowing for nearly arbitrary random vectors ξ. This comes at the cost of a formula less favourable for numerical evaluation. The second work allows g to be difference-of-convex in the second argument, but with polyhedral convex functions. In that setting non-transversality can be avoided by leveraging on underlying linear structure.

1.3 Organisation of the work

This work is organized as follows. Section 2 provides background material, lays down notation and provides further scope for the investigation. Since our investigation makes use of generalized differentiation, we will briefly provide the relevant definitions and some of the properties frequently employed. We also provide the definition of elliptical random vectors and the associated implications. One of these implications being that the probability function φ can be represented as an integral over the Euclidian unit-sphere of a “radial” probability function, labelled e. This radial probability function is in fact none other than the probability of a certain parameter dependent union of intervals. The latter union in itself results as the intersection of similar unions, immediately related to each component g_j, $j = 1, \dots, k$ of the nominal data function g. The structure of this union of intervals, and in particular conditions under which only finitely many intervals can be studied, are examined in Section 3. We also carefully examine conditions under which the parameter dependent boundary points of each interval in these unions are differentiable in a generalized sense. Continuity of the radial probability function e is also examined. Section 4 is dedicated to the study of generalized differentiation of the radial probability function e and most importantly to the generalized differentiability of the probability function φ itself. The results are derived under a series of assumptions specified in terms of properties for each component $j = 1, \dots, k$, g_j of the function g itself. Section 5 shows that these assumptions hold under various structural requirements on the form of each component function g_j. The special structural cases of non-convex quadratic, affine and convex in the second argument are examined. As a corollary of the investigation the case of being able to combine these various structures results. Section 6 provides an illustration of a numerical example showing the interest of the here developed formulæ.

2 Background material

2.1 Notation

Throughout this work we will rely on tools from variational analysis. We will mainly employ standard notation, but for the convenience of the reader we will recall some of the used concepts and relations briefly. Let us begin by recalling the following cones, related to the study of the geometry of non-convex closed sets (e.g., [20, 23]):

▶ **Definition 1** (Tangent and normal cones.). *Let $X \subseteq \mathbb{R}^n$ be a closed set and $\bar{x} \in X$ be given. The Bouligand/Severi tangent or contingent cone and Fréchet normal cone to X at $\bar{x}$ are respectively defined as:*

$$\begin{aligned}
\mathrm{T}_X(\bar{x}) &:= \left\{ d \in \mathbb{R}^n : \exists\, t_n \downarrow 0, \exists\, x_n \in X, \lim_{n\to\infty} t_n^{-1}(x_n - \bar{x}) = d \right\} \\
\mathrm{N}^{\mathtt{F}}_X(\bar{x}) &:= \{x^* \in \mathbb{R}^n : \langle x^*, d\rangle \le 0 \;, \forall\, d \in \mathrm{T}_X(\bar{x})\} \\
&= \left\{ x^* \in \mathbb{R}^n : \limsup_{X \ni x' \to \bar{x}} \langle x^*, x' - \bar{x}\rangle \, \|x' - \bar{x}\|^{-1} \le 0 \right\}.
\end{aligned}$$

The Mordukhovich or limiting normal cone to X at $\bar{x}$ is defined as:

$$\mathrm{N}^{\mathtt{M}}_X(\bar{x}) := \left\{ x^* \in \mathbb{R}^n : \exists\, (x_n, x_n^*) \to (\bar{x}, x^*),\, x_n \in X, x_n^* \in \mathrm{N}^{\mathtt{F}}_X(x_n) \right\}.$$

With the help of these cones, we can define two notions of subdifferentials through the usual construction involving normal cones to epigraphs. This is done as follows:

▶ **Definition 2.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a given lower semi-continuous function, then at $\bar{x} \in \mathbb{R}^n$, its Fréchet subdifferential and Mordukhovich or limiting subdifferential are defined as:*

$$\partial^{\mathtt{F}} f(\bar{x}) = \left\{ x^* : \, (x^*, -1) \in \mathrm{N}^{\mathtt{F}}_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x})) \right\} \tag{3a}$$

$$\partial^{\mathtt{M}} f(\bar{x}) = \left\{ x^* : \, (x^*, -1) \in \mathrm{N}^{\mathtt{M}}_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x})) \right\}, \tag{3b}$$

where epi f *refers to the epigraph of f.*

For a function $f : \mathbb{R}^n \to \mathbb{R}$ that is locally Lipschitzian, Clarke subdifferential enjoys favorable calculus rules. Among the many alternative definitions, let us provide the following characterization, valid in finite dimensions:

$$\partial^{\mathtt{C}} f(\bar{x}) := \mathrm{Co} \left\{ \lim_{\ell\to\infty} \nabla f(x_\ell) : \, x_\ell \to \bar{x}, f \text{ is differentiable at } x_\ell \right\}, \tag{4}$$

resulting from [5, Theorem 2.5.1]. Here Co stands for the convex hull of a given set. With such a function f, we can associate the Clarke directional derivative of f at a point $x \in \mathbb{R}^n$ along a direction d in the following way:

$$f^\circ(x; d) := \limsup_{y \to x,\, t \downarrow 0} \frac{f(y + td) - f(y)}{t}.$$

It can be shown that $f^\circ(x; d) = \max_{s \in \partial^{\mathtt{C}} f(x)} \langle d, s\rangle$ [5], which is actually rather true by definition.

Finally, let us recall the Dini-Hadamard derivative (see, e.g., [23, Chapter 8] for more details). Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, the Dini-Hadamard derivative at u in the direction h is given by

$$f'(u; h) := \liminf_{s \to 0^+, h' \to h} \left(\frac{f(u + sh') - f(u)}{s} \right).$$

The Dini-Hadamard derivative is related to the Fréchet subdifferential in a fashion analogously to the relation between Clarke subdifferential and directional derivative. In fact it holds that:

$$\partial^{\mathtt{F}} f(\bar{x}) = \{x^* : \, \langle x^*, h\rangle \le f'(\bar{x}; h) \; \forall\, h \in \mathbb{R}^n\}$$

and if f is moreover differentiable at u, then:

$$f'(u; h) = \langle \nabla f(u), h\rangle, \text{ for all } h \in \mathbb{R}^n. \tag{5}$$

2.2 Elliptical Distributions

Let us formally define elliptically symmetrically distributed random vectors as follows:

▶ **Definition 3.** *We say that the random vector $\xi \in \mathbb{R}^m$ is elliptically symmetrically distributed with mean $\mathfrak{m}$, positive definite covariance-like matrix Σ and generator $\theta : \mathbb{R}_+ \to \mathbb{R}_+$, which is denoted by $\xi \sim \mathcal{E}(\mathfrak{m}, \Sigma, \theta)$ if its density $f_\xi : \mathbb{R}^m \to \mathbb{R}_+$ is given by*

$$f_\xi(z) = \big(\det \Sigma\big)^{-1/2} \theta\Big((z - \mathfrak{m})^{\mathsf{T}} \Sigma^{-1} (z - \mathfrak{m}) \Big), \tag{6}$$

where the generator function $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ must satisfy

$$\int_0^\infty t^{\frac{m}{2}} \theta(t)\, \mathrm{d}t < \infty.$$

Now consider L as the matrix arising from the Cholesky decomposition of Σ, i.e., $\Sigma = LL^\mathsf{T}$, it can be shown that ξ admits a representation as

$$\xi = \mathfrak{m} + \mathcal{R}L\zeta. \tag{7}$$

Where ζ has a uniform distribution over the Euclidean m-dimensional unit sphere $\mathbb{S}^{m-1} := \{z \in \mathbb{R}^m : \sum_{i=1}^m z_i^2 = 1\}$ and $\mathcal{R}$ possesses a density, which is given by

$$f_{\mathcal{R}}(r) := \begin{cases} \frac{2\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2})} r^{m-1}\theta(r^2) & \text{if } r > 0 \\ 0 & \text{if } r \leq 0. \end{cases}$$

Throughout this work, we will assume this latter density function to be continuous (e.g., θ is assumed to be continuous). The associated distribution function will be denoted $F_{\mathcal{R}}$, and thus in particular $F_{\mathcal{R}}(0) = 0$ holds true. The associated measure will be denoted $\mu_{\mathcal{R}}$, i.e., for any Borel measurable set $A \subseteq \mathbb{R}$, $\mu_{\mathcal{R}}(A) = \int_A f_{\mathcal{R}}(s)\,\mathrm{d}s$.

The family of elliptically symmetric random vectors includes many classical families: for instance, Gaussian random vectors and Student random vectors are elliptical with the respective generators

$$\theta^{\text{Gauss}}(t) = \exp(-t/2)/(2\pi)^{m/2}$$
$$\theta^{\text{Student}}(t) = \frac{\Gamma\left(\frac{m+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}(\pi\nu)^{-m/2}\left(1 + \frac{t}{\nu}\right)^{-\frac{m+\nu}{2}},$$

where Γ is the usual gamma-function. Other examples, such as logistic or exponential power random vectors, are considered in the literature; see e.g. [9, 18].

Finally, since Σ is regular, we may actually assume without loss of generality that $\mathfrak{m} = 0$ and $\Sigma = R$, i.e., is a correlation matrix. Although this latter fact will never be really required (see [29, Section 2.3], [31, Remark 3.2]).

2.3 Spherical-Radial Representation of the Probability Function

As an immediate consequence of the spherical radial decomposition (7), we can provide an alternative representation of $\varphi(x)$. Indeed, for all $x \in \mathbb{R}^n$, we have that

$$\varphi(x) = \int_{v\in\mathbb{S}^{m-1}} \mu_{\mathcal{R}}\left(\{r \geq 0 : g(x, rLv) \leq 0\}\right) \mathrm{d}\mu_\zeta = \int_{v\in\mathbb{S}^{m-1}} e(x, v)\,\mathrm{d}\mu_\zeta \tag{8}$$

where the *radial probability function* is

$$e(x, v) := \mu_{\mathcal{R}}\left(\{r \geq 0 : g(x, rLv) \leq 0\}\right) \quad \forall\, x \in \mathbb{R}^n \;\; \forall\, v \in \mathbb{S}^{m-1}. \tag{9}$$

It will also be convenient to introduce the set valued mapping $R : \mathbb{R}^n \times \mathbb{S}^{m-1} \rightrightarrows \mathbb{R}_+$ as

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

for any $(x, v) \in \mathbb{R}^n \times \mathbb{S}^{m-1}$. As a consequence for any $(x, v) \in \mathbb{R}^n \times \mathbb{S}^{m-1}$ we have $e(x, v) = \mu_{\mathcal{R}}(R(x, v))$. When it comes to numerical evaluation, it can be observed (e.g., [31, (1.5)]), that representation (8) alone, already provides a reduction of sample variance w.r.t. immediately sampling from the nominal representation (1).

More generally, our first endeavour will be to establish continuity of the mapping e, so that we can discuss generalized subdifferentials of it. To this end, let us introduce the set-valued map $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, given by

$$M(x) = \{z \in \mathbb{R}^m :\; g(x, z) \leq 0\}, \tag{10}$$

and observe that M is closed-valued, since g is continuous. Of course the probability function can be rewritten as:

$$\varphi(x) = \mathbb{P}[\xi \in M(x)].$$

Furthermore, the set valued map R also satisfies:

$$R(x, v) = \{r \geq 0 \;:\; rLv \in M(x)\} = \{r \geq 0 \;:\; g(x, rLv) \leq 0\}. \tag{11}$$

It is also clear that if one defines for each $j = 1, \dots, k$, the set-valued map $M_j : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, given by

$$M_j(x) = \{z \in \mathbb{R}^m : g_j(x, z) \leq 0\}, \tag{12}$$

and R_j given by

$$R_j(x, v) = \{r \geq 0 : rLv \in M_j(x)\} = \{r \geq 0 : g_j(x, rLv) \leq 0\}, \tag{13}$$

that $R(x, v) = \bigcap_{j=1}^k R_j(x, v)$ and $M(x) = \bigcap_{j=1}^k M_j(x)$. It is our intention to entail properties of e, M, R and ultimately φ from properties established for each "component" j. We will start our investigation with the study of continuity of e. Throughout this work, we will investigate differentiability at a given "trial point" $\bar{x} \in \mathbb{R}^n$, we will assume given an appropriate neighbourhood U of $\bar{x}$, or "construct" it by shrinking the initially given neighbourhood further if needed. The existence of such a neighbourhood U will be carefully discussed.

3 Continuity of e

In our first investigation concerning the continuity of e, the specific nature of M given through g does not matter much and only properties of M (and through it R) are exploited. This section is dedicated to formally establishing that R can be given as a countable union of intervals. We also study the dependency of the bounds of these intervals on (x, v) at an abstract level. This abstract level will be the one useful for the numerical evaluation of the first order information of φ.

3.1 Interval calculus

Let us first provide some characterization of R as a union of intervals. For this result we will require an assumption that is not very restrictive, ruling out too many zeros of g in the second argument.

► **Lemma 4.** *Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a continuous function and consider $(x, v) \in \mathbb{R}^n \times \mathbb{S}^{m-1}$ such that the potential solutions $r > 0$ to $g(x, rLv) = 0$ do not have accumulation points. Then, the set $R(x, v)$ is a countable union of intervals. More precisely, there exists $p \in \mathbb{N} \cup \{+\infty\}$ and there are points $a_i, b_i \geq 0$, $i = 1, \dots, p$, such that $a_0 \leq b_0 \leq a_1 \leq \cdots \leq b_p$ and*

$$R(x, v) = \bigcup_{i=0}^{p} [a_i, b_i], \tag{14}$$

where we use the convention $[\,\cdot, +\infty] = [\,\cdot, +\infty)$. Moreover, for all $i \geq 0$, $g(x, a_i Lv) = 0$ and whenever $b_i \in \mathbb{R}$, we have $g(x, b_i Lv) = 0$.

Proof. Let $(x, v) \in \mathbb{R}^n \times \mathbb{S}^{m-1}$ be as in the statement of the result. Since, g is continuous, the set

$$W := \{r > 0 : g(x, rLv) > 0\}$$

is open and, so it can be written as an countable union of open disjoint intervals, let us write:

$$W = \bigcup_{i \in I} (c_i, d_i),$$

where the index set I is countable and $c_i, d_i \in \mathbb{R}$ for $i \in I$. Here, we notice that $c_i, d_i \notin W$, hence $g(x, c_i Lv) \leq 0$ and $g(x, d_i Lv) \leq 0$. Moreover, by continuity of g, in the second argument, we have that $g(x, c_i Lv) \geq 0$ and $g(x, d_i Lv) \geq 0$, and consequently $g(x, c_i Lv) = 0 = g(x, d_i Lv)$.

Now, we define the function $f : I \to \mathbb{N}$, by, $f(i) := \#\{j \in I : c_j \leq c_i\}$, the function is well-defined due to the fact that $\{c_j\}_{j \in I}$ does not have accumulation points and it is one-to-one. Indeed for any given $i \in I$ and c_i, the set $[0, c_i]$ can only contain finitely many other c_j. Consequently the inverse of f, let us write it as $f^{-1} : N \to I$, where $N := f(I)$, is well-defined. Therefore, we can assume that there are $q \in \mathbb{N} \cup \{\infty\}$ and (abusing the notation) disjoint intervals (c_i, d_i) with $d_i \leq c_{i+1}$ for all i such that in fact:

$$W = \bigcup_{i=0}^{q} (c_i, d_i). \tag{15}$$

Now, we distinguish two cases:

- First, consider the situation wherein $g(x,0) \leq 0$. Then, if $q = +\infty$, we set $p = q$, $a_0 := 0$, and $b_i = c_i$, $a_{i+1} := d_i$ for all $i \geq 0$. In the case that q is finite we define $a_0 := 0$, and $b_i = c_i$, $a_{i+1} := d_i$ similarly, except for the last element:

$$b_q = c_q \text{ and } p = q \text{ if } d_q = +\infty$$
$$b_{q+1} = +\infty \text{ and } p = q+1 \text{ if } d_q < \infty.$$

- Second, consider the situation wherein $g(x,0) > 0$. Then, if $q = +\infty$, we set $p = q$ in this case we define $a_i = d_i$, $b_i = c_{i+1}$ for all $i \geq 0$. When q is finite, the definitions are similar except for the last element:

$$b_{q-1} = c_q \text{ and } p = q-1 \text{ if } d_q = +\infty$$
$$b_q = +\infty \text{ and } p = q \text{ if } d_q < \infty.$$

Finally, we have to check (14). Indeed, $\supseteq$ holds by construction in (14), since any $w \in \bigcup_{i=0}^{p} [a_i, b_i]$ belongs to the complement of W, i.e., is such that $g(x, wLv) \leq 0$.

Now, assume that $w \in R(x,v)$, if $w = 0$, i.e., $g(x,0) \leq 0$, we have $a_0 = w = 0$. We may thus assume $w > 0$, then if $w \notin W$, we have that by (15) and the fact that $\{c_j\}_{j\in I}$ and $\{d_j\}_{j\in I}$ do not have accumulation points that there exist $i \in \mathbb{N}$ such that $d_i \leq w \leq c_{i+1}$, and consequently $w \in \bigcup_{i=0}^{p} [a_i, b_i]$. ◀

▶ Remark 5. It is worth mentioning that the property that the solutions $r > 0$ such that $g(x, rLv) = 0$ do not have accumulation points is stable under maximum operation. Indeed, if $\{g_j : j = 1, \ldots, k\}$ is a finite family of functions such that for each $j = 1, \ldots, k$ the solutions $r > 0$ to $g_j(x, rLv) = 0$ do not have accumulation points, then the solutions $r > 0$ to $\max_{j=1,\ldots,k} g_j(x, rLv) = 0$ do not have accumulation points either. This is important in our analysis, because the earlier said property holds for analytic functions, and thus consequently for (finite) maxima of analytic functions.

▶ **Corollary 6.** *Assume that the set-valued mapping M is closed-valued and continuous and that the set of boundary points (in $\mathbb{R}$) of the rays $\mathbb{R}_+ Lv \cap M(x)$ do not admit cluster points.*

Then for each $(x,v) \in \mathbb{R}^n \times \mathbb{S}^{m-1}$, the set $R(x,v)$ is a countable union of intervals that we will denote with:

$$R(x,v) = \bigcup_{i=1}^{\infty} [a_i(x,v), b_i(x,v)], \tag{16}$$

where the union is considered finite if for some i, $b_i(x,v) = \infty$ holds true.

Proof. Let us define $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ as $g(x,z) = d(z, M(x))$, where d is based on the underlying (Euclidian) distance. Then as a result of [23, Proposition 5.11] the continuity of g follows. The desired result now follows from Lemma 4. ◀

The first issue we should deal with is ensuring that R is a finite union of (potentially degenerate) intervals. The exact number of such intervals possibly varies and we will require some additional assumptions in view of this. The following Lemma will prove a valuable tool.

▶ **Lemma 7.** *Assume that for each $j = 1, \ldots, k$, there exists K_j such that for each $(x,v) \in \mathbb{R}^n \times \mathbb{S}^{m-1}$, $R_j(x,v)$ as given akin to (16) is the union of at most K_j intervals. Then $R(x,v)$ is the union of at most $\prod_{j=1}^{k} K_j$ intervals.*

Define furthermore $s_1, \ldots, s_k, s : \mathbb{R}^n \times \mathbb{S}^{m-1} \to \mathbb{N}$:

$$s_j(x,v) = \sup\left\{ i = 1, \ldots, K_j : a_i^j(x,v) < \infty \right\},$$

so that $R_j(x,v) = \bigcup_{i=1}^{s_j(x,v)} [a_i^j(x,v), b_i^j(x,v)]$ and likewise s but with respect to R.

Should $(\bar{x}, \bar{v})$ be given such that $s_1, \ldots, s_k$ are constant near $(\bar{x}, \bar{v})$, then the same is true for s.

Proof. Let $(x,v) \in \mathbb{R}^n \times \mathbb{S}^{m-1}$ be given but arbitrary. For this proof it is sufficient to assume $k = 2$. Hence let us assume that for $j = 1, 2$, the following representation holds:

$$R_j(x,v) = \bigcup_{i=1}^{K_j} [a_i^j(x,v), b_i^j(x,v)],$$

and observe now that

$$R_1(x,v) \cap R_2(x,v) = \bigcup_{i=1}^{K_1} \bigcup_{\ell=1}^{K_2} ([a_i^1(x,v), b_i^1(x,v)] \cap [a_\ell^2(x,v), b_\ell^2(x,v)]) \tag{17}$$

as a result of the rule $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, for arbitrary sets A, B, C.

Now, for arbitrary $i = 1, \dots, K_1$, $\ell = 1, \dots, K_2$:

$$[a_i^1(x,v), b_i^1(x,v)] \cap [a_\ell^2(x,v), b_\ell^2(x,v)] = [\max\left\{a_i^1(x,v), a_\ell^2(x,v)\right\}, \min\left\{b_i^1(x,v), b_\ell^2(x,v)\right\}],$$

with the evident rule that should $\max\left\{a_i^1(x,v), a_\ell^2(x,v)\right\} > \min\left\{b_i^1(x,v), b_\ell^2(x,v)\right\}$, then the set is empty. As a result $R_1 \cap R_2$ admits the representation given in Corollary 6 but with at most $K_1 K_2$ terms.

Let us now first make the observation that the mappings s_j are well defined for each $j = 1, \dots, k$. We can also observe that only $b^j_{s_j(x,v)}(x,v)$ can be equal to infinity. As a result, in the representation (17) of $R_1 \cap R_2$ as a union of intervals, only the one corresponding to $i = s_1(x,v)$ and $\ell = s_2(x,v)$ can be of the form $[\,\cdot\,, \infty)$, which happens moreover only if $b^j_{s_j(x,v)}(x,v) = \infty$ for both $j = 1, 2$.

Should (x,v) now be such that locally neither s_1 nor s_2 change, then the representation (17) can be arranged to have locally exactly $s_1(x,v)s_2(x,v)$ intervals (some of which might be empty) and consequently s is locally constant as well. ◀

One may of course wonder if the given assumptions of Lemma 7 (existence of K_j, local constance of s_j) are reasonable. This can be ensured whenever g_j has a polynomial dependency (of fixed degree) in the second argument. Moreover, whenever g_j is (non-convex) quadratic in the second argument, it can be shown that s_j as given in the previous Corollary is indeed locally constant (e.g., [37]), at least on a sufficiently large open set. This will be largely discussed in Section 5.

Based on our representation of R_j, let us present a set of assumptions that will be useful in the sequel. The purpose of the assumption is to rule out a too erratic variation in the number of intervals that compose R_j, as well as to ensure that the total number is bounded. We also request that most directions are transversal. We will postpone however for the time being giving a concrete example, which will be done in Section 5.

▶ **Assumption 8.** *Let $\bar{x} \in \mathbb{R}^n$ along with a neighbourhood U be given. For each $j = 1, \dots, k$, we may find an open set $\mathcal{O}_j \subseteq U \times \mathbb{S}^{m-1}$, such that $\mathcal{O}_j$ is of $\lambda_n \otimes \mu_\zeta$ full measure on $U \times \mathbb{S}^{m-1}$ and for each $(x,v) \in \mathcal{O}_j$, it holds that*

$$\langle \nabla_z g_j(x, rLv), Lv \rangle \neq 0 \text{ for all } r \text{ s.t. } g_j(x, rLv) = 0. \tag{18}$$

Moreover the mapping s_j defined in Lemma 7 is assumed locally constant on $\mathcal{O}_j$ and bounded from above by K_j on $U \times \mathbb{S}^{m-1}$.

Let us observe that Assumption 8 can be entailed under a somewhat simpler condition as follows:

▶ Remark 9. Let $\bar{x} \in \mathbb{R}^n$ along with a neighbourhood U be given. Assume that for all $x \in U$, the set

$$\overline{\mathcal{O}}_j(x) := \{v \in \mathbb{S}^{m-1} : \langle \nabla_z g_j(x, rLv), Lv \rangle \neq 0 \text{ for all } r \text{ s.t. } g_j(x, rLv) = 0\} \tag{19}$$

has μ_ζ full measure on $\mathbb{S}^{m-1}$. Then $\mathcal{O}_j := \operatorname{gph} \overline{\mathcal{O}}_j \cap (U \times \mathbb{S}^{m-1})$ satisfies the properties of Assumption 8 as the result of an application of [8, Korollar V.1.6].

3.2 Fine characterization of R

Under Assumption 8, we may write for some $K > 0$ on $U \times \mathbb{S}^{m-1}$:

$$R(x,v) = \bigcup_{i=1}^{K} [a_i(x,v), b_i(x,v)]. \tag{20}$$

Now let us first provide a formal statement showing how certain of these interval boundary points are differentiable and how their sub-gradients relate to the nominal original data.

Let us designate a_i for some i as an entry point and b as an exit point. We can formalise this as follows:

$$\Lambda^{\texttt{entry}}(x,v) = \{a_i(x,v) : i = 1, \dots, K \text{ s.t. } 0 < a_i(x,v) < \infty\} \tag{21a}$$

$$\Lambda^{\texttt{exit}}(x,v) = \{b_i(x,v) : i = 1, \dots, K \text{ s.t. } b_i(x,v) < \infty\}. \tag{21b}$$

▶ **Lemma 10.** *Let $x \in \mathbb{R}^n$ be given and $v \in \mathbb{S}^{m-1}$ be arbitrary. Let $\bar{r} \geq 0$ be such that $\bar{r} \in \Lambda^{\texttt{entry}}(x,v) \cap \Lambda^{\texttt{exit}}(x,v)$, then there exists $i = 1, \ldots, k$ such that $\langle \nabla_z g_i(x, \bar{r}Lv), Lv \rangle = 0$.*

Proof. Let $i = 1, \ldots, K+1$ be such that $\bar{r} = a_i(x,v) = b_i(x,v)$, the existence of which follows by assumption. It follows from $\bar{r} = a_i(x,v)$ and the definition of the latter that for some $\delta > 0$ sufficiently small, $g^m(x, r'Lv) > 0$ for all $r' \in (\bar{r} - \delta, \bar{r})$ holds true. Indeed, should this not be the case, then for all $\delta > 0$, we may find some $r' \in (\bar{r} - \delta, \bar{r})$ at which $g^m(x, r'Lv) \leq 0$ holds true. Hence, by continuity of g^m in the second argument, there exists $r'' \in [r', \bar{r})$ such that $g^m(x, r''Lv) = 0$. This would contradict that K is finite and/or imply that $\bar{r}$ is a cluster point of "solutions". Likewise (by shrinking δ if needed), it follows from $\bar{r} = b_i(x,v)$ and definition of the latter that $g^m(x, r'Lv) > 0$, $r' \in (\bar{r}, \bar{r} + \delta)$. We claim that we may assume the existence of j^* such that $g_{j^*}(x, r'Lv) > 0$ for $\bar{r} \neq r' \in (\bar{r} - \delta, \bar{r} + \delta)$, and $g_{j^*}(x, \bar{r}Lv) = 0$. Should this not hold, then for any active index j such that $g_j(x, \bar{r}Lv) = 0$,

1. either $g_j(x, .Lv) < 0$ on a given neighbourhood (of $\bar{r}$),
2. or $g_j(x, r_\ell Lv) = 0$ for a sequence $r_\ell \to \bar{r}$.

Should the former situation, i.e., 1. hold for all j, then on an appropriate neighbourhood of $\bar{r}$, $g^m(x, .Lv) < 0$, which contradicts the established properties of $g^m(x, .Lv)$ on the identified neighbourhood $(\bar{r} - \delta, \bar{r} + \delta)$. The latter situation, i.e., 2., must however imply through [36, Lemma 3.1 (1)] that $\langle \nabla_z g_j(x, r_\ell Lv), Lv \rangle = 0$. Indeed if not, we may find some appropriate neighbourhood W of $\bar{r}$, and neighbourhoods U', V' of (x,v), such that [36, Lemma 3.1 (1)] holds. However, for ℓ sufficiently large $r_\ell \in W$, thus leading to the contradiction $r_\ell \neq \rho_{\bar{x}}^{x,v}(x,v) = \bar{r}$. The claim has thus been shown.

If j^* is now as claimed, then the map $r \mapsto g_{j^*}(x, rLv)$ attains a local minimum at $\bar{r}$, thus implying $\langle \nabla_z g_{j^*}(x, \bar{r}Lv), Lv \rangle = 0$ as a first order optimality condition. ◀

We can now establish:

▶ **Lemma 11.** *Let $x \in \mathbb{R}^n$ be such that $g_j(x, 0) \neq 0$ for all $j = 1, \ldots, k$ where g_j denotes the j-th component of the mapping g. Let $v \in \mathbb{S}^{m-1}$ and r^* be such that $g^m(x, r^*Lv) = 0$. Assuming that $\langle \nabla_z g_i(x, r^*Lv), Lv \rangle \neq 0$ for all $i = 1, \ldots, k$, there exist neighbourhoods U of x, V of v and $W \subseteq \mathbb{R}_+$ of r^* as well as a Lipschitz function $\rho_{r^*}^{x,v} : U \times V \to W$ with the following properties:*

1. *For all $(x', v', r') \in U \times V \times W$ the equivalence $\max_{j=1,\ldots,k} g_j(x', r'Lv') = 0 \Leftrightarrow r' = \rho_{r^*}^{x,v}(x', v')$ holds true.*
2. *We have $r^* \in \Lambda^{\texttt{entry}}(x,v) \cup \Lambda^{\texttt{exit}}(x,v)$. Moreover if $r^* \in \Lambda^{\texttt{exit}}(x,v)$ ($\Lambda^{\texttt{entry}}(x,v)$) there exist neighbourhoods U^e, V^e, W^e of x, v, r^* respectively such that for all $(x', v', r') \in U^e \times V^e \times W^e$ with $g^m(x', r'Lv') = 0$ we have $r' \in \Lambda^{\texttt{exit}}(x', v')$ ($\Lambda^{\texttt{entry}}(x,v)$ respectively).*
3. *There exist neighbourhoods $\widetilde{U}$, $\widetilde{V}$ and of x, v and r^* respectively such that for all $(x', v') \in \widetilde{U} \times \widetilde{V}$ one has the characterization*

$$\rho_{r^*}^{x,v}(x', v') = \begin{cases} \min_{j \in \mathcal{J}^{x,v,r^*}} \rho_{r^*,j}^{x,v}(x', v'), & \text{if } r^* \in \Lambda^{\texttt{exit}}(x,v) \\ \max_{j \in \mathcal{J}^{x,v,r^*}} \rho_{r^*,j}^{x,v}(x', v'), & \text{if } r^* \in \Lambda^{\texttt{entry}}(x,v) \end{cases}, \tag{22}$$

   $\rho_{r^*,j}^{x,v}$ is the mapping of [36, Lemma 3.1], and $\mathcal{J}^{x,v,r^*}$ is defined as

$$\mathcal{J}^{x,v,r^*} = \{ j \in \{1, \ldots, k\} : g_j(x, r^*Lv) = g^m(x, r^*Lv) \}. \tag{23}$$

4. *For all $(x', v') \in U \times V$, the partial Clarke subdifferential of $\rho_{r^*}^{x,v}$ is given by*

$$\partial_x^{\mathsf{C}} \rho_{r^*}^{x,v}(x', v') = \mathrm{Co} \left\{ \nabla_x \rho_{r^*,j}^{x,v}(x', v') : j \in \mathcal{J}^{x,v,r^*}(x', v') \right\}, \tag{24}$$

   where Co denotes the convex hull and $\mathcal{J}^{x,v,r^*}(x', v')$ is the subset of $\mathcal{J}^{x,v,r^*}$ where equality holds in equation (22). If the set in equation (24) reduces to a singleton, $\rho_{r^*}^{x,v}$ is continuously differentiable at (x', v').

Proof. We begin by remarking that each component g_j of the mapping g, $j = 1, \ldots, k$ is Clarke regular as a continuously differentiable map. The mapping defined as $g^m : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $g^m(x', z') = \max_{j=1,\ldots,k} g_j(x', z')$ is therefore also Clarke regular according to [5, Proposition 2.3.12]. Since continuously differentiable maps are locally Lipschitz by the mean-value Theorem it is also easily observed that g^m is locally Lipschitz as well. It therefore follows that the Clarke subdifferential of g^m exists and is a non-empty convex compact set. Moreover from [5, Proposition 2.3.15] it follows that

$$\partial^{\mathsf{C}} g^m(x', z') \subseteq \partial_x^{\mathsf{C}} g^m(x', z') \times \partial_z^{\mathsf{C}} g^m(x', z') \quad \forall\, x' \in \mathbb{R}^n, z' \in \mathbb{R}^m. \tag{25}$$

By [5, Proposition 2.3.12] it follows that

$$\partial_z^{\mathsf{C}} g^m(x', z') = \mathrm{Co}\left\{\nabla_z g_j(x', z') : j \in I(x', z')\right\},$$

where $I(x', z') = \{j = 1, \ldots, k : g_j(x', z') = g^m(x', z')\}$.

Let x, v, r^* now be as in the statement of the Lemma and set $z = r^* Lv$. Let $\pi_z \partial^{\mathsf{C}} g^m(x, z)$ denote the projection of $\partial^{\mathsf{C}} g^m(x, z)$ on $\partial_z^{\mathsf{C}} g^m(x, z)$. Let $s \in \pi_z \partial^{\mathsf{C}} g^m(x, z)$ be arbitrary. It follows by definition that one can find $s_x \in \mathbb{R}^n$ such that $(s_x, s) \in \partial^{\mathsf{C}} g^m(x, z)$. But this means that $s \in \partial_z g^m(x, z)$ by (25). Moreover for appropriate $\lambda_1, \ldots, \lambda_k$ elements of the k dimensional unit simplex, we have

$$s = \sum_{j=1}^{k} \lambda_j \nabla_z g_j(x, r^* Lv)$$

with $\lambda_j = 0$ for $j \notin I(x, z)$. Moreover, we can find $h \neq 0$ sufficiently small such that $g^m(x, (r^* + h)Lv) < 0$. Should such h not exist, we may find for every $h \neq 0$, some index $i_h \in \{1, \ldots, k\}$ such that $g_{i_h}(x, (r^* + h)Lv) = 0$, since g^m is a max function over a finite index set. Hence we may identify a sequence $h_n \downarrow 0$ as well as a fixed index i^* such that $g_{i^*}(x, (r^* + h_n)Lv) = 0$ for all n. However this contradicts [36, Lemma 3.1 (1)] since for n sufficiently large and some neighbourhood W of r^* it must hold $(r^* + h_n) \in W$, and thus $r^* = \rho_{r^*}^{x;v}(x, v) \neq r^* + h_n$ is not possible.

Now, for a given $j \in I(x, z)$, as a consequence of Taylor's theorem, we have

$$g_j(x, (r^* + h)Lv) = g_j(x, r^* Lv) + h \left\langle \nabla_z g_j(x, r^* Lv), Lv \right\rangle + o(|h|), \tag{26}$$

which shows that the family of scalars $\{\langle \nabla_z g_j(x, r^* Lv), Lv \rangle\}_{j \in I(x,z)}$ all share the same sign. Consequently, $\langle s, Lv \rangle = \sum_{j=1}^{k} \lambda_j \left\langle \nabla_z g_j(x, r^* Lv), Lv \right\rangle \neq 0$.

Since $w \in \mathbb{R}^m \mapsto \langle s, Lw \rangle$ is continuous this means that s is of full-rank. As s was arbitrary: $\pi_z \partial^{\mathsf{C}} g^m(x, z)$ is of maximal rank. We can therefore apply Clarke Implicit Function theorem ([5, p. 256]) to the equation $g^m(x, r^* Lv) = 0$ to derive the existence of neighbourhoods U of x, V of v and W of r^* along with a Lipschitz function $\rho_{r^*}^{x;v} : U \times V \to W$ such that the equivalence:

$$g^m(x', r' Lv') = 0, (x', v', r') \in U \times V \times W \iff r' = \rho_{r^*}^{x;v}(x', v'),\ (x', v') \in U \times V \tag{27}$$

holds true.

In order to derive item 2, we remark first that by the assumption $\langle \nabla_z g_j(x, r^* Lv), Lv \rangle \neq 0$ for all $j = 1, \ldots, k$ and this condition continues to hold locally near (x, v). Negating $r^* \in \Lambda^{\texttt{entry}}(x, v) \cup \Lambda^{\texttt{exit}}(x, v)$ implies together with $r^* \in R(x, v)$, i.e., $r^* \in (a_i(x, v), b_i(x, v))$ for some i^*, that

- there exists some $\delta > 0$ such that $g^m(x, r' Lv) < 0$ for all $r^* \neq r' \in (r^* - \delta, r^* + \delta) \subseteq (a_{i^*}(x, v), b_{i^*}(x, v))$. Thus in particular for some $i^* \in \{1, \ldots, k\}$, we have $g_{i^*}(x, r' Lv) < 0$, for all $r^* \neq r' \in (r^* - \delta, r^* + \delta)$ and $g_{i^*}(x, r^* Lv) = 0$. Therefore r^* is a local maximum for the map $r \mapsto h(r) = g_{i^*}(x, r Lv)$, i.e., $h'(r^*) = 0$, but $h'(r) = \langle \nabla_z g_{i^*}(x, r Lv), Lv \rangle$ which contradicts the assumption.
- There is a sequence $r_\ell \to r^*$ such that $g^m(x, r_\ell Lv) = 0$, $r_\ell \in (a_i(x, v), b_i(x, v))$. Thus by moving to a subsequence if required, we can identify some i^* such that too $g_{i^*}(x, r_\ell Lv) = 0$. However this contradicts [36, Lemma 3.1 (1)] as already argued above.

Any of these situations, thus lead to a contradiction, so that $r^* \in \Lambda^{\texttt{entry}}(x, v) \cup \Lambda^{\texttt{exit}}(x, v)$ must hold. We have established the first part of item 2. In order to establish the second part, first of all, by continuity of g as well as its first order derivatives, appropriate neighbourhoods U^1, V^1, W^1 of x, v, r^* can be found such that $g_j(x', 0) \neq 0$ for all $j = 1, \ldots, k$, $x' \in U^1$, as well as $\langle \nabla_z g_j(x', r' Lv'), Lv' \rangle \neq 0$ for all $(x', v', r') \in U^1 \times V^1 \times W^1$. Now, let us assume that $r^* \in \Lambda^{\texttt{exit}}(x, v)$ yet the assertion does not hold. Then there exist sequences, $x_\ell \to x$, $v_\ell \to v$ and $r_\ell \to r$ with $g^m(x_\ell, r_\ell L v_\ell) = 0$ as well as $r_\ell \in \Lambda^{\texttt{entry}}(x_\ell, v_\ell)$. Indeed, we may assume $(x_\ell, v_\ell, r_\ell) \in U^1 \times V^1 \times W^1$ so then by the first part of this item, $r_\ell \in \Lambda^{\texttt{entry}}(x_\ell, v_\ell) \cup \Lambda^{\texttt{exit}}(x_\ell, v_\ell)$ and the negation of the assertion yields $r_\ell \in \Lambda^{\texttt{entry}}(x_\ell, v_\ell)$. By Lemma 10, it follows that $r^* = b_{i^*}(x, v)$ for some i^* and too the existence of some $\delta > 0$ such that $a_{i^*}(x, v) < r^* - \delta < b_{i^*}(x, v) = r^*$. But also that $g^m(x, (r^* - \delta)Lv) < 0$ must hold. By item 1, we can find neighbourhoods U^2, V^2, W^2 of x, v, r^* as well as a Lipschitz mapping $\rho_{r^*}^{x;v} : U^2 \times V^2 \to W^2$, satisfying (27). For k large enough we must have $(x_\ell, v_\ell, r_\ell) \in U^2 \times V^2 \times W^2$ as well, hence (27) yields $\rho_{r^*}^{x;v}(x_\ell, v_\ell) = r_\ell$.

We now have on the one hand that $g^m(x, r' Lv) < 0$ for all $r' \in (r^* - \delta, b_{i^*}(x, v))$ and on the other hand by continuity of g^m that for each fixed such r' we can find L such that $g^m(x_\ell, r' L v_\ell) < 0$ for $\ell \geq L$ as well. However since $r' < b_i(x,v) = \rho^{x;v}_{r^*}(x,v)$ and $\rho^{x_\ell,v_\ell}_{r^*} = r_\ell \in \Lambda^{\texttt{entry}}(x_\ell,v_\ell)$, we have $r' < \rho^{x_\ell,v_\ell}_{r^*} = r_\ell$ for ℓ sufficiently large. Hence, by $r_\ell \in \Lambda^{\texttt{entry}}(x_\ell,v_\ell)$, either $g^m(x_\ell, r'Lv_\ell) > 0$ or there exists r'_ℓ with $g^m(x_\ell, r'_\ell Lv_\ell) = 0$. By employing a diagonal argument, we may thus arrange the sequence r'_ℓ such that $r'_\ell \to r^*$ as well. But then for k sufficiently large we arrive at the contradiction $r'_\ell < r_\ell = \rho^{x_\ell,v_\ell}_{r^*}$ with (27), since $(x_\ell, v_\ell, r'_\ell) \in U^2, V^2, W^2$. We have completed the proof of item 2, since the case $r^* \in \Lambda^{\texttt{entry}}(x,v)$ follows analogously.

Let us now move to the proof of item 3. Following the assumptions, for any $j \in \mathcal{J}^{x,v,r^*}$ there exist neighbourhoods U_j of x, V_j of v and W_j of r^* as well as continuously differentiable mappings $\rho^{x;v}_{r^*,j} : U_j \times V_j \to W_j$ by [36, Lemma 3.1] such that for all $(x', v', r') \in U_j \times V_j \times W_j$ we have the equivalence $g_j(x', r'Lv') = 0$ if and only if $r' = \rho^{x;v}_{r^*,j}(x',v')$. Moreover for $j \notin \mathcal{J}^{x,v,r^*}$ we have $g_j(x, r^*Lv) < 0$, so that appropriate neighbourhoods U_j, V_j, W_j of x, v, r^* exist such that $g_j(x', r'Lv') < 0$ for all $(x', v', r') \in U_j \times V_j \times W_j$. By item 1, we are given further neighbourhoods U, V, W of x, v, r^* and a Lipschitz mapping $\rho^{x;v}_{r^*} : U \times V \to W$ satisfying (27). We now define $\widetilde{U} = \bigcap_{j=1}^k U_j \cap U$, $\widetilde{V} = \bigcap_{j=1}^k V_j \cap V$, $\widetilde{W} = \bigcap_{j=1}^k W_j \cap W$, neighbourhoods of x, v, r^* respectively such that both the left and right-hand side expressions in (22) are meaningfully defined. Moreover, since $\rho^{x;v}_{r^*}$ as well as $\rho^{x;v}_{r^*_j}$ are continuous, we may assume that $\widetilde{U}$ and $\widetilde{V}$ are defined such that $\rho^{x;v}_{r^*}(x',v') \in \widetilde{W}$, $\rho^{x;v}_{r^*_j}(x',v') \in \widetilde{W}$ for all $(x',v') \in \widetilde{U} \times \widetilde{V}$. Following item 2, we may assume

- $\rho^{x;v}_{r^*}(x,v) \in \Lambda^{\texttt{exit}}(x,v) \Longrightarrow \rho^{x;v}_{r^*}(x',v') \in \Lambda^{\texttt{exit}}(x',v')$ for all $(x',v') \in \widetilde{U} \times \widetilde{V}$,
- $\rho^{x;v}_{r^*}(x,v) \in \Lambda^{\texttt{entry}}(x,v) \Longrightarrow \rho^{x;v}_{r^*}(x',v') \in \Lambda^{\texttt{entry}}(x',v')$ for all $(x',v') \in \widetilde{U} \times \widetilde{V}$.

It remains to establish the asserted equality in (22).

To this end, let us assume that $r^* \in \Lambda^{\texttt{exit}}(x,v)$ holds. Assume moreover that for a given $(x',v') \in \widetilde{U} \times \widetilde{V}$, we have $\rho^{x;v}_{r^*}(x',v') < \min_{j \in \mathcal{J}^{x,v,r^*}} \rho^{x;v}_{r^*,j}(x',v')$. Then for any given but fixed $j \in \mathcal{J}^{x,v,r^*}$ we have $g_j(x', \rho^{x;v}_{r^*}(x',v')Lv') \le g^m(x', \rho^{x;v}_{r^*}(x',v')Lv') = 0$. However if $g_j(x', \rho^{x;v}_{r^*}(x',v')Lv') = 0$, the fact that $\rho^{x;v}_{r^*}(x',v') \in \widetilde{W} \subseteq W_j$ implies by the equivalence of [36, Lemma 3.1] that $\rho^{x;v}_{r^*}(x',v') = \rho^{x;v}_{r^*,j}(x',v')$, yet by assumption $\rho^{x;v}_{r^*}(x',v') < \rho^{x;v}_{r^*,j}(x',v')$. Consequently, $g_j(x', \rho^{x;v}_{r^*}(x',v')Lv') < 0$. However, the latter inequality holds too whenever $j \notin \mathcal{J}^{x,v,r^*}$ since $\rho^{x;v}_{r^*}(x',v') \in W_j$ then too. Consequently, $\max_{j=1,\ldots,k} g_j(x', \rho^{x;v}_{r^*}(x',v')Lv') < 0$, which contradicts the definition of g^m.

Let us now assume that for a given $(x',v') \in \widetilde{U} \times \widetilde{V}$, we have $\rho^{x;v}_{r^*}(x',v') > \min_{j \in \mathcal{J}^{x,v,r^*}} \rho^{x;v}_{r^*,j}(x',v')$. Then there exists a $\iota \in \mathcal{J}^{x,v,r^*}$ such that $\rho^{x;v}_{r^*}(x',v') > \rho^{x;v}_{r^*,\iota}(x',v')$. Since by definition, $g_\iota(x', \rho^{x;v}_{r^*,\iota}(x',v')Lv') = 0$, we deduce that $g^m(x', \rho^{x;v}_{r^*,\iota}(x',v')Lv') \ge 0$. Now if $g^m(x', \rho^{x;v}_{r^*,\iota}(x',v')Lv') = 0$, from $\rho^{x;v}_{r^*,\iota}(x',v'), \rho^{x;v}_{r^*}(x',v') \in \widetilde{W}$, the equivalence (27) leads to the contradiction $\rho^{x;v}_{r^*,\iota}(x',v') = \rho^{x;v}_{r^*}(x',v')$. Consequently we must have: $g^m(x', \rho^{x;v}_{r^*,\iota}(x',v')Lv') > 0$. However we have $\rho^{x;v}_{r^*}(x',v') \in \Lambda^{\texttt{exit}}(x',v')$, which by continuity of g^m entails the existence of $r' \in (\rho^{x;v}_{r^*,\iota}(x',v'), \rho^{x;v}_{r^*}(x',v')) \subseteq \widetilde{W}$ with $g^m(x', r'Lv') = 0$, again contradicting (27). Therefore (22) must hold true.

Let us now assume that $r^* \in \Lambda^{\texttt{entry}}(x,v)$ holds. Assume moreover that for a given $(x',v') \in \widetilde{U} \times \widetilde{V}$, we have $\rho^{x;v}_{r^*}(x',v') > \max_{j \in \mathcal{J}^{x,v,r^*}} \rho^{x;v}_{r^*,j}(x',v')$. Then for any given but fixed $j \in \mathcal{J}^{x,v,r^*}$ we have $g_j(x', \rho^{x;v}_{r^*}(x',v')Lv') \le g^m(x', \rho^{x;v}_{r^*}(x',v')Lv') = 0$. However if $g_j(x', \rho^{x;v}_{r^*}(x',v')Lv') = 0$, the fact that $\rho^{x;v}_{r^*}(x',v') \in \widetilde{W} \subseteq W_j$ implies by the equivalence of [36, Lemma 3.1] that $\rho^{x;v}_{r^*}(x',v') = \rho^{x;v}_{r^*,j}(x',v')$, yet by assumption $\rho^{x;v}_{r^*}(x',v') > \rho^{x;v}_{r^*,j}(x',v')$. Consequently, $g_j(x', \rho^{x;v}_{r^*}(x',v')Lv') < 0$. However, the latter inequality holds too whenever $j \notin \mathcal{J}^{x,v,r^*}$ since $\rho^{x;v}_{r^*}(x',v') \in W_j$ then too. Consequently, $\max_{j=1,\ldots,k} g_j(x', \rho^{x;v}_{r^*}(x',v')Lv') < 0$, which contradicts the definition of g^m.

Let us now assume that for a given $(x',v') \in \widetilde{U} \times \widetilde{V}$, we have $\rho^{x;v}_{r^*}(x',v') < \max_{j \in \mathcal{J}^{x,v,r^*}} \rho^{x;v}_{r^*,j}(x',v')$. Then there exists a $\iota \in \mathcal{J}^{x,v,r^*}$ such that $\rho^{x;v}_{r^*}(x',v') < \rho^{x;v}_{r^*,\iota}(x',v')$. Since by definition, $g_\iota(x', \rho^{x;v}_{r^*,\iota}(x',v')Lv') = 0$, we deduce that $g^m(x', \rho^{x;v}_{r^*,\iota}(x',v')Lv') \ge 0$. Now if $g^m(x', \rho^{x;v}_{r^*,\iota}(x',v')Lv') = 0$, from $\rho^{x;v}_{r^*,\iota}(x',v'), \rho^{x;v}_{r^*}(x',v') \in \widetilde{W}$, the equivalence (27) leads to the contradiction $\rho^{x;v}_{r^*,\iota}(x',v') = \rho^{x;v}_{r^*}(x',v')$. Consequently we must have: $g^m(x', \rho^{x;v}_{r^*,\iota}(x',v')Lv') > 0$. However we have $\rho^{x;v}_{r^*}(x',v') \in \Lambda^{\texttt{entry}}(x',v')$, which by continuity of g^m entails the existence of $r' \in (\rho^{x;v}_{r^*}(x',v'), \rho^{x;v}_{r^*,\iota}(x',v')) \subseteq \widetilde{W}$ with $g^m(x', r'Lv') = 0$, again contradicting (27). Therefore (22) must hold true.

Item 4, follows immediately from [5, Proposition 2.3.12] in case $r^* \in \Lambda^{\texttt{entry}}(x,v)$ and otherwise since the relation

$$-\rho^{x;v}_{r^*}(x',v') = \max_{j \in \mathcal{J}^{x,v,r^*}} -\rho^{x;v}_{r^*,j}(x',v'),$$

holds true. It allows us to derive:

$$\partial^{\mathrm{C}}(-\rho^{x;v}_{r^*}(x',v')) = \mathrm{Co}\left\{-\nabla_x \rho^{x;v}_{r^*,j}(x',v') : \, j \in \mathcal{J}^{x,v}(x',v')\right\}. \tag{28}$$

On the other hand, [5, Proposition 2.3.1] states that $\partial^{\mathsf{C}}(-\rho_{r^*}^{x;v}(x',v')) = -\partial^{\mathsf{C}}\rho_{r^*}^{x;v}(x',v')$ which allows us to derive (24) since $\mathrm{Co}(-A) = -\mathrm{Co}(A)$ for an arbitrary set $A \subseteq \mathbb{R}^n$. ◀

3.3 Observations regarding continuity

We can now present continuity results of two different forms. First, we can present local Lipschitzian continuity of the composition of $F_{\mathcal{R}}$ and boundary points whenever they are locally finite, i.e., Proposition 12 just below. This already allows us to move a significant step in the direction of showing that e is locally Lipschitzian. This would be a trivial consequence of Proposition 12, if all boundary points are finite (locally in x, but uniformly in v), e.g., when $\{z : g(x,z) \le 0\}$ is compact. This does not happen already when for a given $j = 1, \ldots, k$, g_j is (non-convex) quadratic in the second argument. The gap thus lies in how s, the total number of intervals affects e, and how behaviour is locally near an "infinite" boundary point. For this reason we are only able to show plain continuity of e, in Proposition 13 below, for the time being.

▶ **Proposition 12.** *Let Assumption 8 hold true. Then, for each $j = 1, \ldots, k$, and each $i = 1, \ldots, K_j$, the mappings $a_i^j, b_i^j : \mathcal{O}_j \to \mathbb{R}_+ \cup \{\infty\}$ appearing in the representation* (16) *of R_j, are continuous in the topology of the extended real line. Moreover for each $(x,v) \in \mathcal{O}_j$ at which $a_i^j(x,v) < \infty$ ($b_i^j(x,v) < \infty$ respectively), a_i^j is locally Lipschitzian at (x,v) (likewise b_i^j). Finally at any $(x,v) \in \mathcal{O}_j$, the mapping $(x,v) \mapsto F_{\mathcal{R}}(a_i^j(x,v))$ is locally Lipschitzian (likewise for b_i^j whenever it is finite).*

Proof. Under Assumption 8 we may employ Lemma 11, which shows that the mappings a_i^j, b_i^j are locally Lipschitzian whenever finite. This property is moreover preserved when combined with the continuously differentiable mapping $r \mapsto F_{\mathcal{R}}(r)$, see e.g., [5, Theorem 2.3.9 (ii)]. Convergence in the topology of the extended real line follows along the lines of [31,]. ◀

We have now prepared all elements to show continuity of e, a result that we will gather in the form of a proposition.

▶ **Proposition 13.** *Let Assumption 8 hold true. Then the mapping $e : U \times \mathbb{S}^{m-1} \to$ defined in* (9)*, can be represented as*

$$e(x,v) = \sum_{i=1}^{s(x,v)} \max\left\{F_{\mathcal{R}}(b_i(x,v)) - F_{\mathcal{R}}(a_i(x,v)), 0\right\}, \tag{29}$$

where $a_i, b_i : \mathcal{O} \to \mathbb{R}_+ \cup \{\infty\}$ are locally Lipschitzian functions (whenever finite) and moreover the mapping s is locally constant on $\mathcal{O}$ given by

$$\mathcal{O} := \bigcap_{j=1}^{k} \mathcal{O}_j, \tag{30}$$

which has $\lambda_n \otimes \mu_\zeta$ full-measure over $U \times \mathbb{S}^{m-1}$. As a consequence e is continuous on $\mathcal{O}$.

Proof. It is an immediate result of Assumption 8 that $\mathcal{O}$ as defined has $\lambda_n \otimes \mu_\zeta$ full-measure over $U \times \mathbb{S}^{m-1}$. As for the representation of e, we may once again assume w.l.o.g. that $k = 2$. Lemma 7 provides us with the following representation of R on $\mathcal{O}$:

$$R(x,v) = \bigcup_{i=1}^{s(x,v)} [a_i(x,v), b_i(x,v)], \tag{31}$$

where $a_i(x,v) = \max\left\{a_{i_1}^1(x,v), a_{i_2}^2(x,v)\right\}$ and $b_i(x,v) = \min\left\{b_{i_1}^1(x,v), b_{i_2}^2(x,v)\right\}$ for appropriate indices i_1, i_2. Now observe furthermore that $F_{\mathcal{R}}$ is an increasing function and consequently

$$F_{\mathcal{R}}(a_i(x,v)) = \max\left\{F_{\mathcal{R}}(a_{i_1}^1(x,v)), F_{\mathcal{R}}(a_{i_2}^2(x,v))\right\} \tag{32a}$$

$$F_{\mathcal{R}}(b_i(x,v)) = \min\left\{F_{\mathcal{R}}(b_{i_1}^1(x,v)), F_{\mathcal{R}}(b_{i_2}^2(x,v))\right\}, \tag{32b}$$

where the latter equality can be extended if $b_{i_1}^1(x,v) = b_{i_2}^2(x,v) = \infty$. By Assumption 8 and Lemma 7 the mapping s is locally constant. The intervals defining each R_j are disjoint, and consequently also this is true for the above representation. It thus follows that:

$$e(x,v) = \mu_{\mathcal{R}}(R(x,v)) = \sum_{i=1}^{s(x,v)} \mu_{\mathcal{R}}([a_i(x,v), b_i(x,v)]).$$

Now observe that $b_i(x,v) \le a_i(x,v)$ may happen, but in which case $\max\{F_{\mathcal{R}}(b_i(x,v)) - F_{\mathcal{R}}(a_i(x,v)), 0\} = 0 = \mu_{\mathcal{R}}([a_i(x,v), b_i(x,v)])$. The representation (29) has thus been shown. We may invoke Proposition 12 to entail continuity of e on $\mathcal{O}$. ◀

Now that continuity of e has been established, we can investigate local Lipschitz continuity of e. As mentioned, this will require a specific assumption on the potential nature of $b_{s(x,v)}(x,v)$:

▶ **Assumption 14.** *With $\mathcal{O}$ the open set of Proposition 13. For any $(x,v) \in \mathcal{O}$ at which $b_{s(x,v)}(x,v) = \infty$ we can find appropriate neighbourhoods on U', V' of (x,v) and $r > 0$ such that*

$$|F_{\mathcal{R}}(b_i(x_1,v')) - F_{\mathcal{R}}(b_i(x_2,v'))| \le r\|x_1 - x_2\|, \text{ for all } x_1, x_2 \in U' \text{ and } v' \in V',$$

where $i = s(\cdot,\cdot)$ and s is locally constant on $U' \times V'$.

It is worth mentioning that the above assumption holds under the following stronger, but simpler to verify assumption.

▶ **Assumption 15.** *With $\mathcal{O}$ the open set of Proposition 13. For any $(x,v) \in \mathcal{O}$ at which $b_{s(x,v)}(x,v) = \infty$ we can find appropriate neighbourhoods on U', V' of (x,v) such that $b_{s(x,v)}(x',v') = \infty$ for all $(x',v') \in U' \times V'$ and $s(x',v') = s(x,v)$.*

▶ Remark 16. Although Assumption 14 is formulated in the form involving the global structure and not each individual component, we can leverage the validity of Assumption 14 from each individual component. To that end we may assume, w.l.o.g., that $k = 2$. We may observe from the proof of Proposition 13, that any $(x,v) \in \mathcal{O}$, would belong to $\mathcal{O}_1$ and $\mathcal{O}_2$ and that should $b_{s(x,v)} = \infty$ hold, that if both components are active, that then $b^1_{i_1}(x,v) = \infty = b^2_{i_2}(x,v)$ holds. Then as a consequence of (32), and Assumption 14 holding for component 1 and 2, the mapping $F_{\mathcal{R}}(b_i(.,v'))$ is the minimum of two locally Lipschitzian mappings, and therefore locally Lipschitzian too.

▶ **Proposition 17.** *Let Assumption 8 and 14 hold true. Then the mapping $e : U \times \mathbb{S}^{m-1} \to$ defined in (9) is locally Lipschitzian in the first argument on $\mathcal{O}$.*

Proof. This follows as a result of the representation given in Proposition 13 (whenever b_i, a_i are finite), and combined with Assumption 14 (when $b_i(x,v) = \infty$). On the other hand, under Assumption 8, the same analysis can be done for almost every $x \in U$ and all $v \in \cap_j^k \mathcal{O}_j(x)$. ◀

4 Generalized differentiation

Now that we have established that the mapping e is locally Lipschitzian in the first argument, the main idea will be to take advantage of representation (8) of φ, by moving Clarke subdifferentiation through the integral. The following points make this not evident:

- Our analysis considers a set $\mathcal{O}$, of full-measure, but not containing strictly all points.
- The mappings ν_j, though uniformly integrable, are not bounded. In other words, elements of the subdifferential of e could become arbitrarily large.

We can thus not take advantage of classic results allowing the interchange of subdifferentiation and integration, but need to construct a novel analysis in order to justify this. This is the topic of the current section, even though it's main result Theorem 31 will not be surprising in form. Our first investigation, in Section 4.1 will consist in controlling the growth of these subdifferential elements of e. Then we will study certain Fubini-like formulæ that will allow us to represent the growth of the probability function φ in a given direction h as an integral. This will be achieved in Section 4.2. It is this representation that will allow us to justify the interchange of subdifferentiation and integration in Section 4.3 that also contains the main result.

4.1 Bounding growth

We will require an important assumption on the growth of the compositions $F_{\mathcal{R}} \circ a_i$ (respectively $F_{\mathcal{R}} \circ b_i$). For our results we will require the notion of uniform integrability, of which we recall the Definition:

▶ **Definition 18** ([2, Definition 4.5.1]). *A set of functions $\mathcal{F} \subseteq L^1(\mu)$ is said to be uniformly integrable if*

$$\lim_{\varepsilon \to 0+} \sup_{f \in \mathcal{F}} \int_{\{|f| > \varepsilon^{-1}\}} |f| \, \mathrm{d}\mu = 0.$$

We will use this result with regard to the atomless measure μ_ζ and so in our setting the just given definition of uniform integrability is equivalent to other definitions (e.g., [26, p. 133]) by [2, Proposition 4.5.3].

We can now present our key assumption, again formulated in terms of the individual component functions.

► **Assumption 19.** *With $\mathcal{O}$ given as a result of Assumption 8 and $(x,v) \in \mathcal{O}$ arbitrarily. For each $j = 1,\dots,k$, and each $i = 1,\dots,s_j(x,v)$, any $x^* \in \partial^{\mathsf{C}} F_{\mathcal{R}}(a_i^j(x,v)) \cup \partial^{\mathsf{C}} F_{\mathcal{R}}(b_i^j(x,v))$ satisfies:*

$$\|x^*\| \le \nu_j(x,v), \tag{33}$$

for a mapping $\nu_j : U \times \mathbb{S}^{m-1} \to \mathbb{R}_+ \cup \{\infty\}$ having the properties:

a. *$\nu_j(x,\cdot)$ is integrable over $\mathbb{S}^{m-1}$ w.r.t. μ_ζ for all $x \in U$*

b. *For all $x \in U$, $\nu_j(\cdot,v)$ is continuous at x for almost all $v \in \mathbb{S}^{m-1}$.*

c. *The family of functions $\{\nu_j(x,\cdot)\}_{x\in U}$ is uniformly integrable.*

As a result of these assumptions, this structure carries over to the bounds involved in the definition of R. This is the purpose of the following proposition.

► **Proposition 20.** *Let Assumptions 8, 14 and 19 hold and assume that R is represented as in* (20)*. Then there exists a mapping ν with properties analogous to those of Assumption 19, and such that* (33) *holds for $x^* \in \partial^{\mathsf{C}} F_{\mathcal{R}}(a_i(x,v)) \cup \partial^{\mathsf{C}} F_{\mathcal{R}}(b_i(x,v))$ and $i = 1,\dots,s(x,v)$.*

Proof. For the purpose of this proof, we may assume that $k = 2$. We first recall (32) resulting from the representation derived in Lemma 7. Since the subdifferential of the maxima and minima are included in the convex envelope of the subdifferentials (of the active elements) (see, e.g., [5, Proposition 2.3.12]), we have for $x^* \in \partial_x^{\mathsf{C}} F_{\mathcal{R}}(a_i(x,v))$ that

$$\|x^*\| \le \max_{\lambda\in[0,1]} \left\|\lambda \partial_x^{\mathsf{C}} F_{\mathcal{R}}(a_{i_1}^1(x,v)) + (1-\lambda)\partial_x^{\mathsf{C}} F_{\mathcal{R}}(a_{i_2}^2(x,v))\right\| \le \nu_1(x,v) + \nu_2(x,v), \tag{34}$$

so that upon defining $\nu(x,v) := \nu_1(x,v) + \nu_2(x,v)$ the desired mapping ν is found. The desired properties, highlighted in Assumption 19 evidently carry over to sums. The proof goes along similar lines for b_i, if we recall moreover Assumption 14, which if $b_i(x,v) = \infty$, also implies that the function $F_{\mathcal{R}}(b_i(x,v))$ is locally Lipschitzian, and if we also recall that the same calculus rules for maximum/minimum functions hold. ◄

We can now employ this mapping ν, to establish a key bound for the partial directional derivatives of the mapping e.

► **Lemma 21.** *In the setting of Proposition 20, there exists a constant $M > 0$, such that for all $(x,v) \in \mathcal{O}$ and all $h \in \mathbb{R}^n$, the following estimate holds true:*

$$|e'_x(x,v;h)| \le |e^\circ_x(x,v;h)| \le M\nu(x,v)\,\|h\|\,, \tag{35}$$

where e°_x and e'_x refer to the partial Clarke and Dini-Hadamard directional derivatives in direction h, that is,

$$e^\circ_x(x,v;h) := \limsup_{x'\to x\ s\to 0^+} \left(\frac{e(x'+sh,v) - e(x',v)}{s}\right). \tag{36}$$

$$e'_x(x,v;h) := \liminf_{s\to 0^+, h'\to h} \left(\frac{e(x+sh',v) - e(x,v)}{s}\right). \tag{37}$$

Furthermore, e°_x and e'_x are $\mathcal{B}(\mathcal{O}) \otimes \mathcal{B}(\mathbb{R}^n)$-measurable functions, and e°_x is upper-semicontinuous as a function of (x,h) (for v fixed) and sublinear in h (for x,v fixed), e'_x is lower-semicontinuous and positively homogeneous on h (for x,v fixed).

Proof. For an arbitrary $(x,v) \in \mathcal{O}$, the representation (29) holds true. Furthermore as a result of Proposition 17, the mapping e is locally Lipschitzian on $\mathcal{O}$ with respect to x and therefore the Clarke directional derivative e° is well defined, in particular near the (x,v) set aside. Moreover in this representation $s(x,v)$ (upon picking a sufficiently small neighbourhood of (x,v)) can be assumed to be constant, i.e., $s(x,v) = s \le K \le \prod_{j=1}^k K_j$, where K is as in (20) and K_j as in Lemma 7. Finally, in representation (29), should $b_s(x,v) = \infty$ occur, then as a result of Assumption 14, $F_{\mathcal{R}}(b_s(\cdot,v))$ is Lipschitz near x and therefore also the same is true of e as a finite sum of Lipschitzian functions. Setting $M := K$ and as a result of the sum rule ([5, p. 38]): for all $(x,v) \in \mathcal{O}$

$$\begin{aligned} e^\circ_x(x,v;h) &\le \sum_{i=1}^{s} \max\left\{\beta_i^\circ(x,v;h) + \alpha_i^\circ(x,-v;h), 0\right\} \\ &\le M\nu(x,v)\|h\| \text{ for all } h \in \mathbb{R}^n, \end{aligned} \tag{38}$$

where, for $i = 1, \dots, s$, $\beta_i(x,v) := F_{\mathcal{R}}(b_i(x,v))$ and $\alpha_i(x,v) := F_{\mathcal{R}}(a_i(x,v))$ and we have also used [5, Proposition 2.1.1(a) and (c)] together with Proposition 20. We furthermore emphasize that M does not depend on the chosen (x,v). Through the sublinearity of $e_x^\circ(x,v;\cdot)$ (see, e.g., [5, Proposition 2.1.1]), we obtain

$$0 = e_x^\circ(x,v;0) \le e_x^\circ(x,v;h) + e_x^\circ(x,v;-h),$$

which yields $-e_x^\circ(x,v;h) \le e_x^\circ(x,v;-h) \le M\nu(x,v)\|h\|$, where in the last inequality we used (38). Now, by [23, Theorem 8.18], we have that $e_x'(x,v;h) \le e_x^\circ(x,v;h)$, which ends the proof of (35).

Now, let us show the measurability of e_x°. Consider a countable enumeration of points with rational components $(u_k, s_k) \subseteq \mathbb{R}^n \times (0,+\infty)$ and a sequence $\varepsilon_j \downarrow 0^+$. Then, for $j,k \in \mathbb{N}$ we define the function from $\mathcal{O} \times \mathbb{R}^n$ to $\mathbb{R}$:

$$\Psi_{j,k}(x,v,h) := \begin{cases} \frac{e(x_k+s_k h,v)-e(x_k)}{s_k} & \text{if } \|x-x_k\|+s_k \le \varepsilon_j \\ -\infty & \text{otherwise,} \end{cases} \tag{39}$$

which is clearly measurable. We claim that for all $(x,v,h) \in \mathcal{O} \times \mathbb{R}^n$

$$e_x^\circ(x,v;h) = \inf_{j\ge 0} \sup_{k\ge 0} \Psi_{j,k}(x,v,h). \tag{40}$$

Indeed, consider $(x,v,h) \in \mathcal{O} \times \mathbb{R}^n$, since $\mathcal{O}$ is open we have that for small ε_j, the inequality $\|x - x_j\| \le \varepsilon_j$ implies that $(x_j, v) \in \mathcal{O}$, then by definition of the Clarke directional derivative equality $\ge$ holds in (40). Now, for the opposite one, we consider a sequence $x_l \to x$ and $s_l \to 0^+$ such that

$$e_x^\circ(x,v;h) = \lim_{l\to\infty} \frac{e(x_l+s_l h,v)-e(x_l,v)}{s_l}.$$

Moreover, since e is locally Lipschitz at (x,v), we can assume that e is K-Lipschitz on an open neighbourhood O of (x,v) and $(x_l,v),(x_l+s_l h,v) \in O$, then we can take (x_{k_l}, s_{k_l}) such that $\|x - x_{k_l}\| + s_{k_l} \to 0$ and

$$\left|\frac{e(x_l+s_l h,v)-e(x_{k_l}+s_{k_l}h,v)}{s_l}\right| + \left|\frac{e(x_l,v)-e(x_{k_l},v)}{s_l}\right| + |1-\frac{s_{k_l}}{s_l}| \le \frac{1}{l}.$$

Therefore,

$$\frac{e(x_l+s_l h,v)-e(x_l,v)}{s_l} \le \sup_{k\ge 0} \Psi_{j,k}(x,v,h) + 2/l$$

which implies that (40) holds.

Similarly, we can prove the measurability of e_x'. The upper-semicontinuity and sublinearity of the directional derivative e_x° follow from [5, Proposition 2.1.1], and the lower semicontinuity of e_x' follows from [23, Theorem 8.18]. ◀

4.2 Study of the variation of the probability function in a given direction

A target of this section is to establish Lemma 24 below. Indeed, we would like to be able to express $\varphi(\bar{x}+th)-\varphi(\bar{x})$ as a certain integral in $\mathbb{R}$. This representation will allow us to build a link between various directional derivatives of φ and similar expressions involving e. A step towards this result is the following Fubini-like theorem, which in spirit is close to [37, Lemma 4.2], but does require new arguments for its proof.

▶ **Lemma 22.** *In the setting of Proposition 20, there exists a neighbourhood U of $\bar{x}$ such that for $x \in U$ there exists $\gamma > 0$ such that for almost all $h \in \mathbb{B}$ the function*

$$[0,\gamma] \times \mathbb{S}^{m-1} \ni (s,v) \to \langle \nabla_x e(x+sh,v), h\rangle \tag{41}$$

is well defined (over a set of full measure) and integrable over $[0,\gamma] \times \mathbb{S}^{m-1}$ with respect to $\lambda_1 \otimes \mu_\zeta$. Moreover, for each $t \in [0,\gamma]$

$$\begin{aligned}\int_{[0,t]\times\mathbb{S}^{m-1}} \langle \nabla_x e(x+th,v), h\rangle \,\mathrm{d}(\lambda_1\otimes\mu_\zeta)(s,v) &= \int_{v\in\mathbb{S}^{m-1}} \left(\int_0^t \langle \nabla_x e(x+sh,v), h\rangle ds\right) \mathrm{d}\mu_\zeta(v) \\ &= \int_0^t \left(\int_{v\in\mathbb{S}^{m-1}} \langle \nabla_x e(x+sh,v), h\rangle \mathrm{d}\mu_\zeta(v)\right) \mathrm{d}\lambda_1(s).\end{aligned} \tag{42}$$

Finally, for almost all $v \in \mathbb{S}^{m-1}$ the section $s \to \langle \nabla_x e(x+sh,v), h\rangle$ is integrable over $[0,\gamma]$.

Proof.
▷ Claim 1. $\nabla_x e(x,v)$ exists for almost all $(x,v) \in U \times \mathbb{S}^{m-1}$ and it is measurable with respect to the completion associated with $\lambda_n \otimes \mu_\zeta$.

As a result of Proposition 17, the mapping e is locally Lipschitzian in the first argument on the set of $\lambda_n \otimes \mu_\zeta$-full measure $\mathcal{O}$. Hence, by [8, Korollar V.1.6], for μ_ζ allmost all $v \in \mathbb{S}^{m-1}$, $x \mapsto e(x,v)$ is locally Lipschitzian. Consequently as a result of Rademacher's Theorem (see, e.g., [23, Theorem 9.60]), for almost all $v \in \mathbb{S}^{m-1}$, the partial gradient $\nabla_x e(x,v)$ exists for almost all $x \in U$. Therefore, again by [8, Korollar V.1.6], we have that $\nabla_x e(x,v)$ exists for almost all $(x,v) \in U \times \mathbb{S}^{m-1}$. Finally, the measurability of $\nabla_x e(x,v)$ follows from the fact that it can be computed as a limit of measurable functions, whenever it exists, in the following form

$$\langle \nabla_x e(x,v), h\rangle = \lim_{j\to\infty} j^{-1}\left(e\left(x + \frac{h}{j}\right) - e(x)\right).$$

▷ Claim 2. Fix $x \in U$ and $\gamma > 0$ such that $\mathbb{B}_\gamma(x) \subseteq U$. Define the measure space $(\mathcal{X}, \mathcal{L}, \lambda_{n+1})$ given by $\mathcal{X} := [0,\gamma] \times \mathbb{B}$, the σ-algebra of Lebesgue measurable sets on $\mathcal{X}$ and λ_{n+1} the $n+1$-dimensional Lebesgue measure. Then, the function

$$\mathcal{X} \times \mathbb{S}^{m-1} \ni (t,h,v) \to \nabla_x e(x+th, v) \tag{43}$$

is measurable, and defined on a set of full measure with respect to $\lambda_{n+1} \otimes \mu_\zeta$.

Let us define $A = \{(t,h,v) : \nabla_x e(x+th,v) \text{ does not exists}\}$. Then, by Claim 1, for all $t \in [0,\gamma]$ the set

$$A(t) := \{(h,v) \in \mathbb{B} \times \mathbb{S}^{m-1} : \nabla_x e(x+th,v) \text{ does not exists}\} = \{(h,v) \in \mathbb{B} \times \mathbb{S}^{m-1} : (t,h,v) \in A\}$$

has null measure. Then, by [8, Korollar V.1.6], we have that the set A has null measure on $\mathcal{X} \times \mathbb{S}^{m-1}$. Consequently, the function defined in (43) is defined almost everywhere, and it is the composition of measurable functions, thanks to Claim 1.
▷ Claim 3. The function $(s,h,v) \to \langle \nabla_x e(x+sh,v), h\rangle$ is integrable over $\mathcal{X} \times \mathbb{S}^{m-1}$ with respect to the product measure $\lambda_{n+1} \otimes \mu_\zeta$.

The measurability of $(s,h,v) \to \langle \nabla_x e(x+sh,v), h\rangle$ comes from Claim 2. As a result of Lemma 21, for almost all $(s,h,v) \in \mathcal{X} \times \mathbb{S}^{m-1}$, we have

$$|\langle \nabla_x e(x+sh,v), h\rangle| \le e_x^\circ(x+sh,v;h) \le M\nu(x+sh,v). \tag{44}$$

Moreover, by Proposition 20 (and Assumption 19 regarding uniform integrability), we know that we may find $C > 0$ such that

$$\int_{v\in\mathbb{S}^{m-1}} \nu(x+sh,v)\,\mathrm{d}\mu_\zeta(v) \le C, \quad \text{as long as } x+sh \in U.$$

Hence, the use of Tonelli's Theorem (see, e.g., [6, Proposition 5.2.1]) gives the identity:

$$\begin{aligned}\int_{\mathcal{X}\times\mathbb{S}^{m-1}} |\langle \nabla_x e(x+th,v), h\rangle|\,\mathrm{d}\lambda_{n+1}\otimes\mu_\zeta &= \int_{\mathcal{X}}\left(\int_{\mathbb{S}^{m-1}} |\langle \nabla_x e(x+th,v), h\rangle|\mathrm{d}\mu_\zeta(v)\right)\mathrm{d}\lambda_{n+1}(t,h)\\ &\le \int_{\mathcal{X}}\left(\int_{\mathbb{S}^{m-1}} M\nu(x+sh,v)\mathrm{d}\mu_\zeta(v)\right)\mathrm{d}\lambda_{n+1}(t,h)\\ &\le C\lambda_{n+1}(\mathcal{X}) < \infty.\end{aligned}$$

▷ Claim 4. For almost all $(h,v) \in \mathbb{B} \times \mathbb{S}^{m-1}$ the section $s \to \langle \nabla_x e(x+sh,v), h\rangle$ is integrable over $[0,\gamma]$. Furthermore, for almost all $h \in \mathbb{B}$ the function defined on (41) is integrable over $[0,\gamma] \times \mathbb{S}^{m-1}$, and the identity (42) holds.

Now, by Fubini's Theorem (see, e.g., [6, Theorem 5.2.2]), on the one hand used over $\mathcal{X} \times \mathbb{S}^{m-1} = [0,\gamma] \times (\mathbb{B} \times \mathbb{S}^{m-1})$ we have that the section $s \to \langle \nabla_x e(x+sh,v), h\rangle$ is integrable. On the other hand using Fubini's Theorem over $\mathbb{B} \times ([0,\gamma] \times \mathbb{S}^{m-1})$, we have that for almost all $h \in \mathbb{B}$, the function $(s,v) \mapsto \langle \nabla_x e(x+sh,v), h\rangle$ defined in (41) is integrable over $[0,\gamma] \times \mathbb{S}^{m-1}$. Moreover, also Fubini's Theorem allows us to compute the integral over iterated integrals as in (42). ◀

In order to precisely justify the exchange of subdifferentiation and integration, we require further regularity captured in the following notion:

▶ **Definition 23.** *Let U be an open convex set, and let $\psi : U \times \mathbb{S}^{m-1} \to \mathbb{R}$ be given. Consider $\bar{x} \in U$, we say that ψ is absolutely continuous at $\bar{x}$ nearly uniformly in v provided that for almost all $h \in \mathbb{B}$ there exists $\delta_h > 0$ such that the function*

$$[0, \delta_h) \ni t \to \psi(\bar{x} + th, v)$$

is absolutely continuous for μ_ζ-almost all $v \in \mathbb{S}^{m-1}$.

Furthermore, we say that ψ is absolutely continuous near $\bar{x}$ nearly uniformly in v provided that there exists a neighbourhood U' of $\bar{x}$ such that ψ is absolutely continuous at x nearly uniformly in v for all $x \in U'$.

▶ **Lemma 24.** *In the setting of Proposition 20 suppose that e is absolutely continuous at $\bar{x}$ nearly uniformly in v. Then, for almost all $h \in \mathbb{B}$ there exists $\delta_{\bar{x},h} > 0$ such that for each $t \in [0, \delta_{\bar{x},h})$*

$$\varphi(\bar{x} + th) - \varphi(\bar{x}) = \int_0^t \left(\int_{v \in \mathbb{S}^{m-1}} \langle \nabla_x e(\bar{x} + sh, v), h \rangle \mathrm{d}\mu_\zeta(v) \right) \mathrm{d}s.$$

Proof. By Lemma 22 we can find $\gamma_{\bar{x}} > 0$ such that for almost all $h \in B_1$ (where B_1 is a set of full measure on $\mathbb{B}$) and for all $t \in [0, \gamma_{\bar{x}}]$ we have:

$$\int_0^t \left(\int_{\mathbb{S}^{m-1}} \langle \nabla_x e(\bar{x} + sh, v), h \rangle \, \mathrm{d}\mu_\zeta \right) \mathrm{d}\lambda_1 = \int_{\mathbb{S}^{m-1}} \left(\int_0^t \langle \nabla_x e(\bar{x} + sh, v), h \rangle \, \mathrm{d}\lambda_1 \right) \mathrm{d}\mu_\zeta.$$

Since e is absolutely continuous at $\bar{x}$, there exists a set $B_2 \subseteq B_1$ of full measure on $\mathbb{B}$, such that for all $h \in B_2$ there exists $\delta_h > 0$ such that for almost all $v \in \mathbb{S}^{m-1}$

$$\int_0^t \langle \nabla_x e(\bar{x} + sh, v), h \rangle \, \mathrm{d}s = e(\bar{x} + th, v) - e(\bar{x}, v). \tag{45}$$

We thus obtain that for almost all $h \in \mathbb{B}$ there exists $\delta_{\bar{x},h} > 0$ such that for all $t \in (0, \delta_{\bar{x},h})$

$$\begin{aligned}
\varphi(\bar{x} + th) - \varphi(\bar{x}) &= \int_{v \in \mathbb{S}^{m-1}} (e(\bar{x} + th, v) - e(\bar{x}, v)) \, \mathrm{d}\mu_\zeta(v) \\
&= \int_{v \in \mathbb{S}^{m-1}} \left(\int_0^t \langle \nabla_x e(\bar{x} + sh, v), h \rangle ds \right) \mathrm{d}\mu_\zeta(v) \\
&= \int_0^t \left(\int_{v \in \mathbb{S}^{m-1}} \langle \nabla_x e(\bar{x} + sh, v), h \rangle \mathrm{d}\mu_\zeta(v) \right) ds.
\end{aligned}$$

Where, we have used (45) and (42) of Lemma 22 to establish the last two equalities. ◀

To conclude this section let us briefly mention how absolute continuity carries over from the individual components to the composition.

▶ **Proposition 25.** *In addition to the Assumptions 8, 14 and 19 suppose that the functions $F_{\mathcal{R}} \circ a_j^i$ and $F_{\mathcal{R}} \circ b_j^i$ are absolutely continuous at $\bar{x}$ nearly uniformly in v. Then, e is absolutely continuous at $\bar{x}$ nearly uniformly in v.*

Proof. Let us first observe the following rather immediate consequences of the definition of absolute continuity (see [6, p. 135]):

- constant functions, in particular the all zero function are absolutely continuous
- affine combinations of absolutely continuous functions are absolutely continuous
- the maximum of two absolutely continuous functions is absolutely continuous since $|\max\{a, b\} - \max\{c, d\}| \leq \max\{|a - c|, |b - d|\}$. The same holds true for the minimum.

The result now follows by recalling Proposition 20, especially representation (29) and (32). ◀

4.3 Subdifferential estimates for the probability function

We now turn our attention to justifying the interchange of subdifferentiation and integration, thus producing an estimate of the subdifferential for the probability function φ. This result will be given in Propositions 20 and 29 below. This Proposition is of it's own interest, since it also provides other differential estimates of the probability function φ, in particular with respect to it's Fréchet subdifferential (inner inclusion). The main result of the paper, i.e., Theorem 31, becomes an immediate consequence of these Propositions after all preparatory material has been worked out. The following technical result is essential and studies a certain regularization of the partial Clarke and Dini-Hadamard directional derivatives of e.

▶ **Lemma 26.** *In the setting of Proposition 20, let us define* $\mathrm{d}_x e, \mathrm{D}_x e : U \times \mathbb{S}^{m-1} \times \mathbb{R}^n \to [-\infty, \infty]$ *as follows:*

$$\mathrm{D}_x e(u, v; h) := \begin{cases} \limsup\limits_{\substack{u' \to u\ h' \to h \\ (u',v) \in \mathcal{O}}} e_x^\circ(u', v; h') & \text{if } (u, v) \in \mathrm{cl}_x \mathcal{O} \\ -\infty & \text{otherwise.} \end{cases} \tag{46}$$

$$\mathrm{d}_x e(u, v; h) := \begin{cases} \liminf\limits_{\substack{u' \to u\ h' \to h \\ (u',v) \in \mathcal{O}}} e_x'(u', v; h') & \text{if } (u, v) \in \mathrm{cl}_x \mathcal{O} \\ +\infty & \text{otherwise.} \end{cases} \tag{47}$$

where $\mathrm{cl}_x \mathcal{O} := \{(x, v) \in U \times \mathbb{S}^{m-1} : \exists\, x_i \to x \text{ with } (x_i, v) \in \mathcal{O}\}$.

Then, the following assertions are all true:

a. $\mathrm{D}_x e$ *and* $\mathrm{d}_x e$ *are* $\mathcal{B}(U) \otimes \mathcal{L}(\mathbb{S}^{m-1}) \otimes \mathcal{B}(\mathbb{R}^n)$*-measurable functions, where* $\mathcal{L}(\mathbb{S}^{m-1})$ *is the Lebesgue completion of* $\mathcal{B}(\mathbb{S}^{m-1})$ *with respect to* μ_ζ*. Moreover,* $\mathrm{D}_x e$ *is upper semicontinuous as a function of* (x, h)*, and sublinear in* h*, and* $\mathrm{d}_x e$ *is lower semicontinuous as a function of* (x, h)*, and positively homogeneous in* h*.*

b. $\mathrm{D}_x e(u, v, h) = e_x^\circ(x, v, h)$ *for all* $(x, v, h) \in \mathcal{O} \times \mathbb{R}^n$

c. *For every* $x \in U$ *and all* $h \in \mathbb{R}^n$

$$\mathrm{D}_x e(x, v, h) \le M\nu(x, v)\|h\|, \text{ for almost all } v \in \mathbb{S}^{m-1}, \tag{48}$$

$$\mathrm{d}_x e(x, v, h) \ge -M\nu(x, v)\|h\|, \text{ for almost all } v \in \mathbb{S}^{m-1}, \tag{49}$$

where M *and* ν *are the constant and the function found in Lemma 21.*

d. *For almost all* $x \in U$*,*

$$\int_{v \in \mathbb{S}^{m-1}} \mathrm{D}_x e(x, v; h)\, \mathrm{d}\mu_\zeta(v) = \int_{v \in \mathbb{S}^{m-1}} e_x^\circ(x, v; h)\, \mathrm{d}\mu_\zeta(v), \text{ for all } h \in \mathbb{R}^n. \tag{50}$$

e. *The functions*

$$U \times \mathbb{R}^n \ni (x, h) \to \int_{v \in \mathbb{S}^{m-1}} \mathrm{D}_x e(x, v; h)\, \mathrm{d}\mu_\zeta(v) \in \mathbb{R} \tag{51a}$$

$$U \times \mathbb{R}^n \ni (x, h) \to \int_{v \in \mathbb{S}^{m-1}} \mathrm{d}_x e(x, v; h)\, \mathrm{d}\mu_\zeta(v) \in \mathbb{R} \tag{51b}$$

are well-defined and upper semi-continuous and lower semi-continuous, respectively.

Proof. Let us observe first that $\mathrm{cl}_x \mathcal{O}$ is a Borel measurable sub-set of $\mathcal{B}(U \times \mathbb{S}^{m-1})$, since we have the identity:

$$\mathrm{cl}_x \mathcal{O} = \bigcap_{k \in \mathbb{N}} \left(\mathcal{O} + k^{-1}\mathbb{B} \times \{0\}\right) \cap \left(U \times \mathbb{S}^{m-1}\right).$$

Now, we consider the extension of e_x° and e_x' to $U \times \mathbb{S}^{m-1} \times \mathbb{R}^n$ given by

$$\overline{e}_x(x, v; h) := \begin{cases} e_x^\circ(x, v; h) & \text{if } (x, v) \in \mathcal{O} \\ -\infty & \text{if } (x, v) \in U \times \mathbb{S}^{m-1} \backslash \mathcal{O}. \end{cases}$$

$$\underline{e}_x(x, v; h) := \begin{cases} e_x'(x, v; h) & \text{if } (x, v) \in \mathcal{O} \\ +\infty & \text{if } (x, v) \in U \times \mathbb{S}^{m-1} \backslash \mathcal{O}. \end{cases}$$

We claim that $\overline{e}_x$ is $\mathcal{B}(U) \otimes \mathcal{B}(\mathbb{S}^{m-1}) \otimes \mathcal{B}(\mathbb{R}^n)$-measurable. Indeed, take $\alpha \in \mathbb{R}$, so

$$G_\alpha := \{(u, v, h) \in U \times \mathbb{S}^{m-1} \times \mathbb{R}^n : \overline{e}_x(u, v; h) \ge \alpha\} = \{(u, v, h) \in \mathcal{O} \times \mathbb{R}^n : e_x^\circ(u, v; h) \ge \alpha\},$$

so due to the measurability of $\overline{e}_x(u, v; h)$, showed in Lemma 21, we get that $G_\alpha \in \mathcal{B}(\mathcal{O}) \times \mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathcal{O} \times \mathbb{R}^n)$. Now, we recall that $\mathcal{B}(\mathcal{O} \times \mathbb{R}^n) = \{A \cap (\mathcal{O} \times \mathbb{R}^n) : A \in \mathcal{B}(U \times \mathbb{S}^{m-1} \times \mathbb{R}^n)\}$, so $G_\alpha \in \mathcal{B}(U) \otimes \mathcal{B}(\mathbb{S}^{m-1}) \otimes \mathcal{B}(\mathbb{R}^n)$. Similarly, we can prove that $\underline{e}_x$ is $\mathcal{B}(U) \otimes \mathcal{B}(\mathbb{S}^{m-1}) \otimes \mathcal{B}(\mathbb{R}^n)$-measurable.

Then, the upper-semicontinuous closure of $\overline{e}_x$, which is given by the formula (46), is measurable with respect to $\mathcal{B}(U) \otimes \mathcal{L}(\mathbb{S}^{m-1}) \otimes \mathcal{B}(\mathbb{R}^n)$ and upper semicontinuous in the variables (x, h). In a similar fashion the lower semicontinuous closure of $\underline{e}_x$, which is given by the formula (47), is measurable with respect to

$\mathcal{B}(U) \otimes \mathcal{L}(\mathbb{S}^{m-1}) \otimes \mathcal{B}(\mathbb{R}^n)$ and lower semicontinuous in the variables (x, h). Moreover, sublinearity in h and/or positively homogeneity are inherited from e_x° and e_x', respectively.

To prove b, we recall that by Lemma 21, the function e° is upper semicontinuous at every $(x, v, h) \in \mathcal{O} \times \mathbb{R}^n$, and therefore at such points the upper-semicontinuous closure coincides with itself.

Now, let us show c. Let us fix $x \in U$ arbitrarily. By Lemma 21 (recall Assumption 19b) there exists a measurable set $S_x \subset \mathbb{S}^{m-1}$ of full-measure such that for all $v \in S_x$, $\nu(\,\cdot\,, v)$ is continuous at x. On the one hand if $(x, v) \notin \operatorname{cl}_x \mathcal{O}$, then the inequalities in (48) and (49) hold trivially. On the other hand, if $(x, v) \in \operatorname{cl}_x \mathcal{O}$, we consider a sequence $x_i \to x$. Then, by definition of $\mathrm{D}_x\, e$ and $\mathrm{d}_x\, e$ in (46) and (47), and recalling (35), indeed (48) and (49) hold, and this concludes the proof of c.

Since, $\mathrm{D}_x\, e(\,\cdot\,, \cdot\,; h) = e_x^\circ(\,\cdot\,, \cdot\,; h)$ are equal on $\mathcal{O}$, we have that the function

$$\Psi(x, v) := \sup_{h \in \mathbb{R}^n} \left| \mathrm{D}_x\, e(x, v; h) - e_x^\circ(x, v; h) \right|,$$

is a version of zero. Hence, Fubini's Theorem implies that for almost all $x \in U$, $\int_{v \in \mathbb{S}^{m-1}} \Psi(x, v)\, \mathrm{d}\mu_\zeta = 0$, which in particular implies that (50) holds.

As a result of (48) and (49), the functions defined in (51) are in principle, functions into the extended real lines $\mathbb{R} \cup \{-\infty\}$ and $\mathbb{R} \cup \{+\infty\}$, respectively. Let us show first the upper semicontinuity of the function (51a). By (48), we have that for all $(x, h) \in U \times \mathbb{R}^n$

$$\mathrm{D}_x\, e(x, v; h) - M\nu(x, v)\|h\| \leq 0, \text{ for almost all } v \in \mathbb{S}^{m-1} \tag{52}$$

In what follows we will refer to the mapping $M\nu(x, v)\|h\|$ as $\kappa(x, v, h)$.

Let $(x, h) \in U \times \mathbb{R}^n$ be arbitrary but fixed and consider arbitrary sequences $x_i \to x$ and $h_i \to h$ with $x_i \in U$. Then, by continuity in the first argument of ν and the fact that the family $\{\kappa(x_i, \cdot, h_i)\}_{i \geq 0}$ is uniformly integrable we obtain the following identity: (see, e.g., [6, Exercise 16, Chapter 4.2, p. 129])

$$\lim_{i \to \infty} \int_{\mathbb{S}^{m-1}} \kappa(x_i, v, h_i)\, \mathrm{d}\mu_\zeta(v) = \int_{\mathbb{S}^{m-1}} \kappa(x, v, h)\, \mathrm{d}\mu_\zeta(v).$$

Thus, by Fatou's Lemma

$$\begin{aligned}
&\limsup_{i \to \infty} \int_{\mathbb{S}^{m-1}} \mathrm{D}_x\, e(x_i, v; h_i) \mathrm{d}\mu_\zeta - \int_{\mathbb{S}^{m-1}} \kappa(x, v, h)\, \mathrm{d}\mu_\zeta \\
&\qquad = \limsup_{i \to \infty} \left(\int_{\mathbb{S}^{m-1}} \mathrm{D}_x\, e(x_i, v; h_i) - \kappa(x_i, v, h_i)\, \mathrm{d}\mu_\zeta(v) \right) \\
&\qquad \leq \int_{\mathbb{S}^{m-1}} \limsup_{i \to \infty} \left(e_x^\circ(x_i, v; h_i) - \kappa(x_i, v, h_i) \right) \mathrm{d}\mu_\zeta \\
&\qquad = \int_{\mathbb{S}^{m-1}} \limsup_{i \to \infty} \mathrm{D}_x\, e(x_i, v; h_i)\, \mathrm{d}\mu_\zeta - \int_{\mathbb{S}^{m-1}} \kappa(x, v, h)\, \mathrm{d}\mu_\zeta \\
&\qquad \leq \int_{\mathbb{S}^{m-1}} \mathrm{D}_x\, e(x, v; h)\, \mathrm{d}\mu_\zeta - \int_{\mathbb{S}^{m-1}} \kappa(x, v, h)\, \mathrm{d}\mu_\zeta,
\end{aligned}$$

where in the last inequality we have used that

$$\limsup_{i \to \infty} \mathrm{D}_x\, e(x_i, v; h_i) \leq \mathrm{D}_x\, e(x, v; h),$$

which results from the upper semicontinuity showed in part a. Consequently,

$$\limsup_{i \to \infty} \int_{\mathbb{S}^{m-1}} \mathrm{D}_x\, e(x_i, v; h_i)\, \mathrm{d}\mu_\zeta \leq \int_{\mathbb{S}^{m-1}} \mathrm{D}_x\, e(x, v; h)\, \mathrm{d}\mu_\zeta.$$

The proof of the lower semicontinuity of function (51b) follows similar arguments. Furthermore, equation (35) can be used to show that the functions defined in (51) are real-valued on a dense set of the domain, then by upper and lower semicontinuity, respectively, they must be real valued over the whole $U \times \mathbb{R}^n$. ◀

▶ **Definition 27.** *With notation as in Lemma 26, we will say that the upper integral qualification condition (UCQ) holds at $\bar{x}$ provided that for all $h \in \mathbb{R}^n$*

$$\int_{\mathbb{S}^{m-1}} \mathrm{D}_x\, e(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v) = \int_{\mathbb{S}^{m-1}} e_x^\circ(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v). \tag{53}$$

Similarly, we will say that the lower integral qualification condition (LCQ) holds at $\bar{x}$ provided that for all $h \in \mathbb{R}^n$

$$\int_{\mathbb{S}^{m-1}} \mathrm{d}_x\, e(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v) = \int_{\mathbb{S}^{m-1}} e'_x(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v). \tag{54}$$

► Remark 28. Here, it is important to notice that as a result of Proposition 17, if (19) holds at $\bar{x}$, then (UCQ) holds at $\bar{x}$, because in that case it can be proved that for all $h \in \mathbb{R}^n$, $\mathrm{D}_x\, e(\bar{x}, v; h) = e^\circ_x(\bar{x}, v; h)$ for almost all $v \in \mathbb{S}^{m-1}$. Moreover, due to (50) the upper integral qualification condition holds almost all $x \in U$.

On the other hand, if for almost all $v \in \mathbb{S}^{m-1}$ the function $e(\,\cdot\,, v)$ is continuously differentiable at $\bar{x}$, then (LCQ) holds at $\bar{x}$ To see that it is enough to notice that for almost all $v \in \mathbb{S}$, and $x_k \to x$, (close enough) we have $e'_x(x_k, v; h_k) = \langle \nabla_x e(x_k, v), v), h_k\rangle \to \langle \nabla_x e(\bar{x}, v), v), h\rangle$.

► **Proposition 29.** *In addition to the assumptions of Proposition 20 assume moreover that e is absolutely continuous at $\bar{x}$ (Definition 23) nearly uniformly in v. Then,*

$$\int_{v\in\mathbb{S}^{m-1}} \mathrm{d}_x\, e(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v) \le \varphi'(\bar{x}; h) \le \int_{v\in\mathbb{S}^{m-1}} \mathrm{D}_x\, e(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v), \ \textit{for all } h \in \mathbb{R}^n, \tag{55}$$

where, $\mathrm{D}_x\, e$ *and* $\mathrm{d}_x\, e$ *were defined in* (46) *and* (47)*. In addition suppose that*

1. *The upper integral qualification condition* (53) *holds at $\bar{x}$. Then, the following inclusion holds*

$$\partial^{\mathsf{F}} \varphi(\bar{x}) \subseteq \int_{v\in\mathbb{S}^{m-1}} \partial^{\mathsf{C}}_x e(\bar{x}, v)\, \mathrm{d}\mu_\zeta(v). \tag{56}$$

2. *The lower integral qualification condition* (54) *holds at $\bar{x}$. Then, the following inclusion holds*

$$\int_{v\in\mathbb{S}^{m-1}} \partial^{\mathsf{F}}_x e(\bar{x}, v)\, \mathrm{d}\mu_\zeta(v) \subseteq \partial^{\mathsf{F}} \varphi(\bar{x}). \tag{57}$$

Particularly, If $e(\,\cdot\,, v)$ is continuously differentiable at $\bar{x}$ for almost all $v \in \mathbb{S}^{m-1}$, then function φ is differentiable at $\bar{x}$ and

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

Proof. First, we will show that (55) holds in a set of full measure. By Lemma 24 we can take set $B_1 \subseteq \mathbb{B}$ of full measure that for all $h \in B_1$ there exists $\delta_{\bar{x},h} > 0$ such that for all $t \in (0, \delta_{\bar{x},h})$ we have that

$$\begin{aligned}
\varphi(\bar{x} + th) - \varphi(\bar{x}) &= \int_0^t \left(\int_{v\in\mathbb{S}^{m-1}} \langle \nabla_x e(\bar{x} + sh, v), h\rangle\, \mathrm{d}\mu_\zeta(v) \right) \mathrm{d}s \\
&\ge \int_0^t \int_{v\in\mathbb{S}^{m-1}} \mathrm{d}_x\, e(\bar{x} + sh, v; h) \mathrm{d}\mu_\zeta(v) ds =: F(t), \\
\varphi(\bar{x} + th) - \varphi(\bar{x}) &= \int_0^t \left(\int_{v\in\mathbb{S}^{m-1}} \langle \nabla_x e(\bar{x} + sh, v), h\rangle\, \mathrm{d}\mu_\zeta(v) \right) \mathrm{d}s \\
&\le \int_0^t \int_{v\in\mathbb{S}^{m-1}} \mathrm{D}_x\, e(\bar{x} + sh, v; h) \mathrm{d}\mu_\zeta(v) ds =: G(t),
\end{aligned}$$

where for the last inequalities we used that for almost all $(s, v) \in [0, t] \times \mathbb{S}^{m-1}$:

$$\mathrm{d}_x\, e(\bar{x} + sh, v; h) \le e'_x(\bar{x} + sh) = \langle \nabla_x e(\bar{x} + sh, v), h\rangle \le e^\circ_x(\bar{x} + sh, v; h) = \mathrm{D}_x\, e(\bar{x} + sh, v; h). \tag{58}$$

In order to derive (58), we first observe that $e'_x \le e^\circ_x$ as a result of the definition of the Clarke directional derivative (see, e.g., [5, p. 27]). The identity $e'_x(\bar{x} + sh) = \langle \nabla_x e(\bar{x} + sh, v), h\rangle$, results from (5) and Lemma 22 providing the information that this equality happens on a set of full measure. Finally, the leftmost inequality and the rightmost equality result from Lemma 26. Now since $G(0) = 0$, by dividing by $t > 0$, we obtain:

$$\frac{\varphi(x + th) - \varphi(x)}{t} \le \frac{G(t) - G(0)}{t}. \tag{59}$$

Now, let us analyse the function $G : [0, \delta_{\bar{x},h}) \to \mathbb{R}$. First, (48) and the arguments used in Lemma 22 allow us to establish that the function

$$s \to H(s) := \int_{v\in\mathbb{S}^{m-1}} \mathrm{D}_x\, e(\bar{x} + sh, v; h)\, \mathrm{d}\mu_\zeta(v)$$

is λ_1-integrable over $[0, \delta_{\bar{x},h})$ and consequently G is well-defined. Furthermore, we claim that

$$\limsup_{t\to 0^+} \left(\frac{G(t) - G(0)}{t} \right) \leq H(0). \tag{60}$$

Indeed, let $\varepsilon > 0$ be given. By Lemma 26, we have that the function H is upper semicontinuous, then there exists $\gamma > 0$ such that

$$H(s) \leq H(0) + \varepsilon, \text{ for all } s \in (0, \gamma).$$

Hence, $\left(\frac{G(t)-G(0)}{t}\right) \leq H(0) + \varepsilon$ for all $t \in (0, \gamma)$ and consequently, $\limsup_{t\to 0^+} \left(\frac{G(t)-G(0)}{t}\right) \leq H(0) + \varepsilon$, since $\varepsilon > 0$ was arbitrary, we conclude that (60) holds.

Now, we can take the $\liminf$ as $t \to 0^+$ and $h' \to h$ and combined with (59) to obtain:

$$\begin{aligned}\varphi'(\bar{x}, h) &= \liminf_{t\to 0^+, h'\to h} \left(\frac{\varphi(\bar{x} + th') - \varphi(\bar{x})}{t} \right) \leq \liminf_{t\to 0^+} \left(\frac{\varphi(\bar{x} + th) - \varphi(\bar{x})}{t} \right) \\ &\leq \liminf_{t\to 0^+} \left(\frac{G(t) - G(0)}{t} \right) \leq \limsup_{t\to 0^+} \left(\frac{G(t) - G(0)}{t} \right) \leq \int_{v\in\mathbb{S}^{m-1}} \mathrm{D}_x\, e(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v),\end{aligned}$$

where in the final inequality we used (60), which concludes (55) for all $h \in B_1$. Since, we have established inequality on a dense set B_1, then for an arbitrary point $h \in \mathbb{B}$, we can take $B_1 \ni h_k \to h$, so by lower and upper semicontinuity of the functions in (55) (see, e.g., [23, Theorem 8.18]), we get

$$\begin{aligned}\varphi(\bar{x}; h) &\leq \liminf_{k\to\infty} \varphi(\bar{x}; h_k) \leq \liminf_{k\to\infty} \int_{v\in\mathbb{S}^{m-1}} \mathrm{D}_x\, e(\bar{x}, v; h_k)\, \mathrm{d}\mu_\zeta(v) \\ &\leq \limsup_{k\to\infty} \int_{v\in\mathbb{S}^{m-1}} \mathrm{D}_x\, e(\bar{x}, v; h_k)\, \mathrm{d}\mu_\zeta(v) \leq \int_{v\in\mathbb{S}^{m-1}} \mathrm{D}_x\, e(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v)\end{aligned}$$

Now, since both functions are positively homogeneous we conclude the right-hand side inequality in (55). A similar analysis can be done for the function F to derive the left-hand side inequality in (55).

Finally, on the one hand, in case 1, i.e., (UCQ), we have that

$$\varphi'(\bar{x}; h) \leq \int_{v\in\mathbb{S}^{m-1}} e_x^\circ(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v), \text{ for all } h \in \mathbb{R}^n. \tag{61}$$

Therefore, in this case we can now reuse the arguments following [5, eq. (2) in the proof of Theorem 2.7.2] to establish the asserted inclusion (56). On the other hand, in case 2, i.e., (LCQ), we have

$$\int_{v\in\mathbb{S}^{m-1}} e_x'(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v) \leq \varphi'(\bar{x}; h), \text{ for all } h \in \mathbb{R}^n. \tag{62}$$

Then, consider $x^* \in \int_{v\in\mathbb{S}^{m-1}} \partial_x^{\mathsf{F}} e(\bar{x}, v)\, \mathrm{d}\mu_\zeta(v)$, following the definition of the Aumann integral, there exists a measurable selection $y^*(v) \in \partial_x^{\mathsf{F}} e(\bar{x}, v)$ a.e., such that $\int_{\mathbb{S}^{m-1}} y^*(v)\, \mathrm{d}\mu_\zeta(v) = x^*$, so for all $h \in \mathbb{R}^n$, we have that $\langle y^*(v), h\rangle \leq e_x'(\bar{x}, v; h)$, which by integration and by (62) imply that

$$\langle x^*, h\rangle \leq \varphi'(\bar{x}; h), \text{ for all } h \in \mathbb{R}^n,$$

which in turn implies that $x^* \in \partial^{\mathsf{F}} \varphi(\bar{x})$. ◀

The next results shows that when e is absolutely continuous, uniformly in v near $\bar{x}$, we can establish a stronger inclusion between the subdifferentials.

► **Proposition 30.** *In addition to the assumptions of Proposition 20 assume moreover that e is locally absolutely continuous near $\bar{x}$ nearly uniformly in v. Then, there exists a neighbourhood U' of $\bar{x}$ such that for all $x \in U'$*

$$\varphi^\circ(x; h) \leq \int_{v\in\mathbb{S}^{m-1}} \mathrm{D}_x\, e(x, v; h)\, \mathrm{d}\mu_\zeta(v), \text{ for all } h \in \mathbb{R}^n.$$

and φ is locally Lipschitz on U. Particularly, if the upper integral qualification condition (53) *holds at $x \in U'$, which happens almost everywhere, then the following inclusion holds.*

$$\partial^{\mathsf{C}} \varphi(x) \subseteq \int_{v\in\mathbb{S}^{m-1}} \partial_x^{\mathsf{C}} e(x, v)\, \mathrm{d}\mu_\zeta(v). \tag{63}$$

Proof. First, by Proposition 29 we have that (55) holds for all x close $\bar{x}$. Then, using (48) and the uniform integrability of ν, we can conclude that there exists $C > 0$ such that for all x close to $\bar{x}$

$$\varphi'(x, h) \leq C\|h\|, \text{ for all } h \in \mathbb{R}^n,$$

which by [23, Theorem 9.13 (c)] shows that φ is locally Lipschitz at $\bar{x}$. Second, by [23, Theorem 8.18] and [23, Exercise 7.3], we have that

$$\varphi^\circ(\bar{x}; h) = \sup_{\varepsilon \to 0} \left(\limsup_{x \to \bar{x}} \left[\inf_{h' \in \mathbb{B}(h,\varepsilon)} \varphi'(x; h') \right] \right).$$

In particular,

$$\varphi^\circ(\bar{x}; h) \leq \limsup_{x \to \bar{x}} \varphi'(x; h) \tag{64}$$

Then (64), (55) and Lemma 26 imply that

$$\begin{aligned} \varphi^\circ(x; h) &\leq \limsup_{x \to \bar{x}} \varphi'(x; h) \leq \limsup_{x \to \bar{x}} \int_{v \in \mathbb{S}^{m-1}} \mathrm{D}_x\, e(x, v; h)\, \mathrm{d}\mu_\zeta(v) \\ &\leq \int_{v \in \mathbb{S}^{m-1}} \mathrm{D}_x\, e(\bar{x}, v; h)\, \mathrm{d}\mu_\zeta(v). \end{aligned}$$

We can conclude by reusing the arguments following [5, eq. (2) in the proof of Theorem 2.7.2] to establish the asserted inclusion (63). ◀

We now gather all the previous results in the form of a formal Theorem, that carefully gathers the various assumptions.

▶ **Theorem 31.** *For $j = 1, \ldots, k$, let $g_j : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be given and assume that these mappings are continuously differentiable and satisfy Assumptions 8, 14 and 19. Assume moreover that the mapping e given in (9) is absolutely continuous near $\bar{x}$ nearly uniformly in v. Let the random vector $\xi \in \mathbb{R}^m$ be elliptically symmetrically distributed, with radial distribution function compatible with Assumption 19.*

Then the probability function $\varphi : \mathbb{R}^n \to [0, 1]$ given by $\varphi(x) = \mathbb{P}[g_j(x, \xi) \leq 0, j = 1, \ldots, k]$ is locally Lipschitzian at $\bar{x}$ and it holds

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

where e admits representation (29) and it's partial Clarke subdifferential can be evaluated from (29) through Clarke sum / chain rule and upon recalling Lemma 11.

5 Illustrative examples

We will now carefully illustrate how each of our assumptions hold when the mappings g_j composing g are of special structure.

5.1 Quadratic situation

Let $j \in \{1, \ldots, k\}$ be fixed but arbitrary and assume that the mapping g_j is given by the following formula:

$$g_j(x, z) = \frac{1}{2} z^\mathsf{T} Q_j(x) z + c_j(x)^\mathsf{T} z + d_j(x),$$

where Q_j, d_j, c_j are continuously differentiable functions from $\mathbb{R}^n$ to the respective appropriate spaces. Then we define furthermore the mapping A_j as follows

$$A_j(x) := L^\mathsf{T} \left(c_j(x) c_j(x)^\top - 2 d_j(x) Q_j(x) \right) L.$$

It is then first of all immediately observed that the constant K_j mentioned in Lemma 7 can be set as $K_j = 2$. Moreover let $\bar{x} \in \mathbb{R}^n$ and neighbourhood U of $\bar{x}$ be given such that:

$$\begin{aligned} &Q_j(x) \neq 0, \quad A_j(x) \neq 0, \quad d_j(x) \neq 0 \quad \text{and} \quad c_j(x) \neq 0 \\ &g_j(x, \mathfrak{m}) \neq 0, \end{aligned} \tag{65}$$

for all $x \in U$ holds true. Then we may define the set

$$\mathcal{O}_j = \left\{(x,v) \in U \times \mathbb{S}^{m-1} : v^\top L^\top Q_j(x) L v \neq 0 \vee c_j(x)^\top L v \neq 0 \vee v^\top A_j(x) v \neq 0\right\}.$$

This set has $\lambda_n \otimes \mu_\zeta$ full measure over $U \times \mathbb{S}^{m-1}$ as a result of using [37, Lemma 3.5]. Moreover the mapping s_j as defined in Lemma 7 can be seen to be locally constant on $\mathcal{O}_j$, when it is understood that $\mathcal{O}_j$ reunites the open conditions $C_1, \dots, C_7$ (see [37, Definition 3.1]) given in [37, p. 248]. We can even go further and observe that [37, Proposition 3.1] yields that at any solution r to $g_j(x, rLv) = 0$, we have $\langle \nabla_z g_j(x, rLv), Lv\rangle = \pm\sqrt{v^\top A_j(x) v}$. A necessary condition for such a solution to exist is $v^\top A_j(x) v > 0$, (the case $v^\top A_j(x) v = 0$ being excluded on $\mathcal{O}_j$). It thus follows that Assumption 8 can be shown to hold true for this situation.

Looking further into the conditions C_5, C_6, C_7, all being open, we can see that Assumption 15 also holds true.

Let us now focus on Assumption 19, which will require the identification of a given mapping ν_j. In order to support the construction here, we will require a mild growth condition, which is in fact a requirement for the radial density function $f_{\mathcal{R}}$. Let us assume that the latter is continuous and satisfies

$$\sup_{r \geq 0} r^2 f_{\mathcal{R}}(r) < \infty. \tag{66}$$

Then, we can define, for an appropriate constant $C > 0$,

$$\nu_j(x,v) = \begin{cases} \frac{C}{\sqrt{v^\top A_j(x) v}} & \text{if } v^\top A_j(x) v > 0, \\ +\infty & \text{if } v^\top A_j(x) v = 0, \\ 0 & \text{if } v^\top A_j(x) v < 0. \end{cases}$$

It follows from [37, Lemma 3.6] that for some neighbourhood $\widehat{U}$ of $\bar{x}$ and if constant C is large enough, that the function ν_j satisfies (33). It thus remains to verify if ν_j has the appropriate properties a)-c). As a result of continuity of A_j, it is clear that ν_j satisfies b), i.e., is continuous in the first argument. As for a) and c), these properties result from an application of [37, Proposition 3.2].

It is also worth mentioning that when the function g_j is quadratic in z and analytic in x, then each of the functions $F_{\mathcal{R}} \circ a_j^i$ and $F_{\mathcal{R}} \circ b_j^i$ are absolutely continuous at $\bar{x}$ nearly uniformly in v (see [37, Proposition 4.1]).

As a result of the just given investigation, we may present the following corollary to Theorem 31.

▶ **Corollary 32.** *For $j \in \{1, \dots, k\}$ assume that the mappings g_j are given by the following formula:*

$$g_j(x,z) = \frac{1}{2} z^\top Q_j(x) z + c_j(x)^\top z + d_j(x),$$

where Q_j, d_j, c_j are continuously differentiable and analytical functions. Let $\bar{x}$ along with a neighbourhood U be given such that (65) *holds true. Finally assume that the radial density function is continuous and satisfies the growth condition* (66)*.*

Then the probability function $\varphi : \mathbb{R}^n \to [0,1]$ given by $\varphi(x) = \mathbb{P}[g_j(x,\xi) \leq 0, j = 1, \dots, k]$ is locally Lipschitzian at $\bar{x}$ and it holds

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

where e admits representation (29) *and it's partial Clarke subdifferential can be evaluated from* (29) *through Clarke sum / chain rule and upon recalling Lemma 11.*

5.2 Linear case

Let $j \in \{1, \dots, k\}$ be fixed but arbitrary and assume that the mapping g_j is given by the following formula:

$$g_j(x,z) = c_j(x)^\top z - d_j(x),$$

where c_j, d_j are both continuously differentiable functions. Then in this case too we may carefully verify the stated assumptions of this work. First clearly, there can at most be one solution in r to $g(x, rLv) = 0$, so that

$K_j = 1$ can be set. We may pick $\bar{x}$ and a neighbourhood U of $\bar{x}$ in such a way that $c_j(x) \neq 0$ for all $x \in U$ and consequently define

$$\mathcal{O}_j = \left\{(x,v) \in U \times \mathbb{S}^{m-1} : c_j(x)^\mathsf{T} Lv \neq 0\right\}.$$

Then it turns out that $\mathcal{O}_j$ is of $\lambda_n \otimes \mu_\zeta$ full measure in $U \times \mathbb{S}^{m-1}$ as a result of combining [35, Lemma 2.2] and [8, Korollar V.1.6], which in fact state that the complement of $\mathcal{O}_j$ is a null set.

Moreover, as analyzed in [34], the mapping s_j is indeed locally constant on $\mathcal{O}_j$. The latter analysis is a result of identifying the nature of the set $\{r \geq 0 : g_j(x, rLv) \leq 0\}$, which is in fact an interval, potentially degenerate, of which the specific form relies on the sign of $d_j(x)$ and $c_j(x)^\mathsf{T} Lv$.

Since for an arbitrary r and $(x,v) \in U \times \mathbb{S}^{m-1}$, $\langle \nabla_z g_j(x, rLv), Lv\rangle = c_j(x)^\mathsf{T} Lv$, it thus becomes clear that Assumption 8 holds true. As for assumption 15, the above hinted upon analysis readily shows that this assumption is also true on $\mathcal{O}_j$.

Let us now focus on Assumption 19, which will require the identification of a given mapping ν_j. This identification will once again require a certain growth condition related to the radial density function, as well as continuity of the latter. For the sake of simplicity let us weaken (66) to the following form instead:

$$\lim_{r\to\infty} r^2 f_\mathcal{R}(r) = 0. \tag{67}$$

Then, we can define, for an appropriate constant $C > 0$, the mapping ν_j to be the constant C. Since μ_ζ is the uniform measure over the (compact) Euclidian unit sphere, this mapping readily satisfies conditions a)-c). As for (33), this is the result of [34, Lemma 3.3].

Finally, since both c_j and d_j are continuously differentiable in x, then on $\mathcal{O}_j$ so is $x \mapsto F_\mathcal{R}(a_1(x,v))$ or $x \mapsto F_\mathcal{R}(b_1(x,v))$. Consequently on $\mathcal{O}_j$ these mappings are locally Lipschitzian and therefore the same is true for $t \mapsto F_\mathcal{R}(b_1(\bar{x}+th, v))$ (likewise with a). The last map is thus absolutely continuous in x for nearly all v as required.

We can thus come to the following corollary:

► **Corollary 33.** *For $j \in \{1,\ldots,k\}$ assume that the mappings g_j are given by the following formula:*

$$g_j(x,z) = c_j(x)^\mathsf{T} z - d_j(x),$$

where d_j, c_j are continuously differentiable functions. Let $\bar{x}$ along with a neighbourhood U be given such that $c_j(x) \neq 0$ holds true. Finally assume that the radial density function is continuous and satisfies the growth condition (67).

Then the probability function $\varphi : \mathbb{R}^n \to [0,1]$ given by $\varphi(x) = \mathbb{P}[g_j(x,\xi) \leq 0, j = 1,\ldots,k]$ is locally Lipschitzian at $\bar{x}$ and it holds

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

where e admits representation (29) *and it's partial Clarke subdifferential can be evaluated from* (29) *through Clarke sum / chain rule and upon recalling Lemma 11.*

5.3 Convex case

Let $j \in \{1,\ldots,k\}$ be fixed but arbitrary and assume that the mapping g_j is continuously differentiable and convex in the second argument. Let us also assume that the mean vector $\mathfrak{m}$ of ξ is such that $g_j(\bar{x},\mathfrak{m}) < 0$ at $\bar{x}$, and in fact on an appropriate neighbourhood U of $\bar{x}$. In this particular situation, we may once more pick $K_j = 1$, since due to convexity, there can be at most one solution to $g(x, \mathfrak{m} + rLv) = 0$ for $x \in U$. Following [31], we can also set $\mathcal{O}_j = U \times \mathbb{S}^{m-1}$ and on $\mathcal{O}_j$, $s_j(x,v) = 1$ can be taken. Assumption 8 thus clearly holds true.

Moreover in this case, under an appropriate growth condition, involving $\nabla_x g_j$ and the radial density function, assumption 14 will turn out to hold true. Indeed, let $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing function such that

$$\lim_{r\to\infty} f_\mathcal{R}(r) r\theta(r) = 0. \tag{68}$$

Then assume that a neighbourhood U of $\bar{x}$ can be found, along with constants $C > 0$, δ, such that

$$\|\nabla_x g_j(x,z)\| \leq \delta\theta(\|z\|), \tag{69}$$

for all z such that $\|z\| \geq C$. Under this growth condition, Assumption 14 holds true as a result of [32, Corollary 3.5] (we can also refer to the gluing lemma in [34]).

In order to identify an appropriate integrability function ν_j, we may refer to [31, Theorem 3.10], showing that in fact we may pick a constant $C > 0$ large enough thanks to the just given growth condition. It thus follows that assumption 19 also holds true. Finally, the observation that $x \mapsto F_{\mathcal{R}}(b_1(x,v))$ is Lipschitz continuous (near $\bar{x}$), entails that the same is true for $t \mapsto F_{\mathcal{R}}(b_1(\bar{x}+th,v))$ and thus consequently that the mapping is absolutely continuous uniformly in all v.

The following corollary of Theorem 31 is just a restatement of [32, Theorem 3.6], but we provide it for completeness:

▶ **Corollary 34.** *For $j \in \{1,\ldots,k\}$ assume that the mappings g_j are continuously differentiable and convex in the second argument. Assume moreover that the mean vector $\mathfrak{m}$ of ξ and $\bar{x}$ are such tat $g_j(\bar{x},\mathfrak{m}) < 0$ and that moreover the growth condition* (69) *holds true for all $j = 1,\ldots,k$.*

Then the probability function $\varphi : \mathbb{R}^n \to [0,1]$ given by $\varphi(x) = \mathbb{P}[g_j(x,\xi) \leq 0, j = 1,\ldots,k]$ is locally Lipschitzian at $\bar{x}$ and it holds

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

where e admits representation (29) *and it's partial Clarke subdifferential can be evaluated from* (29) *through Clarke sum / chain rule and upon recalling Lemma 11.*

5.4 Concluding observations

As a result of our construction, we can now combine the various structures and obtain the following result, which by no means covers all possible cases exhaustively.

▶ **Corollary 35.** *For $j \in \{1,\ldots,k\}$ assume that the mappings g_j are of either one of the following forms:*

- *Quadratic as in Corollary 32.*
- *Linear as in Corollary 33*
- *Continuously differentiable and convex in the second argument.*

Let $\bar{x} \in \mathbb{R}^n$ be given and let a neighbourhood U of $\bar{x}$ be available such that for any given $j = 1,\ldots,k$, depending on the structure above, one of the followings holds:

- *in the quadratic case, for all $x \in U$, $Q_j(x) \neq 0$, $c_j(x) \neq 0$, $d_j(x) \neq 0$, $A_j(x) \neq 0$, $g_j(x,\mathfrak{m}) \neq 0$*
- *in the linear case, for all $x \in U$ $c_j(x) \neq 0$.*
- *in the convex case, for all $x \in U$, $g_j(x,\mathfrak{m}) < 0$.*

Let furthermore an increasing mapping $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ be given, such that $\theta(r) \geq r$ for all r sufficiently large and such that (68) *holds for the continuously differentiable radial density function $f_{\mathcal{R}}$. For $j = 1,\ldots,k$ associated with the convex case assume moreover that the growth condition* (69) *holds true.*

Then the probability function $\varphi : \mathbb{R}^n \to [0,1]$ given by $\varphi(x) = \mathbb{P}[g_j(x,\xi) \leq 0, j = 1,\ldots,k]$ is locally Lipschitzian at $\bar{x}$ and it holds

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

where e admits representation (29) *and it's partial Clarke subdifferential can be evaluated from* (29) *through Clarke sum / chain rule and upon recalling Lemma 11.*

6 Numerical Illustration

In order to show that the formulæ developed in this work can be concretely employed, we will place ourselves in the context of Corollary 32. The purpose of this numerical illustration will be to illustrate the usefulness of the formulæ, but we will not consider applications from practice. The main ingredient of the numerical illustration will be the probability constraint, all other elements will be relatively simple.

The problem will be of the following form:

$$\min_{x\in\mathbb{R}^n} \quad c^\mathsf{T} x$$

$$\text{s.t.} \quad \mathbb{P}\left[\frac{1}{2}\xi^\mathsf{T} Q_j(x)\xi + c_j(x)^\mathsf{T}\xi + d_j(x) \leq 0, j = 1,\dots,k\right] \geq p \tag{70}$$

$$\underline{x} \leq x \leq \bar{x}. \tag{71}$$

6.1 Description of the data

Concretely we will consider the following data:

$$Q_1(x) = \begin{bmatrix} 3(x_1-1) & -x_2 \\ -x_2 & 3(x_1-1) \end{bmatrix}, \quad Q_2(x) = \begin{bmatrix} -2x_2 & x-1-1 \\ x_1-1 & -2x_2 \end{bmatrix}$$

as well as

$$c_1(x) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}(x_1-1), \quad c_2(x) = \begin{bmatrix} 1 \\ 4 \end{bmatrix} x_2$$

$d_1(x) = -1, d_2(x) = -2$. We will also pick $p = 0.7$ together with $c = (-1,-1)$, $\underline{x} = (-2,-2)$, $\bar{x} = (2,2)$.

The random vector is taken to be multivariate Gaussian with mean vector 0 and covariance matrix

$$\Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

6.2 Implementation of the formulæ

We have implemented the formulæ for the gradient of the considered probability function in MatLab. The key ingredient for doing so is to make use of interval calculus. Indeed for a given $v \in \mathbb{S}^{m-1}$ and $x \in \mathbb{R}^n$, we can analytically compute and evaluate r such that $g_j(x, rLv) = 0$. This allows us to identify the union of intervals that make up R_j in (13) in a very simple way. The union of intervals R can be immediately constructed by employing (17) and upon simplifying the representation consecutively. In this way we can ensure that each interval making up the union is non-empty (and non-degenerate). The use of formula (29) then leads already to an evaluation of the value of the probability function. Furthermore the latter formula simplifies since the maximum operation is not required. With Corollary 32 firmly established we can indeed employ Lemma 11 to compute the contribution to the gradient of φ at this (x, v). A Monte-Carlo scheme of drawing elements on $\mathbb{S}^{m-1}$ allows us to obtain an estimate of both $\varphi(x)$ and $\nabla\varphi(x)$. We can moreover enhance the procedure by using antithetic drawing and by continuing to draw samples until the 99% confidence interval is smaller than a given threshold (or a maximum number of samples is exceeded). The last idea is close in spirit to what is in place in Genz' code for multivariate Gaussian distribution functions (e.g., [11]). When using a 5000 size sample, evaluating the probability and gradient takes no more than 3 seconds. The actual evaluation time depends on x, since the sample variance depends on x.

6.3 Solution algorithm

For the sake of simplicity we have implemented a projected gradient algorithm working on the penalized problem:

$$\min_{x\in[\underline{x},\bar{x}]} c^\mathsf{T} x + \bar{\mu}\left(\max\{p - \varphi(x), 0\}\right)^2. \tag{72}$$

In our case the projection onto the box is explicit. We have moreover used a fixed step size of $\rho = 0.001$, $\bar{\mu} = 50$. The "costly" step in the algorithm is the computation of the value and gradient of the probability function, both done "at once" and through the procedure described above. Since the problem is highly non-convex, and even the probability function quite steep we let the algorithm run for a maximum number of iterations and record the best feasible solution.

As an alternative to the procedure here, we can also rely on using sample average approximations of the problem (70). In the specific case of the here given data, this leads to a MILP with as many auxiliary binary variables as we pick scenarios. The resulting problem also relies on a "big-M" constant, which allows one to effectively deactivate the package of constraints for a given scenario. We observe that for a sample size of N, the resulting MILP has $2 + N$ variables, of which N are binary, and $2N + 1$ constraints. The resulting $2N + 1 \times N + 2$ constraint matrix is nonetheless sparse and has no more than $7N$ non zeros.

6.4 On the results

For the projected gradient algorithm we have set a maximum of 250 iterations. The algorithm identifies the vector $\widehat{x} = [1.0306, -0.0134)$ as the best found solution. The total computation time is around 316 seconds. We should of course account for the fact that this is a suboptimal implementation in MatLab.

The competing sample average problems where generated in MatLab and processed through CPLEX 12.8.0. Only the solution times in CPLEX are given and this time does not consider the time of generating or saving the problem to disk. For a sample size of $N = 1000$, CPLEX can solve the problem in roughly 8 seconds, but identifies $x = (0.385051, 1.488422)$ as the optimal solution. However the latter vector only has a probability value of around 0.01. We have solved a total of 10 versions of this problem (as they depend on the sample) and found little difference in CPU time. The found optimal vector x differs, but none is, in fact, feasible. When $N = 5000$, CPLEX manages, after 2 hours of computing, to find solutions not better than 1.82% optimality gap. The best found solution was identified to be $(0.687215, 1.160998)$, but once more this solution only has a probability value of around 0.06. This thus shows that a significantly larger sample size is needed. However the latter would not be solvable in a reasonable amount of time.

All together this instance shows that there can be great use of the here developed formulæ for more improved algorithms for optimization problems under probability constraints.

References

1 Denis Adelhütte, Dnis Aßmann, Tatiana Gonzàlez-Gradòn, Martin Gugat, Holger Heitsch, René Henrion, Frauke Liers, Sabrina Nitsche, Rüdiger Schultz, Michael Stingl, and David Wintergerst. Joint model of probabilistic (probust) constraints with application to gas network optimization. *Vietnam J. Math.*, Online:1–34, 2020.

2 Vladimir I. Bogachev. *Measure theory. Vol. I and II.* Springer, 2007.

3 Ingo Bremer, René Henrion, and Andris Möller. Probabilistic constraints via SQP solver: Application to a renewable energy management problem. *Comput. Manag. Sci.*, 12(3):435–459, 2015.

4 Alexander Bruhns, Gilles Deurveilher, and Jean-Sébastien. Roy. A non-linear regression model for mid-term load forecasting and improvements in seasonality, 2005. PSCC 2005 Luik.

5 Frank H. Clarke. *Optimisation and Nonsmooth Analysis.* Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1987.

6 Donald L. Cohn. *Measure theory.* Birkhäuser Advanced Texts. Basler Lehrbücher. Birkhäuser, second edition, 2013.

7 Darinka Dentcheva. Optimisation models with probabilistic constraints. In *Lectures on Stochastic Programming. Modeling and Theory*, volume 9 of *MPS/SIAM Series on Optimization*, pages 87–154. Society for Industrial and Applied Mathematics, 2009.

8 Jürgen Elstrodt. *Maß und Integrationstheorie.* Springer-Lehrbuch. Springer, 7th edition, 2011.

9 Kai-Tai Fang, Samuel Kotz, and Kai-Wang Ng. *Symmetric multivariate and related distributions*, volume 36 of *Monographs on Statistics and Applied Probability.* Chapman & Hall, 1990.

10 Josselin Garnier, Abdennebi Omrane, and Youssef Rouchdy. Asymptotic formulas for the derivatives of probability functions and their Monte Carlo estimations. *Eur. J. Oper. Res.*, 198:848–858, 2009.

11 Alan Genz. Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Stat.*, 1:141–149, 1992.

12 Claudia Gotzes, Holger Heitsch, René Henrion, and Rüdiger Schultz. On the quantification of nomination feasibility in stationary gas networks with random loads. *Math. Methods Oper. Res.*, 84(2):427–457, 2016.

13 Abderrahim Hantoute, René Henrion, and Pedro Pérez-Aros. Subdifferential characterization of continuous probability functions under Gaussian distribution. *Math. Program.*, 174(1-2):167–194, 2019.

14 Holger Heitsch. On probabilistic capacity maximization in a stationary gas network. *Optimization*, 69(3):575–604, 2020.

15 René Henrion. Optimierungsprobleme mit Wahrscheinlichkeitsrestriktionen: Modelle, Struktur, Numerik, 2016.

16 René Henrion and Werner Römisch. Lipschitz and differentiability properties of quasi-concave and singular normal distribution functions. *Ann. Oper. Res.*, 177:115–125, 2010.

17 Andrey Kibzun and Stanislav Uryasev. Differentiability of Probability function. *Stochastic Anal. Appl.*, 16:1101–1128, 1998.

18 Zinoviy M. Landsman and Emiliano A. Valdez. Tail Conditional Expectations for Elliptical distributions. *N. Am. Actuar. J.*, 7(4):55–71, 2003.

19 Kurt Marti. Differentiation formulas for probability functions: The transformation method. *Math. Program.*, 75(2):201–220, 1996.

20 Boris S. Mordukhovich. *Variational Analysis and Applications*. Springer Monographs in Mathematics. Springer, 2018.

21 András Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, 1995.

22 András Prékopa. Probabilistic programming. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*, pages 267–351. Elsevier, 2003.

23 R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer, 3rd edition, 2009.

24 Johannes O. Royset and Elijah Polak. Implementable algorithm for stochastic optimization using sample average approximations. *J. Optim. Theory Appl.*, 122(1):157–184, 2004.

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

26 Walter Rudin. *Real and complex analysis*. McGraw-Hill Series in Higher Mathematics. McGraw-Hill, 3rd edition, 1987.

27 Stanislav Uryasev. Derivatives of probability and Integral functions: General Theory and Examples. In *Encyclopedia of Optimization*, pages 658–663. Springer, 2nd edition, 2009.

28 Wim van Ackooij. A discussion of probability functions and constraints from a variational perspective. *Set-Valued Var. Anal.*, 28(4):585–609, 2020.

29 Wim van Ackooij, Ivana Aleksovska, and M. Munoz-Zuniga. (Sub-)Differentiability of probability functions with elliptical distributions. *Set-Valued Var. Anal.*, 26(4):887–910, 2018.

30 Wim van Ackooij, Irène Danti Lopez, Antonio Frangioni, Fabrizio Lacalandra, and Milad Tahanan. Large-scale Unit Commitment under uncertainty: an updated literature survey. *Ann. Oper. Res.*, 271(1):11–85, 2018.

31 Wim van Ackooij and René Henrion. Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions. *SIAM J. Optim.*, 24(4):1864–1889, 2014.

32 Wim van Ackooij and René Henrion. (Sub-)gradient formulae for probability functions of random inequality systems under Gaussian distribution. *SIAM/ASA J. Uncertain. Quantif.*, 5(1):63–87, 2017.

33 Wim van Ackooij, René Henrion, and Pedro Pérez-Aros. Generalized gradients for probabilistic/robust (probust) constraints. *Optimization*, 69(7-8):1451–1479, 2020.

34 Wim van Ackooij, Paul Javal, and Pedro Pérez-Aros. Derivatives of probability functions acting on parameter dependent unions of polyhedra. *Set-Valued Var. Anal.*, pages 1–33, 2021.

35 Wim van Ackooij and Jérôme Malick. Eventual convexity of probability constraints with elliptical distributions. *Math. Program.*, 175(1-2):1–627, 2019.

36 Wim van Ackooij and Pedro Pérez-Aros. Generalized differentiation of probability functions acting on an infinite system of constraints. *SIAM J. Optim.*, 29(3):2179–2210, 2019.

37 Wim van Ackooij and Pedro Pérez-Aros. Gradient formulae for nonlinear probabilistic constraints with non-convex quadratic forms. *J. Optim. Theory Appl.*, 185(1):239–269, 2020.

38 Wim van Ackooij and Pedro Pérez-Aros. Generalized differentiation of probability functions: parameter dependent sets given by intersections of convex sets and complements of convex sets. working paper, 2021.