Set-Valued and Variational Analysis (2022) 30:487–519
https://doi.org/10.1007/s11228-021-00598-w

Derivatives of Probability Functions: Unions of Polyhedra and Elliptical Distributions

Wim van Ackooij[1] · **Paul Javal**[2] · **Pedro Pérez-Aros**[3]

Received: 12 June 2020 / Accepted: 26 June 2021 / Published online: 17 July 2021

Abstract
In many practical applications models exhibiting chance constraints play a role. Since, in practice one is also typically interesting in numerically solving the underlying optimization problems, an interest naturally arises in understanding analytical properties, such as differentiability, of probability functions. However in order to build nonlinear programming methods, not only knowledge of differentiability, but also explicit formulæ for gradients are important. Unfortunately, differentiability of probability functions cannot be taken for granted. In this paper, motivated by applications from energy management, wherein we face a variety of non-linear transforms of underlying Elliptical distributions, we investigate probability functions acting on decision dependent union of polyhedra. Union of polyhedra naturally occur as soon as one approaches the components of "difference-of-convex" (DC) functions with their respective cutting plane models. In this work, we will establish that the probability functions are locally Lipschitzian and exhibit explicit formulæ for "the" Clarke sub-gradients, under very mild conditions. We also highlight, on a numerical example, that the formulæ can be put to use "in practice".

Keywords Probability functions · Sub-differentiability · Difference-of-convex

Mathematics Subject Classification (2010) MSC 90C15

✉ Wim van Ackooij
wim.van-ackooij@edf.fr

Paul Javal
paul.javal@edf.fr

Pedro Pérez-Aros
pedro.perez@uoh.cl

1 EDF R, D. OSIRIS, 7, Boulevard Gaspard Monge, F-91120 Palaiseau, France

2 EDF R, D. MIRE, 7, Boulevard Gaspard Monge, F-91120 Palaiseau, France

3 Instituto de Ciencias de la Ingeniería, Universidad de O'Higgins, Rancagua, Chile

1 Introduction

Handling uncertainty affecting constraints is a key feature of many optimization problems from practice. In a great deal of these applications, it is meaningful to consider the decision structure to be such that decisions have to be taken prior to observing uncertainty. As a consequence, the effect of both the decision and uncertainty vector impact the underlying inequality systems of the optimization problem at hand. It thus only becomes known a posteriori if the resulting inequalities are met. This raises the question of what meaning should be given to this underlying "random inequality system". Here, probabilistic constraints provide an intuitive tool. One requests that the underlying inequalities hold true with sufficiently large probability level. The requested probability level becomes a parameter that the user can adjust, within reasonable range, to reflect a certain risk aversion. Likely, because of the intuitive strength, probabilistic constraints have become very popular, which becomes apparent from the vast literature describing applications. We care to mention one such field: energy management. Therein one can be interested in "cost"-optimally handling various generation assets. Uncertainty manifests itself through intermittent generation (wind turbines, solar panels) and random fluctuating load. The decisions are subject to constraints that reflect to some degree physical restrictions, or even best practice rules (e.g., avoiding premature ageing of components). It is also frequent, within a certain scope, that some key decisions have to be taken ahead of time. As an example of a possible probabilistic constraint in this context, one can think of ensuring, with large probability, sufficient generation to meet load. We refer to [1–4] for various introductory texts on probabilistic constraints. We recall that in a general setting a probabilistic constraint is of the following form

$$\mathbb{P}[g(x,\xi) \leq 0] \geq p,$$

wherein $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ is a continuous map, p defines a confidence level, and ξ a random vector defined on an appropriate probability space. In current literature chance-constraints are either studied as "conventional" constraints $\varphi(x) \geq p$, with $\varphi(x) := \mathbb{P}[g(x,\xi) \leq 0]$ or by using scenario approaches, with a trade-off between numerical efficiency and approximation quality. Although in this work we will study probability functions from the "non-linear" programming viewpoint thus following the first path, we also briefly mention a wide variety of other approaches, mostly numerical, to handle probability functions. Indeed, popular methods for dealing with probabilistic constraints are sample-based approximations, e.g., [5–11] with various strenghtening procedures, e.g., [12] ; recent trends involve a combination of sampling and smoothing and exhibit promise, e.g., [13, 14] ; boolean approaches, e.g., [15, 16] ; p-efficient point based concepts, e.g., [17–23] ; robust optimization [24, 25] ; penalty approach [26] ; scenario approximation [27, 28] ; convex approximation [29] or yet other approximations [30–32].

In this paper, our motivation stems from an energy application. We are for instance interested in accounting for uncertainty jointly in load, wind and inflows. Then, as it turns out, the underlying inequality system on which the probabilistic constraint acts is of the "difference-of-convex" type in the argument represented by the random vector (see Example 1.1 for an explicit description). When approximating the underlying (convex) component functions with their cutting plane models, we actually end up with unions of polyhedra. Therefore, in this paper we will be interested in investigating (generalized) differentiability of probability functions $\varphi : \mathbb{R}^n \to [0,1]$ of the following form

$$\varphi(x) = \mathbb{P}[\xi \in \bigcup_{j=1}^{\ell} M_j(x)], \tag{1}$$

where $M_j(x) = \left\{z \in \mathbb{R}^m : A_j(x)z \le b_j(x)\right\}$, A_j, b_j are a continuous matrix and vector valued maps respectively and $\xi \in \mathbb{R}^m$ is assumed to be elliptically symmetrically distributed. We are interested in considering finite unions, i.e., $\ell < \infty$ is implicitly assumed.

Although [30] also consider probability functions and DC optimization, their approach is quite different. Indeed an approximation of the indicator function is constructed using a DC function. In this work however we fully study the probability function and its properties, thus providing quite a different approach.

The following example makes precise what happens when g is "difference-of-convex" in the second argument.

Example 1.1 Let us consider the following situation $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ where g admits the following representation

$$g(x,z) = g_1(x,z) - g_2(x,z), \tag{2}$$

where g_1 and g_2 are both convex in the second argument. We refer to such functions as DC in their second argument. From [33, Proposition 4.1], it is well known that finite maxima of DC functions are also DC, and we here recall an explicit DC formulation. Should $g(x,z) = \max_{i\in\mathcal{I}}(h_i(x,z)) = \max_{i\in\mathcal{I}}(h_{1,i}(x,z) - h_{2,i}(x,z))$ hold for a finite index set $\mathcal{I}$ with $h_{1,i}$, $h_{2,i}$ both convex in their second argument, then $g_1(x,z) = \max_{i\in\mathcal{I}}(h_{1,i}(x,z) + \sum_{j\in\mathcal{I}\setminus\{i\}} h_{2,j}(x,z))$ and $g_2(x,z) = \sum_{i\in\mathcal{I}} h_{2,i}(x,z)$ is a valid DC decomposition. As a consequence, we may restrict our attention to the situation wherein g is a scalar function without loss of generality.

Let us moreover assume that both g_1 and g_2 are polyhedral in the second argument and so uniformly (in the first). This assumption can be naturally encountered in the field of DC optimization where some algorithms make use of polyhedral approximation of the first and second components, as in "cutting-plane models". Then for some $\ell_1, \ell_2 \ge 1$ it follows:

$$g_1(x,z) = \max_{i=1,\dots,\ell_1} \left\{a_1^i(x) + \left\langle s_1^i(x), z\right\rangle\right\}, \tag{3}$$

and

$$g_2(x,z) = \max_{j=1,\dots,\ell_2} \left\{a_2^j(x) + \left\langle s_2^j(x), z\right\rangle\right\}, \tag{4}$$

for appropriate scalar and vector valued functions $x \mapsto a_k^i(x)$, $x \mapsto s_k^i(x)$, $i = 1, \dots, \ell_k$, $k = 1, 2$.

Now when looking at the constraint $g(x,z) \le 0$ and the impact on z, thanks to the previously exhibited structure, we can rewrite this as:

$$\min_{j=1,\dots,\ell_2} \max_{i=1,\dots,\ell_1} a_1^i(x) + \left\langle s_1^i(x), z\right\rangle - a_2^j(x) - \left\langle s_2^j(x), z\right\rangle \le 0, \tag{5}$$

or when introducing, for $j = 1, \dots, \ell_2$, the set-valued maps $M_j : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ defined as:

$$M_j(x) := \left\{z \in \mathbb{R}^m : (s_1^i(x) - s_2^j(x))^{\mathsf{T}} z \le a_2^j(x) - a_1^i(x), i = 1, \dots, \ell_1\right\} \tag{6}$$

it becomes clear that $\{z \in \mathbb{R}^m : g(x,z) \le 0\} = \bigcup_{j=1}^{\ell_2} M_j(x)$: a finite union of polyhedra.

It thus holds that

$$\mathbb{P}[g(x,\xi) \leq 0] := \mathbb{P}[\xi \in \bigcup_{j=1}^{\ell_2} M_j(x)],$$

which is exactly the structure we will investigate in this work.

In our targeted applications, the non-linear mapping φ of (1), will be part of a constraint of the form $\varphi(x) \geq p$, where $p \in [0,1]$ is a user defined safety level. Should one actually seek to solve such an application with numerical software, then one requires knowledge of properties of φ. Primordial, th0en would be knowledge of "first-order" information, but evidently also some numerical procedure to efficiently compute such information. Since understanding differentiability is relatively natural, it is not surprising that this property has indeed received significant attention. Differentiability, however, can not be asserted without additional regularity properties. One form of such a regularity condition is for instance the assumption that the set

$$K := \left\{z \in \mathbb{R}^m \ : g(x,z) \leq 0\right\}$$

is bounded near or at a point of interest $\bar{x}$. Alternatively one can assume that the density function of the random vector ξ has compact support. A condition of either form can be found in [34–42]. While the just given condition is one among other conditions needed to obtain smoothness of φ, this particular one is interesting to highlight as its relaxation brings more complexity to the analysis and may result in non-smoothness.

To continue our historical account, let us briefly mention that a first formula, in the form of an involved surface integral, for the gradient of a probability function was derived in [38]. It turned out that this formula could in fact be generalized by employing a combination of a surface and volume integral (e.g., [41, 42]). The transformation to a volume integral is interesting in the respect that it ensures that the integration domain will no longer depend on the decision vector (e.g., see [36], who was first to make this observation). As a result, the above mentioned papers provide gradient formulæ for relatively general classes of distributions. Although the results are very general, the mentioned surface/volume integrals are not immediate to transpose to numerical computer code. This is especially true when the mapping g is non-linear. We refer to [1, p. 207], [40, p.3], where this observation is also made. Furthermore, beyond this first point, the above stipulated compactness condition on K (e.g., [41, p. 200, Assumptions (A2)], [40, Assumption 2.2 (i)],[37, p. 902]) is also somewhat restrictive. For instance it rules out the consideration of distribution functions. Nonetheless, let us mention shortly [35, Remark 4.6] wherein an abstract integrability condition is suggested as replacement for the compactness condition. By taking a slightly different stance, and upon assuming the random vector ξ to belong to a specific class of distributions, it turns out that some of the "abstract conditions" just mentioned, can in fact be tied in with nominal properties of the underlying data. Such underlying properties can be related to verifiable conditions on the mapping g. The rich class of elliptically symmetric random vectors has turned out quite promising. In particular it has allowed a relaxation of the compactness condition previously mentioned. We refer to [43–49] for various works in this line of research. In these previously cited works, the inequality system g is assumed to be convex in the second argument. The pioneering works [39, 40] request rather a "star-shaped" structure and append a regularity condition (Assumption 2.2(iii) [40]) to avoid a certain degeneracy. It was also possible to allow x to be an element of an appropriate Banach space [47]. For as far as we are aware, the only work not requiring convexity in the second argument is [50]. The latter paper carefully discusses the potential issues caused by relaxing this requirement. In our situation too, we will relax this convexity request in the second argument. Our contribution therefore resides in being able to cover significantly more situations than our prior analyses allowed. We are also currently working on a different intermediate loosely related approach, [51] wherein we will be able to handle constraints that contain either the "convex" or "concave" part of a "DC" inequality, but not both. The extension to that setting, while not having polyhedrality as here, requires a completely different analysis from what is investigated here. Moreover the random vector ξ can be relatively arbitrary whereas it is elliptically symmetric here. In a situation of possible overlap however, the formulae presented in this work is more efficient computationally.

Let us also observe that the use of tools from non-smooth analysis provide a rather powerful methodology for the analysis of probabilistic constraints. They typically allow us to obtain sub-differential inclusions and even assert that the probability function itself is locally Lipschitzian. Additional "qualification" conditions then allow us to assert that in fact the probability function was smooth at a given trial point. We expect this to be the regular situation. In the current work we present a readily verifiable condition of such kind. The subdifferential inclusions can themselves also be put to use in deriving enlarged optimality conditions. Designing dedicated robust optimization algorithms for the general non-smooth and non-convex situation is an open research topic for future developments.

This work is organized as follows. Section 2 provides background material, carefully presenting the setting. In order to properly study (1), we will analyse in Section 3 the case where only one polyhedron is at hand, which amounts to setting $\ell = 1$. A novel analysis is required, since a previously made assumption on the position of the mean vector of ξ is relaxed. This entails the need to carefully partition the space into mutually disjoint open and closed sets. Re-uniting the results on the whole space requires in turn the study of specific limits.

Section 4 builds upon our study of the single polyhedral case to provide efficient estimates for the sub-differential of φ when dealing with a finite union of polyhedra.

Finally Section 5 highlights on a numerical example how the formulæ can be put to use in practice.

Our notation is essentially standard. We therefore mention only the use of ∂^{C} to denote Clarke's sub-differential.

2 Discussion of Background Material

2.1 Elliptical Distributions

Since elliptically symmetric random vectors play a vital role in our analysis, we will lay down in this section, not only the definition, but also some of the heavily employed results.

Definition 2.1 A random vector $\xi \in \mathbb{R}^m$ is said to be elliptically symmetrically distributed with mean $\mathfrak{m}$, covariance-like[1] matrix Σ and generator $\theta : \mathbb{R}_+ \to \mathbb{R}_+$, and this will be denoted by $\xi \sim \mathcal{E}(\mathfrak{m}, \Sigma, \theta)$, if and only if its density $f_\xi : \mathbb{R}^m \to \mathbb{R}_+$ is given by

$$f_\xi(z) = \big(\det \Sigma\big)^{-1/2}\theta\Big((z-\mathfrak{m})^{\mathsf{T}}\Sigma^{-1}(z-\mathfrak{m})\Big), \tag{7}$$

[1]It is the covariance matrix up to a scalar factor, see, e.g., Theorem 2.17 [52]

where the generator function $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ must satisfy
$\int_0^\infty t^{\frac{m}{2}}\theta(t)dt < \infty$.

The covariance-like matrix Σ will be assumed to be positive definite and throughout this manuscript the matrix L will designate the resulting Choleski factor, i.e., $\Sigma = LL^\mathsf{T}$. With this notation in place, we can show that ξ can be represented alternatively as follows:

$$\xi = \mathfrak{m} + \mathcal{R}L\zeta. \tag{8}$$

In this decomposition, the random vector ζ has a uniform distribution over the Euclidean m-dimensional unit sphere $\mathbb{S}^{m-1} := \{z \in \mathbb{R}^m : \sum_{i=1}^m z_i^2 = 1\}$. The random variable $\mathcal{R}$, which is independent of ζ, possesses a density, by virtue of ξ having one, and the corresponding density function has the following analytical form:

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

In our work we will make the mild blanket assumption that θ is continuous, which entails continuity of $f_\mathcal{R}$. Moreover, with $\mathcal{R}$ we associate the distribution function $F_\mathcal{R}$. We can observe that $F_\mathcal{R}(0) = 0$ holds true.

The just described family of elliptical symmetric random vectors is quite rich. As prominent examples, we can state that it contains Gaussian and Student random vectors, but also logistic or exponential power random vectors (see e.g. [52] and [53]). The Gaussian and Student random vectors fall in this class upon considering respectively the following generators

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

where Γ is the usual gamma-function.

2.2 Representing the probability function through the spherical radial decomposition

As a consequence of (8), for any Lebesgue measurable set $M \subseteq \mathbb{R}^m$ the following identity holds:

$$\mathbb{P}[\xi \in M] = \int_{v \in \mathbb{S}^{m-1}} \mu_\mathcal{R}\left(\{r \geq 0 : \mathfrak{m} + rLv \cap M \neq \emptyset\}\right) d\mu_\zeta(v), \tag{9}$$

where $\mu_\mathcal{R}$ and μ_ζ are the laws of $\mathcal{R}$ and ζ, respectively. Therefore, in particular for any $j = 1, ..., \ell$, it follows that

$$\mathbb{P}[\xi \in M_j(x)] = \int_{v \in \mathbb{S}^{m-1}} \mu_\mathcal{R}\left(\{r \geq 0 : r(A_j(x)Lv) \leq (b_j(x) - A_j(x)\mathfrak{m})\}\right) d\mu_\zeta(v), \tag{10}$$

and consequently (1) can be written in the following form:

$$\varphi(x) = \int_{v \in \mathbb{S}^{m-1}} \mu_\mathcal{R}\left(\{r \geq 0 : \exists j = 1, ..., \ell \ : r(A_j(x)Lv) \leq (b_j(x) - A_j(x)\mathfrak{m})\}\right) d\mu_\zeta(v). \tag{11}$$

There is an immediate keen advantage to representation (11) when it comes to numerical evaluation. As observed (e.g., [43, eq. (1.5)]), representation (11) allows for a reduction in sample variance when compared directly with sampling from the "native" representation (1).

2.3 Continuity of the probability function

Before discussing differentiability of the probability function at all, let us first try to come to insights regarding continuity. In our argumentation, we will require the upcoming classic result. Let us briefly recall that the result is easily shown through application of induction and elementary properties of measures, e.g., [54, Satz 1.7(b)].

Lemma 2.1 (Inclusion-Exclusion Formula) *Consider a finite measure space* $(\Theta, \mu, \mathcal{A})$ *and let a finite family* $B_i \in \mathcal{A}$, $i \in I$ *be given. Then:*

$$\mu(\bigcup_{i \in I} B_i) = \sum_{J \subseteq I, J \neq \emptyset} (-1)^{|J|+1} \mu(\bigcap_{j \in J} B_j). \tag{12}$$

We can thus already observe that φ in (11) can be expressed as a finite, involved, sum of terms akin to those of (10), involving various intersections of polyhedra.

For these latter polyhedra, i.e., for an arbitrary $J \subseteq \{1, ..., \ell\}$, we can assume without loss of generality that $M_J(x) := \left\{z \in \mathbb{R}^m \; : A_j(x)z \leq b_j(x),\, j \in J\right\}$ has non-empty interior. Indeed should that be true on a set of strictly positive measure of $\mathbb{R}^n$, using continuity arguments, this would hold on $\mathbb{R}^n$: we could discard the corresponding "empty" polyhedra from the union of polyhedra. We may thus assume that x is such that $M_J(x)$ has non-empty interior. Moreover, it is well known that $M_J(x) := \left\{z \in \mathbb{R}^m \; : A_j(x)z \leq b_j(x),\, j \in J\right\}$ satisfies the regularity condition:

$$\mathrm{bd}\left\{z \in \mathbb{R}^m \; : A_j(x)z \leq b_j(x),\, j \in J\right\} \subseteq \left\{z \in \mathbb{R}^m \; : \; \exists j \in J, \max\left\{A_j(x)z - b_j(x)\right\} = 0\right\},$$

whenever $M_J(x)$ is not empty, and none of the rows of $A_j(x)$ are the null vector. Observe under these assumptions that the set $\left\{z \in \mathbb{R}^m \; : A_j(x)z = b_j(x)\right\}$ is of at most dimension $m-1$ and thus a Lebesgue null set (in $\mathbb{R}^m$). Therefore, since ξ admits, by assumption, a density with respect to the Lebesgue measure, it follows, from well known arguments that $x \mapsto \mathbb{P}[\xi \in M_J(x)]$ is continuous (see, e.g., Propositions 2 and 3 [55]). It thus follows from Lemma 2.1 that φ is also continuous. It should be observed here that it is not immediate that φ possesses further regularity! To this end, let us briefly recall an example that we have already cited in many earlier works, but care to provide once more for the convenience of the reader.

Example 2.1 (Example 1 in [56, 57]) Let $\xi \sim \mathcal{N}(0, 1)$ and $B = \begin{pmatrix}1 & 1\end{pmatrix}^{\mathsf{T}}$ be given. Now observe that the identity $\varphi(x) = \mathbb{P}[B\xi \leq x] = \mathbb{P}[\xi \leq \min\{x_1, x_2\}]$ holds true. Observe also that this example fits the structure (1), by setting $\ell = 1$, $M_1(x) = \{z \in \mathbb{R} \; : Bz \leq x\}$. Since the mapping φ consists of the evaluation of the 1-dimensional Gaussian distribution function on $\min\{x_1, x_2\}$, it fails to be differentiable on the line $x_1 = x_2$.

2.4 Growth condition

Much like continuous differentiability that can not be taken for granted, it is also not necessarily true that φ is locally Lipschitzian. The latter, will require the somewhat involved analysis of Section 3, but also a certain growth condition that we will now introduce.

Definition 2.2 (First-order polynomial growth condition) Let $\xi \in \mathbb{R}^m$ be elliptically symmetrically distributed with associated radial density function $f_{\mathcal{R}} : \mathbb{R}_+ \to \mathbb{R}_+$. Then we say that ξ or $f_{\mathcal{R}}$ is compatible with the first order polynomial growth condition if and only if:

$$\lim_{r\to\infty} f_{\mathcal{R}}(r)r^2 = 0. \tag{13}$$

Observe also that the growth condition can be immediately expressed in terms of the generator:

$$\lim_{r\to\infty} r^{m+1}\theta(r^2) = 0.$$

Remark 2.1 Table 1 in [45] provides a large family of radial distributions compatible with this first order polynomial growth condition. When ξ is multivariate Gaussian random, the condition is satisfied. It also holds for multivariate Student random vectors ξ, whenever the related degrees of freedom ν is larger than 1.

3 Polyhedra in General Position

In order to work out the full situation considered in (1), let us first focus on the case $\ell = 1$, that is a probability function of the form

$$\varphi(x) = \mathbb{P}[g(x, \xi) \leq 0], \tag{14}$$

wherein

$$g(x, z) = A(x)z - b(x),$$

and ξ elliptical symmetric.

Before starting our analysis, let us briefly mention already well understood cases:

- The case of Eq. 14 around a point $\bar{x}$ with $g(\bar{x}, \mathfrak{m}) < 0$. This situation is studied in [44] and [45].
- The situation wherein A and b are constant, ξ is multivariate Gaussian, which are covered by [56, 58].
- The situation wherein A is of full rank locally around a point $\bar{x}$ [59], once more with ξ multivariate Gaussian.

The position of the polyhedra is characterized as *general* as we abandon the assumption $g(\bar{x}, \mathfrak{m}) < 0$, and only discard some troublesome cases at the margin that we will discuss.

This assumption was convenient, since it allowed us to assert a certain non-degeneracy of vital terms in our analysis. The general difficulty is well illustrated in [50] and we will thus require a fundamentally extended non-trivial analysis. This will be the topic of the current section. Prior to this topic, we care to mention that we make extensive use of the spherical-radial representation of ξ (8) which offers a didactic geometric interpretation. As described by (10), computing φ amounts to computing the spherical integral of measures along directions which are functions of v. It is readily seen that the integrand values are dependent on the position of the mean vector $\mathfrak{m}$, on direction v, and on the polyhedron state for a given x value. Our first steps will then be to carefully partition space accordingly. We will then analyse the variation of the integrand on each of the partition subsets, before providing a concluding result over the whole initial set.

3.1 Partitioning the space

A polyhedron is the intersection of hyperplanes. We'll start by looking at a single row $i = 1, ..., p$ of $A(x)$, where p denotes the number of rows of this matrix. Studying the behaviour of $g_i(x, \mathfrak{m}+rLv)$ will naturally lead to the partition of space of variables (x, v), as we show in this subsection.

Should a neighbourhood U of $\bar{x}$ be given on which for $x \in U$, $a_i(x)^\mathsf{T}\mathfrak{m} > b_i(x)$ holds true, then

$$\{r \geq 0 \ : g_i(x, \mathfrak{m} + rLv) \leq 0\} = \left[\frac{b_i(x) - a_i(x)^\mathsf{T}\mathfrak{m}}{a_i(x)^\mathsf{T}Lv}, \infty\right), \tag{15}$$

for all $v \in \mathbb{S}^{m-1}$ for which $a_i(x)^\mathsf{T}Lv < 0$. For all $v \in \mathbb{S}^{m-1}$, with $a_i(x)^\mathsf{T}Lv \geq 0$, it holds that $\{r \geq 0 \ : g_i(x, \mathfrak{m} + rLv) \leq 0\} = \emptyset$. Both facts are immediately checked through straightforward computation.

Should we be given a neighbourhood U of $\bar{x}$ on which, for $x \in U$, $a_i(x)^\mathsf{T}\mathfrak{m} < b_i(x)$ holds true, then

$$\{r \geq 0 \ : g_i(x, \mathfrak{m} + rLv) \leq 0\} = \left[0, \frac{b_i(x) - a_i(x)^\mathsf{T}\mathfrak{m}}{a_i(x)^\mathsf{T}Lv}\right], \tag{16}$$

for all $v \in \mathbb{S}^{m-1}$ for which $a_i(x)^\mathsf{T}Lv > 0$. For all $v \in \mathbb{S}^{m-1}$, with $a_i(x)^\mathsf{T}Lv \leq 0$, it holds that $\{r \geq 0 \ : g_i(x, \mathfrak{m} + rLv) \leq 0\} = [0, \infty)$.

Let us lay down formally our trial point of interest:

Assumption 1 Let $\bar{x} \in \mathbb{R}^n$ be given such that $a_i(\bar{x})^\mathsf{T}\mathfrak{m} \neq b_i(\bar{x})$ for all $i = 1, ..., p$.

Let us thus distinguish and separate the index set $\{1, ..., p\} = I^+ \cup I^-$ as well as fix the neighbourhood U of $\bar{x}$ with $a_i(x)^\mathsf{T}\mathfrak{m} < b_i(x)$ for all $i \in I^+$ and $x \in U$; $a_i(x)^\mathsf{T}\mathfrak{m} > b_i(x)$ for all $i \in I^-$ and $x \in U$. As already argued, we will assume the existence of $i = 1, ..., p$ such that $a_i(x)^\mathsf{T}\mathfrak{m} > b_i(x)$ holds on an appropriate neighbourhood U of the point of interest $\bar{x}$. This condition translates as I^- being not empty. It is still however possible for I^+ to be empty. Observe that we rule out that $\bar{x}$ is such that $a_i(\bar{x})^\mathsf{T}\mathfrak{m} = b_i(\bar{x})$ for some $i = 1, ..., p$. This condition, combined with definition of U, that requires the whole neighbourhood to verify strict inequalities, ensures the local stability with respect to x of our index sets I^- and I^+. Following the geometric interpretation, one could notice that I^- corresponds to the set of hyperplane indices that (strictly) do not include mean vector $\mathfrak{m}$ for $x \in U$; symmetrically, I^+ corresponds to hyperplane indices that strictly include $\mathfrak{m}$. Additionally, one could see that the intersection of the closed half-line emanating from $\mathfrak{m}$ and of direction v and the hyperplane defined by row i of A and b is intrinsically related to the sign of $a_i(x)^\mathsf{T}Lv$. Figure 1a provides a visualization of these index sets as well as the general context.

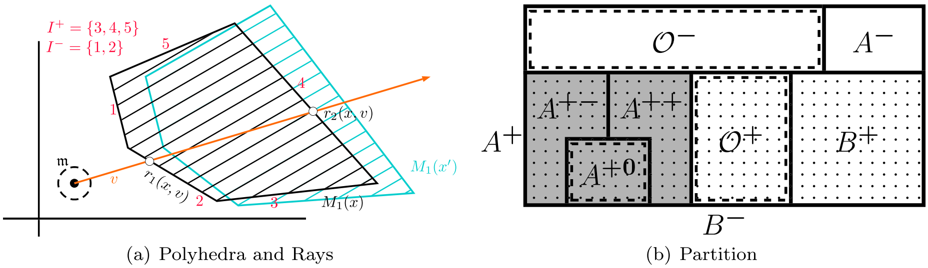

Fig. 1 **a** Illustration of the partition I^+, I^- as well as the resolvant mappings r_1, r_2 for a case where $(x, v) \in B^-$. **b** Illustration of the partition of $U \times \mathbb{S}^{m-1}$. The dotted set is B^-, while A^+ is the darker set. Sets with dashed inner linings are closed, the other ones being open

These observations will allow us to partition the set $U \times \mathbb{S}^{m-1}$ usefully, with the help of the following sets:

Definition 3.1 Let $\bar{x} \in \mathbb{R}^n$ be given as by Assumption 1 and let U be the neighbourhood of $\bar{x}$ on which the index sets I^-, I^+ are stable. Define the following subsets of $U \times \mathbb{S}^{m-1}$:

$$B^- = \left\{(x, v) \in U \times \mathbb{S}^{m-1} \ : a_i(x)^\mathsf{T} Lv < 0, \ \forall i \in I^-\right\}, \tag{17a}$$

$$\mathcal{O}^- = \left\{(x, v) \in U \times \mathbb{S}^{m-1} \ : \ a_i(x)^\mathsf{T} Lv \leq 0, \ \forall i \in I^- \ \wedge \ \exists i \in I^- \mid a_i(x)^\mathsf{T} Lv = 0\right\} \tag{17b}$$

$$A^- = \left\{(x, v) \in U \times \mathbb{S}^{m-1} \ : \ \exists i \in I^- \mid a_i(x)^\mathsf{T} Lv > 0\right\}, \tag{17c}$$

$$B^+ = \left\{(x, v) \in B^- \ : a_i(x)^\mathsf{T} Lv < 0, \ \forall i \in I^+\right\} \tag{17d}$$

$$\mathcal{O}^+ = \left\{(x, v) \in B^- \ : \ a_i(x)^\mathsf{T} Lv \leq 0, \ \forall i \in I^+ \ \wedge \ \exists i \in I^+ \mid a_i(x)^\mathsf{T} Lv = 0\right\} \tag{17e}$$

$$A^+ = \left\{(x, v) \in B^- \ : \ \exists i \in I^+ \mid a_i(x)^\mathsf{T} Lv > 0\right\} \tag{17f}$$

The following lemma shows the structure of Definition 3.1and backs up Fig. 1b

Lemma 3.1 *With notation as in Definition 3.1, it holds that the sets $B^+, \mathcal{O}^+, A^+$ form a partition of B^- (i.e., are mutually disjoint and their union makes up the whole set). Furthermore the sets $B^+, \mathcal{O}^+, A^+, \mathcal{O}^-, A^-$ form a partition of $U \times \mathbb{S}^{m-1}$.*

Proof It is clear from the definition that B^+, $\mathcal{O}^+$, A^+ are mutually disjoint, even when I^+ is empty in which case $\mathcal{O}^+$, A^+ are empty and $B^+ = B^-$. Otherwise, for an arbitrary $i \in I^+$ and an arbitrary $(x, v) \in B^-$, we must have $a_i(x)^\mathsf{T} Lv < 0$, $a_i(x)^\mathsf{T} Lv = 0$ or $a_i(x)^\mathsf{T} Lv > 0$. In the third case $(x, v) \in A^+$. We may thus assume that $a_i(x)^\mathsf{T} Lv \leq 0$ for all $i \in I^+$, in which case should there be i with $a_i(x)^\mathsf{T} Lv = 0$, we deduce $(x, v) \in \mathcal{O}^+$, if not, it follows that $(x, v) \in B^+$.

For the second claim it now suffices to show that B^-, $\mathcal{O}^-$, A^- form a partition of $U \times \mathbb{S}^{m-1}$. The argumentation is analogous to the one just given, except that we recall that I^- is assumed to not be empty. □

For any $i = 1, ..., p$, we can now define $\mathfrak{r}_i : B^- \to \mathbb{R} \cup \{-\infty, \infty\}$ as

$$\mathfrak{r}_i(x, v) = \frac{b_i(x) - a_i(x)^\mathsf{T} \mathfrak{m}}{a_i(x)^\mathsf{T} Lv}, \tag{18}$$

where division "by zero" is interpreted as leading to $\pm\infty$. Our first observation is that for $i \in I^-$, and $(x, v) \in B^-$, $\mathfrak{r}_i(x, v) \in (0, \infty)$ always. Furthermore, since B^- is open, $\mathfrak{r}_i$ is evidently continuously differentiable on this set. For any $i \in I^+$ and any $(x, v) \in A^+ \cup B^+$, we observe that $\mathfrak{r}_i(x, v)$ is finite valued. Moreover $\mathfrak{r}_i$ is continuously differentiable on $A^+ \cup B^+$, the latter being an open set.

With the help of these mappings, we now define $r_1 : B^- \to \mathbb{R}_+$ as

$$r_1(x, v) = \max_{i \in I^-} \mathfrak{r}_i(x, v) = \max_{i \in I^-} \frac{b_i(x) - a_i(x)^\top \mathfrak{m}}{a_i(x)^\top L v} > 0. \tag{19}$$

Since r_1 is the maximum of continuously differentiable maps on the open set B^-, it is locally Lipschitz there. Figure 1a shows how at a specific point (x, v), $r_1(x, v) = \mathfrak{r}_2(x, v)$, $\mathfrak{r}_1(x, v)$ being strictly smaller at this (x, v).

We can also define $r_2 : B^- \to \mathbb{R}_+ \cup \{\infty\}$ as follows:

$$r_2(x, v) = \begin{cases} \sup_t & t \\ \text{ s.t. } & t a_i(x)^\top L v \leq b_i(x) - a_i(x)^\top \mathfrak{m}, i \in I^+ \end{cases}$$

It now follows that $\mathcal{D}om(r_2) = A^+$ and with $\mathcal{I}(x, v) = \left\{ i \in I^+ \ : a_i(x)^\top L v > 0 \right\}$ (being locally stable), it follows that $r_2(x, v) = \min_{i \in \mathcal{I}(x,v)} \mathfrak{r}_i(x, v) = \min_{i \in \mathcal{I}(x,v)} \frac{b_i(x) - a_i(x)^\top \mathfrak{m}}{a_i(x)^\top L v} > 0$ and this mapping is once more locally Lipschitzian on its domain (being open). This follows from Lemma 3.1 [44], when considering a reduced inequality system to the index set I^+.

From a practical point of view, one could have the following interpretations:

- at a given $(x, v) \in B^-$, $r_1(x, v)$ is the largest distance $r \geq 0$ on the half-line from $\mathfrak{m}$ in direction v before reaching the intersection of all hyperplanes defined by rows of I^-.
- At a given $(x, v) \in A^+$, $r_2(x, v)$ is the smallest distance $r \geq 0$ on the half-line from $\mathfrak{m}$ in direction v before leaving any hyperplane defined by rows of I^+.

Now it is readily visible in (11) that the value of the integrand depends on the order between $r_1(x, v)$ and $r_2(x, v)$. Depicted in Fig. 1a is the particular case where $r_1(x, v) < r_2(x, v)$ but obviously this does not hold in general. We may therefore partition further the set A^+ as follows:

$$A^{+-} = \left\{ (x, v) \in A^+ \ : r_1(x, v) > r_2(x, v) \right\} \tag{20a}$$

$$A^{+0} = \left\{ (x, v) \in A^+ \ : r_1(x, v) = r_2(x, v) \right\} \tag{20b}$$

$$A^{++} = \left\{ (x, v) \in A^+ \ : r_1(x, v) < r_2(x, v) \right\} \tag{20c}$$

$$\tag{20d}$$

and it follows, from the previous observations, that A^{+-}, A^{++} are both open (since r_1, r_2 are continuous). Based on this partition, we will now provide an explicit representation of the integrand occurring in the spherical radial representation (10) that will be called e.

3.2 Studying and representing the inner radial probability function: *e*

We now define the mapping $e : U \times \mathbb{S}^{m-1} \to [0, 1]$ as

$$e(x, v) = \mu_{\mathcal{R}} \left(\{ r \geq 0 : r(A(x) L v) \leq (b(x) - A(x)\mathfrak{m}) \} \right), \tag{21}$$

and recall that $\varphi(x) = \int_{v \in \mathbb{S}^{m-1}} e(x, v) d\mu_\zeta(v)$ as a result of (10).

Due to the preparatory material above, we can now provide a full partition-wise description of e:

Lemma 3.2 *The mapping* $e : U \times \mathbb{S}^{m-1} \to [0, 1]$ *specified in* (21) *is also given by:*

$$e(x, v) = \begin{cases} 0 & \text{if } (x, v) \in A^- \cup \mathcal{O}^- \cup A^{+-} \cup A^{+0} \\ 1 - F_{\mathcal{R}}(r_1(x, v)) & \text{if } \quad (x, v) \in B^+ \cup \mathcal{O}^+ \\ F_{\mathcal{R}}(r_2(x, v)) - F_{\mathcal{R}}(r_1(x, v)) & \text{if } \quad (x, v) \in A^{++} \end{cases} . \tag{22}$$

Proof It holds:

$$\{r \geq 0 \ : g(x, \mathfrak{m} + rLv) \leq 0\} = \bigcap_{j=1}^{p} \left\{r \geq 0 \ : g_j(x, \mathfrak{m} + rLv) \leq 0\right\}. \tag{23}$$

Observe that on $\mathcal{O}^- \cup A^-$, there exists $j \in I^-$ for which $\left\{r \geq 0 \ : g_j(x, \mathfrak{m} + rLv) \leq 0\right\} = \emptyset$. Hence, $e(x, v) = 0$ on these sets. Moreover for all $(x, v) \in B^-$ and $j \in I^-$, (15) holds true. It thus follows that

$$\bigcap_{j \in I^-} \left\{r \geq 0 \ : g_j(x, \mathfrak{m} + rLv) \leq 0\right\} = [r_1(x, v), \infty).$$

For $(x, v) \in \mathcal{O}^+ \cup B^+$, $r_2(x, v) = \infty$ and hence

$$\{r \geq 0 \ : g(x, \mathfrak{m} + rLv) \leq 0\} = \bigcap_{j \in I^-} \left\{r \geq 0 \ : g_j(x, \mathfrak{m} + rLv) \leq 0\right\} = [r_1(x, v), \infty),$$

so that $e(x, v) = 1 - F_{\mathcal{R}}(r_1(x, v))$.

Finally on A^+, (16) holds true and recalling also Lemma 3.2 [44], we have

$$\bigcap_{j \in I^+} \left\{r \geq 0 \ : g_j(x, \mathfrak{m} + rLv) \leq 0\right\} = [0, r_2(x, v)).$$

It thus remains to work out the intersection of $[r_1(x, v), \infty)$ and $[0, r_2(x, v)]$, which is the interval $[r_1(x, v), r_2(x, v)]$ on $A^{++} \cup A^{+0}$ (reduced to a singleton on the last set) and empty on A^{+-}. The proof is thus complete. □

3.3 Useful partition-wise limits

The main focus of the lemmas in this section is to characterize the asymptotic behaviour of the sub-differential of e over certain subsets of our partition. These lemmas will be useful to go from an initial subset-wise study to the unified result over $U \times \mathbb{S}^{m-1}$. In particular, asymptotic studies are necessary along sequences from within an open subset of our partition to a neighbouring closed subset. Difficulties arise from the fact

mapping r_1 and r_2 can, in principle, reach arbitrarily large values: we will need to carefully study the behaviour of "their" sub-differentials under such limits. This will be the topic of the current section. First, we will concentrate on r_2, for which the result follows relatively analogously to earlier established results.

Lemma 3.3 *Let* $f_{\mathcal{R}}$ *be compatible with a 1st order polynomial growth condition (see Definition 2.2). Let* $(x_k, v_k) \in A^+$ *be a sequence such that* $(x_k, v_k) \to (\bar{x}, \bar{v}) \in \mathcal{O}^+$*, then it holds that*

i) $r_2(x_k, v_k) \to \infty$

ii)

$$\lim_{k \to \infty} f_{\mathcal{R}}(r_2(x_k, v_k)) \partial_x^{\mathrm{C}} r_2(x_k, v_k) = \{0\},$$

where the last limit is to be understood in the Painlevé-Kuratowski sense.

Proof We may observe here that $r_2(x, v)$ on A^+ is the unique solution (in r) to the equation

$$\max_{j \in I^+} g_j(x, \mathfrak{m} + rLv) = 0,$$

where g is convex in the second argument (it is actually linear therein) and $g(x, \mathfrak{m}) < 0$ holds true. The first item thus follows from Lemma 2.4 [44], which can be applied when restricting the analysis to the subsystem defined by the index set I^+. The second item follows from Lemma 3.4 [44]. □

The analogue of Lemma 3.3 for r_1 is novel and requires carefully working out several estimates.

Lemma 3.4 *Let $f_{\mathcal{R}}$ be compatible with a 1st order polynomial growth condition. Let $(x_k, v_k) \in B^-$ be a sequence such that $(x_k, v_k) \to (\bar{x}, \bar{v}) \in \mathcal{O}^-$, then it holds that*

i) $r_1(x_k, v_k) \to \infty$

ii)

$$\lim_{k \to \infty} f_{\mathcal{R}}(r_1(x_k, v_k))\partial_x^{\mathrm{C}} r_1(x_k, v_k) = \{0\},$$

where the last limit is to be understood in the Painlevé-Kuratowski sense.

Proof The first item is an immediate consequence of the fact that along any sequence $(x_k, v_k) \to (\bar{x}, \bar{v}) \in \mathcal{O}^-$, there must exist an index $i \in I^-$ such that $a_i(x_k)^{\mathsf{T}} L v_k \to 0$. Thus from the definition of r_1, it follows immediately that $r_1(x_k, v_k) \to \infty$.

As for the second point, observe first that for all $(x, v) \in B^-$,

$$\partial_x^{\mathrm{C}} r_1(x, v) = \mathrm{Co}\left\{\nabla_x \mathfrak{r}_i(x, v)\ : i \in I^-(x, v)\right\},$$

with $I^-(x, v) := \left\{i \in I^-\ : \mathfrak{r}_i(x, v) = r_1(x, v)\right\}$ being the active index set. Indeed, each mapping $\mathfrak{r}_i$ is continuously differentiable on B^- and has gradient:

$$\nabla_x \mathfrak{r}_i(x, v) = \frac{(b_i(x) - a_i(x)^{\mathsf{T}}\mathfrak{m})\nabla a_i(x) L v - (a_i(x)^{\mathsf{T}} L v)\nabla b_i(x)}{(a_i(x)^{\mathsf{T}} L v)^2}.$$

We claim that there exists a constant $C > 0$, such that $\|\nabla_x \mathfrak{r}_i(x, v)\| \leq C\mathfrak{r}_i(x, v)^2$. This can be seen as follows,

$$\begin{aligned}
\|\nabla_x \mathfrak{r}_i(x, v)\| &\leq \frac{\left|b_i(x) - a_i(x)^{\mathsf{T}}\mathfrak{m}\right| \|\nabla a_i(x) L v\|}{(a_i(x)^{\mathsf{T}} L v)^2} + \frac{\left|a_i(x)^{\mathsf{T}} L v\right| \|\nabla b_i(x)\|}{(a_i(x)^{\mathsf{T}} L v)^2} \\
&= \frac{\left|b_i(x) - a_i(x)^{\mathsf{T}}\mathfrak{m}\right|^2}{(a_i(x)^{\mathsf{T}} L v)^2} \left(\frac{\|\nabla a_i(x) L v\|}{\left|b_i(x) - a_i(x)^{\mathsf{T}}\mathfrak{m}\right|} + \frac{\|\nabla b_i(x)\|}{\left|b_i(x) - a_i(x)^{\mathsf{T}}\mathfrak{m}\right|^2}\right) \\
&= (\mathfrak{r}_i(x, v))^2 \left(\frac{\|\nabla a_i(x) L v\|}{\left|b_i(x) - a_i(x)^{\mathsf{T}}\mathfrak{m}\right|} + \frac{\|\nabla b_i(x)\|}{\left|b_i(x) - a_i(x)^{\mathsf{T}}\mathfrak{m}\right|^2}\right).
\end{aligned}$$

With the fact that we can assume the neighbourhood U to be compact, $(b_i(x) - a_i(x)^{\mathsf{T}}\mathfrak{m})$ to be non vanishing, we may assume that $\left|b_i(x) - a_i(x)^{\mathsf{T}}\mathfrak{m}\right| > \delta$ for all $x \in U$. The matrix L is regular, i.e., $Lv \neq 0$, and the continuous maps $x \mapsto \|\nabla a_i(x)\|$, $x \mapsto \nabla b_i(x)$ and $x \mapsto \|a_i(x)\|$ can be assumed to have an upper bound on U. We may thus indeed assume given, a constant $C > 0$, such that $\|\nabla_x \mathfrak{r}_i(x, v)\| \leq C\mathfrak{r}_i(x, v)^2$ as claimed.

Consequently,

$$\lim_{k\to\infty} \| f_{\mathcal{R}}(\mathfrak{r}_i(x_k, v_k))\nabla_x \mathfrak{r}_i(x_k, v_k)\| \le \lim_{k\to\infty} f_{\mathcal{R}}(\mathfrak{r}_i(x_k, v_k)) \|\nabla_x \mathfrak{r}_i(x_k, v_k)\| \tag{24}$$

$$\le \lim_{k\to\infty} C f_{\mathcal{R}}(\mathfrak{r}_i(x_k, v_k))(\mathfrak{r}_i(x_k, v_k))^2 = 0, \tag{25}$$

where the last limit results from item 1 combined with Definition 2.2.

Now any $s_k \in f_{\mathcal{R}}(r_1(x_k, v_k))\partial^{\mathrm{C}} r_1(x_k, v_k)$ is of the form:

$$s_k = \sum_{j\in I^-} \lambda_j^k f_{\mathcal{R}}(\mathfrak{r}_j(x_k, v_k))\nabla_x \mathfrak{r}_j(x_k, v_k),$$

with λ_j^k simplex weights such that $\lambda_j^k = 0$ if $j \notin I^-(x_k, v_k)$. Since λ_j^k belongs to the unit simplex of dimension I^-, it must admit a cluster point, moreover all right hand side terms, admit zero as limit. It thus follows that $s_k \to 0$ and the result is shown. □

Prior to continuing we now need to establish that certain limits, within our partition can not exist:

Lemma 3.5 *There can not exist any sequence $(x_k, v_k) \in A^{+0} \cup A^{+-}$ with limit $(\bar{x}, \bar{v}) \in \mathcal{O}^+$ (likewise with B^+).*

Proof Any such sequence belongs to A^+, which is disjoint from B^+, both sets are open. Continuity of the data shows that one has to move through $\mathcal{O}^+$, thus the statement holds for the case B^+. Observe now that on B^-, r_1 is locally Lipschitz and finite valued. Furthermore, as a result of Lemma 3.3 any sequence in A^+ tending to a limit in $\mathcal{O}^+$ must have $r_2(x_k, v_k) \to \infty$. Now let us assume by contradiction the existence of some sequence in $A^{+0} \cup A^{+-}$ tending to $(\bar{x}, \bar{v}) \in \mathcal{O}^+$ and let W be the neighbourhood of $(\bar{x}, \bar{v})$ on which r_1 is locally Lipschitzian. We may in particular pick W to be compact and assume given some $M > 0$ such that $r_1(x, v) < M$ for all $(x, v) \in W$. Of course for k sufficiently large $(x_k, v_k) \in W$. But for k (possibly even larger), sufficiently large as well $r_2(x_k, v_k) > M$ must hold true equally. Then for such k we have $r_1(x_k, v_k) \le M < r_2(x_k, v_k)$ so that by definition $(x_k, v_k) \in A^{++}$ must hold true, which contradicts our assumption. □

3.4 On the partial subgradients of *e*

Now that some of the preparatory material has been laid out, we turn our attention to the identification of the sub-differential of e. First, on part of the set $U \times \mathbb{S}^{m-1}$, the partial sub-differential of e with respect to x is easily identified.

Lemma 3.6 *The mapping e, given in (21), is locally Lipschitzian on the open set $A^- \cup B^+ \cup A^{++} \cup A^{+-}$ and has partial Clarke derivative, w.r.t. x satisfying:*

$$\partial_x^{\mathrm{C}} e(x, v) \subseteq \begin{cases} \{0\} & \textit{if } (x, v) \in A^{+-} \cup A^- \\ -f_{\mathcal{R}}(r_1(x, v))\partial_x^{\mathrm{C}} r_1(x, v) & \textit{if } (x, v) \in B^+ \\ f_{\mathcal{R}}(r_2(x, v))\partial_x^{\mathrm{C}} r_2(x, v) - f_{\mathcal{R}}(r_1(x, v))\partial_x^{\mathrm{C}} r_1(x, v) & \textit{if } (x, v) \in A^{++} \end{cases} \tag{26}$$

Proof The sets A^-, A^{+-}, A^{++} and B^+ are all open as already commented. In particular on $A^{+-} \cup A^-$, $e = 0$ so that e is continuously differentiable there and thus also locally Lipschitzian with immediately identified Clarke subgradient.

On B^+ as a subset of B^-, the mapping r_1 is locally Lipschitzian, the mapping $F_{\mathcal{R}}$ is continuously differentiable, by assumption, and so since $e(x,v) = 1 - F_{\mathcal{R}}(r_1(x,v))$ is locally Lipschitzian too, Clarke's chain rule, e.g., [60, Theorem 2.3.9(ii)] yields the desired formula.

On the open set A^{++}, both the mapping r_1 and r_2 are locally Lipschitzian. Moreover on this set we have $e(x,v) = F_{\mathcal{R}}(r_2(x,v)) - F_{\mathcal{R}}(r_1(x,v))$ and so e is, as the sum of two locally Lipschitzian maps also locally Lipschitzian. The Clarke sum rule, [60, Proposition 2.3.3] and chain rule [60, Theorem 2.3.9(ii)] give the indicated formula. □

As a final observation, we note that the generalized derivatives of r_1 and r_2, as finite extrema are also immediately identified, and we will not make them fully explicit.

Our next idea is to identify a set-valued mapping which is a good candidate for "being" $\partial_x^{\mathrm{C}} e$. Observe however that for the time being we have not yet established that e is locally Lipschitz on $U \times \mathbb{S}^{m-1}$. We therefore define the map $e_x : U \times \mathbb{S}^{m-1} \rightrightarrows \mathbb{R}^n$ as follows

$$e_x(x,v) = \begin{cases} \{0\} & \text{if } (x,v) \in A^{+-} \cup \mathcal{O}^- \cup A^- \\ -f_{\mathcal{R}}(r_1(x,v))\partial_x^{\mathrm{C}} r_1(x,v) & \text{if } (x,v) \in \mathcal{O}^+ \cup B^+ \\ \mathrm{Co}(\{0\} \cup f_{\mathcal{R}}(r_2(x,v))\partial_x^{\mathrm{C}} r_2(x,v) - f_{\mathcal{R}}(r_1(x,v))\partial_x^{\mathrm{C}} r_1(x,v)) & \text{if } (x,v) \in A^{+0} \\ f_{\mathcal{R}}(r_2(x,v))\partial_x^{\mathrm{C}} r_2(x,v) - f_{\mathcal{R}}(r_1(x,v))\partial_x^{\mathrm{C}} r_1(x,v) & \text{if } (x,v) \in A^{++} \end{cases} \tag{27}$$

and observe that this map coincides with the upper-estimation found in Eq. 3.6 over the open sets on which e was shown to be locally Lipschitz (recall Lemma 3.6). Our first endeavour will be to establish that e_x is outer semi-continuous (o.s.c.). To this end, we require the following result:

Lemma 3.7 *Let $f_{\mathcal{R}}$ be compatible with a 1st order polynomial growth condition. Along any sequence $(x_k, v_k) \in A^{++} \cup \mathcal{O}^+ \cup B^+$ converging to $(\bar{x}, \bar{v}) \in \mathcal{O}^-$ it holds*

the last limit is to be understood in the Painlevé-Kuratowski sense.

Proof For the subsequence (x_k, v_k) contained in $\mathcal{O}^+ \cup B^+ \subset B^-$ the result follows from Lemma 3.4, item 2. We may thus focus on the subsequence (x_k, v_k) contained in A^{++}, but then we may simply invoke the triangular inequality and use both Lemmas 3.4 and 3.3. □

Lemma 3.8 *Let $f_{\mathcal{R}}$ be compatible with a 1st order polynomial growth condition. Along any sequence $(x_k, v_k) \in A^{++} \cup A^{+-}$ converging to $(\bar{x}, \bar{v}) \in A^{+0}$ it holds that* $\limsup_{k \to \infty} e_x(x_k, v_k) \subseteq e_x(\bar{x}, \bar{v})$.

Proof Observe first that on $A^+ = A^{++} \cup A^{+-} \cup A^{+0}$ both r_2 and r_1 are well defined, finite and locally Lipschitz. Moreover on A^+, e_x is nothing other than $\partial_x^{\mathrm{C}} h(x,v)$, with

$$h(x,v) = \max\{F_{\mathcal{R}}(r_2(x,v)) - F_{\mathcal{R}}(r_1(x,v)), 0\},$$

which is clearly locally Lipschitzian on A^+. Moreover $\partial_x^{\mathrm{C}} h(x,v)$ is outer semi-continuous as is known, thus yielding the claim. □

We can now gather all are findings into the following proposition:

Proposition 3.1 *Let $f_{\mathcal{R}}$ be compatible with a 1st order polynomial growth condition. Then the mapping e_x defined in (27) is outer semi-continuous on $U \times \mathbb{S}^{m-1}$ and locally bounded.*

Proof The sets $\mathcal{O}^-$, $\mathcal{O}^+$, A^{+0} are closed, all the other sets open. The cases are therefore

- Convergence from within A^- or A^{+-} into $\mathcal{O}^-$, which is evident.
- Convergence from $A^{++} \cup \mathcal{O}^+ \cup B^+$ into $\mathcal{O}^-$, which is the object of Lemma 3.7.
- Convergence to $\mathcal{O}^-$ can not occur from within the closed set A^{+0}.
- Convergence from within A^+ to $\mathcal{O}^+$, which only can occur from within A^{++} due to Lemma 3.5 and is thus the result of Lemma 3.3.
- Convergence from within A^{+-} or A^{++} into A^{+0} and is the result of Lemma 3.8.

□

3.5 Gluing the pieces together: the main result

In order to show that e is locally Lipschitzian on $U \times \mathbb{S}^{m-1}$, we require a technical result that allows us to bridge, the closed sets that interspace A^-, B^-, A^{++} and A^{+-}. We thus need to glue together in a way, the various open sets.

Lemma 3.9 (Gluing lemma) *Let $U \subseteq \mathbb{R}^n$ be an open set, U be partitioned into U' open and $\mathcal{O}$ (closed). Assume given the following objects:*

i) $f : U \to [0, 1]$ *a continuous mapping that is locally Lipschitzian on U', constant on $\mathcal{O}$;*
ii) $\partial f : U \to \mathbb{R}^n$ *a set-valued mapping such that $\partial f = \partial^{\mathrm{C}} f$ on U' that is moreover outer semi-continuous (o.s.c.) and locally bounded.*

Then f is locally Lipschitzian on U and $\partial^{\mathrm{C}} f \subseteq Co\,\partial f$ if ∂f contains 0 on $\mathcal{O}$ and f is extremal on $\mathcal{O}$ (i.e., takes the values 0 or 1).

Proof First let us show that f is locally Lipschitzian on U. Indeed, let $x \in U$ be arbitrary and fixed. We may of course assume that $x \in \mathcal{O}$, since the case $x \in U'$ follows by assumption. Now, since ∂f is locally bounded we can find a constant $M \geq 0$ and a open convex neighbourhood W of x such that

$$\|\partial f(w)\| \leq M, \;\forall w \in W.$$

We will show that f is Lipschitzian on W, i.e., for all $y, z \in W$:

$$|f(y) - f(z)| \leq M\|y - z\|. \tag{28}$$

Let $y, z \in W$ be fixed. If both elements belong to $\mathcal{O}$, then (28) holds trivially. Let us assume that $y \in U'$, and consider $\epsilon > 0$ to be such that $x : (-\epsilon, 1 + \epsilon) \to U$ given by $x(t) := (1 - t)y + tz \in W$ for all t.

Define the set $A := \{t \in (-\epsilon, 1 + \epsilon) : x(t) \in U'\}$, since A is an open set in $\mathbb{R}$, it can be written as a disjoint union of a sequence intervals (a_n, b_n), that is, $A = \bigcup_{n\in\mathbb{N}}(a_n, b_n)$.

Let $n \in \mathbb{N}$ and points $t_1, t_2 \in (a_n, b_n)$ be fixed but arbitrary. Then by Lebourg's mean value theorem (Theorem 1.7 [60]) there exists $t^* \in [t_1, t_2]$ and $x^* \in \partial^{\mathrm{C}} f(x(t^*))$ such that

$$|f(x(t_1)) - f(x(t_2))| \leq \|x^*\|\|x(t_1) - x(t_2)\| \leq M\|x(t_1) - x(t_2)\| = M\,|t_1 - t_2|\,\|y - z\|\,,$$

where we have used the fact that $x(t^*) \in W$.

By continuity of f we conclude that for every $n \in \mathbb{N}$ and every $t_1, t_2 \in [a_n, b_n]$

$$|f(x(t_1)) - f(x(t_2))| \leq M\|x(t_1) - x(t_2)\| = M|t_1 - t_2| \cdot \|y - z\| \tag{29}$$

Finally, let us show (28). By our assumptions we have that $0 \in A$ (recall $y \in U'$), which implies that $0 \in (a_{n_0}, b_{n_0})$ for some $n_0 \in \mathbb{N}$. Let us distinguish two cases:

i) $1 \notin A$. Then, $f(x(b_{n_0})) = f(z)$ (because in this case $x(b_{n_0}), z \in \mathcal{O}$) and by (29)

$$|f(y) - f(z)| \leq |f(y) - f(x(b_{n_0}))| + |f(x(b_{n_0})) - f(z)| \leq Mb_{n_0}\|y - z\| \leq M\|y - z\|.$$

Here we have used that $0 < b_{n_0} < 1$ holds true since $1 \notin A$.

ii) $1 \in (a_k, b_k)$ for some $k \in \mathbb{N}$. First, if $k = n_0$, then the inequality follows from (29). Second, let us assume that $k \neq n_0$, then, without loss of generality we can assume that $0 < b_{n_0} < a_k < 1$, so using (29) and the fact that $f(x(b_{n_0})) = f(x(a_k))$ we have that

$$\begin{aligned} |f(y) - f(z)| &\leq |f(y) - f(x(b_{n_0}))| + |f(x(b_{n_0})) - f(x(a_k))| + |f(x(a_k)) - f(z)| \\ &\leq M\left(b_{n_0} + 1 - a_k\right)\|y - z\| \leq M\|y - z\|. \end{aligned}$$

This shows that (28) holds and consequently f is locally Lipschitz on U.

Now, let us show that $U \subseteq \overline{D}$, where $D := U' \cup \operatorname{int}(\mathcal{O})$. Indeed, consider $x \in U$, then there exists $r > 0$ such that $\mathbb{B}(x, r) \subseteq U$. Now, if there exists $\epsilon \in (0, r)$ such that $\mathbb{B}(x, \epsilon) \subseteq \mathcal{O}$ we have that $x \in \operatorname{int}(\mathcal{O})$, otherwise for every $\epsilon \in (0, r)$, it holds that $\mathbb{B}(x, \epsilon) \cap U' \neq \emptyset$, which implies that $x \in \overline{U} \subseteq \overline{D}$.

Now, by [61, Proposition 2.2] we have that for every $x \in U$:

$$\partial^{\mathrm{C}} f(x) \subseteq \operatorname{cl}\operatorname{Co}\left(\limsup_{D \ni y \xrightarrow{f} x} \partial^{\mathrm{C}} f(y)\right)$$

Furthermore, from the assumptions we have that

$$\partial^{\mathrm{C}} f(y) \subseteq \partial f(y), \ \forall y \in D.$$

Indeed recall that on $\operatorname{int}\mathcal{O}$, f is extremal and with f moreover being constant on $\mathcal{O}$, it must follow that $\partial^{\mathrm{C}} f = \{0\}$ on $\operatorname{int}\mathcal{O}$. Consequently, by using the o.s.c. of ∂f, it follows that:

$$\partial^{\mathrm{C}} f(x) \subseteq \operatorname{cl}\operatorname{Co}\left(\limsup_{D \ni y \xrightarrow{f} x} \partial f(y)\right) \subseteq \operatorname{cl}\operatorname{Co}\left(\partial f(x)\right).$$

Finally since the space is finite dimensional the above closure operation can be omitted. □

We can now put to use all these results to establish that the mapping e itself, is locally Lipschitz, on the whole set $U \times \mathbb{S}^{m-1}$.

Proposition 3.2 *Let $f_{\mathcal{R}}$ be compatible with a 1st order polynomial growth condition. The mapping $e : U \times \mathbb{S}^{m-1} \to [0, 1]$ defined in (22) is locally Lipschitzian. Moreover, for any $(x, v) \in U \times \mathbb{S}^{m-1}$, the following inclusion holds $\partial_x^{\mathrm{C}} e(x, v) \subseteq e_x(x, v)$, where e_x is as in (27).*

Proof Let us first establish that e is locally Lipschitzian on the open set A^+. Indeed on this set, $e(x, v) = \max\{F_{\mathcal{R}}(r_2(x, v)) - F_{\mathcal{R}}(r_1(x, v)), 0\}$, with $(x, v) \in A^{++}$ if and only if $F_{\mathcal{R}}(r_2(x, v)) - F_{\mathcal{R}}(r_1(x, v)) > 0$, $(x, v) \in A^{+-}$ if and only if $F_{\mathcal{R}}(r_2(x, v)) - F_{\mathcal{R}}(r_1(x, v)) < 0$ and $(x, v) \in A^{+0}$ if and only if $F_{\mathcal{R}}(r_2(x, v)) - F_{\mathcal{R}}(r_1(x, v)) = 0$. Moreover both $F_{\mathcal{R}}(r_2(x, v)) - F_{\mathcal{R}}(r_1(x, v))$ and 0 are locally Lipschitzian (and finite valued) on A^+. It is thus also true that e is locally Lipschitzian on A^+ and Proposition 2.3.12 [60] gives the identification $\partial_x^C e(x, v) = e_x(x, v)$ on A^+.

Let us now turn our attention to the situation of establishing that e is locally Lipschitzian on the open set B^-, which as we recall is partitioned into the open sets B^+, A^+ and closed set $\mathcal{O}^+$. It is already clear that e is locally Lipschitz on A^+ and B^+. Moreover the mapping $(x, v) \mapsto F_{\mathcal{R}}(r_1(x, v))$ is locally Lipschitz on B^- (and finite valued). By Lemma 3.5, it is sufficient to concentrate on the open set $A^{++} \cup B^+$, since neighbourhoods of $\mathcal{O}^+$ can be made sufficiently small to not intersect with $A^{+0} \cup A^{+-}$. Let us define $f : A^{++} \cup B^+ \cup \mathcal{O}^+ \to [0, 1]$ as

$$f(x, v) = \begin{cases} 1 & \text{if } (x, v) \in B^+ \cup \mathcal{O}^+ \\ F_{\mathcal{R}}(r_2(x, v)) & \text{if } (x, v) \in A^{++} \end{cases}$$

then it follows that $e(x, v) = f(x, v) - F_{\mathcal{R}}(r_1(x, v))$ on $A^{++} \cup B^+ \cup \mathcal{O}^+$. Moreover f is evidently extremal on $\mathcal{O}^+$ and constant (taking the maximal value 1). Together with Proposition 3.1, we may thus apply Lemma 3.9 to conclude that f is locally Lipschitz on $A^{++} \cup B^+ \cup \mathcal{O}^+$, and therefore so too is e as the sum of two Lipschitz mappings.

We have thus shown that e is Lipschitz on B^- and it is clearly so on A^-. Moreover e takes the extremal value 0 on $\mathcal{O}^-$ and is constant there. Once more through Proposition 3.1 and Lemma 3.9 we can conclude that e is locally Lipschitz on $U \times \mathbb{S}^{m-1}$ and the asserted inclusion also follows. □

We can now gather all previously established material and show that the probability function φ itself is locally Lipschitzian. We can also obtain an outer estimate of its Clarke sub-differential:

Theorem 3.1 *Let $A : \mathbb{R}^n \to \mathbb{R}^{p \times m}$ be a continuously differentiable matrix valued map, and let $b : \mathbb{R}^n \to \mathbb{R}^p$ be continuously differentiable. Let the random vector $\xi \in \mathbb{R}^m$ be elliptically symmetric with mean $\mathfrak{m}$, covariance-like matrix Σ and generator θ. Let ξ be compatible with the 1st order polynomial growth condition.*

Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ be defined as $g(x, z) = A(x)z - b(x)$. Let $\bar{x}$ and the neighbourhood U of $\bar{x}$ be such that $g_i(x, \mathfrak{m}) \neq 0$ for all $i = 1, ..., p$ and all $x \in U$. Then, the probability function $\varphi : \mathbb{R}^n \to [0, 1]$ defined as

$$\varphi(x) = \mathbb{P}[A(x)\xi \leq b(x)], \tag{30}$$

is locally Lipschitz on U and has Clarke sub-differential satisfying:

$$\partial^C \varphi(x) \subseteq \int_{v \in \mathbb{S}^{m-1}} e_x(x, v) d\mu_\zeta(v), \tag{31}$$

where μ_ζ is the uniform measure over the sphere $\mathbb{S}^{m-1} = \left\{z \in \mathbb{R}^m \ : \|z\|^2 = 1\right\}$, and e_x is as in (27).

Proof With e being continuous, even locally Lipschitzian as a result of Proposition 3.2, for any $x \in U$, the map $v \mapsto e(x, v)$ is measurable.

Let us define $\alpha : U \times \mathbb{S}^{m-1} \to \mathbb{R}$ as

$$\alpha(x, v) = \max\left\{\|s\| \ : s \in \partial^C e(x, v)\right\}.$$

Then as a result of Proposition 3.2, first of all e is well defined and in particular finite valued. Moreover, α is upper semi-continuous on $U \times \mathbb{S}^{m-1}$. Should this not hold, then at some $(\bar{x}, \bar{v}) \in U \times \mathbb{S}^{m-1}$ and for some $\varepsilon > 0$ we may find a sequence $(x_k, v_k) \to (\bar{x}, \bar{v})$ such that $\alpha(x_k, v_k) \geq \alpha(\bar{x}, \bar{v}) + \varepsilon$. We can thus find $s_k \in \partial^C e(x_k, v_k)$, with $\|s_k\| \geq \alpha(\bar{x}, \bar{v}) + \frac{\varepsilon}{2}$

for all k. However, s_k must admit a cluster point s^*, since e is locally Lipschitzian at $(\bar{x}, \bar{v})$ (see Proposition 2.1.5(d) [60]). In particular, since Clarke's sub-differential is o.s.c, $s^* \in \partial^{\mathsf{C}} e(\bar{x}, \bar{v})$. But, then we arrive at the contradiction $\|s^*\| \geq \alpha(\bar{x}, \bar{v}) + \frac{\varepsilon}{2}$, so that α must be upper semi-continuous (u.s.c.).

We may well assume that U is compact (or restrict U further if needed). So then it follows that α as an u.s.c. map on $U \times \mathbb{S}^{m-1}$ must reach a maximum, say M. Moreover in Proposition 3.2 we have shown that e is locally Lipschitz on $U \times \mathbb{S}^{m-1}$. We may thus at any $x, y \in U$ and for an arbitrary v employ Lebourg's mean-value theorem to entail that

$$|e(x, v) - e(y, v)| \leq \|s\| \, \|x - y\| \leq M \, \|x - y\| \, ,$$

with $s \in \partial^{\mathsf{C}}_x e(u, v)$ for some $u \in [x, y]$. We can invoke Clarke's Theorem of the interchange of integration and sub-differentiation [60, Thm 2.7.2] (see also [62] for further extensions) to establish that φ is locally Lipschitz on U and to obtain that

$$\partial^{\mathsf{C}} \varphi(x) \subseteq \int_{v \in \mathbb{S}^{m-1}} \partial^{\mathsf{C}}_x e(x, v) d\mu_\zeta(v) \subseteq \int_{v \in \mathbb{S}^{m-1}} e_x(x, v) d\mu_\zeta(v), \tag{32}$$

where the last inequality follows from Proposition 3.2. □

Let us briefly comment on the assumption of the placement of x and the neighbourhood, excluding $a_i(x)^{\mathsf{T}} \mathfrak{m} = b_i(x)$ for some $i = 1, ..., p$ and x in this neighbourhood:

Example 3.1 Consider the following data:

$$A(x)^{\mathsf{T}} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 1 & 1 & -1 \end{bmatrix}$$

and $b(x) = (1, 1, 1 + x, -x)$. Consider $\bar{x} = 0$, and say $\xi \sim \mathcal{N}(0, I)$, i.e., $\mathcal{R}$ is a χ distribution with $m - 1 = 1$ degree of freedom. Now for any $v = (v_1, v_2) \in \mathbb{S}^{m-1}$, with $v_1 < 0$, $e(\bar{x}, v) = 0$, yet $e(\bar{x}, (0, 1)) = F_{\mathcal{R}}(1) = 0.6827$. Moreover for any arbitrary $x > 0$, $e(x, (0, 1)) = 0$. It thus immediately follows that e is not continuous at $(\bar{x}, (0, 1))$, neither in x, nor in v. It can therefore not be locally Lipschitz and our argument to interchange sub-differentiation and integration not used in their current form. This example thus shows the importance of the assumptions $a_i(x)^{\mathsf{T}} \mathfrak{m} \neq b_i(x)$ at the point of interest.

4 Union of Polyhedra

We can now turn our attention to the case of probability functions given in the form of (1), and drop the assumption $\ell = 1$ we had in Section 3. Let us first justify the interchange of sub-differentiation and integration, prior to giving formulæ for the sub-differential of the probability function.

Theorem 4.1 *Consider the probability function* (1)*, wherein the random vector $\xi \in \mathbb{R}^m$ is elliptically symmetric with mean $\mathfrak{m}$, covariance-like matrix Σ and generator θ. For each $j = 1, ..., \ell$, let $A_j : \mathbb{R}^n \to \mathbb{R}^{p_j \times m}$ be a continuously differentiable matrix valued map, and let $b_j : \mathbb{R}^n \to \mathbb{R}^{p_j}$ be continuously differentiable. Assume moreover that ξ is compatible with the 1st order polynomial growth condition.*

Let $\bar{x} \in \mathbb{R}^n$ be given along with a neighbourhood of $\bar{x}$ such that $a_{ij}(x)^{\mathsf{T}} \mathfrak{m} - b_{ij}(x) \neq 0$ for all $i = 1, ..., p_j$, $j = 1, ..., \ell$ and all $x \in U$ holds true. Here a_{ij} denotes the ith row of *matrix A_j. Then φ is (locally) Lipschitz on U and*

$$\partial^{\mathrm{C}}\varphi(x) \subseteq \int_{v\in\mathbb{S}^{m-1}} \partial_x^{\mathrm{C}} e(x,v)d\mu_\zeta(v), \tag{33}$$

where μ_ζ is the uniform measure over the sphere $\mathbb{S}^{m-1} = \left\{z \in \mathbb{R}^m \ : \|z\|^2 = 1\right\}$, and $e : U \times \mathbb{S}^{m-1} \to [0,1]$ is also locally Lipschitz and given by:

$$e(x,v) = \mu_{\mathcal{R}}(\left\{r \geq 0 \ : \mathfrak{m} + rLv \in \bigcup_{j=1}^{\ell} M_j(x)\right\}). \tag{34}$$

Proof It follows immediately from (9) and (1) that

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

with e as in (34). Alternatively let us write $\{\mathfrak{m} + rLv : r \geq 0\} = \mathfrak{m} + \mathbb{R}_+ Lv =: R$, so that $e(x,v) = \mu_{\mathcal{R}}(R \cap \bigcup_{j=1}^{\ell} M_j(x))$. We recall that $R \cap \bigcup_{j=1}^{\ell} M_j(x) = \bigcup_{j=1}^{\ell}(R \cap M_j(x))$ and we now apply Lemma 2.1 to obtain

$$e(x,v) = \sum_{J\subseteq\{1,\ldots,\ell\}, J\neq\emptyset} (-1)^{|J|+1}\mu_{\mathcal{R}}(R \cap \bigcap_{j\in J} M_j(x)).$$

For any given index set J above, define $e_J : U \times \mathbb{S}^{m-1} \to [0,1]$ as $e_J(x,v) = \mu_{\mathcal{R}}(R \cap \bigcap_{j\in J} M_j(x))$.

Observe moreover that

$$\bigcap_{j\in J} M_j(x) = \left\{z \in \mathbb{R}^m \ : A_j(x)z \leq b_j(x),\ j \in J\right\}.$$

Consequently, as a result of Proposition 3.2, the mapping e_J is locally Lipschitz. Hence, so is e as a finite sum of locally Lipschitzian mappings. With this having been established, we may argue completely analogously as in Theorem 3.1 to justify the interchange of subdifferentiation and integration and to obtain the asserted formula. □

Remark 4.1 (On subdifferential inclusion) If we make the additional assumption that the mapping e is regular at x (see [60, Definition 2.4.10]), then the inclusion (33) is in fact an equality at x (see [60, Thm 2.7.2]).

Moreover, a closer look at the proof of [60, Thm 2.7.2] gives us that the inclusion

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

should also hold under the assumptions of Theorem 4.1. Here ∂^{F} denotes Fréchet's subdifferential. Finally, it is worth mentioning that other upper-estimations can be shown using the Mordukhovich/limiting subdifferential instead of Clarke subdifferential (see, e.g., [62]).

In general we could make use of these subdifferential inclusions in "relaxed optimality conditions". One can think here of a problem exhibiting a first order condition of the type $0 \in \partial^{\mathrm{C}}\varphi(\bar{x})$. The situation is very similar when φ appears in constraints, when an appropriate qualification condition holds (e.g., when applying Propositions 2.1 and Theorem 2.3 [63]) in which case an appropriate scalar multiple of $\partial^{\mathrm{C}}\varphi(\bar{x})$ appears. In this case the subdifferential inclusion given allows us to formulate relaxed conditions that would hold for every local solution. These conditions thus allow us, in principle, to identify favourable candidates. The use of inclusions is typical in MPECs (see the discussion following Theorem 4.1 in [64]). The restricted inner Fréchet characterization would allow us to formulate restrictive conditions. However it is well possible that these restrictive conditions allow no solution.

We will also formulate a readily numerically verifiable condition in Theorem 4.2 below under which φ turns out to be continuously differentiable. This condition holds true, except on a set of zero Lebesgue measure. The just given formula could thus in principle be put to use in (smooth) first order optimization methods, wherein an additional check of the condition given in Theorem 4.2 ensures proper use. It can not be excluded that a local (or global) optimal solution lies in a point of non-differentiability of the probability function. We have not observed this in the experiments conducted in this work. However, we believe this to be a very good reason to investigate dedicated solvers for nonsmooth and nonconvex optimization.

The issue with the expanded form of e, through the inclusion-exclusion formula, is that it does not help produce an efficient estimate of $\partial_x^{\mathsf{C}} e(x,v)$. We will illustrate this in Example 4.1 below. First however we will introduce notation that will help us provide a more elegant representation of e and $\partial_x^{\mathsf{C}} e(x,v)$. To this end, let us introduce the set

$$\mathcal{O}_{ij} = \left\{(x,v) \in U \times \mathbb{S}^{m-1} \ : a_{ij}(x)^{\mathsf{T}} L v = 0\right\}, \tag{35}$$

for $i = 1, ..., p_j$ and $j = 1, ..., \ell$ and $\mathcal{O} = \bigcup_{j=1}^{\ell} \cup_{i=1}^{p_j} \mathcal{O}_{ij}$ as well as $\mathcal{Z} = U \times \mathbb{S}^{m-1} \setminus \mathcal{O}$. Then we can observe that the sets $\mathcal{O}_{ij}$ are all of zero measure (the latter fact follows from Lemma 2.2 [65] and Korollar V.1.6 [54]). Consequently $\mathcal{O}$ is also of zero measure (in $\lambda \otimes \mu_\zeta$).

Next for each $j = 1, ..., \ell$ we define the following sets:

$$\begin{aligned}
I_j^- &= \left\{i = 1, ..., p_j \ : a_{ij}(x)^{\mathsf{T}} \mathfrak{m} > b_{ij}(x), \ \forall x \in U\right\}\\
I_j^+ &= \left\{i = 1, ..., p_j \ : a_{ij}(x)^{\mathsf{T}} \mathfrak{m} < b_{ij}(x), \ \forall x \in U\right\}\\
B_j^- &= \left\{(x,v) \in \mathcal{Z} \ : a_{ij}(x)^{\mathsf{T}} L v < 0, \ \forall i \in I_j^-\right\}\\
A_j^- &= \left\{(x,v) \in \mathcal{Z} \ : \ \exists i \in I_j^- \ : a_{ij}(x)^{\mathsf{T}} L v > 0\right\},
\end{aligned}$$

where implicitly it is assumed that U is sufficiently small, for the index sets I_j^- and I_j^+ to form a partition of $\{1, ..., p_j\}$. Observe that this can be achieved as a result of continuity and the standing assumption at $\bar{x}$.

Next we may observe that $B_j^- \cup A_j^- = \mathcal{Z}$, and that both B_j^- and A_j^- are open sets. Moreover, should $I_j^- = \emptyset$, then simply $B_j^- = \mathcal{Z}$. Moreover define for all $i = 1, ..., p_j$, $j = 1, ..., \ell$, the map $\mathfrak{r}_{ij} : \mathcal{Z} \to \mathbb{R}$ as

$$\mathfrak{r}_{ij}(x,v) = \frac{b_{ij}(x) - a_{ij}(x)^{\mathsf{T}} \mathfrak{m}}{a_{ij}(x)^{\mathsf{T}} L v} \tag{36}$$

and observe that this map is well defined and continuously differentiable on $\mathcal{Z}$.

Now for each $j = 1, ..., \ell$, we define $r_1^j : B_j^- \to \mathbb{R}_+$ as

$$r_1^j(x,v) = \begin{cases} 0 & \text{if} \quad I_j^- = \emptyset. \\ \max_{i \in I_j^-} \mathfrak{r}_{ij}(x,v) & \text{otherwise} \end{cases}$$

and we define $r_2^j : B_j^- \to \mathbb{R}_+ \cup \{\infty\}$ as

$$r_2^j(x,v) = \begin{cases} \sup_t & t \\ \text{s.t.} & t a_{ij}(x)^\mathsf{T} L v \le b_{ij}(x) - a_{ij}(x)^\mathsf{T} \mathfrak{m}, \, i \in I_j^+ \end{cases}.$$

Now at any $(x,v) \in \mathcal{Z}$, let us define the index set $F(x,v) = \left\{ j = 1, ..., \ell \ : (x,v) \in B_j^- \right\}$. Then, we can observe that F is locally stable since B_j^- is open and that

$$e(x,v) = \mu_{\mathcal{R}} \left(\bigcup_{j \in F(x,v)} [r_1^j(x,v), \max\left\{ r_2^j(x,v), r_1^j(x,v) \right\}] \right), \tag{37}$$

since $\left\{ r \ge 0 \ : \mathfrak{m} + rLv \in M_j(x) \right\} = \emptyset$ if $(x,v) \in A_j^-$. In order to further simplify the formula, let us introduce for any $(x,v) \in \mathcal{Z}$, the index set $F'(x,v)$, defined as follows:

$$F'(x,v) = \left\{ j \in F(x,v) \ : r_2^j(x,v) \ge r_1^j(x,v) \right\}.$$

Hence, formula (37) becomes:

$$e(x,v) = \mu_{\mathcal{R}} \left(\bigcup_{j \in F'(x,v)} [r_1^j(x,v), r_2^j(x,v)] \right). \tag{38}$$

We have already hinted on the potential issue, in representing economically $\partial_x^{\mathrm{C}} e(x,v)$ when employing the inclusion-exclusion formula. The problem is fully illustrated in the next example:

Example 4.1 We consider a trial point $x \in \mathbb{R}^n$ and two polyhedra $M_1(x)$ and $M_2(x)$, as well as a direction v at which we have $r_1^1(x,v) < r_1^2(x,v)$, $r_2^1(x,v) < r_2^2(x,v)$ and $r_1^2(x,v) < r_2^1(x,v)$. The inclusion-exclusion based formula from Theorem 4.1 yields $e(x,v) = F_{\mathcal{R}}(r_2^2(x,v)) - F_{\mathcal{R}}(r_1^2(x,v)) + F_{\mathcal{R}}(r_2^1(x,v)) - F_{\mathcal{R}}(r_1^1(x,v)) - (F_{\mathcal{R}}(r_2^1(x,v)) - F_{\mathcal{R}}(r_1^2(x,v)))$, which in turn provides a non-economic representation of its sub-differential:

$$\begin{aligned} \partial_x^{\mathrm{C}} e(x,v) \subseteq [\ & f_{\mathcal{R}}(r_2^2(x,v)) \partial_x^{\mathrm{C}} r_2^2(x,v) - f_{\mathcal{R}}(r_1^2(x,v)) \partial_x^{\mathrm{C}} r_1^2(x,v) + f_{\mathcal{R}}(r_2^1(x,v)) \partial_x^{\mathrm{C}} r_2^1(x,v) - \\ & f_{\mathcal{R}}(r_1^1(x,v)) \partial_x^{\mathrm{C}} r_1^1(x,v) - (f_{\mathcal{R}}(r_2^1(x,v)) \partial_x^{\mathrm{C}} r_2^1(x,v) - f_{\mathcal{R}}(r_1^2(x,v)) \partial_x^{\mathrm{C}} r_1^2(x,v))]. \end{aligned}$$

One can of course observe that in most cases, (38), through potentially tedious computations, will allow us to identify swiftly the most economic representation of e. One would simply have to work out, what the union of intervals amounts to and, unless some local degeneracy occurs, this will immediately provide the appropriate estimate of $\partial_x^{\mathrm{C}} e(x,v)$. Degeneracy will occur as soon as any two interval bounds match and locally "cross". The subsequent result provides a unified formula, rendering this interval calculus unnecessary in the above situation. We will also illustrate, that typically, an efficient estimate of the sub-differential is obtained. Let us just briefly mention, prior to stating the result, that $F_{\mathcal{R}}$ is increasing in increasing arguments. Therefore, minimum/maximum operations may be pulled through $F_{\mathcal{R}}$ at will.

Proposition 4.1 *At any $(x, v) \in \mathcal{Z}$, the following identity holds true:*

$$e(x,v) = \max\left\{F_{\mathcal{R}}(\max_{j\in F'(x,v)} r_2^j(x,v)) - F_{\mathcal{R}}(\min_{j\in F'(x,v)} r_1^j(x,v)), 0\right\} - \sum_{S,S'\subseteq F'(x,v), S\cap S'=\emptyset, S\cup S'=F'(x,v), S,S'\neq\emptyset} \max\left\{F_{\mathcal{R}}(\min_{i\in S'} r_1^i(x,v)) - F_{\mathcal{R}}(\max_{i\in S} r_2^i(x,v)), 0\right\}, \quad (39)$$

where $F_{\mathcal{R}}(\infty) = 1$ is to be understood.

Proof The proof will be performed by induction. To this end, let us fix an arbitrary $(\bar{x}, \bar{v}) \in \mathcal{Z}$, and first of all observe that on an appropriate neighbourhood of $(\bar{x}, \bar{v})$, $F(\bar{x}, \bar{v})$ remains unchanged. We may thus, without loss of generality assume that $F(\bar{x}, \bar{v}) = \{1, ..., \ell\}$.

When $\ell = 1$, the stipulated formula is immediately seen to be identical to the formula of (22).

Let us thus assume that the formula is valid for a given $\ell' \geq 1$, we will then show that the formula is valid for $\ell' + 1$. First of all, we may assume that for all $j = 1, ..., \ell' + 1$, it holds that $r_2^j(\bar{x}, \bar{v}) \geq r_1^j(\bar{x}, \bar{v})$. Indeed, should this not be the case, i.e., should there exist $j = 1, ..., \ell' + 1$ for which $r_2^j(\bar{x}, \bar{v}) < r_1^j(\bar{x}, \bar{v})$ holds true, then as a result of continuity, this continues to hold true on an appropriate neighbourhood of $(\bar{x}, \bar{v})$, therefore, for (x, v), near $(\bar{x}, \bar{v})$, $j \notin F'(x, v)$. Re-arranging indices if needed, we may assume $j = \ell' + 1$ and hence locally,

$$e(x,v) = \mu_{\mathcal{R}}(\left\{r \geq 0 \ : \mathfrak{m} + rLv \in \bigcup_{j=1}^{\ell'} M_j(x)\right\}).$$

We may thus use the induction assumption to obtain the formula for the case ℓ'. Since $\ell' + 1 \notin F(x, v)$ for (x, v) near $(\bar{x}, \bar{v})$, the formulae (39) with ℓ' and $\ell' + 1$ do indeed coincide.

From now on we will thus assume that $F'(x, v) = \{1, ..., \ell'\}$.

Next, let us assume the existence of $k = 1, ..., \ell'$ such that $r_1^k(\bar{x}, \bar{v}) \leq r_1^{\ell'+1}(\bar{x}, \bar{v}) \leq r_2^{\ell'+1}(\bar{x}, \bar{v}) \leq r_2^k(\bar{x}, \bar{v})$. We may employ the induction assumption to derive the formula (39), since indeed (38) holds true also when eliminating the $\ell' + 1$th term. It is moreover clear that:

$$\begin{aligned}
\max_{j\in F'(x,v)} F_{\mathcal{R}}(r_2^j(x,v)) &= \max_{j\in F'(x,v)\cup\{\ell'+1\}} F_{\mathcal{R}}(r_2^j(x,v)) \\
\min_{j\in F'(x,v)} F_{\mathcal{R}}(r_1^j(x,v)) &= \min_{j\in F'(x,v)\cup\{\ell'+1\}} F_{\mathcal{R}}(r_1^j(x,v)).
\end{aligned}$$

Let us now pick an arbitrary but fixed partition S', S of $F'(x, v) \cup \{\ell' + 1\}$, then, an exhaustive case distinction yields the identity:

$$\max\left[F_{\mathcal{R}}\left(\min_{i\in S'} r_1^i(\bar{x},\bar{v})\right) - F_{\mathcal{R}}\left(\max_{i\in S} r_2^i(\bar{x},\bar{v})\right), 0\right] =$$

$$= \begin{cases}
\max\left[F_{\mathcal{R}}\left(\min_{i\in S'\setminus\{\ell'+1\}} r_1^i(\bar{x},\bar{v})\right) - F_{\mathcal{R}}\left(\max_{i\in S} r_2^i(\bar{x},\bar{v})\right), 0\right] & \text{if } k, \ell'+1 \in S' \\
\max\left[F_{\mathcal{R}}\left(\min_{i\in S'} r_1^i(\bar{x},\bar{v})\right) - F_{\mathcal{R}}\left(\max_{i\in S\setminus\{\ell'+1\}} r_2^i(\bar{x},\bar{v})\right), 0\right] & \text{if } k, \ell'+1 \in S \\
0 & \text{if } k \in S \text{ and } \ell'+1 \in S' \\
0 & \text{if } \ell'+1 \in S \text{ and } k \in S'
\end{cases}$$

and thus consequently, the sum over all partitions in (39) is unaltered when moving from partitions of $F'(\bar{x}, \bar{v})$ to all partitions of $F'(\bar{x}, \bar{v}) \cup \left\{\ell' + 1\right\}$. We have thus shown the formula to extend to $\ell' + 1$ in this particular situation.

For the remainder we may, upon reordering the indices if needed, assume that $r_1^{\ell'+1}(\bar{x}, \bar{v}) \leq r_1^i(\bar{x}, \bar{v})$ for all $i \in F'(\bar{x}, \bar{v})$. Now two further cases appear

– The situation wherein $r_2^{\ell'+1}(\bar{x}, \bar{v}) < r_1^i(\bar{x}, \bar{v})$ for all $i \in F'(\bar{x}, \bar{v})$
– Or the existence of $i \in F'(\bar{x}, \bar{v})$ such that $r_1^i(\bar{x}, \bar{v}) \leq r_2^{\ell'+1}(\bar{x}, \bar{v})$.

The first case corresponds to adding a disjoint interval to the union of intervals. It thus immediately follows that:

$$\begin{aligned} e(\bar{x}, \bar{v}) = \; & F_{\mathcal{R}}(r_2^{\ell'+1}(\bar{x}, \bar{v})) - F_{\mathcal{R}}(r_1^{\ell'+1}(\bar{x}, \bar{v})) + \\ & \max\left\{\max_{j \in F'(x,v)} F_{\mathcal{R}}(r_2^j(x, v)) - \min_{j \in F'(x,v)} F_{\mathcal{R}}(r_1^j(x, v)), 0\right\} \\ & - \sum_{S,S' \subseteq F'(x,v), S \cap S' = \emptyset, S \cup S' = F'(x,v), S,S' \neq \emptyset} \max\left\{F_{\mathcal{R}}(\min_{i \in S'} r_1^i(x, v)) - F_{\mathcal{R}}(\max_{i \in S} r_2^i(x, v)), 0\right\}, \end{aligned}$$

upon employing the indication hypothesis for the ℓ' remaining intervals. Observe now that on $F'(\bar{x}, \bar{v})$, $\max_{j \in F'(x,v)} F_{\mathcal{R}}(r_2^j(x, v)) - \min_{j \in F'(x,v)} F_{\mathcal{R}}(r_1^j(x, v)) \geq 0$, so that the first max operation can be removed. We can now identify:

$$\begin{aligned} \max_{j \in F'(\bar{x},\bar{v}) \cup \{\ell'+1\}} F_{\mathcal{R}}(r_2^j(\bar{x}, \bar{v})) &= \max_{j \in F'(\bar{x},\bar{v})} F_{\mathcal{R}}(r_2^j(\bar{x}, \bar{v})) \\ F_{\mathcal{R}}(r_1^{\ell'+1}(\bar{x}, \bar{v})) &= \min_{j \in F'(\bar{x},\bar{v}) \cup \{\ell'+1\}} F_{\mathcal{R}}(r_1^j(\bar{x}, \bar{v})) \\ x F_{\mathcal{R}}(r_2^{\ell'+1}(\bar{x}, \bar{v})) - \min_{j \in F'(\bar{x},\bar{v})} F_{\mathcal{R}}(r_1^j(\bar{x}, \bar{v})) &= -\max\left\{F_{\mathcal{R}}(\min_{j \in S'} r_1^j(\bar{x}, \bar{v})) - F_{\mathcal{R}}(\max_{j \in S} r_2^j(\bar{x}, \bar{v})), 0\right\}, \end{aligned}$$

with $S' = F'(\bar{x}, \bar{v})$, $S = \left\{\ell' + 1\right\}$. Now let S, S' be an arbitrary partition of $F'(\bar{x}, \bar{v}) \cup \left\{\ell' + 1\right\}$. Then as soon as $\ell' + 1 \in S'$, it follows that: $\max\left\{F_{\mathcal{R}}(\min_{j \in S'} r_1^j(\bar{x}, \bar{v})) - F_{\mathcal{R}}(\max_{j \in S} r_2^j(\bar{x}, \bar{v})), 0\right\} = 0$. If $|S| \geq 2$ and $\ell' + 1 \in S$, then, clearly:

$$\begin{aligned} & \max\left\{F_{\mathcal{R}}(\min_{j \in S'} r_1^j(\bar{x}, \bar{v})) - F_{\mathcal{R}}(\max_{j \in S} r_2^j(\bar{x}, \bar{v})), 0\right\} \\ & = \max\left\{F_{\mathcal{R}}(\min_{j \in S'} r_1^j(\bar{x}, \bar{v})) - F_{\mathcal{R}}(\max_{j \in S \setminus \{\ell'+1\}} r_2^j(\bar{x}, \bar{v})), 0\right\} \end{aligned}$$

Combining these derivations with the already derived formula for $e(\bar{x}, \bar{v})$, we thus establish the validity of Eq. 39 for $\ell' + 1$.

In the last situation, we may thus assume the existence of some i for which $r_1^i(\bar{x}, \bar{v}) \leq r_2^{\ell'+1}(\bar{x}, \bar{v})$. We may assume w.l.o.g. that $i = \ell'$. Then setting $\bar{r}_2^{\ell'} = \max\left\{r_2^{\ell'}(\bar{x}, \bar{v}), r_2^{\ell'+1}(\bar{x}, \bar{v})\right\}$ and $\bar{r}_1^{\ell'} = \min\left\{r_1^{\ell'}(\bar{x}, \bar{v}), r_1^{\ell'+1}(\bar{x}, \bar{v})\right\}$, and $\bar{r}_{1,2}^i(\bar{x}, \bar{v}) =$

$r^i_{1,2}(\bar{x}, \bar{v})$ for $i < \ell'$, we may employ the induction assumption to derive (39) holds for $e(\bar{x}, \bar{v})$ but involving $\bar{r}^i_{1,2}$. It is however clear that:

$$\max_{j \in F'(\bar{x},\bar{v}) \cup \{\ell'+1\}} F_{\mathcal{R}}(r^j_2(\bar{x}, \bar{v})) = \max_{j \in F'(\bar{x},\bar{v})} F_{\mathcal{R}}(\bar{r}^j_2(\bar{x}, \bar{v}))$$
$$\min_{j \in F'(\bar{x},\bar{v}) \cup \{\ell'+1\}} F_{\mathcal{R}}(r^j_1(\bar{x}, \bar{v})) = \min_{j \in F'(\bar{x},\bar{v})} F_{\mathcal{R}}(\bar{r}^j_1(\bar{x}, \bar{v})).$$

Once again, let S, S' be an arbitrary partition of $F'(\bar{x}, \bar{v}) \cup \{\ell' + 1\}$. Now should $\ell' \in S'$, $\ell' + 1 \in S$ or vice-versa, then it follows that: $\max\left\{F_{\mathcal{R}}(\min_{j \in S'}(r^j_1(\bar{x}, \bar{v}))) - F_{\mathcal{R}}(\max_{j \in S}(r^j_2(\bar{x}, \bar{v}))), 0\right\} = 0$. For the latter term to contribute it must thus hold that $\ell', \ell' + 1 \in S'$ or $\ell', \ell' + 1 \in S'$, and thus by construction:

$$F_{\mathcal{R}}(\min_{j \in S'}(r^j_1(\bar{x}, \bar{v}))) = F_{\mathcal{R}}(\min_{j \in S' \setminus \{\ell'+1\})}(\bar{r}^j_1(\bar{x}, \bar{v})))$$
$$F_{\mathcal{R}}(\max_{j \in S}(r^j_2(\bar{x}, \bar{v}))) = F_{\mathcal{R}}(\max_{j \in S \setminus \{\ell'+1\}}(\bar{r}^j_2(\bar{x}, \bar{v}))),$$

so that upon combining these derivations, the formula (39) is shown for $\ell' + 1$.

The proof is thus completed, by induction over ℓ'.

□

Let us now revisit our earlier example 4.1 and show that an efficient estimate is obtained.

Example 4.2 (Example 4.1 revisited) By definition $F'(x, v) = \{1, 2\}$. Using notations from formula (39), either $S' = \{2\}$ and $S = \{1\}$, or $S' = \{1\}$ and $S = \{2\}$. Consequently, the formula (39) yields:

$$\begin{aligned} e(x, v) &= \max\left\{\max_{j \in \{1,2\}} F_{\mathcal{R}}(r^j_2(x, v)) - \min_{j \in \{1,2\}} F_{\mathcal{R}}(r^j_1(x, v)), 0\right\} \\ &\quad - \max\left\{F_{\mathcal{R}}(r^1_1(x, v)) - F_{\mathcal{R}}(r^2_2(x, v)), 0\right\} - \max\left\{F_{\mathcal{R}}(r^2_1(x, v)) - F_{\mathcal{R}}(r^1_2(x, v)), 0\right\} \\ &= F_{\mathcal{R}}(r^2_2(x, v)) - F_{\mathcal{R}}(r^1_1(x, v)), \end{aligned}$$

which immediately offers an economic estimate of the sub-differential $\partial^C_x e(x, v)$.

Although up until now we have fully left open the possibility for the probability function φ to be non-smooth, we will briefly mention some conditions under which the probability function will turn out smooth after all.

Theorem 4.2 *Additional to the assumptions of Theorem 4.1, assume that the point $\bar{x} \in \mathbb{R}^n$ is such that the following qualification condition holds:*

$$(b_{i_1,j_1}(\bar{x}) - a_{i_1,j_1}(\bar{x})^{\mathsf{T}}\mathfrak{m})a_{i_2,j_2}(\bar{x})^{\mathsf{T}} \neq (b_{i_2,j_2}(\bar{x}) - a_{i_2,j_2}(\bar{x})^{\mathsf{T}}\mathfrak{m})a_{i_1,j_1}(\bar{x})^{\mathsf{T}}, \quad (40)$$

for all $j_1, j_2 = 1, ..., \ell$ and all $i_1 = 1, ..., p_{j_1}$, $i_2 = 1, ..., p_{j_2}$, such that $(i_1, j_2) \neq (i_2, j_2)$. Then there exists a neighbourhood U of $\bar{x}$ on which φ is continuously differentiable and formula (33) holds as an equality, since at any $x \in U$: $\partial^C_x e(x, v)$ is a singleton for μ_ζ almost all $v \in \mathbb{S}^{m-1}$.

Proof First, let us make the observation that condition (40) can be assumed to hold locally, due to continuity of the underlying data. The neighbourhood exhibited in Theorem 4.1, may thus be assumed to encompass this condition. As a result, our claim will be that for any $(i_1, j_1), (i_2, j_2)$ selected as in the statement of condition (40), it holds, for all $x \in U$ fixed but arbitrary, that:

$$\mu_\zeta(\left\{v \in \mathbb{S}^{m-1} \ : \mathfrak{r}_{i_1,j_1}(x,v) = \mathfrak{r}_{i_2,j_2}(x,v)\right\}) = 0.$$

Indeed, the set of $v \in \mathbb{S}^{m-1}$ at which $\mathfrak{r}_{i_1,j_1}(x,v) = \mathfrak{r}_{i_2,j_2}(x,v)$ holds is

$$M = \left\{v \in \mathbb{S}^{m-1} \ : c_{(i_1,j_1),(i_2,j_2)}(x)^\mathsf{T} L v = 0\right\},$$

with

$$c_{(i_1,j_1),(i_2,j_2)}(x) := (b_{i_1,j_1}(x) - a_{i_1,j_1}(x)^\mathsf{T}\mathfrak{m})a_{i_2,j_2}(x) - (b_{i_2,j_2}(x) - a_{i_2,j_2}(x)^\mathsf{T}\mathfrak{m})a_{i_1,j_1}(x).$$

We may recall (36) to see this. Now by assumption (40), $c_{(i_1,j_1),(i_2,j_2)}(x) \neq 0$ and the claim thus follows from Lemma 2.2 [65] regarding hyperspherical caps. As an immediate consequence, for any $j = 1, ..., \ell$ and μ_ζ almost all $v \in \mathbb{S}^{m-1}$, both $r_1^j(., v)$ and $r_2^j(., v)$ are in fact smooth around x, since they are the finite maximum (minimum respectively) of mappings $\mathfrak{r}_{ij}$, of which the active index set is a singleton. Furthermore also, the set of $v \in \mathbb{S}^{m-1}$ at which $r_2^j(x,v) = r_1^j(x,v)$ for some $j = 1, ..., \ell$ is of μ_ζ zero measure. In formula (39), we may thus rather consider $F'(x,v) = \left\{j \in F(x,v) \ : r_2^j(x,v) > r_1^j(x,v)\right\}$, which as a result of continuity is locally stable. Moreover, in light of the above discussion it is clear that all maxima/minima in (39) have for μ_ζ almost all $v \in \mathbb{S}^{m-1}$ a singleton active set, thus ensuring that e is in fact differentiable with respect to x. Since x was arbitrary, this is also true on the neighbourhood of $\bar{x}$. Now, since $\partial^\mathsf{C}\varphi$ is nonempty by local Lipschitz continuity of φ, on the one hand [60, Proposition 2.1.2], and is contained in the single-valued integral, by (33) and the above, it follows that $\partial^\mathsf{C}\varphi(x)$ coincides with the integral. In particular, $\partial^\mathsf{C}\varphi(x)$ is single-valued, and hence φ is Fréchet differentiable [60, Proposition 2.2.4]. Since this holds for $x \in U$ arbitrarily, continuous differentiability follows from [60, corollary to Proposition 2.2.4].

□

Remark 4.2 Under the assumptions of Theorem 4.2, the computation of $\partial_x^\mathsf{C} e(x,v)$ amounts to the computation of $\partial_x r_1^j(x,v)$ and $\partial_x r_2^j(x,v)$, $j = 1, \ldots, \ell$ (which are singletons). If these assumptions are not met, computing an element of $\partial_x^\mathsf{C} e(x,v)$ requires more work. In case there is a single polyhedron j at hand, computing the values of $r_1^j(x,v)$ and $r_2^j(x,v)$ amounts to computing a maximum and a minimum of list of values obtained from (18) thus resulting in $O(p_j)$ operations. In the general case with multiple polyhedra, there are $2^\ell - 2$ possible S, S' sets in the sum of (39), that verify $|S| + \left|S'\right| \leq \ell$. As a consequence the worst-case complexity of computing (39) belongs to $O(2^\ell \sum_{j=1}^{\ell} p_j)$. Obtaining an element of $\partial_x^\mathsf{C} e(x,v)$ is then straightforward thanks to the preliminary material of Section 2.1. The last operation is the spherical integral over $\mathbb{S}^{m-1}$, of which the complexity depends on m. This part could be treated by Monte-Carlo or Quasi Monte-Carlo sampling, e.g., [66]. We also refer to the discussion in Section 5.1 of [67] on the efficiency of such a scheme and to [68] for an extensive between Monte-Carlo and Quasi Monte-Carlo based sampling.

5 Application

In order to show the relevance of our derived result, we will consider an application taking inspiration from energy management systems. We will consider a system on the discretized time horizon $t = 1, ..., T$. The stylized system contains a water reservoir subject to random inflows (e.g., [69]). The water level in the reservoir is updated as follows:

$$\ell_{t+1} = \ell_t - q_t + a_t,$$

where ℓ_t denotes the level of the reservoir at time t (in m^3), q_t is the amount of water released through the turbine, thus producing electricity, and a_t the random amount of inflows (both are in m^3/s). The previous equation thus has to be understood in an almost sure sense. The water level in the reservoir is subject to bounds, $\underline{\ell}_t$ and $\bar{\ell}_t$, and ideally one has $\ell_t \in [\underline{\ell}_t, \bar{\ell}_t]$ for all $t = 1, ..., T$. We will however assume that the decisions $q_1, ..., q_T$ have to be taken prior to observing $a_1, ..., a_T$ so that one has to give an appropriate meaning to the inequalities. Introducing the vectors $\underline{\ell} = (\underline{\ell}_1, ..., \underline{\ell}_T)$, $\bar{\ell} = (\bar{\ell}_1, ..., \bar{\ell}_T)$ and likewise letting a, q denote the appropriate vectors, the matrix:

$$C = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ 1 & 1 & \cdots & 1 \end{pmatrix},$$

cumulating entries, as well as the vector $\ell_0 = (\ell_0, ..., \ell_0)$ gathering the initial volumes, the system can be written in compact form as:

$$\underline{\ell} - \ell_0 \leq Ca - Cq \leq \bar{\ell} - \ell_0.$$

This random inequality system, will constitute an important element of our, to be specified, probabilistic constraint.

We have yet to comment on the model taken for a. It appears reasonable to consider that inflows are modelled using a log-normal distribution, i.e., $a_t = \exp(\xi_t)$, where ξ_t is a Gaussian random variable (see e.g., comments in [70]). In our case this leads to $a = \exp(\xi)$, where $\xi \sim \mathcal{N}(\mathfrak{m}_a, \Sigma_a)$ so that we can also account for a correlation structure.

The amount of released water q_t interacts with load d_t, energy generated by a wind turbine w_t and conventionally generated energy y_t in the following energy balance equation:

$$d_t - y_t - \rho q_t - w_t \leq 0, t = 1, ..., T$$

thus ensuring that sufficient energy is generated to meet load d_t. The efficiency ρ in MWs/m^3, represents in a stylized fashion, the hydro production function. Furthermore, both d_t and w_t are also subject to uncertainty. Following [71], it appears reasonable to model d_t immediately as Gaussian, and following [72], for an appropriate constant $c > 0$, $w_t = c \max\{\eta_t, 0\}^{\frac{3}{\theta}}$, where $\theta > 0$ is an appropriate parameter and η_t once again Gaussian. Altogether we can thus gather the random elements $(a, d, w) = h(\xi)$, where $\xi \sim \mathcal{N}(\mathfrak{m}, \Sigma)$ is a multivariate Gaussian random vector and $h : \mathbb{R}^{3T} \to \mathbb{R}^{3T}$ will have typical DC structure. The conventional generator has an upper bound on the amount that can be generated $\bar{y}$ and a convex cost function $f : \mathbb{R}^T \to \mathbb{R}_+$.

The full, stylized, model is thus given by:

$$\begin{aligned} \min_{y,q \in \mathbb{R}^T_+} \quad & f(y) \\ \text{s.t.} \quad & y \leq \bar{y}, q \leq \bar{q} \\ & \mathbb{P}[\underline{\ell} - \ell_0 \leq Ca - Cq \leq \bar{\ell} - \ell_0, d_t - y_t - \rho q_t - w_t \leq 0, t = 1, ..., T] \geq p. \end{aligned}$$

The main idea is thus to minimize the costs of conventional generation, while ensuring that, simultaneously the bounds of the reservoir are met, and sufficient energy is generated, all with high probability. The typical DC structure in the argument ξ within this probabilistic constraint is immediately apparent. Indeed, this is the case for w but also for a, that despite being represented by an exponential function in ξ, is involved in bilateral inequalities! For the "w-part", we observe that despite being a convex function in η, w_t is preceded by a minus sign and hence $d_t - w_t$ is of DC type. To fit the setting of our work, we will approximate the various convex mappings with a cutting plane model. We refer to Example 1.1 for the relation to the studied structure.

It is also worthwhile to highlight that the model can account for any possible correlation structure between the various underlying sources of uncertainty. In particular, we thus have a consistent set of inflows, wind generation and load levels!

5.1 Data

For a set of breakpoints $\hat{y}_1, ..., \hat{y}_k$ and increasing proportional costs $c_1 < ... < c_k$, the mapping f is given as follows:

$$f(y) = \sum_{t=1}^{T} \max_{i=1,...,k} c_i y + b_i,$$

with $b_i = \sum_{j=1}^{i-1}(c_j - c_{j+1})\hat{y}_j$. The chosen vector $c = (5, 35, 50, 70, 120)$ and $\hat{y} = (8, 12, 15, 20)$. The maximum generation level $\bar{y} = 25$ and $\bar{q} = 15$. The initial, minimum and maximum reservoir level are respectively $1.5e5$, 0 and $3.5e5$. The time step is set to be 2 hours in length and we consider 12 time steps, i.e., one day. The efficiency coefficient $\rho = 1.4$ and the "wind" coefficients are $c = 2.2/1000$ and $\theta = 0.7$. The latter setting will correspond, roughly, to the consideration of a 4 MW wind turbine. The Gaussian random vectors underlying inflows, wind and load have a persistency effect modelled in their correlation structure: the correlations between adjacent time steps is $0.8, 0.7, 0.6$ respectively. Furthermore the cross-correlations between inflows and "wind" are set to be 0.4, between inflows and load 0.3 and between "wind" and load they are set to 0.1. The mean vector for "inflows" is $(1.3, ..., 1.3)$ and the variance vector is $(0.1, ..., 0.1)$. For wind this is $(4, ..., 4)$ and $(0.5, ..., 0.5)$ respectively. Finally for load the mean vector is $(18, 18, 18, 18, 22.5, 27, 25.2, 21.6, 27, 27, 21.6, 18)$ and the variance is 10% of this mean vector. The latter thus shows the daily variability of load as a result of the economic cycle. It can also be noted that wind, in terms of installed capacity has a relatively large share. Altogether, the underlying Gaussian random vector is in dimension 36. Finally, we set $p = 0.9$ as our target probability level.

5.2 Numerical experiment

For simplicity we have implemented the model in MatLab, and use the basic approximate projected gradient method described in [67, Section 6]. This latter approach requires a feasible point, in order to "project" back the intermediate solutions onto the feasible set. In our particular situation, this feasible point is relatively easy to obtain, since it suffices to consider large generation levels y, having a beneficial effect on satisfying the load inequality. As the turbining levels are concerned, these can be approximately set, by looking at a set of scenarios. A more extensive description of this algorithm can be found in Appendix A.

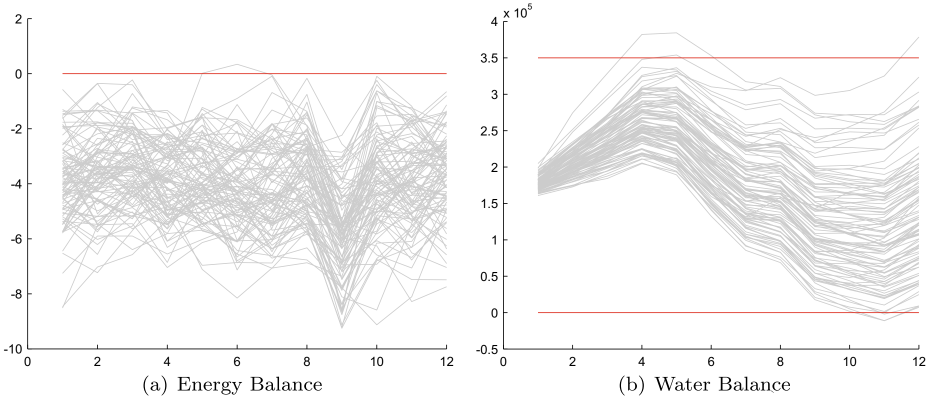

Fig. 2 Illustration of the inequalities within the probabilistic constraint on 100 simulated scenariosx

As a stopping criterion we have set a condition on the distance between the next and current iterate, that they differ less 0.01 MW. The solver thus stops after 5 iterations, producing the local solution $y^* = (20.59, 20.14, 19.90, 21.01, 18.79, 16.42, 15.78, 17.81, 18.21, 23.16, 18.06, 20.55)$,$q^* = (0.131, 0.274, 0.281, 0.027, 4.819, 9.955, 9.095, 4.844, 9.804, 4.939, 4.772, 0.077)$ with objective function cost $1.5524e4$.

Let us finally show the energy balance and water level balance, when generating 100 scenarios a posteriori. This is done on Fig. 2.

One can also evaluate, a posteriori the probability level under these 100 scenarios. It turns out that empirically, under these 100 scenarios, the level is 0.92. This empirically computed probability level is of course linked to the randomly generated 100 scenarios. Evaluating the given rounded solution, through the computation of 50 batches of 1000 scenarios each, allows us to evaluate the empirical probability level at 0.89. This value is confirmed when performing a similar experiment with 10000 scenarios per batch.

This experiment is intended to show that the given formulæ can be readily put to work in otherwise fairly simple software. Of course, should computational efficiency be a concern, then one should orient oneself to a C++ implementation. Evidently, more sophisticated algorithms would also be needed and likewise general purpose approaches to find good starting points. Exploring those themes will be left for future research, and goes largely beyond the scope of the current work.

6 Concluding Remarks

In this paper we have provided insights in generalized differentiation of probabilistic constraints acting on union of polyhedra. Such underlying structure naturally arises when investigating certain energy applications.

Acknowledgements The third author was partially supported by: CONICYT GRANTS: FONDECYT REGULAR 1190110 AND FONDECYT REGULAR 1200283 AND PROGRAMA REGIONAL MATHAMSUD 20-MATH-08 CODE: MATH190003

Appendix

A An approximate projected gradient algorithm

We present in this Appendix a generic algorithm, similar to the one used in [67, Section 6] for solving the following problem:

$$\begin{array}{cc} \min\limits_{x \in X} & f(x) \\ \text{s.t.} & \varphi(x) \geq p, \end{array}$$

where f is assumed to be affine, and φ is as defined in (1). We also assume that a point $x_0 \in X' := \{x \in X \ : \mathbb{P}[g(x,\xi) \leq 0] \geq p\}$ is known (and therefore this set is not empty). Let us define $\tau_{X'}^{x_0} : X \mapsto [0,1]$ as follows:

$$\tau_{X'}^{x_0}(x) = \begin{cases} \min_{t \in [0,1]} & t \\ \quad \text{s.t.} & \mathbb{P}[g((1-t)x + tx_0, \xi) \leq 0] \geq p \end{cases}$$

and let us define $P_{X'}^{x_0} : X \to X'$ as $P_{X'}^{x_0}(x) = (1-\tau_{X'}^{x_0}(x))x + \tau_{X'}^{x_0}(x)x_0$. Since X' is a closed set, $\tau_{X'}^{x_0}$ and thus $P_{X'}^{x_0}$ are well defined. Since, it is not a proper projection, this algorithm is sometimes referred to as an approximate projected gradient algorithm. Convergence of such an algorithm is studied for example in [73]. Obtaining a numerical value for $\tau_{X'}^{x_0}$ can be done using a bisection method, having at each iteration k to compute $\mathbb{P}[g((1-t_k)x + t_k x_0, \xi) \leq 0]$ and stopping when it reaches p up to a given tolerance.

The formulæ (33) can be exploited to compute $\nabla\varphi(x_k)$ at a given trial point upon verifying condition (40).

Algorithm 1 An approximate projected gradient algorithm.

Step 0: Initialization. Let $k = 0$, $x_0 \in X'$, and Tol > 0 a given tolerance.
Step 1: Oracle call. Compute $f(x_k)$, $s_k^f = \nabla f(x_k)$, and $s_k^g = \nabla\varphi(x_k)$.
If $f(x_k) \leq f^{\text{best}}$, then $x^{\text{best}} \leftarrow x_k$ and $f^{\text{best}} \leftarrow f(x_k)$.
Step 2: Descent step. If $\|s_k^g\| = 0$ then define $d_k = -s_k^f$, else define $d_k = -s_k^f + \dfrac{\left\langle s_k^f, s_k^g \right\rangle}{\|s_k^g\|^2} s_k^g$.
Step 3: Line Search. Perform a line search to find θ such that $x'_{k+1} := x_k + \theta d_k \in X$ and achieves descent.
Step 4: Projection step. $x_{k+1} \leftarrow P_{X'}^{x_0}(x'_{k+1})$.
Step 5: Stopping test. If $\|x_{k+1} - x_k\| \leq$ Tol return x^{best} and f^{best} and terminate the algorithm. Else proceed to step 6.
Step 6: Loop. Set $k := k + 1$ and go back to Step 1.

References

1. Prékopa, A.: Stochastic programming. Kluwer, Dordrecht (1995)
2. Prékopa, A.: Probabilistic programming. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, Handbooks in Operations Research and Management Science, vol. 10, pp. 267–351. Elsevier, Amsterdam (2003)

3. Dentcheva, D.: Optimisation models with probabilistic constraints. In: Shapiro, A., Dentcheva, D., Ruszczyński, A. (eds.) Lectures on Stochastic Programming. Modeling and Theory, MPS-SIAM series on optimization, vol. 9, pp. 87–154. SIAM and MPS, Philadelphia (2009)
4. Henrion, R.: Optimierungsprobleme mit wahrscheinlichkeitsrestriktionen: Modelle, struktur, numerik. Lect. Notes, 1–53 (2016)
5. Luedtke, J., Ahmed, S.: A sample approximation approach for optimization with probabilistic constraints. SIAM J. Optim. **19**, 674–699 (2008)
6. Luedtke, J.: An integer programming and decomposition approach to general chance-constrained mathematical programs. In: Eisenbrand, F., Shepherd, F.B. (eds.) Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science, vol. 6080, pp. 271–284. Springer (2010)
7. Luedtke, J., Ahmed, S., Nemhauser, G.L.: An integer programming approach for linear programs with probabilistic constraints. Math. Program. **122**(2), 247–272 (2010)
8. Luedtke, J.: A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. Math. Program. **146**(1-2), 219–244 (2014)
9. Liu, X., Küçükyavuz, S., Luedtke, J.: Decomposition algorithm for two-stage chance constrained programs. Math. Programm. Ser. B **157**(1), 219–243 (2016). https://doi.org/10.1007/s10107-014-0832-7
10. Pagnoncelli, B., Ahmed, S., Shapiro, A.: Sample average approximation method for chance constrained programming: Theory and applications. J. Optim. Theory Appl **142**, 399–416 (2009)
11. van Ackooij, W., Frangioni, A., de Oliveira, W.: Inexact stabilized Benders' decomposition approaches: with application to chance-constrained problems with finite support. Comput. Optim. Appl. **65**(3), 637–669 (2016). https://doi.org/10.1007/s10589-016-9851-z
12. Küçükyavuz, S.: On mixing sets arising in chance-constrained programming. Math. Program. **132**(1-2), 31–56 (2012)
13. Kannan, R., Luedtke, J.: A stochastic approximation method for approximating the efficient frontier of chance-constrained nonlinear programs. Math. Program. Comput., 1–47. https://doi.org/10.1007/s12532-020-00199-y (2020)
14. n-Ordieres, A.P., Luedtke, J.R., Wächter, A.: Solving chance-constrained problems via a smooth sample-based nonlinear approximation. SIAM J. Optim. **30**(3), 2221–2250 (2020)
15. Lejeune, M.A.: Pattern-based modeling and solution of probabilistically constrained optimization problems. Oper. Res. **60**(6), 1356–1372 (2012)
16. Lejeune, M., Margot, F.: Solving chance-constrained optimization problems with stochastic quadratic inequalities. Oper. Res. **64**(4), 939–957 (2016)
17. Dentcheva, D., Prékopa, A., Ruszczyński, A.: Concavity and efficient points for discrete distributions in stochastic programming. Math. Program. **89**, 55–77 (2000)
18. Dentcheva, D., Lai, B., Ruszczyński, A.: Dual methods for probabilistic optimization problems. Math. Methods Oper. Res. **60**(2), 331–346 (2004)
19. Dentcheva, D., Martinez, G.: Augmented lagrangian method for probabilistic optimization. Ann. Oper. Res. **200**(1), 109–130 (2012)
20. Dentcheva, D., Martinez, G.: Regularization methods for optimization problems with probabilistic constraints. Math. Programm. (Ser. A) **138**(1-2), 223–251 (2013)
21. Lejeune, M.A.: Pattern definition of the p-efficiency concept. Ann. Oper. Res. **200**, 23–36 (2012)
22. Lejeune, M.A., Noyan, N.: Mathematical programming approaches for generating p-efficient points. Eur. J. Oper. Res. **207**(2), 590–600 (2010)
23. van Ackooij, W., Berge, V., de Oliveira, W., Sagastizábal, C.: Probabilistic optimization via approximate p-efficient points and bundle methods. Comput. Oper. Res. **77**, 177–193 (2017). https://doi.org/10.1016/j.cor.2016.08.002
24. Ben-Tal, A., Nemirovski, A.: Robust solutions of linear programming problems contaminated with uncertain data. Math. Programm. Ser. A **88**, 411–424 (2000)
25. Ben-Tal, A., Ghaoui, L.E., Nemirovski, A.: Robust optimization. Princeton University Press (2009)
26. Ermoliev, Y.M., Ermolieva, T.Y., Macdonald, G.J., Norkin, V.I.: Stochastic optimization of insurance portfolios for managing exposure to catastrophic risk. Ann. Oper. Res. **99**, 207–225 (2000)
27. Calafiore, G.C., Campi, M.C.: The scenario approach to robust control design. IEEE Trans. Automat. Control **51**, 742–753 (2006)
28. Campi, M.C., Garatti, S.: A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality. J. Optim. Theory Appl. **148**(2), 257–280 (2011)
29. Nemirovski, A., Shapiro, A.: Convex approximations of chance constrained programs. SIAM J. Optim. **17**(4), 969–996 (2006)
30. Hong, L.J., Yang, Y., Zhang, L.: Sequential convex approximations to joint chance constrained programed: A monte carlo approach. Oper. Res. **3**(59), 617–630 (2011)

31. Geletu, A., Hoffmann, A., Klöppel, M., Li, P.: A tractable approximation of non-convex chance constrained optimization with non-gaussian uncertainties. Eng. Optim. **47**(4), 495–520 (2015)
32. Hu, Z., Hong, L.J., Zhang, L.: A smooth monte carlo approach to joint chance-constrained programs. IIE Trans. **45**(7), 716–735 (2013)
33. Tuy, H.: Convex analysis and global optimization, Nonconvex Optimization and Its Applications, vol. 22. Springer (2016)
34. Garnier, J., Omrane, A., Rouchdy, Y.: Asymptotic formulas for the derivatives of probability functions and their Monte Carlo estimations. Eur. J. Oper. Res. **198**, 848–858 (2009). https://doi.org/10.1016/j.ejor.2008.09.026
35. Kibzun, A.I., Uryas'ev, S.: Differentiability of probability function. Stoch. Anal. Appl. **16**, 1101–1128 (1998). https://doi.org/10.1080/07362999808809581
36. Marti, K.: Differentiation of probability functions : The transformation method. Comput. Math. Appl. **30**, 361–382 (1995). https://doi.org/10.1016/0898-1221(95)00113-1
37. Pflug, G., Weisshaupt, H.: Probability gradient estimation by set-valued calculus and applications in network design. SIAM J. Optim. **15**, 898–914 (2005). https://doi.org/10.1137/S1052623403431639
38. Raik, E.: The differentiability in the parameter of the probability function and optimization of the probability function via the stochastic pseudogradient method (russian). Izvestiya Akad. Nayk Est. SSR, Phis. Math. **24**(1), 3–6 (1975)
39. Royset, J.O., Polak, E.: Implementable algorithm for stochastic optimization using sample average approximations. J. Optim. Theory Appl. **122**(1), 157–184 (2004). https://doi.org/10.1023/B:JOTA.0000041734.06199.71
40. Royset, J.O., Polak, E.: Extensions of stochastic optimization results to problems with system failure probability functions. J. Optim. Theory Appl. **133**(1), 1–18 (2007). https://doi.org/10.1007/s10957-007-9178-0
41. Uryas'ev, S.: Derivatives of probability functions and integrals over sets given by inequalities. J. Comput. Appl. Math. **56**(1-2), 197–223 (1994). https://doi.org/10.1016/0377-0427(94)90388-3
42. Uryas'ev, S.: Derivatives of probability functions and some applications. Ann. Oper. Res. **56**, 287–311 (1995). https://doi.org/10.1007/BF02031712
43. van Ackooij, W., Henrion, R.: Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions. SIAM J. Optim. **24**(4), 1864–1889 (2014). https://doi.org/10.1137/130922689
44. van Ackooij, W., Henrion, R.: (Sub-) Gradient formulae for probability functions of random inequality systems under Gaussian distribution. SIAM J. Uncertain. Quant. **5**(1), 63–87 (2017). https://doi.org/10.1137/16M1061308
45. van Ackooij, W., Aleksovska, I., Zuniga, M.M.: (sub-)differentiability of probability functions with elliptical distributions. Set Valued Var. Anal. **26**(4), 887–910 (2018). https://doi.org/10.1007/s11228-017-0454-3
46. van Ackooij, W., Malick, J.: Second-order differentiability of probability functions. Optim. Lett. **11**(1), 179–194 (2017). https://doi.org/10.1007/s11590-016-1015-7
47. Hantoute, A., Henrion, R., Pérez-Aros, P.: Subdifferential characterization of continuous probability functions under Gaussian distribution. Math. Program. **174**(1-2), 167–194 (2019). https://doi.org/10.1007/s10107-018-1237-9
48. van Ackooij, W., Pérez-Aros, P.: Generalized differentiation of probability functions acting on an infinite system of constraints. SIAM J. Optim. **29**(3), 2179–2210 (2019)
49. van Ackooij, W., Henrion, R., Pérez-Aros, P.: Generalized gradients for probabilistic/robust (probust) constraints. Optimization **69**(7-8), 1451–1479 (2020). https://doi.org/10.1080/02331934.2019.1576670
50. van Ackooij, W., Pérez-Aros, P.: Gradient formulae for nonlinear probabilistic constraints with non-convex quadratic forms. J. Optim. Theory Appl. **185**(1), 239–269 (2020). https://doi.org/10.1007/s10957-020-01634-9
51. van Ackooij, W., Pérez-Aros, P.: Generalized differentiation of probability functions: parameter dependent sets given by intersections of convex sets and complements of convex sets. Working paper (2021)
52. Fang, K., Kotz, S., Ng, K.W.: Symmetric multivariate and related distributions, 1st edn. Monographs on Statistics and Applied Probability, vol. 36. Springer-Science (1990)
53. Landsman, Z.M., Valdez, E.A.: Tail conditional expectations for elliptical distributions. North Amer. Actuarial J. **7**(4), 55–71 (2013). https://doi.org/10.1080/10920277.2003.10596118
54. Elstrodt, J.: Maß und integrationstheorie, 7th edn. Springer (2011)
55. Farshbaf-Shaker, M.H., Henrion, R., Hömberg, D.: Properties of chance constraints in infinite dimensions with an application to pde constrained optimization. Set Valued Var. Anal. **26**(4), 821–841 (2018). https://doi.org/10.1007/s11228-017-0452-5

56. van Ackooij, W., Minoux, M.: A characterization of the subdifferential of singular Gaussian distribution functions. Set Valued Var. Anal. **23**(3), 465–483 (2015). https://doi.org/10.1007/s11228-015-0317-8
57. Henrion, R., Römisch, W.: Lipschitz and differentiability properties of quasi-concave and singular normal distribution functions. Ann. Oper. Res. **177**, 115–125 (2010). https://doi.org/10.1007/s10479-009-0598-0
58. Henrion, R., Möller, A.: A gradient formula for linear chance constraints under Gaussian distribution. Math. Oper. Res. **37**, 475–488 (2012). https://doi.org/10.1287/moor.1120.0544
59. van Ackooij, W., Henrion, R., Möller, A., Zorgati, R.: On joint probabilistic constraints with Gaussian Coefficient Matrix. Oper. Res. Lett. **39**, 99–102 (2011). https://doi.org/10.1016/j.orl.2011.01.005
60. Clarke, F.H.: Optimisation and nonsmooth analysis, Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (1987)
61. Jofre, A., Thibault, L.: D-representation of subdifferentials of directionally lipschitz functions. Proc. Amer. Math. Soc. **110**(1), 117–123 (1990)
62. Correa, R., Hantoute, A., Pérez-Aros, P.: Subdifferential calculus rules for possibly nonconvex integral functions. SIAM J. Control Optim. **58**(1), 462–484 (2020). https://doi.org/10.1137/18M1176476
63. Henrion, R.: Calmness as a constraint qualification for M-stationarity conditions in MPECs. In: Aussel, D., Latitha, C. (eds.) Generalized nash equilibrium problems, bilevel programming and MPEC, Forum for Interdisciplinary Mathematics, pp. 21–41. Springer Singapore (2017)
64. Surowiec, T.M.: Explicit stationarity conditions and solution characterization for equilibrium problems with equilibrium constraints. Ph.D. Thesis, Humboldt-Universität zu Berlin (2010)
65. van Ackooij, W., Malick, J.: Eventual convexity of probability constraints with elliptical distributions. Math. Program. **175**(1), 1–27 (2019). https://doi.org/10.1007/s10107-018-1230-3
66. Kuo, F.Y., Sloan, I.H.: Quasi-Monte Carlo methods can be efficient for integration over products of spheres. J. Complex. **21**, 196–210 (2005)
67. Heitsch, H.: On probability capacity maximization in a stationary gas network. Optimization **69**(3), 575–604 (2020). https://doi.org/10.1080/02331934.2019.1625353
68. Gotzes, C., Heitsch, H., Henrion, R., Schultz, R.: On the quantification of nomination feasibility in stationary gas networks with random loads. Math. Methods Oper. Res. **84**, 427–457 (2016). https://doi.org/10.1007/s00186-016-0564-y
69. van Ackooij, W., Henrion, R., Möller, A.s., Zorgati, R.: Joint chance constrained programming for hydro reservoir management. Optim. Eng. **15**, 509–531 (2014). https://doi.org/10.1007/s11081-013-9236-4
70. Shapiro, A., Tekaya, W., da Costa, J.P., Soares, M.P.: Risk neutral and risk averse stochastic dual dynamic programming method. Eur. J. Oper. Res. **224**(2), 375–391 (2013). https://doi.org/10.1016/j.ejor.2012.08.022
71. Bruhns, A., Deurveilher, G., Roy, J.S.: A non-linear regression model for mid-term load forecasting and improvements in seasonality. PSCC 2005 Luik (2005)
72. Bremer, I., Henrion, R., Möller, A.: Probabilistic constraints via SQP solver: Application to a renewable energy management problem. Comput. Manag. Sci. **12**, 435–459 (2015)
73. Hager, W.W., Phan, D.T., Zhu, J.: Projection algorithms for nonconvex minimization with application to sparse principal component analysis. J. Glob. Optim. **65**(4), 657–676 (2016). https://doi.org/10.1007/s10898-016-0402-z