LIPSCHITZ GRADIENT GUARANTEES FOR PROBABILITY FUNCTIONS AND A NEW ALGORITHM FOR PROBABILITY MAXIMIZATION

WIM VAN ACKOOIJ, CAROLINA CHIU, WELINGTON DE OLIVEIRA, AND PEDRO PÉREZ-AROS

Abstract. This work studies probability functions that appear in stochastic programming models. Although their differentiability has been widely investigated, the Lipschitz continuity of their gradients, crucial for the design and analysis of modern optimization algorithms, has received little attention. We develop a general framework that ensures differentiability and gradient Lipschitz continuity under practical conditions. Our framework unifies and extends existing results and applies to a broad class of continuous distributions. Building on this theory, we propose a specialized method for a class of probability maximization problems. Our approach handles general distributions and enjoys convergence guarantees. Numerical experiments against state-of-the-art methods demonstrate clear gains in accuracy, efficiency, and scalability.

1. Introduction

Let $G\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ be a given differentiable mapping, and $\xi \in \mathbb{R}^m$ be a random vector with continuous probability measure $\mathbb{P}$. In this work we are concerned with the associated *probability function*

$$\varphi(x) := \mathbb{P}\big(G(x,\xi) \leq 0\big), \tag{1}$$

which quantifies the probability that the decision vector $x \in \mathbb{R}^n$ jointly satisfies the random system of p inequalities $G(x,\xi) \leq 0$. Many decision-making problems from operations research, economics, and finance can be cast as optimization problems involving one or more probability functions of the form (1); see for instance the textbooks [27, 32, 41]. From a practical standpoint, incorporating a probability function into an optimization model provides a convenient and effective way to account for uncertainties when reliability is a key concern. This approach is particularly relevant when no recourse is available, that is, when no corrective actions on a decision x can be taken after the realization of the random event. In such settings, the decision must be made before the uncertainty is revealed, and its reliability becomes essential. Consequently, probability functions offer a natural and powerful modeling strategy for decision-making under uncertainty.

Depending on the application and decision-maker's preference, the probability function may appear either in the objective or in the constraints of the underlying optimization model. Using it as objective function leads to *probability maximization problems* [27, § 8.2],[14, 7, 3], whereas incorporating it into the constraints gives rise to the well known class of *chance-constrained problems* [27, § 8.2], [4, 18]. Both probability maximization and chance-constrained optimization models are well established classes of problems that have been studied since the early days of stochastic programming [4].

From an algorithmic viewpoint, explicit knowledge of the mathematical properties of φ such as differentiability is pivotal. For instance, several off-the-shelf and dedicated solvers for probability

2020 *Mathematics Subject Classification.* 49J53, 90C15, 49J52, 90C31, 65K10.

Key words and phrases. Stochastic optimization, Variational analysis, Probability functions, Probability Maximization.

maximization and chance-constrained problems [13, 17, 7, 3, 23, 24] require at least that φ be differentiable. For this reason, several works [28, 33, 34, 27, 22, 20, 30, 26, 31, 15] investigate differentiability of probability functions. In many of these contributions, gradient formulæ are expressed through involved surface and/or volume integrals. To the best of our knowledge, numerical implementations of these formulæ have not been documented.

A different and fruitful line of investigation for obtaining workable gradient formulæ relies on the so-called spherical-radial decomposition of the underlying random vector ξ; see [37]. This approach sacrifices some generality with respect to the distribution of ξ, yet gains generality on other aspects and leads to several practical gradient expressions. Indeed, under the assumption that G is convex in the second argument and ξ elliptically symmetric, differentiability of φ was obtained under a mild growth condition on ∇G [35, 37]. Similar subdifferential formulæ have also been established in comparable settings; see [36, 16]. More recent contributions have sought to relax the convexity assumption and have developed a spherical-radial-like decomposition for nearly arbitrary random vectors, thus extending the scope of available gradient formulæ, e.g., [39, 40]. In some cases, it is even possible to establish second-order expressions for φ, potentially enabling Newton-type methods [35], although these expressions may be computationally demanding in high dimensions.

Despite extensive investigations on (sub)differentiability and the various (sub)gradient formulæ being available for probability functions, the specific question of the Lipschitz continuity of the gradient of φ has, to the best of our knowledge, not received dedicated attention in the literature. We highlight that investigating whether φ is continuously differentiable with a Lipschitz continuous gradient (abbreviated as $C^{1,1}$) is not merely of theoretical interest, but a practical requirement for deploying efficient and reliable optimization algorithms for both probability-maximization and chance-constrained problems. Indeed, several optimization methods, ranging from (proximal) gradient schemes and coordinate-descent algorithms to Frank-Wolfe and quasi-Newton methods, require the involved function to be $C^{1,1}$. Convergence guarantees for many of these schemes hinge on sufficient decrease conditions obtained via the classical *descent lemma*, which requires differentiability with a Lipschitz continuous gradient (at least locally). This requirement also appears in certain specialized algorithms, such as the probability-maximization method introduced in [3]. Interestingly, although the method in [3] is derivative-free (and thus does not require computing gradients of the probability function), its convergence analysis still relies on the assumption that φ is $C^{1,1}$.

In this work, we provide a framework that ensures that φ is differentiable and has a Lipschitz continuous gradient under readily verifiable assumptions on the constraint mapping G and on the distribution of ξ. Our results unify and extend existing differentiability and subdifferential formulæ, cover a wide range of continuous distributions beyond the Gaussian case, and yield concrete Lipschitz bounds that can be used in algorithm design and complexity analysis.

A second major contribution of this work is the development of a new specialized algorithm for probability-maximization problems in which the uncertainties appear exclusively in the right-hand side of the system of inequalities. In contrast to most specialized methods [14, 13, 12, 7, 17], which require the probability to be related to a (multivariate) Gaussian distribution, our approach accommodates more general distributions. Under the assumption that φ is locally $C^{1,1}$, we establish convergence guarantees to stationary points without requiring the explicit value of the corresponding Lipschitz constant. Our method builds upon a local yet accurate model of φ that directly exploits the marginal distributions of ξ and incorporates the full structure of the mapping G without resorting to linearizations. This yields a model that is both faithful to the underlying probability function and computationally tractable, allowing the algorithm to make efficient search steps. We complement the theoretical results with a numerical study comparing our approach against state-of-the-art methods, including sequential quadratic programming (SQP), Frank-Wolfe, projected BFGS variants, and other established optimization algorithms. The numerical evidence highlights the efficiency and robustness of our method, demonstrating superior performance in terms of accuracy, computational cost, and scalability. These results further emphasize the practical importance of the $C^{1,1}$ regularity in enabling the design of specialized algorithms for optimization problems involving probability functions.

The rest of the paper is organized as follows. Section 2 recalls some growth conditions and constraint qualifications associated with the probability function (1). The new theoretical results concerning the $C^{1,1}$ regularity of this probability function are presented in Section 3. The technical details and proofs supporting these results are collected in Section 4, which begins by reviewing several known facts that are instrumental for our analysis. This section may be skipped by readers who prefer to bypass the more involved details, without compromising the flow of the paper. Section 5 introduces our new algorithm for probability maximization problems, together with its convergence analysis and convergence rate guarantees. Finally, Section 6 closes the work with numerical experiments that compare our algorithm with standard nonlinear optimization methods.

Notation and Terminology. We work in the Euclidean space $\mathbb{R}^n$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. For $x \in \mathbb{R}^n$ and $r > 0$, the open and closed balls are defined as $\mathbb{B}_r(x) := \{z \in \mathbb{R}^n : \|z - x\| < r\}$ and $\mathbb{B}_r[x] := \{z \in \mathbb{R}^n : \|z - x\| \leq r\}$. The unit sphere in $\mathbb{R}^m$ is $\mathbb{S}^{m-1} := \{z \in \mathbb{R}^m : \|z\| = 1\}$. We use the notation ξ to represent a (continuous) random vector in $\mathbb{R}^m$. Its density function is denoted by $f_\xi : \mathbb{R}^m \to [0, \infty)$. A function $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ has an epigraph $\operatorname{epi} f = \{(x, \alpha) \in \mathbb{R}^{n+1} \mid f(x) \leq \alpha\}$. It is lower semicontinuous (lsc) if $\operatorname{epi} f$ is a closed subset of $\mathbb{R}^n \times \mathbb{R}$ and it is convex if $\operatorname{epi} f$ is convex. It is proper if $\operatorname{epi} f$ is nonempty and $f(x) > -\infty$ for all $x \in \mathbb{R}^n$. A function f is said to be $C^{1,1}$ on a set U if it is continuously differentiable on U and its gradient is Lipschitz continuous; if this holds on $\mathbb{R}^n$, we simply call f a $C^{1,1}$ function. If U is a neighbourhood of a point $\bar{x}$, then f is said to be $C^{1,1}$ around or near $\bar{x}$. For $h \colon \mathbb{R}^n \times \mathbb{S}^{m-1} \to \mathbb{R}$ that is $C^{1,1}$ in x, the modulus of Lipschitz continuity of $\nabla_x h$ with respect to x at $(\bar{x}, \bar{v})$ is $\operatorname{lip} \nabla_x h(\bar{x}, \bar{v}) := \limsup_{x, x', v} \frac{\|\nabla h_x(x', v) - \nabla h_x(x, v)\|}{\|x' - x\|}$, where the lim sup is taken with $x, x' \to \bar{x}, x \neq x'$, and $v \to \bar{v}$. The Fréchet normal cone to $X \subset \mathbb{R}^n$ at $x \in X$ is denoted by $\widehat{N}_X(x)$ and consists of those $x^* \in \mathbb{R}^n$ satisfying $\langle x^*, x' - x \rangle \leq o(\|x' - x\|)$ for all $x' \in X$. The Mordukhovich (or limiting) normal cone to X at $x \in X$ is $N_X(x) = \{x^* \in \mathbb{R}^n \mid \exists v^\nu \to x^*, x^\nu \in X \to x \text{ with } v^\nu \in \widehat{N}_X(x^\nu)\}$. These cones are empty when $x \notin X$. The subdifferential of a lsc function $f : \mathbb{R}^n \to \mathbb{R}$ at a point x is $\partial f(x) = \{v \in \mathbb{R}^n \mid (v, -1) \in N_{\operatorname{epi} f}(x, f(x))\}$. Finally, the symmetric subdifferential of f at x is $\partial^0 f(x) := \partial f(x) \cup [-\partial(-f)(x)]$.

2. Background: growth conditions and constraint qualification

Here we present some basic definitions that will be useful to understand the following sections. In what follows, the mapping G appearing in (1) will need to be well-behaved with respect to the density f_ξ of ξ and/or qualified. To this end, we will recall the growth condition, which is essential to control the behaviour of the partial derivative $\nabla_x G$ for large values of the argument z. A variety of growth conditions can be presented, depending on the structural properties of the function G and the distribution of the random variable ξ. For example, in the case of Gaussian-like distributions, the authors of [37] employ a polynomial growth condition, while in [36], an exponential growth condition is shown to be sufficient. The point being that a balance has to be struck between the ease of presenting the growth condition and its generality.

Definition 1 (Growth condition). *Let $\eta\colon \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing function such that*

$$\lim_{\|z\|\to\infty} f_\xi(z)\,\|z\|^m\,\eta(\|z\|) = 0. \tag{2}$$

We say that a family of mappings $\{h_i\}_{i=1}^p$ satisfies the η-first-order growth condition at $\bar{x}$ if for some $C > 0$ and neighbourhood U of $\bar{x}$ it holds that

$$\|\nabla_x h_i(x,z)\| \le \eta(\|z\|),\ \text{for all } z \text{ such that } \|z\| \ge C, \text{for all } i = 1,\dots,p. \tag{3}$$

When there is no risk of confusion, we simply say that $\{h_i\}_{i=1}^p$ satisfies the first-order growth condition. *Moreover, if $\{h_i\}_{i=1}^p$ reduces to a single function h, we simply say that h satisfies the* first-order growth condition *around $\bar{x}$.*

Remark 1 (Exponential growth condition). *If $\xi \sim \mathcal{N}(\mu,\Sigma)$ is multivariate Gaussian, it is sufficient to adopt a growth function of the form $\eta(t) = a\exp(bt)$, for some $a, b > 0$. This requirement is referred to as the* exponential growth condition *(see, e.g.,* [37, 36]*).*

Moreover, we impose a growth condition on the gradient, formulated in terms of Lipschitz continuity.

Definition 2 ((1,1)-smooth growth condition). *Let $\eta\colon \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing function satisfying* (2) *and such that, for any $\delta > 0$,*

$$\lim_{\|z\|\to\infty} \|\nabla f_\xi(z)\|\,\|z\|^{m+1}\,\eta^2(\|z\|) = 0. \tag{4}$$

Assume moreover that there exists a second increasing, upper semicontinuous function $\tilde{\eta}\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lim_{\|z\|\to\infty} f_\xi(z)\,\|z\|^{m+i}\,\eta^i(\|z\|)\,\tilde{\eta}^j(\|z\|) = 0, \tag{5}$$

for all $i = 0, 1, 2$ and $j = 0, 1$. Let $\{h_i\}_{i=1}^p$ be a family of mappings of class $C^{1,1}$. We say that $\{h_i\}_{i=1}^p$ satisfies the $(\eta,\tilde{\eta})$-$(1,1)$-smooth growth condition around $\bar{x}$ if for each $i = 1,\dots,p$, h_i satisfies the first-order growth condition involving η and for some neighbourhood U of $\bar{x}$ it holds that for all $x_1, x_2 \in U$ and all $z_1, z_2 \in \mathbb{R}^m$

$$\|\nabla h_i(x_1,z_1) - \nabla h_i(x_2,z_2)\| \le \tilde{\eta}(\max\{\|z_1\|,\|z_2\|\})\|(x_1,z_1)-(x_2,z_2)\|, \text{ for all } i = 1,\dots,p. \tag{6}$$

When there is no risk of confusion about the roles of η and $\tilde{\eta}$, we simply say that h satisfies the $(1,1)$-smooth growth condition *around $\bar{x}$. Moreover, if $\{h_i\}_{i=1}^p$ reduces to a single function h, we simply say that h satisfies the* $(1,1)$-smooth growth condition *around $\bar{x}$.*

Remark 2 (Exponential $(1,1)$-smooth growth condition). *A concrete example of functions η and $\tilde{\eta}$ satisfying the above requirements when $\xi \sim \mathcal{N}(\mu,\Sigma)$ is multivariate Gaussian is $\eta(t) = \tilde{\eta}(t) = a\exp(bt)$, for some constants $a, b > 0$. We refer to this property as the* Exponential $(1,1)$-smooth growth condition.

The final tool we will need in the following section is a simple condition ensuring differentiability of the joint probability function (1). This is the so-called *Rank-2 Constraint Qualification (R2CQ)*, originally presented in [36] and subsequently used in several works as a condition for differentiability of joint probability functions.

Definition 3 (Rank-2 Constraint Qualification). *We say that the inequality system $G(x,z) \le 0$ satisfies the Rank-2 Constraint Qualification (R2CQ) at $x \in \mathbb{R}^n$ if*

(R2CQ)

$$\operatorname{rank}\{\nabla_z G_j(x,z),\, \nabla_z G_i(x,z)\} = 2, \quad \forall\, i \ne j \in \mathcal{I}(x,z), \quad \forall\, z \in \mathbb{R}^m \text{ with } \max_{i=1,\dots,p} G_i(x,z) = 0,$$

where, for $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$, we define

$$\mathcal{I}(x,z) := \left\{\, i \in \{1,\dots,p\} \;\middle|\; G_i(x,z) = 0 \,\right\}. \tag{7}$$

Moreover, we say that R2CQ holds around x *if there exists a neighbourhood U of x such that R2CQ holds at every point $y \in U$.*

3. New insights on $C^{1,1}$ regularity of probability functions

In this section, we present the main theoretical contributions of the paper, namely the Lipschitz continuity of the gradient of the probability function. We focus on the case of joint probability functions as defined in (1). Our results are established under the following assumptions on the underlying nominal data.

Assumption 1. *The constraint mapping $G\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ and random vector $\xi \in \mathbb{R}^m$ satisfy:*

i) for each $i = 1,\dots,p$, the mapping $G_i\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is of class $C^{1,1}$;
ii) for every $x \in \mathbb{R}^n$ and every $i = 1,\dots,p$, the function $G_i(x,\cdot)$ is convex;
iii) the random vector ξ admits a continuously differentiable density f_ξ.

The following main differentiability result can be derived under these assumptions, the proof of which can be found in Section 4.2.2.

Theorem 3.1 (Main result)**.** *In addition to Assumption 1, suppose that for a fixed $\bar{x} \in \mathbb{R}^n$:*

i) $G_i(\bar{x},0) < 0$ for all $i = 1,\dots,p$;
ii) $\{G_i\}_{i=1}^p$ satisfies the $(1,1)$*-smooth growth condition* (6) *around $\bar{x}$;*
iii) $\{G_i\}_{i=1}^p$ satisfies (R2CQ) at $\bar{x}$ (see Definition 3).

Then the probability function φ, given in (1)*, is differentiable and its gradient is Lipschitz continuous around $\bar{x}$.*

It is important to notice that the second hypothesis of the theorem can be replaced by a compactness assumption instead, as proven in Subsection 4.2.2.

Corollary 3.2. *The conclusion of Theorem 3.1 remains true if the* $(1,1)$*-smooth growth condition is replaced by the condition that the set $\{z : G(\bar{x},z) \le 0\}$ is bounded.*

We further present two relevant settings in which the Slater condition involving zero in Theorem 3.1 can be relaxed. The first corresponds to the case where the random vector follows a Gaussian distribution and an exponential growth condition is imposed.

Corollary 3.3. *Under Assumption 1, suppose that $\xi \sim \mathcal{N}(\mu,\Sigma)$. Let $\bar{x} \in \mathbb{R}^n$ be such that:*

i) there exists $z_0 \in \mathbb{R}^m$ with $G_i(\bar{x},z_0) < 0$ for all $i = 1,\dots,p$;
ii) the family $\{G_i\}_{i=1}^p$ satisfies the exponential $(1,1)$*-smooth growth condition at $\bar{x}$;*
iii) the family $\{G_i\}_{i=1}^p$ satisfies (R2CQ) around $\bar{x}$.

Then φ is differentiable at $\bar{x}$, and its gradient $\nabla\varphi$ is Lipschitz continuous in a neighbourhood of $\bar{x}$.

Proof. Let $z_0 \in \mathbb{R}^m$ that satisfies $G_i(\bar{x},z_0) < 0$ for all $i = 1,\dots,p$ be given. By continuity, there exists a neighbourhood U of $\bar{x}$ such that $G_i(x,z_0) < 0$ for all $x \in U$ and all $i = 1,\dots,p$. Now, let us set $\tilde{\xi} = \xi - z_0$ (which is also a Gaussian) and define $\tilde{G}_i(x,z) := G_i(x,z+z_0)$ with $i = 1,\dots,p$ and $\tilde{G}(x,z) = G(x,z+z_0)$, we have that $\varphi(x) = \mathbb{P}(G(x,\xi) \le 0) = \mathbb{P}(\tilde{G}(x,\tilde{\xi}) \le 0)$. Moreover, $\tilde{G}_i(\bar{x},0) < 0$ for all $i = 1,...,p$. Now from the triangle inequality and the fact that the exponential function is increasing, we observe that for any $z \in \mathbb{R}^m$, we have $a\exp(b\,\|z - z_0\|) \le (a\exp(b\,\|z_0\|)\exp(b\,\|z\|) = \tilde{a}\exp(b\,\|z\|)$. Now if G satisfies condition (3) with the exponential growth condition, then the argument above shows that the same holds true for $\tilde{G}$ with a modified constant a. This however does not impact condition (2). The same analysis can be made for the exponential $(1,1)$-smooth growth condition. And clearly (R2CQ) carries over. Then, the result follows by Theorem 3.1. □

The second case is particularly relevant for the forthcoming analysis and corresponds to the separable case, i.e., probability function with right-hand side uncertainty.

Corollary 3.4. *Consider* (1) *with $G(x,\xi) = Q\xi - h(x)$, where $Q \in \mathbb{R}^{p\times m}$, ξ has a continuously differentiable density f_ξ, and $h\colon \mathbb{R}^n \to \mathbb{R}^p$ is a mapping. Assume that any two distinct rows of Q are linearly independent and that*

$$\lim_{\|z\|\to\infty} \|\nabla f_\xi(z)\|\, \|z\|^{m+1} = 0 \quad \textit{and} \quad \lim_{\|z\|\to\infty} f_\xi(z)\, \|z\|^{m+2} = 0. \tag{8}$$

Let $\bar{x} \in \mathbb{R}^n$ be such that there exists $z_0 \in \mathbb{R}^m$ with $Qz_0 < h(\bar{x})$ (understood for each component), and suppose that h is $C^{1,1}$ around $\bar{x}$. Then φ is $C^{1,1}$ around $\bar{x}$.

Proof. By a standard localization argument, we may construct a function $\tilde{h}$ that coincides with h on a neighbourhood of $\bar{x}$ and is $C^{1,1}$ on $\mathbb{R}^n$. Replacing h by $\tilde{h}$ if necessary, we may therefore assume without loss of generality that h is $C^{1,1}$ on $\mathbb{R}^n$. With z_0 as in the assumptions, let $\tilde{\xi} := \xi - z_0$ and define $G(x,z) := Q(z+z_0) - h(x) \in \mathbb{R}^p$. Then $\varphi(x) = \mathbb{P}\big(G(x,\tilde{\xi}) \le 0\big)$. We verify the assumptions of Theorem 3.1. First, since $h \in C^{1,1}$, the mapping G is $C^{1,1}$ on $\mathbb{R}^n \times \mathbb{R}^m$. For each fixed x, the map $z \mapsto G(x,z)$ is affine (hence convex). The density of $\tilde{\xi}$ is $f_{\tilde{\xi}}(z) = f_\xi(z+z_0)$, which is continuously differentiable because f_ξ is. Next, at the point $(\bar{x},0)$ we have $G(\bar{x},0) = Qz_0 - h(\bar{x}) < 0$ componentwise, so the Slater condition holds. Moreover, $\nabla_z G(x,z) = Q$, and therefore the regularity condition (R2CQ) in Definition 3 holds at any x by the linear independence of the rows of Q. Particularly, (R2CQ) holds around $\bar{x}$. Because $h \in C^{1,1}$, there exist a neighbourhood U of $\bar{x}$ and a constant $\kappa > 0$ such that, for all $x_1, x_2 \in U$ and all $z_1, z_2 \in \mathbb{R}^m$, $\|\nabla_x G(x,z)\| \le \kappa$ and $\|\nabla G(x_1,z_1) - \nabla G(x_2,z_2)\| \le \kappa\, \|(x_1,z_1) - (x_2,z_2)\|$. Hence the $(1,1)$-smooth growth condition holds on U with $\eta(t) = \tilde{\eta}(t) \equiv \kappa$. Since $f_{\tilde{\xi}}(z) = f_\xi(z+z_0)$, the tail limits assumed for f_ξ and ∇f_ξ transfer to $f_{\tilde{\xi}}$ and $\nabla f_{\tilde{\xi}}$. All the hypotheses of Theorem 3.1 are thus satisfied in a neighbourhood of $\bar{x}$, and the conclusion follows: φ is $C^{1,1}$ around $\bar{x}$. □

4. Technical results and proofs of the $C^{1,1}$ regularity

In this section, we present the technical results underlying Theorem 3.1 stated in the previous section. We first develop the analytical tools required in the scalar case $p = 1$, and then extend the arguments to the vector-valued setting $p > 1$ for the mapping $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$.

4.1. **Differential properties of φ – scalar case.**

We now study the differential structure of the probability function φ defined in (1) in the special case $p = 1$, i.e., when the constraint mapping reduces to a scalar function $G\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. The results established here will be used in Subsection 4.2 as preliminary tools for analyzing the Lipschitz continuity of the gradient of φ in (1), under multiple random inequalities.

This subsection is organized as follows. We begin by recalling some technical tools and some known results on the differentiability of probability functions. Then, we establish new properties of the involved objects that will be instrumental in Subsection 4.2.

4.1.1. *Known results.* Next we recall the spherical-radial decomposition used in the study of probability functions. For further details, the reader is referred to Chapter 5 of the book [41]. Let us recall that the continuous random vector ξ admits a density f_ξ with respect to the Lebesgue measure λ_m on $\mathbb{R}^m$. Consequently, the probability function $\varphi(x)$ (with scalar G) can be written as

$$\varphi(x) = \int_{\{z\in\mathbb{R}^m : G(x,z)\le 0\}} f_\xi(z) d\lambda_m(z).$$

Following the construction of the *spherical–radial like decomposition* for general distributions presented in [40], the probability function can be further expressed as

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

where $\mu_\zeta\colon \mathcal{B}(\mathbb{S}^{m-1}) \to [0,1]$ is the probability measure given by

$$\mu_\zeta(A) = \frac{m\Gamma\left(\frac{m}{2}\right)}{2\pi^{\frac{m}{2}}} \cdot \lambda_m\big(\{rv \in \mathbb{R}^m : r \in [0,1], v \in A\}\big),$$

where $\mathcal{B}(\mathbb{S}^{m-1})$ denotes the Borel σ-algebra on $\mathbb{S}^{m-1}$. Moreover $e\colon \mathbb{R}^n \times \mathbb{S}^{m-1} \to \mathbb{R}\cup\{\infty\}$ is the *radial probability-like function* defined by

$$e(x,v) = \frac{2\pi^{\frac{m}{2}}|\det(L)|}{\Gamma\left(\frac{m}{2}\right)} \int_{\{r\ge 0: G(x,rLv)\le 0\}} r^{m-1} f_\xi(rLv) dr, \tag{9}$$

with L a nonsingular matrix. Following [40], one can also introduce a *density-like* function θ defined as

$$\theta(r,v) := \frac{2\pi^{\frac{m}{2}}|\det(L)|}{\Gamma\left(\frac{m}{2}\right)} r^{m-1} f_\xi(rLv). \tag{10}$$

We define the sets of finite and infinite directions of $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ respectively by

$$F(x) := \{v \in \mathbb{S}^{m-1} : \exists r \ge 0,\ G(x,rLv) = 0\}, \quad \text{and} \quad I(x) := \{v \in \mathbb{S}^{m-1} : \forall r \ge 0,\ G(x,rLv) < 0\}.$$

Next, we introduce the *radial function* $\rho\colon \mathbb{R}^n \times \mathbb{S}^{m-1} \to [0,\infty]$ defined by

$$\rho(x,v) := \sup\{r \ge 0 : G(x,rLv) \le 0\}, \tag{11}$$

where, by convention, $\sup\emptyset = 0$. It is important to note that, without further assumptions, this radial function is not necessarily continuous. The relation between this radial function ρ, the density-like function θ and the probability function φ is established in the next lemmas. We direct the interested reader to [37, 39], and [40] for a detailed exposition and proofs.

Lemma 4.1. ([40, Lemma 3.3]) *Let Assumption 1 ($p = 1$) hold, and let $x \in \mathbb{R}^n$ be such that $G(x,0) < 0$. Then the following statements are true:*

a) a direction $v \in \mathbb{S}^{m-1}$ belongs to $I(x)$ if and only if $G(x,rLv) < 0$ for all $r > 0$;
b) the sets of finite and infinite directions partition the unit sphere, that is, $F(x)\cup I(x) = \mathbb{S}^{m-1}$;
c) for every $v \in F(x)$, let $r > 0$ be such that $g(x,rLv) = 0$. Then $\langle \nabla_z G(x,rLv), Lv\rangle \ge -\frac{G(x,0)}{r}$.

As stated earlier, the objective is to analyze the differentiability of φ and to derive a representation for its gradient $\nabla\varphi$. This analysis relies on the gradient of the radial solution given in (11), whose definition is recalled below. The proof can be found in [40, Lemma 3.4].

Lemma 4.2. *([40, Lemma 3.4]) Suppose Assumption 1 holds ($p = 1$), and let $(\bar{x}, \bar{v})$ be such that $G(\bar{x}, 0) < 0$ and $\bar{v} \in F(\bar{x})$. Then, there exist neighbourhoods U of $\bar{x}$ and V of $\bar{v}$ such that for all $(x, v) \in U \times V$ one has the gradient formula*

$$\nabla_x \rho(x, v) = -\frac{1}{\langle \nabla_z G(x, \rho(x, v)Lv), Lv \rangle} \nabla_x G(x, \rho(x, v)Lv). \tag{12}$$

The next result provides a local representation of the radial probability-like function e, defined in (9), in terms of the density-like function θ, given in (10), and the radial function ρ, defined in (11). This formulation leads to the following lemma, originally stated in [40, Lemma 3.2].

Lemma 4.3. *([40, Lemma 3.2]) If Assumption 1 holds ($p = 1$), and $\bar{x} \in \mathbb{R}^n$ is such that $G(\bar{x}, 0) < 0$. Then the following statements are true:*

a) there exists a neighbourhood U of $\bar{x}$ such that, for all $(x, v) \in U \times \mathbb{S}^{m-1}$, $e(x, v) = \int_0^{\rho(x,v)} \theta(r, v)\, dr$, where e is the function defined in (9), θ is the density-like function given in (10), and ρ is the radial function defined in (11);

b) if $v \in I(x)$, then $\rho(x_k, v_k) \to \infty$ for any sequence $(x_k, v_k) \to (x, v)$ with $v_k \in F(x_k)$.

In particular, Lemma 4.3 characterizes the behaviour of the radial function in both finite and infinite directions. It can be applied to establish the continuity of the radial probability-like function e given in (9).

We now recall a complete characterization of the gradient of e. The next result follows directly from [37, Corollary 3.8] and [37, Corollary 3.9], so its proof is omitted.

Corollary 4.4. *Under Assumption 1 ($p = 1$), and with $\bar{x}$ such that $G(\bar{x}, 0) < 0$, assume furthermore that G satisfies the first-order growth condition around $\bar{x}$. Then, for any $\bar{v} \in \mathbb{S}^{m-1}$, the mapping $\nabla_x e$ is continuously differentiable around $(\bar{x}, \bar{v})$, and it is given by*

$$\nabla_x e(\bar{x}, \bar{v}) = \begin{cases} -\dfrac{\theta(\rho(\bar{x}, \bar{v}), \bar{v})}{\langle \nabla_z G(\bar{x}, \rho(\bar{x}, \bar{v})L\bar{v}), L\bar{v} \rangle} \nabla_x G(\bar{x}, \rho(\bar{x}, \bar{v})L\bar{v}) & \text{if } \bar{v} \in F(\bar{x}), \\ 0 & \text{if } \bar{v} \in I(\bar{x}). \end{cases}$$

Finally, we can establish the differentiability of φ, as claimed, in the following theorem, where the result follows directly from [35, Corollary 1].

Theorem 4.5. *In addition to Assumption 1 ($p = 1$), let $\bar{x} \in \mathbb{R}^n$ be such that $G(\bar{x}, 0) < 0$, and suppose that G satisfies the first-order growth condition (3) at $\bar{x}$. Then there exists a neighbourhood U' of $\bar{x}$ such that φ is continuously differentiable on U', and its gradient is given by*

$$\nabla \varphi(x) = -\int_{\mathbb{S}^{m-1}} \nabla_x e(\bar{x}, v) d\mu_\zeta(v) \quad \text{for all } x \in U'. \tag{13}$$

4.1.2. *New insights: Lipschitz continuity of $\nabla_x e$.* We now build on the recalled expression for $\nabla \varphi$ to obtain explicit bounds on the Lipschitz modulus of the gradient of the radial probability-like function. These bounds will play a key role in showing that φ is of class $C^{1,1}$. First, the next proposition yields estimates for the Lipschitz modulus of this gradient along finite directions on the m-dimensional sphere.

Proposition 4.6. *In addition to Assumption 1 ($p = 1$), let $\bar{x} \in \mathbb{R}^n$ be such that $G(\bar{x}, 0) < 0$, and let $\bar{v} \in F(\bar{x})$. Then $\nabla_x e$ is Lipschitz around $(\bar{x}, \bar{v})$. Moreover, if G satisfies the $(1, 1)$-smooth growth condition (as given in Definition 2) at $\bar{x}$, then the following estimate holds for* *the Lipschitz constant:*

$$\operatorname{lip} \nabla_x e(\bar{x}, \bar{v}) \leq C(\bar{x}) \Big(A(\bar{x}, \bar{v}) + \sum_{i=1}^{4} B_i(\bar{x}, \bar{v}) \Big), \tag{14}$$

where

$$\begin{aligned}
A(\bar{x}, \bar{v}) &= \|\nabla f_\xi(\rho(\bar{x}, \bar{v}) L\bar{v})\| \, \eta^2\big(\|L\bar{v}\| \rho(\bar{x}, \bar{v})\big) \, \rho(\bar{x}, \bar{v})^{m+1}, \\
B_1(\bar{x}, \bar{v}) &= f_\xi(\rho(\bar{x}, \bar{v}) L\bar{v}) \, \tilde{\eta}\big(\|L\bar{v}\| \rho(\bar{x}, \bar{v})\big) \, \rho(\bar{x}, \bar{v})^{m}, \\
B_2(\bar{x}, \bar{v}) &= f_\xi(\rho(\bar{x}, \bar{v}) L\bar{v}) \, \tilde{\eta}\big(\|L\bar{v}\| \rho(\bar{x}, \bar{v})\big) \, \eta\big(\|L\bar{v}\| \rho(\bar{x}, \bar{v})\big) \, \rho(\bar{x}, \bar{v})^{m+1}, \\
B_3(\bar{x}, \bar{v}) &= f_\xi(\rho(\bar{x}, \bar{v}) L\bar{v}) \, \tilde{\eta}\big(\|L\bar{v}\| \rho(\bar{x}, \bar{v})\big) \, \eta^2\big(\|L\bar{v}\| \rho(\bar{x}, \bar{v})\big) \, \rho(\bar{x}, \bar{v})^{m+2}, \\
B_4(\bar{x}, \bar{v}) &= f_\xi(\rho(\bar{x}, \bar{v}) L\bar{v}) \, \eta^2\big(\|L\bar{v}\| \rho(\bar{x}, \bar{v})\big) \, \rho(\bar{x}, \bar{v})^{m}, \\
C(\bar{x}) &= (m-1) \, \frac{2\pi^{m/2} |\det(L)|}{\Gamma(m/2)} \max_{i=1,2,3} \left\{ \frac{1}{|G(\bar{x}, 0)|^i} \right\} \max_{i=0,1,2} \{\|L\|^i\}.
\end{aligned} \tag{15}$$

Proof. By Corollary 4.4, there exists a neighbourhood $\widetilde{U} \times \widetilde{V}$ of $(\bar{x}, \bar{v})$ such that

$$\nabla_x e(x, v) = -\frac{\theta(\rho(x, v), v) \, \nabla_x G(x, \rho(x, v) Lv)}{\langle \nabla_z G(x, \rho(x, v) Lv), Lv \rangle}, \quad (x, v) \in \widetilde{U} \times \widetilde{V}.$$

Then, it is easy to see that $\nabla_x e$ is Lipchitz around $(\bar{x}, \bar{v})$. Therefore, we proceed to show (14) under the $(1,1)$-smooth growth condition. For convenience, set

$$\kappa(x, v) := \langle \nabla_z G(x, \rho(x, v) Lv), Lv \rangle, \text{ for all } x, y \in \widetilde{U} \text{ and } v \in \widetilde{V}.$$

Now, for $x, y \in \widetilde{U}$ and $v \in \widetilde{V}$ we then have

$$\|\nabla_x e(x, v) - \nabla_x e(y, v)\| \leq a_1(x, y, v) + a_2(x, y, v) + a_3(x, y, v),$$

where

$$\begin{aligned}
a_1(x, y, v) &:= \frac{|\theta(\rho(x, v), v) - \theta(\rho(y, v), v)|}{|\kappa(x, v)|} \|\nabla_x G(x, \rho(x, v) Lv)\|, \\
a_2(x, y, v) &:= \frac{|\theta(\rho(x, v), v)|}{|\kappa(x, v) \kappa(y, v)|} \|\nabla_x G(x, \rho(x, v) Lv)\| \cdot |\kappa(x, v) - \kappa(y, v)|, \\
a_3(x, y, v) &:= \frac{|\theta(\rho(y, v), v)|}{|\kappa(y, v)|} \|\nabla_x G(x, \rho(x, v) Lv) - \nabla_x G(y, \rho(y, v) Lv)\|.
\end{aligned}$$

and let us also define the constant

$$c(m, L) := \frac{2\pi^{m/2} |\det(L)|}{\Gamma(m/2)},$$

so that $\theta(r, v) = c(m, L) \, r^{m-1} \, f_\xi(rLv)$. Moreover, for convenience of notation, let us set

$$\tilde{r}(x, y, v) := \|Lv\| \max\{\rho(x, v), \rho(y, v)\}. \tag{16}$$

Now, let us estimate $a_3(x, y, v)$. Since G satisfies the $(1,1)$-smooth growth condition, we have

$$\|\nabla_x G(x, \rho(x, v) Lv) - \nabla_x G(y, \rho(y, v) Lv)\| \leq \tilde{\eta}(\tilde{r}(x, y, v)) \Big(\|x - y\| + \|Lv\| |\rho(x, v) - \rho(y, v)| \Big),$$

where $\tilde{r}$ is defined in (16). Now, by the Mean Value Theorem, there exists $z_{xy} \in (x, y)$ such that

$$|\rho(x,v) - \rho(y,v)| \leq \frac{\|\nabla_x G(z_{xy}, \rho(z_{xy}, v)Lv)\|}{|\kappa(z_{xy}, v)|} \|x - y\|.$$

Moreover, Lemma 4.1 implies $\left|\frac{1}{\kappa(y,v)}\right| \leq \frac{\rho(y,v)}{|G(y,0)|}$, which holds for all $y \in \widetilde{U}$. Substituting these estimates yields

$$\begin{aligned} a_3(x,y,v) \leq c(m,L)\, \rho(y,v)^m f_\xi(\rho(y,v)Lv)\, \frac{1}{|G(y,0)|}\, \tilde{\eta}(\tilde{r}(x,y,v)) \\ \times \left(1 + \|Lv\| \frac{\|\nabla_x G(z_{xy}, \rho(z_{xy}, v)Lv)\|}{|\kappa(z_{xy}, v)|}\right) \|x - y\|. \end{aligned}$$

Using again the growth condition and the bound on α, we obtain

$$\begin{aligned} \frac{a_3(x,y,v)}{\|x - y\|} \leq c(m,L)\, \rho(y,v)^m f_\xi(\rho(y,v)Lv)\, \frac{1}{|G(y,0)|}\, \tilde{\eta}(\tilde{r}(x,y,v)) \\ \times \left(1 + \frac{\|Lv\|\rho(z_{xy}, v)}{|G(z_{xy}, 0)|}\, \eta(\rho(z_{xy}, v)\|Lv\|)\right). \end{aligned}$$

Passing to the limit we conclude

$$\begin{aligned} \limsup_{\substack{x,y\to\bar{x} \\ v\to\bar{v}}} \frac{a_3(x,y,v)}{\|x - y\|} &\leq c(m,L)\, \rho(\bar{x},\bar{v})^m f_\xi(\rho(\bar{x},\bar{v})L\bar{v}) \frac{1}{|G(\bar{x},0)|} \\ &\quad \times \tilde{\eta}(\|L\bar{v}\|\rho(\bar{x},\bar{v})) \left(1 + \frac{\|L\bar{v}\|\rho(\bar{x},\bar{v})}{|G(\bar{x},0)|}\, \eta(\rho(\bar{x},\bar{v})\|L\bar{v}\|)\right) \\ &\leq C(\bar{x})\big[B_1(\bar{x},\bar{v}) + B_2(\bar{x},\bar{v})\big]. \end{aligned} \tag{17}$$

Now, let us estimate $a_2(x,y,v)$. We note that

$$\begin{aligned} |\kappa(x,v) - \kappa(y,v)| &= |\langle \nabla_z G(x, \rho(x,v)Lv) - \nabla_z G(y, \rho(y,v)Lv), Lv\rangle| \\ &\leq \|\nabla_z G(x, \rho(x,v)Lv) - \nabla_z G(y, \rho(y,v)Lv)\| \cdot \|Lv\| \\ &\leq \|x - y\| \tilde{\eta}(\tilde{r}(x,y,v)) \\ &\quad \times \left(1 + \|Lv\| \frac{\rho(z_{xy}, v)}{|G(z_{xy}, 0)|}\, \eta(\rho(z_{xy}, v)\|Lv\|)\right) \cdot \|Lv\|, \end{aligned}$$

where the last step uses the $(1,1)$-smooth growth condition and $\tilde{r}$ is defined in (16). Substituting into a_2 gives

$$\begin{aligned} \frac{a_2(x,y,v)}{\|x - y\|} \leq c(m,L)\, \rho(x,v)^m\, f_\xi(\rho(x,v)Lv)\, \frac{\|Lv\|\rho(y,v)}{|G(x,0)||G(y,0)|} \\ \times \tilde{\eta}(\tilde{r}(x,y,v))\eta(\rho(x,v)\|Lv\|) \\ \times \left(1 + \|Lv\| \frac{\rho(z_{xy}, v)}{|G(z_{xy}, 0)|}\, \eta(\rho(z_{xy}, v)\|Lv\|)\right). \end{aligned}$$

Taking the limit, we obtain

$$\begin{aligned}
\limsup_{\substack{x,y\to\bar{x}\\ v\to\bar{v}}} \frac{a_2(x,y,v)}{\|x-y\|} &\le c(m,L)\,\rho(\bar{x},\bar{v})^m f_\xi(\rho(\bar{x},\bar{v})L\bar{v}) \frac{\|L\bar{v}\|\rho(\bar{x},\bar{v})}{|G(\bar{x},0)|^2}\\
&\qquad \times \tilde{\eta}(\|L\bar{v}\|\rho(\bar{x},\bar{v}))\eta(\rho(\bar{x},\bar{v})\|L\bar{v}\|)\left(1+\|L\bar{v}\|\frac{\rho(\bar{x},\bar{v})}{|G(\bar{x},0)|}\,\eta(\rho(\bar{x},\bar{v})\|L\bar{v}\|)\right)\\
&\le C(\bar{x})\big[B_2(\bar{x},\bar{v})+B_3(\bar{x},\bar{v})\big]. \qquad (18)
\end{aligned}$$

Finally, let us check the case $a_1(x,y,v)$. Lemma 4.2 and the first-order growth condition give

$$\|\nabla_x\rho(x,v)\| \le \eta(\rho(x,v)\|Lv\|)\frac{\rho(x,v)}{|G(x,0)|}, \quad x\in\widetilde{U}.$$

Define $\mathfrak{m}(x,v) := \rho(x,v)^{m-1} f_\xi(\rho(x,v)Lv)$. By the Mean Value Theorem, there exists $z_{xy}\in(x,y)$ such that $\mathfrak{m}(x,v)-\mathfrak{m}(y,v) = \langle \nabla_x\mathfrak{m}(z_{xy},v), x-y\rangle$. Computing the gradient,

$$\begin{aligned}
\nabla_x\mathfrak{m}(x,v) &= (m-1)\rho(x,v)^{m-2}\nabla_x\rho(x,v)\, f_\xi(\rho(x,v)Lv)\\
&\qquad + \rho(x,v)^{m-1}\langle \nabla f_\xi(\rho(x,v)Lv), Lv\rangle\, \nabla_x\rho(x,v).
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{\|\mathfrak{m}(x,v)-\mathfrak{m}(y,v)\|}{\|x-y\|} &\le \frac{\rho(z_{xy},v)^{m-1}}{|G(z_{xy},0)|}\,\eta(\rho(z_{xy},v)\|Lv\|)\\
&\qquad \times \Big((m-1)f_\xi(\rho(z_{xy},v)Lv) + \rho(z_{xy},v)\|Lv\|\|\nabla f_\xi(\rho(z_{xy},v)Lv)\|\Big).
\end{aligned}$$

This implies

$$\begin{aligned}
\limsup_{\substack{x,y\to\bar{x}\\ v\to\bar{v}}} \frac{a_1(x,y,v)}{\|x-y\|} &\le c(m,L)\,\eta(\rho(\bar{x},\bar{v})\|L\bar{v}\|)\frac{\rho(\bar{x},\bar{v})}{|G(\bar{x},0)|}\frac{\rho(\bar{x},\bar{v})^{m-1}}{|G(\bar{x},0)|}\,\eta(\rho(\bar{x},\bar{v})\|L\bar{v}\|)\\
&\qquad \times \Big((m-1)f_\xi(\rho(\bar{x},\bar{v})L\bar{v}) + \rho(\bar{x},\bar{v})\|L\bar{v}\|\,\|\nabla f_\xi(\rho(\bar{x},\bar{v})L\bar{v})\|\Big)\\
&\le C(\bar{x})\big[A(\bar{x},\bar{v})+B_4(\bar{x},\bar{v})\big], \qquad (19)
\end{aligned}$$

with A and B_4 as defined in (15). Combining (17), (18), and (19), we obtain (14). This establishes the local Lipschitz continuity of $\nabla_x e$. □

Having established that $\nabla_x e$ is locally Lipschitz continuous in a neighbourhood of any point (x,v) with $v\in F(x)$, we now aim to extend this property to the whole domain. To do so, we first analyze the behaviour of $\nabla_x e$ when approaching infinite directions. The next lemma shows that the Lipschitz modulus of $\nabla_x e$ vanishes along sequences converging to infinite directions.

Lemma 4.7. *Under the assumptions of Proposition 4.6, let $(\bar{x},\bar{v})$ be such that $\bar{v}\in I(\bar{x})$, and let $(x_k,v_k)\to(\bar{x},\bar{v})$ with $v_k\in F(x_k)$. Then* $\lim_{k\to\infty}\operatorname{lip}\nabla_x e(x_k,v_k)=0$.

Proof. The proof follows by combining the estimate in (14) with the limits appearing in the definitions of the growth conditions (see Definitions 1 and 2), evaluated at $\|z\| = \|Lv\|\,\rho(x,v)$. □

In the next technical lemma, we provide a lower bound for the Lipschitz modulus in terms of subgradients. To that end, recall that the symmetric subdifferential is defined at the end of the Introduction.

Lemma 4.8. *Let $\psi\colon\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be a function. If $x^*\in\partial^0\psi(u)$, then* $\operatorname{Lip}\psi(u)\ge\|x^*\|$.

Proof. To this end, let $\ell > 0$ be arbitrary but fixed and such that $\operatorname{Lip}\psi(u) < \ell$. Then there exists $\varepsilon > 0$ such that

$$|\psi(u_1) - \psi(u_2)| \leq \ell\,\|u_1 - u_2\| \quad \text{for all } u_1, u_2 \in \mathbb{B}_\varepsilon(u).$$

Applying [25, Theorem 4.15] to both ψ and $-\psi$, we obtain $\partial\psi(u) \subseteq \ell\,\mathbb{B}$ and $-\partial(-\psi)(u) \subseteq \ell\,\mathbb{B}$, hence $\partial^0\psi(u) \subseteq \ell\,\mathbb{B}$. Since $x^* \in \partial^0\psi(u)$, it follows that $\|x^*\| \leq \ell$. As this holds for an arbitrary ℓ, the result is shown. □

We next establish a lemma ensuring the twice differentiability of e under the imposed growth conditions at infinite directions. This result will be crucial for proving the existence of a global Lipschitz constant for the gradient of the radial probability-like function e.

Lemma 4.9. *In addition to Assumption 1, let $\bar{x} \in \mathbb{R}^n$ be such that $G(\bar{x}, 0) < 0$, and suppose that G satisfies the $(1,1)$-smooth growth condition (Definition 2) at $\bar{x}$. Let $\bar{v} \in I(\bar{x})$. Then $e(\cdot, \bar{v})$ is twice Fréchet differentiable at $\bar{x}$ and $\nabla_x^2 e(\bar{x}, \bar{v}) = 0$.*

Proof. First, let us recall that by Corollary 4.4, we have that $\nabla_x e(\bar{x}, \bar{v}) = 0$. Suppose, for a contradiction, that there exist sequences $h_k \in \mathbb{S}^{n-1}$ and $\beta_k \downarrow 0$, and some $\varepsilon > 0$, such that

$$\|\nabla_x e(\bar{x} + \beta_k h_k, \bar{v}) - \nabla_x e(\bar{x}, \bar{v})\| > \varepsilon\,\beta_k.$$

This implies $\nabla_x e(\bar{x} + \beta_k h_k, \bar{v}) \neq 0$ and so for each $k \in \mathbb{N}$ we set

$$\alpha_k := \sup\left\{t \in [0, \beta_k] : \ \nabla_x e(\bar{x} + t h_k, \bar{v}) = 0\right\}.$$

By continuity of $\nabla_x e(\cdot, \bar{v})$ we have $\alpha_k < \beta_k$ and $\nabla_x e(\bar{x} + \alpha_k h_k, \bar{v}) = 0$. Moreover, by the definition of α_k,

$$\nabla_x e(\bar{x} + t h_k, \bar{v}) \neq 0 \quad \text{for all } t \in (\alpha_k, \beta_k]. \tag{20}$$

By Corollary 4.4, we get that (20) implies that $\bar{v} \in F(\bar{x} + t h_k)$ for all $t \in (\alpha_k, \beta_k]$. Define $\psi_k : [\alpha_k, \beta_k] \to \mathbb{R}$ by

$$\psi_k(t) := \|\nabla_x e(\bar{x} + t h_k, \bar{v})\|.$$

It follows that $\psi_k(\alpha_k) = 0$, and that ψ_k is continuous. By the symmetric subdifferential mean value theorem for continuous functions (see, e.g., [25, Theorem 4.11]), there exist $t_k \in (\alpha_k, \beta_k)$ and $r_k^* \in \partial^0\psi_k(t_k)$ such that

$$\psi_k(\beta_k) - \psi_k(\alpha_k) = r_k^*(\beta_k - \alpha_k).$$

Therefore,

$$r_k^* = \frac{\psi_k(\beta_k) - \psi_k(\alpha_k)}{\beta_k - \alpha_k} \geq \frac{\|\nabla_x e(\bar{x} + \beta_k h_k, \bar{v})\|}{\beta_k} \geq \varepsilon.$$

By Lemma 4.8, $\operatorname{Lip}\psi_k(t_k) \geq \varepsilon$. Moreover, by the definition of ψ_k and the fact that $\|h_k\| = 1$,

$$\operatorname{Lip}\psi_k(t_k) \leq \operatorname{Lip}\nabla_x e(x_k, \bar{v}) \quad \text{with } x_k := \bar{x} + t_k h_k.$$

Hence we have constructed a sequence $x_k \to \bar{x}$ with $\bar{v} \in F(x_k)$ and $\operatorname{Lip}\nabla_x e(x_k, \bar{v}) \geq \varepsilon$, which contradicts Lemma 4.7. This completes the proof. □

Finally, we establish the existence of a global Lipschitz constant by combining the results derived in this section with a compactness argument.

Proposition 4.10. *Under Assumption 1 ($p = 1$), let U be an open set such that:*

i) $G(x, 0) < 0$ for all $x \in U$;
ii) G satisfies the $(1,1)$-smooth growth condition at every $x \in U$.

Then, for all $x \in U$ and $v \in \mathbb{S}^{m-1}$, we have

$$\operatorname{lip} \nabla_x e(x,v) \leq \Psi(x,v), \tag{21}$$

where the mapping $\Psi\colon U \times \mathbb{S}^{m-1} \to \mathbb{R}$ is defined by

$$\Psi(x,v) := \begin{cases} C(\bar{x})\Big(A(x,v) + \sum_{i=1}^{4} B_i(x,v)\Big), & \textit{if } v \in F(x), \\ 0, & \textit{if } v \in I(x), \end{cases}$$

and A, B_i, and C are as in (15).

Proof. Fix $x \in U$ and $v \in \mathbb{S}^{m-1}$. If $v \in F(x)$, the (local) estimate follows directly from Proposition 4.6. If instead $v \in I(x)$, Lemma 4.7 yields $\operatorname{lip} \nabla_x e(x,v) = 0$, so that (14) also holds in this case. Finally, the global bound is obtained by considering a classical compactness argument over $\mathbb{S}^{m-1}$. □

4.2. **Differential properties of φ – vector-valued case.** Under Assumption 1 from Section 3, the spherical–radial decomposition of ξ yields the representation

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

where e is the radial probability-like function associated with a constraint mapping given by the pointwise maximum

$$g(x,z) := \max_{1 \leq i \leq p} G_i(x,z). \tag{22}$$

This reduction rewrites the joint probability function (1) as a probability constraint of the form $\mathbb{P}\big(g(x,\xi) \leq 0\big)$. However, the corresponding constraint function g in (22) is in general nonsmooth, see [36, Example 1.1]. Consequently, the estimates developed in Section 4.1 for $C^{1,1}$ data do not apply directly. Moreover, this issue cannot be circumvented by appealing to Rademacher's theorem: the probability function may fail to be differentiable even when the underlying max-structure is differentiable almost everywhere. To address this difficulty, we employ a smooth approximation of the maximum and derive the desired properties for φ via gradient consistency. Specifically, we employ the classical *log-sum-exponential* (LSE) smoothing of the maximum, defined for $\lambda > 0$ by

$$g_\lambda(x,\xi) := \frac{1}{\lambda} \ln\left(\frac{1}{p} \sum_{i=1}^{p} \exp\big(\lambda\, G_i(x,\xi)\big) \right). \tag{23}$$

The family $\{g_\lambda\}_{\lambda>0}$ induces the smoothed probability functions

$$\varphi_\lambda(x) := \mathbb{P}\big(g_\lambda(x,\xi) \leq 0\big),$$

which we use to approximate the joint probability function φ in (1). We also denote by

$$\rho_\lambda(x,v) := \sup\Big\{ r \geq 0 : g_\lambda(x, rLv) \leq 0 \Big\}$$

the resolvent function associated with g_λ. The remainder of this subsection is divided into two subsections. The first collects preparatory results showing that φ_λ approximates φ as $\lambda \to +\infty$. The second contains the proofs of the main results from Section 3, namely Theorem 3.1 and Corollary 3.2.

4.2.1. *Approximation of φ through φ_λ.* The next proposition collects basic approximation properties of g_λ. See [6] and the references therein for more details.

Lemma 4.11. *Under Assumption 1, the following holds for any arbitrary $\lambda > 0$:*

a) g_λ is of class $C^{1,1}$;
b) $g_\lambda \nearrow g$ pointwise as $\lambda \nearrow \infty$;
c) for every $x \in \mathbb{R}^n$, the mapping $z \mapsto g_\lambda(x,z)$ is convex;
d) $g_\lambda \xrightarrow{e} g$, i.e., g_λ converges epigraphically to g.

We can establish a hierarchy for the resolvant functions:

Corollary 4.12. *For any $\lambda \geq \lambda' > 0$, the following holds for the resolvant functions:*

$$\rho(x,v) \leq \rho_\lambda(x,v) \leq \rho_{\lambda'}(x,v), \quad \textit{for all } x \in \mathbb{R}^n,\ v \in \mathbb{S}^{m-1}. \tag{24}$$

Proof. For any $x \in \mathbb{R}^n$, $v \in \mathbb{S}^{m-1}$ and $r \geq 0$, as a result of Lemma 4.11 b) we get

$$g_{\lambda'}(x, rLv) \leq g_\lambda(x, rLv) \leq g(x, rLv),$$

so that the result follows from the definition of the resolvant maps as suprema over appropriate $r \geq 0$, whenever there exists at all some r' with $g(x, r'Lv) \leq 0$. Should there not be any $r \geq 0$ for which $g(x, rLv) \leq 0$, then by definition $\rho(x,v) = 0$; hence, by extension, the inequalities also hold true. $\square$

We also formally introduce the finite and infinite directions associated with g_λ even though it will turn out that they are identical, at relevant points, to $F(x)$ and $I(x)$ respectively:

$$\begin{aligned} F_\lambda(x) &:= \{v \in \mathbb{S}^{m-1} : \exists\, r \geq 0,\ g_\lambda(x, rLv) = 0\}, \\ I_\lambda(x) &:= \{v \in \mathbb{S}^{m-1} : \forall\, r \geq 0,\ g_\lambda(x, rLv) < 0\}. \end{aligned} \tag{25}$$

The next lemma provides a characterization of finite directions in terms of coercivity of the univariate profile $r \mapsto g(x, rLv)$.

Lemma 4.13. *Under Assumption 1, let $\bar{x} \in \mathbb{R}^n$ be such that $g(\bar{x}, 0) < 0$. Then*

$$v \in F(\bar{x}) \quad \Longleftrightarrow \quad \lim_{r \to +\infty} g(\bar{x}, rLv) = +\infty.$$

Proof. Set $f\colon [0,\infty) \to \mathbb{R}$ by $f(r) := g(\bar{x}, rLv)$. Since $z \mapsto g(\bar{x}, z)$ is convex and $r \mapsto rLv$ is affine, f is convex; moreover $f(0) = g(\bar{x}, 0) < 0$. On the one hand, if $v \in F(\bar{x})$, there exists $r_1 \geq 0$ with $f(r_1) = 0$. Fix any $r_2 > r_1$ and write $r_1 = \alpha r_2 + (1-\alpha)0$ with $\alpha = \frac{r_1}{r_2} \in (0,1)$. By convexity, $0 = f(r_1) \leq \alpha f(r_2) + (1-\alpha) f(0)$, hence

$$f(r_2) \ \geq\ -\frac{1-\alpha}{\alpha}\, f(0) = -\Big(\frac{r_2}{r_1} - 1\Big) f(0) = -\frac{r_2}{r_1} f(0) + f(0).$$

Because $f(0) < 0$, the right-hand side diverges to $+\infty$ as $r_2 \to +\infty$, so $\lim_{r \to +\infty} f(r) = +\infty$. On the other hand, if $\lim_{r \to +\infty} f(r) = +\infty$, there exists $r_0 > 0$ with $f(r_0) > 0$. Since f is continuous on $[0, r_0]$ and $f(0) < 0 < f(r_0)$, the Intermediate Value Theorem yields some $r \in (0, r_0)$ with $f(r) = 0$. Thus $v \in F(\bar{x})$. $\square$

Using the previous lemma, we can now show that the finite and infinite directions of g and those of its log-sum-exponential approximation (23) coincide at points satisfying Slater's condition. Formally, we obtain the following result.

Lemma 4.14. *Under Assumption 1, let $\bar{x} \in \mathbb{R}^n$ be such that $g(\bar{x}, 0) < 0$, where g is the maximum function defined in* (22)*. Denote by $F(\bar{x})$ and $I(\bar{x})$ the sets of finite and infinite directions associated with g, respectively. For any arbitrary but fixed $\lambda > 0$, let $F_\lambda(\bar{x})$ and $I_\lambda(\bar{x})$ be the finite and infinite directions defined in* (25) *with respect to the LSE approximation g_λ. Then it holds that:*

$$F(\bar{x}) = F_\lambda(\bar{x}) \qquad \text{and} \qquad I(\bar{x}) = I_\lambda(\bar{x}).$$

Proof. As a result of Lemma 4.1 item b), we know that $F_\lambda(\bar{x}), I_\lambda(\bar{x})$ and $F(\bar{x}), I(\bar{x})$ form a partition of $\mathbb{S}^{m-1}$. It will thus suffice to show that $F_\lambda(\bar{x}) = F(\bar{x})$. To this end, let $v \in F(\bar{x})$ be given. Then there exist $r > 0$ and some index $i_0 \in \{1, \dots, p\}$ such that $G_{i_0}(\bar{x}, rLv) > 0$. By the definition of g_λ, we have

$$g_\lambda(\bar{x}, rLv) \;\geq\; \frac{1}{\lambda} \ln\Big(\tfrac{1}{p} \exp(\lambda G_{i_0}(\bar{x}, rLv)) \Big) \;=\; G_{i_0}(\bar{x}, rLv) - \tfrac{1}{\lambda} \ln(p).$$

Hence, by Lemma 4.13, we conclude that $v \in F_\lambda(\bar{x})$. Conversely, suppose $v \in F_\lambda(\bar{x})$ given. Then there exists $r \geq 0$ such that $g_\lambda(\bar{x}, rLv) = 0$. Since $g_\lambda \leq g$ for all $\lambda \in \mathbb{N}$, it follows that $g(\bar{x}, rLv) \geq 0$. Recalling that $g(\bar{x}, 0) < 0$, the continuity of g guarantees the existence of some $r_0 > 0$ with $g(\bar{x}, r_0 Lv) = 0$, which shows that $v \in F(\bar{x})$. □

The following lemma establishes that if the family $\{G_i\}_{i=1}^p$ satisfies the $(1,1)$-smooth growth condition, then the LSE approximation also satisfies this condition.

Lemma 4.15. *Consider the components $\{G_i\}_{i=1}^p$ and the LSE approximation g_λ given in* (23)*. If the family $\{G_i\}_{i=1}^p$ satisfies the growth conditions, then the LSE approximation g_λ also satisfies the same growth conditions.*

Proof. Given the approximation g_λ in (23), then its gradient is given by:

$$\nabla g_\lambda(x, z) = \frac{1}{\sum_{i=1}^p \exp(\lambda G_i(x, z))} \sum_{i=1}^p \exp(\lambda G_i(x, z)) \nabla G_i(x, z).$$

Therefore, defining $\alpha_i = \frac{\exp(\lambda G_i(x,z))}{\sum_{i=1}^p \exp(\lambda G_i(x,z))}$, and observing that $\sum_{i=1}^p \alpha_i = 1$, we in fact have that $\nabla g_\lambda(x, z) = \sum_{i=1}^p \alpha_i \nabla G_i(x, z)$, meaning that ∇g_λ is just a convex combination of the gradients of the family $\{\nabla G_i\}_{i=1}^p$. Hence, we have that if the same growth conditions hold for $\{G_i\}_{i=1}^p$, then they will also hold for g_λ. □

Now, the following proposition shows that the gradient of the radial probability-like function e_λ is locally Lipschitz, uniformly in $\lambda > 0$.

Proposition 4.16. *Suppose that Assumption 1 holds, and let $\bar{x} \in \mathbb{R}^n$ satisfy $g(\bar{x}, 0) < 0$. Moreover, assume that the family $\{G_i\}_{i=1}^p$ fulfills the $(1,1)$-smooth growth condition at $\bar{x}$. Then there exist a neighbourhood U of $\bar{x}$ and a constant $\ell > 0$ such that*

$$\big\| \nabla_x e_\lambda(x, v) - \nabla_x e_\lambda(y, v) \big\| \leq \ell \|x - y\| \quad \forall\, x, y \in U,\ \forall\, v \in \mathbb{S}^{m-1},\ \forall\, \lambda > 0. \tag{26}$$

Proof. Let U be a neighbourhood of $\bar{x}$ such that the family $\{G_i\}_{i=1}^p$ satisfies the same $(1,1)$-smooth growth condition at each $x \in U$, and such that $G_i(x, 0) < 0$ for all $x \in U$ and all $i = 1, \dots, p$. By Lemma 4.15, for every $\lambda > 0$ and $x \in U$, one has $g_\lambda(x, 0) < 0$, and g_λ satisfies the $(1,1)$-smooth growth condition at x. Hence, by Proposition 4.10,

$$\operatorname{lip} \nabla e_\lambda(x, v) \;\leq\; \Psi_\lambda(x, v), \qquad \forall\, x \in U,\ v \in \mathbb{S}^{m-1},\ \forall\, \lambda > 0, \tag{27}$$

where

$$\Psi_\lambda(x,v) := \begin{cases} C_\lambda(x)\Big[A_\lambda(x,v) + \sum_{i=1}^4 B_{\lambda,i}(x,v)\Big], & \text{if } v \in F_\lambda(x),\\ 0, & \text{if } v \in I_\lambda(x),\end{cases}$$

and A_λ, $B_{\lambda,i}$ $(i = 1,\dots,4)$, and C_λ are defined as in (15), but with $\rho(x,v)$ replaced by $\rho_\lambda(x,v)$ and g by g_λ.

Claim 1. There exist $\varepsilon > 0$ and $n_0 > 0$ such that

$$\sup\big\{\Psi_\lambda(x,v) : (x,v) \in \mathbb{B}_\varepsilon(\bar{x}) \times \mathbb{S}^{m-1},\ \lambda \geq n_0\big\} < \infty. \tag{28}$$

To prove this, fix $\bar{v} \in \mathbb{S}^{m-1}$ arbitrarily. We show that there exist a neighbourhood $U_{\bar{v}} \subset U$ of $\bar{x}$, a neighbourhood $V_{\bar{v}}$ of $\bar{v}$, and a constant $r_{\bar{v}} > 0$ such that

$$s_{\bar{v}} := \sup\big\{\Psi_\lambda(x,v) : (x,v) \in U_{\bar{v}} \times V_{\bar{v}},\ \lambda > r_{\bar{v}}\big\} < +\infty.$$

First, suppose that $\bar{v} \in F(\bar{x})$. We claim that $\rho_\lambda(x,v)$ remains bounded for $x \in U_{\bar{v}}$, $v \in V_{\bar{v}}$, and $\lambda > n_{\bar{v}}$, for some suitable neighbourhoods $U_{\bar{v}}, V_{\bar{v}}$ and constant $n_{\bar{v}}$. Assume, to the contrary, that there exists a sequence (x_k, v_k, λ_k) such that $(x_k, v_k) \to (\bar{x}, \bar{v})$, $\lambda_k \to \infty$, and $\rho_{\lambda_k}(x_k, v_k) \to +\infty$. For large arbitrary r_0, there is k_0 such that $r_0 \leq \rho_{\lambda_k}(x_k, v_k)$ for all $k \geq k_0$, implying $g_{\lambda_k}(x_k, rLv_k) \leq 0$ for all $r \in [0, r_0]$. By epiconvergence of g_λ (Lemma 4.11 item d)), we obtain

$$0 \ \geq\ \liminf_{k\to+\infty} g_{\lambda_k}(x_k, rLv_k) \ \geq\ g(\bar{x}, rL\bar{v}),$$

which thus holds for all $r_0 \geq r \geq 0$. However since r_0 was arbitrary it in fact holds for all $r \geq 0$. This implies $\bar{v} \in I(\bar{x})$, which is a contradiction. Therefore, following the definition of the various constants and continuity of the involved maps, it follows that $A_\lambda, B_{\lambda,1}, \dots, B_{\lambda,4}$ are also bounded on $U_{\bar{v}} \times V_{\bar{v}}$. There is thus some $M_{\bar{v}} > 0$ such that

$$\Psi_\lambda(x,v) \leq C_\lambda(x) M_{\bar{v}}, \quad \forall (x,v) \in U_{\bar{v}} \times V_{\bar{v}},\ \forall \lambda > 0.$$

From (4)–(5), there exists γ_0 such that for all $\|z\| \geq \gamma_0$,

$$\begin{aligned} &\|\nabla f_\xi(z)\| \|z\|^{m+1} \eta(\|z\|)^2 \leq 1,\\ &f_\xi(z)\, r^{m+i} \eta(\|z\|)^i \tilde{\eta}(\|z\|)^j \leq 1, \quad \forall i \in \{0,1,2\},\ j \in \{0,1\}. \end{aligned} \tag{29}$$

For any $v \in \mathbb{S}^{m-1}$, posing $z = rLv$, observing that L is nonsingular, it follows that we can find $r_0(v)$ large enough such that $\|z\| \geq \gamma_0$ for all $r \geq r_0(v)$.

Now, let us assume that $\bar{v} \in I(\bar{x})$, i.e., $\rho(\bar{x}, \bar{v}) = \infty$. Then by continuity of ρ in the extended real line (e.g., [36, Lemma 2.4]), and the inequality $\rho_\lambda \geq \rho$, we can choose neighbourhoods $U_{\bar{v}} \ni \bar{x}$, $V_{\bar{v}} \ni \bar{v}$ such that

$$\rho_\lambda(x,v) \ \geq\ r_0(\bar{v}), \quad \forall (x,v) \in U_{\bar{v}} \times V_{\bar{v}},\ \forall \lambda > 0.$$

Thus, for $z^\lambda_{x,v} := \rho_\lambda(x,v) Lv$, we have $\|z^\lambda_{x,v}\| \geq \gamma_0$ for all $(x,v) \in U_{\bar{v}} \times V_{\bar{v}}$ and $\lambda > 0$. As a result, we observe by combining (29) with the definition of Ψ_λ, that

$$\Psi_\lambda(x,v) \leq C_\lambda(x) M_{\bar{v}}, \quad \forall (x,v) \in U_{\bar{v}} \times V_{\bar{v}},\ \forall \lambda > 0,$$

with $M_{\bar{v}} = 1$. Now for $x \in U_{\bar{v}}$, possibly shrinking the latter neighbourhood, we may assume the existence of $\eta > 0$ such that $g_\lambda(x,0) \leq g(x,0) < -\eta < 0$. As a result we have $\eta \leq |g(x,0)| \leq |g_\lambda(x,0)|$ and therefore $C_\lambda(x) \leq C(x)$. Moreover it follows that $C(x)$ is bounded from above on $U_{\bar{v}}$ by some constant C.

Finally, since the family of neighbourhoods $\{V_{\bar{v}} : \bar{v} \in \mathbb{S}^{m-1}\}$ covers $\mathbb{S}^{m-1}$, and by compactness of $\mathbb{S}^{m-1}$, we can extract a finite subcover, say

$$\{V_{\bar{v}_i} : i = 1,\dots,q\}, \quad q \in \mathbb{N}.$$

Defining $n_0 := \max\{r_{\bar{v}_i} : i = 1, \ldots, q\}$, and $\varepsilon > 0$ such that $\mathbb{B}_\varepsilon(\bar{x}) \subset W := \bigcap_{i=1}^q U_{\bar{v}_i}$, it follows that (28) holds. *Claim 2.* Inequality (26) holds. Indeed, define

$$\ell := \sup\big\{\Psi_\lambda(x, v) : (x, v) \in \mathbb{B}_\varepsilon(\bar{x}) \times \mathbb{S}^{m-1},\ \lambda \geq n_0\big\},$$

where ε and n_0 are as in Claim 1. By Claim 1, $\ell < \infty$. Moreover, by (27), for each $\lambda \geq n_0$ and $v \in \mathbb{S}^{m-1}$, the function $\nabla e_\lambda(\cdot, v)$ is ℓ-Lipschitz on $\mathbb{B}_{\varepsilon/2}[\bar{x}]$ (see, e.g., [29, Theorem 9.2]). This completes the proof. □

We can now establish gradient consistency.

Lemma 4.17. *Let Assumption 1 hold, and let $\bar{x} \in \mathbb{R}^n$ be such that $g(\bar{x}, 0) < 0$. Assume further that the family $\{G_i\}_{i=1}^p$ satisfies the $(1,1)$-smooth growth condition at $\bar{x}$ and that* (R2CQ) *holds at $\bar{x}$. Then the probability function φ, given in* (1)*, is differentiable at $\bar{x}$, and for every $\lambda_k \to \infty$, and $x_k \to \bar{x}$, we have that*

$$\lim_{k \to +\infty} \nabla \varphi_{\lambda_k}(x_k) = \nabla \varphi(\bar{x}).$$

Proof. Let us denote $\limsup_{\substack{x \to \bar{x} \\ k \to +\infty}} \nabla \varphi_{\lambda_k}(x)$ as the set of all x^* such that there exists $k_j \to \infty$ and $x_j \to \bar{x}$ such that $\lim \nabla \varphi_{\lambda_{k_j}}(x_j) = x^*$. First, let us notice that as a consequence of [38, Proposition 3.6] we have that

$$\partial \varphi(\bar{x}) \subset \limsup_{\substack{x \to \bar{x} \\ k \to +\infty}} \nabla \varphi_{\lambda_k}(x).$$

Now, since g is maximum of $C^{1,1}$ functions, then in particular g is locally Lipschitz and lower-regular on $\mathbb{R}^n \times \mathbb{R}^m$ (see, e.g.,[29, Theorem 10.31]). Hence, by [38, Theorem 3.10] we have that

$$\limsup_{\substack{x \to \bar{x} \\ k \to +\infty}} \partial \varphi_{\lambda_k}(\bar{x}) \subset -\int_{v \in F(\bar{x})} \left\{ \frac{\theta(\rho(\bar{x}, v), v) x^*}{\langle z^*, Lv \rangle} :\ (x^*, z^*) \in \partial g(\bar{x}, \rho(\bar{x}, v) Lv) \right\} d\mu_\zeta(v).$$

Finally, let us notice that [36, Lemma 4.3] gives us that

$$\mu_\zeta(\{v \in \mathbb{S}^{m-1} : \#\mathcal{I}(\bar{x}, v) > 1\}) = 0. \tag{30}$$

This implies that $\mathcal{I}(\bar{x}, v)$ is $\mu_\zeta-$a.e. a singleton, and Clarke's rule (see, e.g., [5, Proposition 2.2.4]) yields

$$\partial g(\bar{x}, \rho(\bar{x}, v) Lv) = \{\nabla G_{i^*}(\bar{x}, \rho(\bar{x}, v) Lv)\},$$

where $\mathcal{I}(\bar{x}, v) = i^*$, and therefore the integrand on the left side is a singleton. Moreover, [36, Theorem 4.1] shows that if (30) holds, then φ is Fréchet differentiable at $\bar{x}$ and the equality follows. □

4.2.2. *Proofs of new differentiability results.* In the final part of this section we present the proofs of Theorem 3.1 and Corollary 3.2.

Proof. of Theorem 3.1. First, note that we can choose a neighbourhood U of $\bar{x}$ such that the assumptions of the theorem hold at every point of U. Second, according to Theorem 4.5,

$$\nabla \varphi_\lambda(x) = \int_{\mathbb{S}^{m-1}} \nabla_x e_\lambda(x, v)\, d\mu_\zeta(v), \qquad \text{for all } x \in U,\ \text{for all } \lambda > 0,$$

where U is a neighbourhood of $\bar{x}$. Now, by (4.10) and by shrinking U if necessary, we can ensure the existence of $\ell > 0$ such that, for every $x_1, x_2 \in U$, the Lipschitz condition (26) holds. Then, for any $x_1, x_2 \in U$, we have

$$\begin{aligned}\|\nabla\varphi_\lambda(x_1) - \nabla\varphi_\lambda(x_2)\| &\leq \int_{\mathbb{S}^{m-1}} \|\nabla_x e_\lambda(x_1, v) - \nabla_x e_\lambda(x_2, v)\| \, d\mu_\zeta(v) \\ &\leq \ell \, \|x_1 - x_2\|,\end{aligned}$$

where in the last inequality we used (26) and the fact that $\mu_\zeta(\mathbb{S}^{m-1}) = 1$. Finally, applying Lemma 4.17, we can pass to the limit as $\lambda \to +\infty$ and conclude that $\nabla\varphi$ is Lipschitz continuous around $\bar{x}$. □

Proof. of Corollary 3.2. Let us denote $M(x) := \{z : \ g(x, z) \leq 0\}$. We divide the proof into claims.

Claim 1: If g is of class $C^{1,1}$, then for every $r \geq 0$ there exists $\ell_r \in \mathbb{R}$ such that

$$\|\nabla g(x_1, z_1) - \nabla g(x_2, z_2)\| \leq \ell_r \, \|(x_1, z_1) - (x_2, z_2)\|, \quad \forall x_1, x_2 \in \tilde{U}, \ \forall z_1, z_2 \leq r.$$

This follows directly from the definition of a $C^{1,1}$ function and the fact that x_1, x_2 and z_1, z_2 are bounded.

Claim 2: There exists a neighbourhood U of $\bar{x}$ and a constant $M > 0$ such that

$$\rho(x, v) < M \quad \text{for all } x \in U, \ v \in \mathbb{S}^{m-1}.$$

Suppose there exists a sequence $\{(x_k, v_k)\}_k$ such that $(x_k, v_k) \to (\bar{x}, \bar{v})$ and $\rho(x_k, v_k) \to +\infty$. Similarly to Claim 2 in the proof of Proposition 4.10, let $r > 0$. Then there exists k_0 such that $r \leq \rho(x_k, v_k)$ for all $k > k_0$. By definition of ρ, we have $g(x_k, rLv_k) \leq 0$ for all $k > k_0$. Passing to the limit as $k \to +\infty$, continuity of g yields $g(\bar{x}, rL\bar{v}) \leq 0$ for all $r > 0$, meaning that $\bar{v} \in I(\bar{x})$. This is a contradiction, since according to [37, Lemma 3.12], $M(x)$ is bounded if and only if $I(x) = \emptyset$, which implies that $F(x) = \mathbb{S}^{m-1}$. Hence, there must exist $M > 0$ such that $\rho(x, v) < M$ for all $(x, v) \in U \times \mathbb{S}^{m-1}$.

Claim 3: There exist a neighbourhood U of $\bar{x}$ and a constant $\tilde{\ell}$ such that

$$\text{Lip } \nabla_x e(x, v) \leq \tilde{\ell} \quad \forall x \in U, \ \forall v \in \mathbb{S}^{m-1}.$$

This follows by repeating the steps of Proposition 4.6, but using the bounds from Claim 1 for $\nabla_x g$ and from Claim 2 for ρ, respectively.

Claim 4: If $\bar{x}$ satisfies $g(\bar{x}, 0) < 0$ and $M(\bar{x})$ is bounded, then the probability function φ is of class $C^{1,1}$. By [37, Theorem 3.14], φ is continuously differentiable when $M(x)$ is bounded. Moreover, the fact that $\nabla\varphi$ is Lipschitz continuous around $\bar{x}$ follows by repeating the proof of Proposition 4.10 using the Lipschitz constant from Claim 3. □

5. Probability Maximization: the maju Algorithm

We introduce an algorithm to solve a class of probability maximization problems for which the probability function is locally $C^{1,1}$. More precisely, our goal is to solve the problem

$$\text{(PMP)} \qquad \max_{x \in X} \ \varphi(x) := \mathbb{P}(\xi \leq h(x)),$$

where $h \colon \mathbb{R}^n \to \mathbb{R}^m$ is a given mapping, $X \subset \mathbb{R}^n$ denotes the feasible set and $\xi \in \mathbb{R}^m$ is a random vector with continuous density. Throughout this section, we impose the following hypotheses on the structure of the maximization problem (PMP):

Assumption 2. *Suppose that:*

i) X is a nonempty closed set;

ii) the marginal functions $F_j(z) := \mathbb{P}(\xi_j \le z)$ *and the constraint mapping* $h\colon \mathbb{R}^n \to \mathbb{R}^m$ *are* $C^{1,1}$ *in a neighbourhood of every point in* X*;*

iii) the random vector ξ *has a continuously differentiable density* f_ξ *satisfying* (8)*.*

In order to solve the probability maximization problem (PMP), we propose an algorithm that, at each iteration, solves a model-approximation subproblem based on the marginal distribution functions of the random vector ξ. More precisely, given a stability center $\hat{x}^\nu \in X$, we approximate the probability function φ with the model

$$M_\nu(x) := \varphi(\hat{x}^\nu) + \sum_{j=1}^{m} w_j^\nu \left[F_j(h_j(x)) - F_j(h_j(\hat{x}^\nu))\right], \tag{31}$$

where $w^\nu \in [0,1]^m$ is a solution (on variable w) of the linear system

$$\sum_{j=1}^{m} w_j f_j(h_j(\hat{x}^\nu))\, \nabla h_j(\hat{x}^\nu) = \nabla\varphi(\hat{x}^\nu). \tag{32}$$

In this notation, f_j denotes the density function of the j-th marginal distribution function F_j of ξ. Since model (31) approximates the joint probability function using marginals, we name it MAJU model, with MAJU standing for *Marginal Assessment of Joint Uncertainty.*

Remark 3 (Solvability of (32))**.** *By* [11, Theorem 1.6.1]*, the linear system* (32) *admits at least one solution of the form*

$$w^\nu = \nabla C\big(F(h(\hat{x}^\nu))\big) \quad \in \quad [0,1]^m,$$

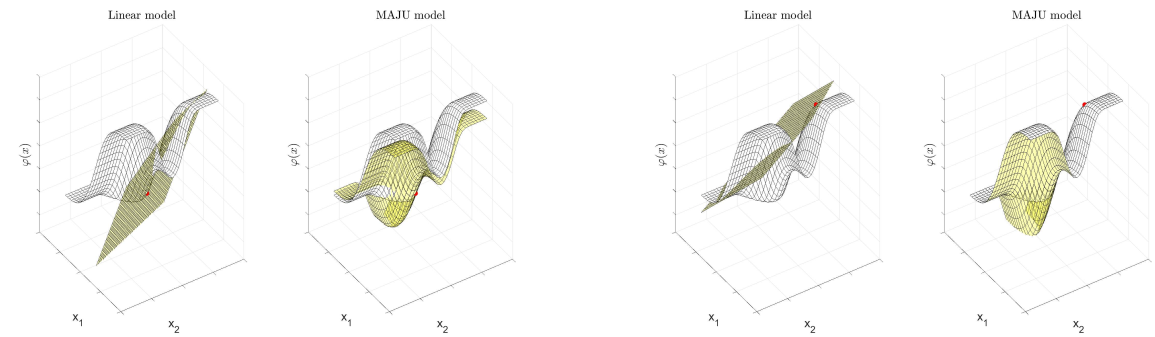
where C *is the copula associated with*[1] $\mathbb{P}$*, and the gradient is evaluated at* $F(h(\hat{x}^\nu))$*. We recall that* C *exists and is unique since each* F_j *is continuous* [41, § 5.3].

This remark clarifies the role of the model M_ν: ideally, w^ν would correspond to the gradient of the associated copula evaluated at $F(h(\hat{x}^\nu))$. Since this gradient (and even the copula itself) may not be available, we compute it using the information at hand through equation (32). Model M_ν is inspired by the upper-model $\overline{M}_\nu$ (respectively lower $\underline{M}_\nu$) proposed in [8], which instead of $w_j^\nu \left[F_j(h_j(x)) - F_j(h_j(\hat{x}^\nu))\right]$ it uses $\max\{F_j(h_j(x)) - F_j(h_j(\hat{x}^\nu)), 0\}$ (respectively $\min\{F_j(h_j(x)) - F_j(h_j(\hat{x}^\nu)), 0\}$). As shown in [8, Lemma 2.1], $\varphi(x) \in [\underline{M}_\nu(x), \overline{M}_\nu(x)]$. Since we have $w^\nu \in [0,1]^m$ in the definition of our model, it follows that $M_\nu(x) \in [\underline{M}_\nu(x), \overline{M}_\nu(x)]$ as well. However, M_ν is not necessarily an upper or lower approximation of φ, as illustrated in Figure 1(a). On the other hand, in contrast to the models of [8], M_ν is differentiable and $\nabla M_\nu(\hat{x}^\nu) = \nabla\varphi(\hat{x}^\nu)$ thanks to (32). Figure 1 illustrates the quality of the MAJU model on a bi-variate Gaussian probability distribution. We compare our model with the linear one $\varphi(\hat{x}) + \langle\nabla\varphi(\hat{x}), x - \hat{x}\rangle$.

Given the model (31), we depict in Algorithm 5 our algorithmic approach: Marginal Assessment of Joint Uncertainty (MAJU) algorithm.

We claim that the iterates of the MAJU Algorithm are well defined. Indeed, by Assumption 2(ii), the model M_ν in (31) is continuous. Furthermore, since $w^\nu \in [0,1]^m$ and F_j are marginal probability distributions, we have that the image of M_ν is contained in a bounded interval: $M_\nu(x) \in [-m,\, 1+m]$ for all $x \in \mathbb{R}^m$ and all ν. Hence, the objective function of the master program (P_M) is continuous and tends to $-\infty$ as $\|x\|$ goes to ∞ (i.e., $-(M_\nu(x) - L_k\|x - \hat{x}^\nu\|^2/2)$ is coercive). Together with Assumption 2(i), this property guarantees that (P_M) has a solution and thus the iterates produced by the algorithm are well defined.

[1] That is, $\mathbb{P}(\xi \le z) = C(F_1(z_1), \ldots, F_m(z_m))$.



(a) (b)

FIGURE 1. Comparison of linear (linearization of φ) and MAJU model at two different points. In this example, $\varphi(x) = \mathbb{P}(\xi \leq h(x))$, with $h(x) = (x, x^2)^\top$ and ξ a bi-variate Gaussian distribution with $\texttt{Cov}(\xi_1, \xi_2) = -0.95$. The function is depicted in gray, and the models in yellow.

Algorithm 1 Marginal Assessment of Joint Uncertainty – MAJU Algorithm

---

1: **Input:** Initial point $\hat{x}^0 \in X$ and parameters $\ell > 0$, $\beta > 1$ and $\gamma \in (0, 1)$
   Choose $L_0 \in (0, \ell)$ and set $\nu = 0$
2: **for** $k = 0, \ldots$ **do**
3: Define M_ν by (31) with $w^\nu \in [0, 1]^m$ solving the linear system (32)
4: Let x^{k+1} be a stationary point of the master problem:

$$\max_{x \in X} \; M_\nu(x) - \frac{L_k}{2}\|x - \hat{x}^\nu\|^2 \tag{P$_M$}$$

5: **if** $x^{k+1} = \hat{x}^\nu$ **then**
6: Stop: Return $\hat{x}^\nu$ and $\varphi(\hat{x}^\nu)$
7: **end if**
8: **if** $\varphi(x^{k+1}) \geq \varphi(\hat{x}^\nu) + \frac{\gamma}{2}\|x^{k+1} - \hat{x}^\nu\|^2$ **then** ▷ Serious step
9: Set $\hat{x}^{\nu+1} \leftarrow x^{k+1}$, $\nu \leftarrow \nu + 1$ and choose $L_{k+1} \in (0, \ell)$
10: **else** ▷ Null step
11: Set $L_{k+1} \leftarrow \beta L_k$ and go back to Step 4
12: **end if**
13: **end for**

Observe that the algorithm uses two distinct index counters, k and ν. The index k tracks the iterations performed by the algorithm, which coincide with the number of times the subproblem (P$_M$) is solved. In contrast, the index ν counts the number of serious steps, which coincides with the number of evaluations of the function and its gradient. Once a new serious step is performed, the new iterate becomes the model's center of stability.

5.1. **Summary of convergence properties.** This section states that if the algorithm stops after finitely many steps, then the last stability center $\hat{x}^\nu$ is stationary for problem (PMP). Otherwise, the algorithm produces an infinite of sequence $(\hat{x}^\nu)_{\nu \in \mathbb{N}}$ whose cluster points are stationary. We say that $\hat{x} \in X$ is a stationary point of problem (PMP) if it satisfies the first-order optimality condition

$$0 \in -\nabla\varphi(\hat{x}) + N_X(\hat{x}).$$

We will need the following mild assumption.

Assumption 3. *Suppose that at every iteration k, the stationary point x^{k+1} of subproblem* (P_M) *is guaranteed to be at least as good as $\hat{x}^\nu$ with respect to the same subproblem, that is, x^{k+1} satisfies:*

$$M_\nu(x^{k+1}) - \frac{L_k}{2}\|x^{k+1} - \hat{x}^\nu\|^2 \geq M_\nu(\hat{x}^\nu). \tag{33}$$

It is worth noting that Assumption 3 is not restrictive. Indeed, since $\hat{x}^\nu \in X$ is available, we can initialize the solver with $\hat{x}^\nu$ when computing the next iterate. Any mathematically sound solver implementing an *ascent algorithm* will return a stationary point that is at least as good as the initial point.

The following theorem asserts that the MAJU Algorithm has convergence gurarantees to stationary points. The proof is presented in Subsection 5.2.1.

Theorem 5.1 (Convergence of cluster points)**.** *Consider the* MAJU *Algorithm applied to problem* (PMP) *and suppose that* Assumptions *2 and 3 hold. If the algorithm stops after finitely many steps, then the last stability center $\hat{x}^\nu$ is a stationary point of problem* (PMP)*. Otherwise, the algorithm loops indefinitely, $\nu \to \infty$, and every cluster point (if any) $\hat{x}$ of the sequence $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ is stationary for problem* (PMP)*. Furthermore,*

$$\min_{\iota=0,\dots,\nu} \|\hat{x}^{\iota+1} - \hat{x}^\iota\| \leq \sqrt{\frac{2}{\gamma(\nu+1)}}.$$

In what follows we state the convergence rate of the sequence generated by Algorithm 1 under the *Polyak–Łojasiewicz–Kurdyka (PLK)* property. Recall that a function $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ satisfies the PLK property[2] at $\bar{x}$ if there exist constants $\eta > 0$ and a continuous, concave function $\theta : [0, \eta] \to [0, \infty)$ such that $\theta(0) = 0$, θ is continuously differentiable on $(0, \eta)$, and $\theta' > 0$, with the following inequality holding:

$$\theta'\big(\psi(\bar{x}) - \psi(x)\big)\, \mathrm{dist}\big(0; \partial(-\psi)(x)\big) \ \geq 1, \tag{34}$$

for all $x \in \mathbb{B}_\eta(\bar{x})$ satisfying $\psi(\bar{x}) > \psi(x) > \psi(\bar{x}) - \eta$. Here, $\mathrm{dist}(\cdot; C)$ denotes the distance to the set C.

Furthermore, we will need the following stronger version of Assumption 2:

Assumption 4. *Suppose that:*

i) X is a nonempty closed set;

ii) the probability function φ, marginal functions $F_j(z) := \mathbb{P}(\xi_j \leq z)$, and the constraint mapping $h\colon \mathbb{R}^n \to \mathbb{R}^m$ are $C^{1,1}$ on X.

Remark 4. *Given the structure in* (PMP)*, the function φ is $C^{1,1}$ on X, for instance, when Assumption 2 holds and X is in addition convex and compact.*

Remark 5 (On updating the Lipschitz estimate L_k)**.** *An alternative strategy for estimating the Lipschitz constant L_k in* (P_M) *is to* keep it unchanged after serious steps*, i.e., to set $L_{k+1} = L_k$, rather than resetting $L_{k+1} \in (0, \ell)$ once a serious step is accepted. In this case, one can show, under Assumption 4, that the Lipschitz estimate eventually stabilizes, and only finitely many null steps can occur (see Proposition 5.6).*

[2] Note that the inequalities have been reversed to align with the maximization framework considered here.

Theorem 5.2 (Convergence of the whole sequence under the PLK condition)**.** *Consider the* MAJU *Algorithm applied to problem* (PMP) *and suppose that* Assumptions *3 and 4 hold. Let $\hat{x}$ be an accumulation point of the sequence $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ produced by the algorithm. Suppose furthermore that the Polyak–Lojasiewicz–Kurdyka property* (34) *holds for the function $\psi(x) = \varphi(x) - \mathbf{i}_X(x)$ at $\hat{x}$, and that the algorithm does not update the Lipschitz estimate L_k after serious steps (Remark 5). Then, $\sum_{\nu=0}^{\infty} \|\hat{x}^{\nu+1} - \hat{x}^\nu\| < +\infty$, and consequently, the whole sequence $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ converges to $\hat{x}$, which is a stationary point of problem* (PMP)*.*

The proof can be found in Section 5.2.2, alongside other needed technical results.

Remark 6 (On weakening Assumption 4)**.** *It is possible to weaken Assumption 4 by reverting instead solely on Assumption 2 all while deriving the same convergence rates. However this comes at the cost of more involved proofs. We have opted for a simpler presentation relying on a global and not local property.*

We note that one widely used version of the PLK condition involves $\theta(t) = Mt^{1-q}$. In this case, we can provide convergence rates that depend on the allowed q value.

Corollary 5.3 (Convergence rates)**.** *Under the setting of Theorem 5.2, suppose that the PLK property holds at $\hat{x}$ with $\theta(t) = Mt^{1-q}$ for some $M > 0$ and $q \in [0, 1)$. The following convergence rates are guaranteed for the sequences $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ and $(\varphi(\hat{x}^\nu))_{\nu\in\mathbb{N}}$:*

a) If $q = 0$, then $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ and $(\varphi(\hat{x}^\nu))_{\nu\in\mathbb{N}}$ terminate at $\hat{x}$ and $\varphi(\hat{x})$ in a finite number of steps.

b) If $q \in (0, \frac{1}{2})$, then either $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ and $(\varphi(\hat{x}^\nu))_{\nu\in\mathbb{N}}$ terminate at $\hat{x}$ and $\varphi(\hat{x})$ in finitely many steps, or they converge superlinearly to $\hat{x}$ and $\psi(\hat{x})$ as $\nu \to \infty$.

c) If $q = \frac{1}{2}$, then $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ and $(\varphi(\hat{x}^\nu))_{\nu\in\mathbb{N}}$ converge linearly to $\hat{x}$ and $\varphi(\hat{x})$ as $\nu \to \infty$.

d) If $q \in (\frac{1}{2}, 1)$, then there exist positive constants η and σ such that $\varphi(\hat{x}) - \varphi(\hat{x}^\nu) \leq \eta\, \nu^{-\frac{1}{2q-1}}$ and $\|\hat{x}^\nu - \hat{x}\| \leq \sigma\, \nu^{-\frac{1-q}{2q-1}}$, for all sufficiently large $\nu \in \mathbb{N}$.

Proof. Let us first observe that, by Theorem 5.2, the sequence $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ eventually enters the ball $\mathbb{B}_r(\hat{x})$ given in Lemma 5.8. Hence, the *relative error* condition is satisfied, and the claim follows by applying [2, Theorem 2 and Theorem 3]. □

5.2. **Analysis and proofs of the main results.** This subsection establishes the convergence of the sequence generated by the MAJU Algorithm. The exposition is divided into two parts. The first part collects the main ingredients needed to analyze subsequence convergence of the iterates, thereby proving Theorem 5.1. The second part addresses convergence of the full sequence stated in Theorem 5.2 and provides convergence-rate results under the Polyak–Łojasiewicz–Kurdyka property.

5.2.1. *Subsequence convergence of the iterates - proof of Theorem 5.1.* Given Assumption 2, the first result shows that the MAJU Algorithm generates points $\hat{x}^\nu$ around which the probability function φ from problem (PMP) is $C^{1,1}$.

Lemma 5.4. *Let φ be the probability function in* (PMP)*, and suppose* Assumption *2 holds. Then φ is locally $C^{1,1}$ in a neighbourhood of the set X.*

Proof. Let $\hat{x} \in X$. Since h is a mapping and $h(\hat{x}) \in \mathbb{R}^m$ some vector, clearly, there exists a vector $z \in \mathbb{R}^m$ having all components strictly smaller, i.e., $z < h(\hat{x})$. Therefore, by Corollary 3.4, we conclude that φ is of class $C^{1,1}$ in a neighbourhood of $\hat{x}$. As $\hat{x} \in X$ is arbitrary, we conclude that φ is locally $C^{1,1}$ in a neighbourhood of the set X. □

In what follows we show that, under our standing assumptions, the model M_ν in (31) provides a quadratically accurate local approximation of φ near the stability center $\hat{x}^\nu$.

Lemma 5.5. *Suppose that* Assumption *2 and 3 hold, and let $\hat{x} \in X$. Then, there exist a neighbourhood W of $\hat{x}$ and a constant $L_M > 0$ such that for every stability center $\hat{x}^\nu \in W$ of the* MAJU *Algorithm, the following inequality holds $|M_\nu(x) - \varphi(x)| \leq \frac{L_M}{2}\|x - \hat{x}^\nu\|^2$ for all $x \in W \cap X$. If, instead of Assumption 2, we impose Assumption 4 then the last bound holds uniformly on X.*

Proof. It follows from lemma 5.4 that φ is $C^{1,1}$ on a neighbourhood of $\hat{x}$. Let us denote W such a neighbourhood and L_φ the associated Lipschitz constant. Moreover, by Assumption 2, the same regularity (but with different constant) holds for each function $x \mapsto F_j(h_j(x))$. Shrinking W if necessary, we denote by L_j the Lipschitz constants associated with $F_j \circ h_j$ on $W \cap X$. Observe that for every $\nu \in \mathbb{N}$, the model M_ν can be equivalently written as:

$$M_\nu(x) = \varphi(\hat{x}^\nu) + \nabla\varphi(\hat{x}^\nu)^\top(x - \hat{x}^\nu) + \sum_{j=1}^m w_j^\nu \left[F_j(h_j(x)) - F_j(h_j(\hat{x}^\nu))\right] - \nabla\varphi(\hat{x}^\nu)^\top(x - \hat{x}^\nu).$$

By expanding the last term using (32) with $w = w^\nu$, we further rewrite $M_\nu(x)$ as:

$$\begin{aligned} M_\nu(x) = {} & \varphi(\hat{x}^\nu) + \nabla\varphi(\hat{x}^\nu)^\top(x - \hat{x}^\nu) \\ & + \sum_{j=1}^m w_j^\nu \Big[F_j(h_j(x)) - \Big(F_j(h_j(\hat{x}^\nu)) + \nabla h_j(\hat{x}^\nu) f_j(h_j(\hat{x}^\nu))^\top (x - \hat{x}^\nu)\Big)\Big], \end{aligned}$$

where f_j denotes the density function corresponding to the j-th marginal. Now, if $\hat{x}^\nu \in W$, let us consider an arbitrary point $x \in W \cap X$. Using the descent lemma (see [19, Lemma A.11]), we establish that:

$$\begin{aligned} |M_\nu(x) - \varphi(x)| \leq {} & \Big|\varphi(\hat{x}^\nu) + \nabla\varphi(\hat{x}^\nu)^\top(x - \hat{x}^\nu) - \varphi(x)\Big| \\ & + \sum_{j=1}^m \Big|w_j^\nu \Big[F_j(h_j(x)) - \Big(F_j(h_j(\hat{x}^\nu)) + \nabla h_j(\hat{x}^\nu) f_j(h_j(\hat{x}^\nu))^\top (x - \hat{x}^\nu)\Big)\Big]\Big| \\ \leq {} & \frac{L_\varphi}{2}\|x - \hat{x}^\nu\|^2 + \sum_{j=1}^m w_j^\nu \frac{L_j}{2}\|x - \hat{x}^\nu\|^2 \leq \frac{L_M}{2}\|x - \hat{x}^\nu\|^2, \end{aligned} \tag{35}$$

with $L_M := \Big(L_\varphi + \sum_{j=1}^m L_j\Big)$. If instead of Assumption 2, we impose Assumption 4, then φ and each $F_j \circ h_j$ are $C^{1,1}$ on X. We can then take W such that $X \subset W$. The proof is completed. □

The next proposition shows that the inner loop described in Steps 4–11 of the MAJU Algorithm must terminate finitely.

Proposition 5.6 (Null steps)**.** *Suppose that* Assumption *2 and 3 hold. Then the* MAJU *Algorithm applied to problem* (PMP) *produces only finitely many consecutive null steps.*
If, instead of Assumption 2, we impose Assumption 4 and additionally require that the algorithm does not update the Lipschitz estimate L_k after serious steps (Remark 5), then only finitely many null steps are generated, and the sequence L_k eventually becomes constant.

Proof. If the algorithm stops at an iteration k with $x^{k+1} = \hat{x}^\nu$ the result holds trivially. In what follows we consider the case in which the algorithm loops indefinitely. We proceed with a proof by contradiction. Suppose that the algorithm produces a last stability center $\hat{x}^\nu$ at iteration $\hat{k}$ and only null steps are produced there after:

$$\nu \text{ is fixed and } \varphi(\hat{x}^\nu) > \varphi(x^{k+1}) - \tfrac{\gamma}{2}\|x^{k+1} - \hat{x}^\nu\|^2 \text{ for all } k > \hat{k}. \tag{36}$$

Then, $L_k \to \infty$ due to the rule to update the Lipschitz estimate after null steps (recall that $\beta > 1$ in the MAJU Algorithm). It follows from Assumption 3 that $M_\nu(x^{k+1}) - \frac{L_k}{2}\|x^{k+1} - \hat{x}^\nu\|^2 \geq M_\nu(\hat{x}^\nu) = \varphi(\hat{x}^\nu)$. Combining this inequality with (36) for $k > \hat{k}$ and recalling that $M_\nu(x) \leq m+1$ for all x we get

$$m + 1 - \frac{L_k}{2}\|x^{k+1} - \hat{x}^\nu\|^2 \geq M_\nu(x^{k+1}) - \frac{L_k}{2}\|x^{k+1} - \hat{x}^\nu\|^2 \geq \varphi(x^{k+1}) - \frac{\gamma}{2}\|x^{k+1} - \hat{x}^\nu\|^2,$$

i.e., $m + 1 - \frac{L_k - \gamma}{2}\|x^{k+1} - \hat{x}^\nu\|^2 \geq 0$. As $L_k \to \infty$, we conclude that $\lim_{k\to\infty} x^{k+1} = \hat{x}^\nu$. Lemma 5.4 ensures the existence of a neighbourhood W of $\hat{x} = \hat{x}^\nu$ such φ is $C^{1,1}$ on W, with constant $L_M > 0$. By taking $k' > \hat{k}$ sufficiently large, we have that $x^{k+1} \in W$ for all $k > \hat{k}$. As a result, Lemma 5.5 gives $M_\nu(x^{k+1}) \leq \varphi(x^{k+1}) + \frac{L_M}{2}\|x^{k+1} - \hat{x}^\nu\|^2$ for all $k \geq k'$. Together with Assumption 3, this inequality yields

$$\varphi(\hat{x}^\nu) \leq M_\nu(x^{k+1}) - \frac{L_k}{2}\|x^{k+1} - \hat{x}^\nu\|^2 \leq \varphi(x^{k+1}) - \frac{L_k - L_M}{2}\|x^{k+1} - \hat{x}^\nu\|^2. \tag{37}$$

For $L_k \geq L_M + \gamma$, the above gives $\varphi(\hat{x}^\nu) + \frac{\gamma}{2}\|x^{k+1} - \hat{x}^\nu\|^2 \leq \varphi(x^{k+1})$, i.e., x^{k+1} satisfies the algorithm ascent step, qualifying the latter as the new serious step. As a result, this contradicts that $\hat{x}^\nu$ is the last stability center. Hence, only finitely many consecutive null steps can occur. If, instead of Assumption 2, we impose Assumption 4 and additionally require that the algorithm does not update the Lipschitz estimate L_k after serious steps, then the sequence $(L_k)_{k\in\mathbb{N}}$ is nondecreasing. After finitely many null steps, the algorithm forces L_k to exceed $L_M + \gamma$, where the constant L_M and the inequality from Lemma 5.5 hold uniformly on X by Assumption 4. Consequently, once $L_k \geq L_M + \gamma$, relation (37) implies that x^{k+1} becomes a stability center. Since L_k is not allowed to decrease, the algorithm subsequently performs only serious steps, and therefore L_k becomes constant. ☐

Proposition 5.7 (Serious steps)**.** *Suppose that* Assumption *2 and 3 hold, and that the* MAJU *Algorithm applied to problem* (PMP) *loops indefinitely. Then $\nu \to \infty$ and every cluster point (if any) of the sequence of stability center $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ is stationary for problem* (PMP)*. Furthermore,*

$$\min_{\iota=0,\dots,\nu} \|\hat{x}^{\iota+1} - \hat{x}^\iota\| \leq \sqrt{\frac{2}{\gamma(\nu+1)}}.$$

Proof. Since the algorithm loops indefinitely, it follows from Proposition 5.6 that the algorithm produces infinitely many serious steps: $\nu \to \infty$. As a result of the ascent test, we have that

$$\frac{\gamma}{2}\|\hat{x}^{\iota+1} - \hat{x}^\iota\|^2 \leq \varphi(\hat{x}^{\iota+1}) - \varphi(\hat{x}^\iota) \quad \text{for all} \quad \iota.$$

Summing over $\iota = 0, \dots \nu$ and using that $\varphi(x) \in [0,1]$, we get

$$\frac{\gamma}{2}(\nu+1) \min_{\iota=0,\dots,\nu} \|\hat{x}^{\iota+1} - \hat{x}^\iota\|^2 \leq \sum_{\iota=0}^{\nu} \frac{\gamma}{2}\|\hat{x}^{\iota+1} - \hat{x}^\iota\|^2 \leq \sum_{\iota=0}^{\nu}[\varphi(\hat{x}^{\iota+1}) - \varphi(\hat{x}^\iota)] \leq \varphi(\hat{x}^{\nu+1}) - \varphi(\hat{x}^0) \leq 1.$$

Thus $\min_{\iota=0,\dots,\nu} \|\hat{x}^{\iota+1} - \hat{x}^\iota\| \leq \sqrt{\frac{2}{\gamma(\nu+1)}}$ and moreover it also holds that $\lim_{\nu\to\infty} \|\hat{x}^{\nu+1} - \hat{x}^\nu\| = 0$. Let $\hat{x} \in X$ be an arbitrary accumulation point of $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$. Then there exists an index set $\mathcal{N} \subset \mathbb{N}$ such that $\lim_{\nu\in\mathcal{N}} \hat{x}^\nu = \hat{x}$. As the difference $\hat{x}^{\nu+1} - \hat{x}^\nu$ vanishes, we also get that $\lim_{\nu\in\mathcal{N}} \hat{x}^{\nu+1} = \hat{x}$. It follows from (31) that $\nabla M_\nu(x) = \sum_{j=1}^m w_j^\nu f_j(h_j(x))\nabla h_j(x)$ for all $x \in X$. By the choice of the w in (32), we have that

$$\nabla M_\nu(\hat{x}^{\nu+1}) = \sum_{j=1}^m w_j^\nu f_j(h_j(\hat{x}^{\nu+1}))\nabla h_j(\hat{x}^{\nu+1}) \;\; \text{and} \;\; \sum_{j=1}^m w_j^\nu f_j(h_j(\hat{x}^\nu))\nabla h_j(\hat{x}^\nu) = \nabla\varphi(\hat{x}^\nu).$$

By adding and subtracting $\nabla\varphi(x^\nu)$ to and from the first equality above, and by using the second inequality, we obtain

$$\nabla M_\nu(\hat{x}^{\nu+1}) = \nabla\varphi(\hat{x}^\nu) + \sum_{j=1}^m w_j^\nu \left[\nabla h_j(\hat{x}^{\nu+1}) f_j(h_j(\hat{x}^{\nu+1})) - \nabla h_j(\hat{x}^\nu) f_j(h_j(\hat{x}^\nu))\right].$$

Boundedness of $(w^\nu)_{\nu\in\mathbb{N}}$ and continuity of the involved functions and gradients ensure that $\lim_{\nu\in\mathcal{N}} \nabla M_\nu(\hat{x}^{\nu+1}) = \nabla\varphi(\hat{x})$. Now, let us introduce an additional notation: $k(\nu)$ is an iteration index at which the stability center $\hat{x}^{\nu+1}$ is produced. In other words, $\hat{x}^{\nu+1}$ is a stationary point of

$$\min_{x\in X}\ M_\nu(x) - \frac{L_{k(\nu)}}{2}\|x-\hat{x}^\nu\|^2.$$

The optimality condition for this problem reads:

$$0 \in -\nabla M_\nu(\hat{x}^{\nu+1}) + N_X(\hat{x}^{\nu+1}) - L_{k(\nu)}(\hat{x}^{\nu+1} - \hat{x}^\nu) \quad \text{for all} \ \ \nu.$$

If there exists an infinite index set $\mathcal{N}' \subset \mathcal{N}$ such that the subsequence $(L_{k(\nu)})_{\nu\in\mathcal{N}'}$ is bounded, then by taking the limit in the above inclusion as $\nu\in\mathcal{N}'$ goes to infinity, and recalling that the normal cone is upper semicontinuous, we obtain $0 \in -\nabla\varphi(\hat{x}) + N_X(\hat{x})$.

Hence, to conclude the proof we need to show that such an index set $\mathcal{N}'\subset\mathcal{N}$ exists. This is clearly the case if the algorithm generates only finitely many null steps: we can take $\mathcal{N}'=\mathcal{N}$ as the whole sequence $(L_k)_{k\in\mathbb{N}}$ is bounded in this case. In what follows we consider the remaining case in which the algorithm generates infinitely many null steps (interchanged with serious steps). To this end, let us denote the index set

$$\mathcal{N}^{\mathtt{n}} = \{\nu\in\mathcal{N} :\, x^{k(\nu)} \text{ is a null iterate}\} \quad = \quad \{\nu\in\mathcal{N} :\, x^{k(\nu)} \neq \hat{x}^{\nu'}\ \forall\,\nu'\},$$

and consider two alternatives: $\mathcal{N}^{\mathtt{n}}$ is finite or infinite.

If $\mathcal{N}^{\mathtt{n}}$ is finite (or even empty), we conclude that $x^{k(\nu)} = \hat{x}^\nu$ for all $\nu\in\mathcal{N}$ large enough. As a result, the rule for updating L_k after a serious step ensures that $L_{k(\nu)} \leq \ell$ for all $\nu\in\mathcal{N}$ large enough. Hence, there exists $\mathcal{N}'\subset\mathcal{N}$ such that $(L_{k(\nu)})_{\nu\in\mathcal{N}'}$ is bounded.

Therefore, we can now move to the case that $\mathcal{N}^{\mathtt{n}}$ is infinite, then for all $\nu\in\mathcal{N}^{\mathtt{n}}$, the point $x^{k(\nu)}$ is a null iterate:

$$\varphi(x^{k(\nu)}) < \varphi(\hat{x}^\nu) + \frac{\gamma}{2}\|x^{k(\nu)} - \hat{x}^\nu\|^2.$$

Assumption 3 ensures that

$$\varphi(\hat{x}^\nu) \leq M_\nu(x^{k(\nu)}) - \frac{L_{k(\nu)-1}}{2}\|x^{k(\nu)} - \hat{x}^\nu\|^2.$$

As in this case $L_{k(\nu)} = \beta L_{k(\nu)-1}$ (null step), combining these two inequalities yields

$$\begin{aligned} 0 \leq \varphi(x^{k(\nu)}) &< \varphi(\hat{x}^\nu) + \frac{\gamma}{2}\|x^{k(\nu)} - \hat{x}^\nu\|^2 \\ &< M_\nu(x^{k(\nu)}) - \frac{L_{k(\nu)-1}-\gamma}{2}\|x^{k(\nu)} - \hat{x}^\nu\|^2 \\ &= M_\nu(x^{k(\nu)}) - \frac{\beta^{-1}L_{k(\nu)}-\gamma}{2}\|x^{k(\nu)} - \hat{x}^\nu\|^2 \\ &\leq m+1 - \frac{\beta^{-1}L_{k(\nu)}-\gamma}{2}\|x^{k(\nu)} - \hat{x}^\nu\|^2, \end{aligned} \tag{38}$$

i.e., $\frac{\beta^{-1}L_{k(\nu)}-\gamma}{2}\|x^{k(\nu)} - \hat{x}^\nu\|^2 \leq m+1$ for all $\nu\in\mathcal{N}^{\mathtt{n}}$. Suppose for the purpose of contradiction that $\mathcal{N}^{\mathtt{n}}$ does not admit an infinite subset $\mathcal{N}'$ such that $(L_{k(\nu)})_{\nu\in\mathcal{N}'}$ is bounded. Therefore, $\lim_{\nu\in\mathcal{N}^{\mathtt{n}}} \beta^{-1}L_{k(\nu)} = \infty$, implying $\lim_{\nu\in\mathcal{N}^{\mathtt{n}}} x^{k(\nu)} = \lim_{\nu\in\mathcal{N}^{\mathtt{n}}} \hat{x}^\nu = \hat{x}$. From Lemmas 5.4 and 5.5

we have that φ is $C^{1,1}$ in a neighbourhood W of $\hat{x}$. Observe that $\hat{x}^\nu$, $x^{k(\nu)} \in W$ for all $\nu \in \mathcal{N}^{\mathtt{n}}$ large enough. As a result, Lemma 5.5 applies: $M_\nu(x^{k(\nu)}) \leq \varphi(x^{k(\nu)}) + \frac{L_M}{2}\|x^{k(\nu)} - \hat{x}^\nu\|^2$ for all $\nu \in \mathcal{N}^{\mathtt{n}}$ large enough. By plugging this inequality into (38) we get

$$\varphi(x^{k(\nu)}) < \varphi(x^{k(\nu)}) - \frac{\beta^{-1}L_{k(\nu)} - \gamma - L_M}{2}\|x^{k(\nu)} - \hat{x}^\nu\|^2,$$

i.e, $\beta^{-1}L_{k(\nu)} - \gamma - L_M < 0$, or equivalently, $L_{k(\nu)} < \beta(L_M + \gamma)$ for all $\nu \in \mathcal{N}^{\mathtt{n}}$ large enough. This contradicts our assumption that $\mathcal{N}^{\mathtt{n}}$ does not admit an infinite subset $\mathcal{N}'$ such that $(L_{k(\nu)})_{\nu \in \mathcal{N}'}$ is bounded. The proof is thus complete. □

We are now in a position to present the proof of Theorem 5.1.

Proof. of Theorem 5.1.
We distinguish two cases, depending on whether the algorithm terminates after finitely many iterations or loops indefinitely. Let us first consider the case in which the algorithm terminates: there exists an index k such that $x^{k+1} = \hat{x}^\nu$. In this situation, given the definition of x^{k+1}, we have that $\hat{x}^\nu$ is stationary for the subproblem (P_M), i.e., $0 \in -\nabla M_\nu(\hat{x}^\nu) + N_X(\hat{x}^\nu)$. By construction of the model M_ν, we have $\nabla M_\nu(\hat{x}^\nu) = \nabla\varphi(\hat{x}^\nu)$, and hence the last stability center is stationary for (PMP): $0 \in -\nabla\varphi(\hat{x}^\nu) + N_X(\hat{x}^\nu)$.
The analysis of the case in which the algorithm loops indefinitely follows directly from Proposition 5.7. □

5.2.2. *Convergence rate - proof of Theorem 5.2.* If the MAJU Algorithm terminates after finitely many steps, then Theorem 5.2 holds trivially. In what follows we assume that the algorithm loops indefinitely. Recall that, by the theorem's assumption, the algorithm does not update the Lipschitz estimate L_k after serious steps. It thus follows from Proposition 5.6 that the algorithm produces only finitely many null steps, and consequently the sequence $(L_k)_{k\in\mathbb{N}}$ becomes eventually constant; that is, $L_k = \hat{L}$ for all k sufficiently large. Consequently, for all sufficiently large k, the algorithm's iterates satisfy

$$\begin{cases} 0 \in -\nabla M_\nu(\hat{x}^{\nu+1}) + N_X(\hat{x}^{\nu+1}) + \hat{L}(\hat{x}^{\nu+1} - \hat{x}^\nu) \\ \varphi(\hat{x}^{\nu+1}) \geq \varphi(\hat{x}^\nu) + \frac{\gamma}{2}\|\hat{x}^{\nu+1} - \hat{x}^\nu\|^2 \\ M_\nu(\hat{x}^{\nu+1}) - \frac{\hat{L}}{2}\|\hat{x}^{\nu+1} - \hat{x}^\nu\|^2 \geq M_\nu(\hat{x}^\nu) \\ \hat{x}^{\nu+1} \neq \hat{x}^\nu. \end{cases} \tag{39}$$

Remark 7. *As we are concerned here with the asymptotic behaviour of the sequence* $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$*, we can, without loss of generality, assume that the conditions in* (39) *hold for all* ν*.*

The convergence properties of the sequence generated by the MAJU Algorithm follow from the results in [1, 2], which establish general convergence and rate results for sequences satisfying two fundamental assumptions. Specifically, these assumptions concern a generic ascent sequence $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ associated with an upper semicontinuous function $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ satisfying, for $a, b > 0$ constants:

- (H1) **Sufficient increase**[3]: for each $\nu \in \mathbb{N}$, $\psi(\hat{x}^{\nu+1}) - a\|\hat{x}^{\nu+1} - \hat{x}^\nu\|^2 \geq \psi(\hat{x}^\nu)$.
- (H2) **Relative error:** for each $\nu \in \mathbb{N}$, there exists $v^{\nu+1} \in \partial\psi(\hat{x}^{\nu+1})$ such that $\|v^{\nu+1}\| \leq b\|\hat{x}^{\nu+1} - \hat{x}^\nu\|$.

[3]In [1, 2], this condition is stated as *sufficient decrease*. Here, it is reformulated to match our maximization setting.

We remind the reader that in our context, we use $\psi(x) = \varphi(x) - \mathbf{i}_X(x)$, with $\mathbf{i}_X(x)$ the indicator function of set X. Given Remark 7, it is clear that sequence $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ satisfies the sufficient increase condition (H1).
It remains to show that the MAJU Algorithm also satisfies the relative error condition (H2).

Lemma 5.8. *Under the assumptions of Theorem 5.2, let $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ the sequence of stability centers produced by the* MAJU *Algorithm. Then there exist a constant $L > 0$ and $u^{\nu+1} \in N_X(\hat{x}^{\nu+1})$ such that*

$$\| - \nabla\varphi(\hat{x}^{\nu+1}) + u^{\nu+1}\| \leq L \,\|\hat{x}^{\nu+1} - \hat{x}^\nu\| \quad \forall\nu. \tag{40}$$

Proof. Given Assumption 4, let L_φ and L_j, $j = 1, \ldots, m$, be the Lipschitz constants of the gradients of φ and $F_j \circ h_j$ on X. It follows from (39) that there exists $u^{\nu+1} \in N_X(\hat{x}^{\nu+1})$ such that $u^{\nu+1} = \nabla M_\nu(\hat{x}^{\nu+1}) - \hat{L}(\hat{x}^{\nu+1} - \hat{x}^\nu)$, or equivalently,

$$-\nabla\varphi(\hat{x}^{\nu+1}) + u^{\nu+1} = \nabla M_\nu(\hat{x}^{\nu+1}) - \nabla\varphi(\hat{x}^\nu) + \nabla\varphi(\hat{x}^\nu) - \nabla\varphi(\hat{x}^{\nu+1}) - \hat{L}(\hat{x}^{\nu+1} - \hat{x}^\nu).$$

As a result,

$$\begin{aligned}
\left\|-\nabla\varphi(\hat{x}^{\nu+1}) + u^{\nu+1}\right\| &\leq \left\|\nabla M_\nu(\hat{x}^{\nu+1}) - \nabla\varphi(\hat{x}^\nu)\right\| + \left\|\nabla\varphi(\hat{x}^\nu) - \nabla\varphi(\hat{x}^{\nu+1})\right\| + \hat{L}\left\|\hat{x}^{\nu+1} - \hat{x}^\nu\right\| \\
&\leq \left\|\nabla M_\nu(\hat{x}^{\nu+1}) - \nabla\varphi(\hat{x}^\nu)\right\| + (L_\varphi + \hat{L})\left\|\hat{x}^{\nu+1} - \hat{x}^\nu\right\| \\
&= \left\|\sum_{j=1}^m w_j^\nu\left[\nabla h_j(\hat{x}^{\nu+1}) f_j(\hat{x}^{\nu+1}) - \nabla h_j(\hat{x}^\nu) f_j(\hat{x}^\nu)\right]\right\| + (L_\varphi + \hat{L})\left\|\hat{x}^{\nu+1} - \hat{x}^\nu\right\| \\
&\leq (\sum_{j=1}^m L_j + L_\varphi + \hat{L})\left\|\hat{x}^{\nu+1} - \hat{x}^\nu\right\|.
\end{aligned}$$

The result hols with $L = \sum_{j=1}^m L_j + L_\varphi + \hat{L}$. ☐

Consequently, (H2) holds, and we can now prove Theorem 5.2.

Proof. of Theorem 5.2.
As already mentioned, if the MAJU Algorithm terminates in finitely many steps, the claim is immediate. In what follows we focus on the case in which the algorithm loops indefinitely. Based on Remark 7, we assume without of generality that (39) holds for all iterations ν.
As already mentioned, $(\hat{x}^\nu)_{\nu\in\mathbb{N}}$ satisfies the sufficient increase condition (H1) for $\psi(x) = \varphi(x) - \mathbf{i}_X(x)$. Condition (H2) is ensured by Lemma 5.8, with $v^{\nu+1} = \nabla\varphi(\hat{x}^{\nu+1}) - u^{\nu+1}$ and $b = L$ (remember $\partial\mathbf{i}_X(x) = \mathrm{N}_X(x)$). Therefore, applying [2, Theorem 1] we get the result. ☐

6. Efficiency of the approach: numerical experiments

We propose a numerical framework to compare the performance of various classic algorithms for solving probability maximization problems under the assumption that the probability function φ is locally $C^{1,1}$. The comparison includes both established methods and the just introduced MAJU Algorithm.
To evaluate the probability function and its gradient within each algorithm, we employ the spherical–radial decomposition technique. Following [27], given a polytope C and linear function f, we reformulate chance-constrained problems $\min_{x\in C} f(x)$ s.t. $\varphi(x) \geq 1 - \alpha$ as PMPs by defining the feasible set $X := \{x \in C : f(x) \leq T\}$, where $T = \tau f(x^0)$ for a given target factor τ and initial point $x^0 \in C$. The initial point x^0 is obtained by solving the simpler individual chance-constrained problem: $\min_{x\in C} f(x)$ s.t. $\mathbb{P}[\xi_i \leq h_i(x)] \geq 0.95$, $i = 1, \ldots, m$.

In all considered test problems, the mapping h is affine: $h(x) = Ax + b$. Therefore, the previous individual chance-constrained problem is a linear programming problem. All in all, the problems considered in this study share the following canonical form:

$$\max_{x \in X} \ \varphi(x) := \mathbb{P}(\xi \leq Ax + b), \tag{41}$$

where $X \subseteq \mathbb{R}^n$ is a polytope, $\xi \in \mathbb{R}^m$ is a random vector with Gaussian distribution, and $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are given matrix and vector, respectively. Following the numerical setup in [18, 21, 9], we consider three test problems of increasing size and structure. For each of these problems we fix six target factors uniformly spaced $\tau \in \{1.000, 1.020, 1.040, 1.060, 1.080, 1.100\}$, which define different feasible sets in the associated PMP. This yields a total of 18 test instances. We use the following methods for solving problem (41) and compare their performance with that of the MAJU Algorithm: Matlab SQP (SQP), Projected gradient method (PGD), Projected BFGS method (BFGS), Nesterov projected gradient (NPG), and Frank-Wolfe (FW). With the exception of Matlab's built-in SQP solver, all remaining solvers were implemented by us in Matlab. The solvers employ different stopping criteria, and their corresponding tolerances were tuned to ensure a fair and meaningful comparison. For the MAJU Algorithm we have employed the stopping test $\left\|x^{k+1} - \hat{x}^k\right\| \leq \texttt{tol}$, with $\texttt{tol} = 10^{-4}\sqrt{n}$. The solvers are compared on the following test problems.

Cash Matching problem. This problem is a PMP variant of the classical chance-constrained formulation in [18]. It concerns selecting a portfolio of $n = 3$ bond types for a pension fund to maximize the probability of meeting $m = 15$ random payments, subject to a minimum required return. The decision variable $x \in \mathbb{R}^n$ denotes the bond allocation, and $\xi \in \mathbb{R}^m$ models the payment obligations. Additional data appear in [18], and results are compared with those in [3].

Probabilistic Transportation problem. This problem is a probabilistic (PMP) reformulation of the stochastic transportation model of [21]. The goal is to maximize the probability that the random demand for products shipped from $S = 30$ suppliers to $m = 30$ customers is met, while satisfying supply-capacity constraints and ensuring that the total shipment cost remains within a prescribed budget. The decision variable is $x \in \mathbb{R}^{Sm}$, representing the amount of product shipped from suppliers to customers. The random vector $\xi \in \mathbb{R}^m$ represents the uncertain demands of the customers.

Plantoy problem. This test family models a two-month planning problem for two hypothetical oil refineries, following [9, Sec. 6.2.1]. The goal is to choose production, storage, and import decisions for two fuel types so as to maximize the probability of meeting the random second-month demand vector ξ. Specifically, we aim to satisfy this random demand while respecting deterministic constraints, including storage limits, first month demand, and a total budget. The decision vector $x \in \mathbb{R}^n$ $(n = 8)$ encodes the refinery operations, and $\xi \in \mathbb{R}^m$ $(m = 2)$ represents the second-month fuel demand.

Table 1 shows that, for considered instances, MAJU Algorithm consistently attains the largest probability and at the same time requires substantially fewer function evaluations and less CPU time. On average, MAJU Algorithm only uses 6 function evaluations per instance, compared to $72 - 85$ evaluations for the projected gradient and quasi-Newton methods and more than 180 evaluations for SQP and Frank-Wolfe. In terms of CPU time, MAJU Algorithm is the fastest method on average and wins 12 out of 18 instances, while SQP and Frank-Wolfe are never the fastest.

Table 1. Academic problems: function values $\varphi(x^k)$, number of function evaluations $\#F$, and CPU time for all solvers.

Problem data		Solvers																	
		SQP			PGD			BFGS			NPG			FW			MAJU		
	τ	$\varphi(x^k)$	$\#F$	CPU	$\varphi(x^k)$	$\#F$	CPU	$\varphi(x^k)$	$\#F$	CPU	$\varphi(x^k)$	$\#F$	CPU	$\varphi(x^k)$	$\#F$	CPU	$\varphi(x^k)$	$\#F$	CPU
Cash matching $n=3,\ m=15$	1.00	0.321	140	313	0.321	51	123	0.320	65	159	0.317	91	214	0.313	60	129	0.321	4	103
	1.02	0.491	211	479	0.491	85	205	0.491	102	250	0.458	80	191	0.473	353	765	0.491	3	83
	1.04	0.648	311	711	0.646	130	312	0.646	101	246	0.593	80	191	0.632	406	874	0.648	4	110
	1.06	0.767	247	570	0.767	130	313	0.767	110	269	0.722	98	232	0.755	373	805	0.771	5	162
	1.08	0.854	307	704	0.854	133	321	0.854	135	331	0.797	97	230	0.843	341	735	0.857	7	242
	1.10	0.912	359	823	0.912	180	433	0.910	148	364	0.829	88	209	0.900	231	494	0.915	11	391
Plantoy $n=8,\ m=2$	1.00	0.935	13	2	0.935	8	1	0.935	11	2	0.935	12	2	0.935	16	3	0.935	4	1
	1.02	0.981	17	3	0.981	15	3	0.981	11	2	0.981	22	2	0.977	587	96	0.981	4	1
	1.04	0.996	13	2	0.996	25	4	0.996	12	2	0.996	40	6	0.994	585	96	0.996	3	1
	1.06	0.999	17	3	0.999	3	0	0.999	3	0	0.999	6	1	0.998	90	15	0.999	6	1
	1.08	0.999	21	3	0.999	21	3	0.999	5	1	0.999	81	13	0.999	23	4	0.999	6	1
	1.10	0.999	21	3	0.999	37	6	0.999	5	1	0.999	62	10	0.999	16	3	0.999	9	2
Transport $n=900,\ m=30$	1.00	0.320	140	316	0.320	51	126	0.320	65	160	0.317	91	217	0.313	60	132	0.321	4	105
	1.02	0.491	211	495	0.491	85	207	0.491	102	253	0.458	80	194	0.473	353	783	0.491	3	85
	1.04	0.648	311	766	0.646	130	317	0.646	101	249	0.593	80	195	0.632	406	900	0.648	4	112
	1.06	0.767	247	605	0.767	130	317	0.767	110	274	0.722	98	236	0.755	373	832	0.771	5	164
	1.08	0.854	307	714	0.854	133	324	0.854	135	337	0.797	97	234	0.843	341	763	0.857	7	244
	1.10	0.912	359	835	0.912	180	437	0.910	148	367	0.829	88	212	0.900	231	515	0.915	11	399

These observations are consistent with the performance and data profiles (see Figures 2 and 3), which identify the MAJU Algorithm as the most robust solver in this test set. Here, in order to compare the algorithms, we have made use of performance profiles as introduced in [10].

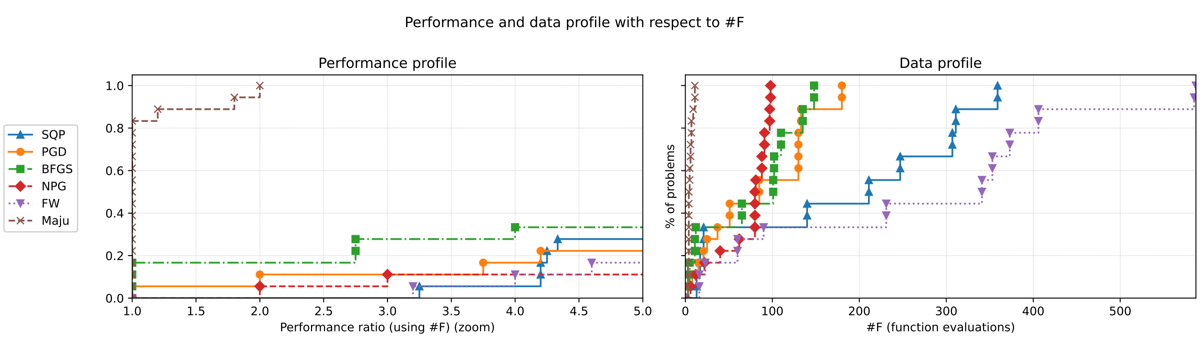


Figure 2. Performance profile (left) and data profile (right) with respect to $\#F$

Figure 2 (left) shows the performance profile with respect to the number of function evaluations $\#F$, that is, the number of times we call the oracle φ. We can see that MAJU Algorithm attains the smallest number of function evaluations on 83% of the instances, while BFGS and PGD are best on at most 17% and 6% of the cases, respectively. Moreover, within a factor 2 of the best $\#F$, MAJU Algorithm already solves 100% of the instances, whereas no other method reaches 50% even within factor 5 of the best.

As for the data profile in Figure 2 (right), it highlights the very low evaluation cost of MAJU Algorithm. With a budget of only 10 function evaluations, MAJU Algorithm already solves about 89% of the instances, whereas the other methods solve at most 17%. With 25 evaluations, MAJU

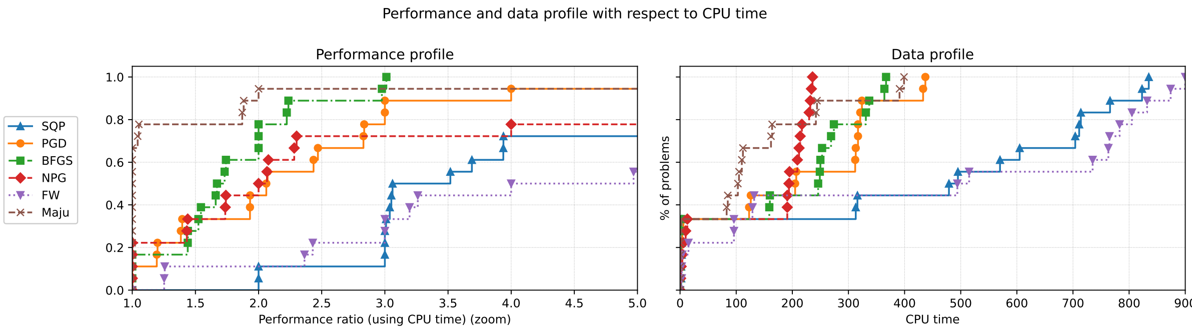


FIGURE 3. Performance profile (left) and data profile (right) with respect to CPU time (in seconds)

Algorithm solves all problems, while the classic algorithms solve at most one third of them. Only when the budget is increased to 200 evaluations do PGD, BFGS and NPG catch up to MAJU Algorithm.

With respect to CPU time, Figure 3 shows us the performance profile on the left, and the data profile on the right. MAJU Algorithm is the fastest method in 67% of the instances and solves 94% of the problems within twice the CPU time of the fastest solver. For larger factors (x-axis), the curves of MAJU Algorithm, BFGS and PGD become quite close, reflecting that these methods eventually solve essentially all problems with comparable CPU times, whereas SQP and FW remain clearly less competitive. The data profile reveals that all methods behave similarly for very small time budgets, since only the easiest instances are solved within a few seconds. As the budget increases, MAJU Algorithm starts to dominate, since with 200 seconds it solves around 78% of the instances, significantly more than the classic algorithms. With a 400 seconds budget, MAJU Algorithm, BFGS and NPG solve all problems, while SQP and FW still solve less than half of the test set.

Acknowledgments. The research of C. Chiu was supported by ANID BECAS/DOCTORADO NACIONAL 2024-21241565. P. Pérez-Aros was supported by ANID (Chile) through Fondecyt Regular grants No. 1220886, No. 1240120, No. 1240335, and No. 1261728; by the CMM BASAL funds for the Center of Excellence FB210005; and by the projects ECOS230027, MATH-AmSud 23-MATH-17 and MATH-AmSud 23-MATH-09.

REFERENCES

[1] Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods. *Mathematical Programming*, 137(1):91–129, 2013.

[2] Glaydston Bento, Boris Mordukhovich, Tiago Mota, and Yurii Nesterov. Convergence of descent optimization algorithms under polyak-łojasiewicz-kurdyka conditions. *Journal of Optimization Theory and Applications*, 207(3):41, 2025.

[3] Emerson Butyn, Elizabeth W. Karas, and Welington de Oliveira. A derivative-free trust-region algorithm with copula-based models for probability maximization problems. *European J. Oper. Res.*, 298(1):59–75, 2022.

[4] A. Charnes and W. Cooper. Chance-constrained programming. *Management Science*, 6:73–79, 1959-1960.

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

[6] Rafael Correa, Marco López, and Pedro Pérez-Aros. Log-exponential approximation in semi-infinite programming: a variational approach. *J. Nonlinear Var. Anal.*, 9(4):539–560, 2025.

[7] E. Csizmás, R. Drenyovszki, T. Szántai, and C. I. Fábián. Random descent steps in a probability maximization scheme. *Journal of Optimization Theory and Applications*, 205(13):1–26, 2025.
[8] W. de Oliveira. Derivative-free approaches for chance-constrained problems with right-hand side uncertainty. *SIAM J. Optim.*, 35(1):1–27, 2025.
[9] Welington de Oliveira. Proximal bundle methods for nonsmooth DC programming. *J. Global Optim.*, 75(2):523–563, 2019.
[10] Elizabeth D. Dolan and Jorge J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91(2, Ser. A):201–213, 2002.
[11] Fabrizio Durante and Carlo Sempi. *Principles of Copula Theory*. Chapman and Hall/CRC, New York, July 2015.
[12] C. I. Fábián. Gaining traction: on the convergence of an inner approximation scheme for probability maximization. *Central European Journal of Operations Research*, 29:491–519, 2021.
[13] C. I. Fábián, E. Csizmás, R. Drenyovszki, T. Vajnai, L. Kovács, and T. Szántai. A randomized method for handling a difficult function in a convex optimization problem, motivated by probabilistic programming. *Annals of Operations Research*, pages 1–32, 2019.
[14] C. I. Fábián, E. Csizmás, R. Drenyovszki, W. van Ackooij, T. Vajnai, L. Kovács, and T. Szántai. Probability maximization by inner approximation. *Acta Polytechnica Hungarica*, 15(1):105–125, 2018.
[15] J. Garnier, A. Omrane, and Y. Rouchdy. Asymptotic formulas for the derivatives of probability functions and their Monte Carlo estimations. *European Journal of Operations Research*, 198:848–858, 2009.
[16] Abderrahim Hantoute, René Henrion, and Pedro Pérez-Aros. Subdifferential characterization of probability functions under Gaussian distribution. *Math. Program.*, 174(1-2, Ser. B):167–194, 2019.
[17] Holger Heitsch. On probabilistic capacity maximization in a stationary gas network. *Optimization*, 69(3):575–604, 2020.
[18] René Henrion. Introduction to chance constraint programming. *Tutorial paper for the Stochastic Programming Community HomePage, http://www.wias-berlin.de/people/henrion/publikat.html*, 01 2004.
[19] Alexey F. Izmailov and Mikhail V. Solodov. *Newton-Type Methods for Optimization and Variational Problems*. Springer, Cham, 2014.
[20] A.I. Kibzun and S. Uryas'ev. Differentiability of probability function. *Stoch. Anal. Appl.*, 16:1101–1128, 1998.
[21] James Luedtke and Shabbir Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM J. Optim.*, 19(2):674–699, 2008.
[22] K. Marti. Differentiation of probability functions : The transformation method. *Computers and Mathematics with Applications*, 30:361–382, 1995.
[23] M. Minoux and R. Zorgati. Convexity of gaussian chance constraints and of related probability maximization problems. *Computational Statistics*, 31(1):387–408, 2016.
[24] M. Minoux and R. Zorgati. Global probability maximization for a gaussian bilateral inequality in polynomial time. *Journal of Global Optimization*, 68(4):879–898, 2017.
[25] B. S. Mordukhovich. *Variational analysis and applications*. Springer Monographs in Mathematics. Springer, Cham, 2018.
[26] G. Pflug and H. Weisshaupt. Probability gradient estimation by set-valued calculus and applications in network design. *SIAM J. Optim.*, 15:898–914, 2005.
[27] András Prékopa. *Programming under Probabilistic Constraint and Maximizing Probabilities under Constraints*, pages 319–371. Springer Netherlands, Dordrecht, 1995.
[28] E. Raik. The differentiability in the parameter of the probability function and optimization of the probability function via the stochastic pseudogradient method (russian). *Izvestiya Akad. Nayk Est. SSR, Phis. Math.*, 24(1):3–6, 1975.
[29] R.T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer Verlag, Berlin, 3rd edition, 2009.
[30] J.O. Royset and E. Polak. Implementable algorithm for stochastic optimization using sample average approximations. *Journal of Optimization Theory and Applications*, 122(1):157–184, 2004.
[31] J.O. Royset and E. Polak. Extensions of stochastic optimization results to problems with system failure probability functions. *Journal of Optimization Theory and Applications*, 133(1):1–18, 2007.
[32] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming*, volume 9 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, second edition, 2014. Modeling and theory.
[33] S. Uryas'ev. Derivatives of probability functions and integrals over sets given by inequalities. *Journal of Computational and Applied Mathematics*, 56(1-2):197–223, 1994.

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

[35] W. van Ackooij, I. Aleksovska, and M. Munoz-Zuniga. (Sub-)differentiability of probability functions with elliptical distributions. *Set-Valued Var. Anal.*, 26(4):887–910, 2018.

[36] W. van Ackooij and R. Henrion. (Sub-) Gradient formulae for probability functions of random inequality systems under Gaussian distribution. *SIAM Journal on Uncertainty Quantification*, 5(1):63–87, 2017.

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

[38] Wim van Ackooij, Diego Morales-Poblete, Pedro Pérez-Aros, and David Villacís. Approximating inequality systems within probability functions: studying implications for problems and consistency of first-order information. 2025.

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

[40] Wim van Ackooij and Pedro Pérez-Aros. Generalized differentiation of probability functions: parameter dependent sets given by intersections of convex sets and complements of convex sets. *Appl. Math. Optim.*, 85(1):Paper No. 2, 39, 2022.

[41] Wim Stefanus van Ackooij and Welington Luis de Oliveira. *Methods of Non-smooth Optimization in Stochastic Programming: From Conceptual Algorithms to Real-World Applications*. International Series in Operations Research & Management Science. Springer, Cham, 1 edition, 2025.

OSIRIS, EDF Lab Paris-Saclay, 7 Boulevard Gaspard Monge, Palaiseau, 91120, France

Departamento de Ingeniería Matemática, Universidad de Chile, Santiago, Chile
Email address: cchiu@dim.uchile.cl

MINES Paris-PSL, CMA – Centre de Mathématiques Appliquées, France

Departamento de Ingeniería Matemática, Universidad de Chile, Santiago, Chile

Centro de Modelamiento Matemático, Universidad de Chile, Santiago, Chile