UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

TECHNIQUES OF VARIATIONAL ANALYSIS: PROBABILITY FUNCTIONS AND ESTIMATORS OF NON-CONVEXITY.

TESIS PARA OPTAR AL GRADO DE DOCTORA EN CIENCIAS DE LA INGENIERÍA, MENCIÓN MODELACIÓN MATEMÁTICA.

CLAUDIA ANDREA SOTO SILVA

PROFESOR GUÍA:
PEDRO PÉREZ AROS

PROFESORES CO-GUÍA:
ARIS DANIILIDIS
ABDERRAHIM HANTOUTE

MIEMBROS DE LA COMISIÓN:
RAFAEL CORREA FONTECILLA
FABIÁN FLORES BAZÁN
JAIME ORTEGA PALMA
EMILIO VILCHES GUTIÉRREZ

Este trabajo ha sido parcialmente financiado por: Agencia Nacional de Investigación y Desarrollo (ANID), Beca de Doctorado Nacional Chile/2017-21170428. Centro de Modelamiento Matemático (CMM) Proyecto ANID BASAL FB210005.

SANTIAGO DE CHILE
2022

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE DOCTORA EN CIENCIAS
DE LA INGENIERÍA,
MENCIÓN MODELACIÓN MATEMÁTICA
POR: **CLAUDIA ANDREA SOTO SILVA**
FECHA: 2022
PROF. GUÍA: PEDRO PÉREZ AROS
PROFES. CO-GUÍA: ARIS DANIILIDIS Y
ABDERRAHIM HANTOUTE

TÉCNICAS DE ANÁLISIS VARIACIONAL: FUNCIONES DE PROBABILIDAD Y ESTIMADORES DE NO-CONVEXIDAD.

Esta tesis tiene como objetivo aplicar técnicas de análisis variacional a dos diferentes temas: el primero es funciones de probabilidad y el segundo la medida de no-convexidad de Cassels.

Comenzamos aproximando dos diferentes formulaciones abstractas de funciones de probabilidad. La primera aproximación es motivada por el hecho que las restricciones en un problema de optimización con incertidumbre pueden resultar ser no-suaves. En este trabajo proponemos una regularización empleando la envoltura de Moreau a una representación escalar de una función de probabilidad que consiste de una desigualdad vectorial, la cual cubre la mayoría de las clases generales de restricciones probabilísticas. Demostramos, bajo leve condiciones, la diferenciabilidad de tal regularización y además su convergencia variacional hacia la función nominal. En consecuencia, cuando consideramos un problema apropiadamente estructurado con restricciones probabilisticas, podemos entonces obtener la convergencia de los minimizadores de los problemas regularizados a los minimizadores del problema original. Finalmente, ilustramos nuestros resultados con ejemplos y aplicaciones en el campo de problemas de optimización chance constrained del tipo joint, semidefinido y probusto. La segunda formulación es una función de probabilidad generada por multifunciones. Aquí nuestro objetivo principal es probar su continuidad del tipo Lipschitz. Para esto, proponemos un enlargement, el cual, via la función distancia, puede ser demostrado que tiene tal propiedad y en consecuencia, por aproximación, obtenemos nuestro objectivo principal. Como aplicación de este resultado, probamos la propiedad de Lipschitz de una función de probabilidad del tipo joint bajo la condición de cuasiconvexidad.

Para el segundo tema, recordamos que el operador proyección sobre conjuntos cerrados y convexos en un espacio de Hilbert es siempre un singleton. La inversa también es cierta en espacios de Hilbert finito-dimensionales, y también para conjuntos débilmente cerrados en cualquier espacio de Hilbert. Esto es el famoso Teorema de Klee. El problema de si tal inversa es cierta para conjuntos cerrados que no son débilmente cerrados está aún sin responder. En esta tesis, aplicamos caracterizaciones variacionales de convexidad a la llamada función de Asplund para obtener una respuesta positiva parcial a este problema mediante una relajación de la proyección. Finalmente, via la medida de Cassels, estimamos la distancia de Hausdorff entre un conjunto y su envoltura convexa en términos de las proyecciones simultáneas hacia el conjunto y hacia su envoltura convexa. En consecuencia, obtenemos una cuantifiación del Teorema de Klee cuando la medida de Cassels es finita.

Esta tesis termina con conclusiones y trabajo a futuro.

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE DOCTORA EN CIENCIAS
DE LA INGENIERÍA,
MENCIÓN MODELACIÓN MATEMÁTICA
POR: **CLAUDIA ANDREA SOTO SILVA**
FECHA: 2022
PROF. GUÍA: PEDRO PÉREZ AROS
PROFES. CO-GUÍA: ARIS DANIILIDIS Y
ABDERRAHIM HANTOUTE

TECHNIQUES OF VARIATIONAL ANALYSIS: PROBABILITY FUNCTIONS AND ESTIMATORS OF NON-CONVEXITY.

This thesis aims to apply techniques of variational analysis to two different subjects: the first one being probability functions and the second one, a particular nonconvexity measure called effective standard deviation.

We approximate two different abstract formulations of probability functions. The first approximation is motivated by the fact that the constraints in optimization problems with uncertainty may result to be nonsmooth. We propose a regularization by applying the Moreau envelope to a scalar representation of a probability function consisting of a vector inequality, which covers most of the general classes of probabilistic constraints. We demostrate, under mild assumptions, the smoothness of such a regularization and that it satisfies a type of variational convergence to the original probability function. Consequently, when considering an appropriately structured problem involving probabilistic constraints, we can thus entail the convergence of the minimizers of the regularized approximate problems to the minimizers of the original problem. Finally, we illustrate our results with examples and applications in the field of (nonsmooth) joint, semidefinite and probust chance constrained optimization problems. The second formulation is a probability function generated by a set-valued mapping. Our main objective is to prove its local Lipschitz continuity. To do so, we propose an inner enlargement that, via the distance function, can be proven to be locally Lipschitz continuous. Subsequently, by approximation, we obtain our main result. As a consequence, we prove the local Lipschitz continuity of a Joint probability function given by a system of inequality constraints with a relaxed convexity assumption.

We recall that the projection operator onto closed convex subsets of Hilbert spaces is single-valued. The converse is also true in finite-dimensional Hilbert spaces, and also for weakly closed sets in any Hilbert space. This is the famous Theorem of Klee. The problem of whether such a converse holds in any Hilbert space for closed sets which are not weakly closed is still unanswered. In this thesis, we apply variational characterizations of convexity results to the Asplund function to obtain a partial positive answer to this problem, provided that the concept of projection is relaxed to the one of weak projections. Finally, via the effective standard deviation measure, we estimate the Hausdorff distance between a set and its closed convex hull in terms of the size of the simultaneous projections on the set and its closed convex hull. Accordingly, we give a quantified version of Klee's theorem provided that the effective standard deviation of the set is finite. This thesis ends with conclusions and future work.

A mi familia.

Agradecimientos

Gracias a todas esas personas que han estado conmigo durante estos años. Desde los que veía semanalmente por las oficinas hasta los que veía una vez al año a kilómetros de distancia.

Gracias en primer lugar a mis profesores guías. Profesor Pedro, gracias por dedicarme tantas horas, por escuchar mis dudas, y por la oportunidad de aprender tanto sobre un tema que nunca pensé que me gustaría tanto. A los profesores Abderrahim y Aris por guiarme y darme la oportunidad de trabajar con ustedes. Ser su alumna es un honor.

Gracias a mis amigos más cercanos del DIM. Diana y Ariel, por esas tardes, noches, y madrugadas de café, estudio, chocolate, y desesperación.

Gracias a todas esas personas del DIM que notan y me llaman la atención cuando desaparezco. Silvia, Karen, Don Oscar, gracias por su preocupación y por su cariño.

Gracias a mis amigos de la UCN por estar ahí a pesar de verlos una vez al año. Diego, Nathy, Fran, espero poder verlos más seguido. También gracias a mis profesores de la UCN, Ricardo Soto y Mario Salas, por su apoyo para continuar mis estudios. En especial a Oscar Rojo, de todo lo que voy aprendiendo siempre hay algo que me recuerda lo que aprendí en sus clases.

Y por supuesto, gracias a mi familia, aunque no entiendan nada de lo que hago y por confiar en mí (pensar que soy la mejor). Muchas gracias.

Contents

Index of Tables

Index of Illustrations

Chapter 1

Introduction

This manuscript is organized as follows: Chapter 2 presents an overview of basic topics, tools from variational analysis, and generalized differentiation, followed by a section dedicated to introduce one of the two main subjects of this thesis: probability functions. Chapter 3, Chapter 4 and Chapter 5 expose the main contributions of this thesis.

Chapter 2: Inner Moreau envelope of probability functions

In this chapter, we consider a probability function $\varphi : \mathcal{H} \to [0, 1]$ formulated by

$$\varphi(x) := \mathbb{P}\left(\omega \in \Omega : \Phi(x, \xi(\omega)) \in -\mathcal{K}\right), \tag{1.1}$$

where $\mathcal{H}$ is a separable Hilbert space, $\xi : \Omega \to \mathbb{R}^m$ is an m-dimensional random vector, $\mathcal{K} \subset \mathcal{Y}$ is a (nonempty) convex cone of a Banach space $\mathcal{Y}$ and $\Phi : \mathcal{H} \times \mathbb{R}^m \to \mathcal{Y}$ is a function.

When solving an optimization problem by applying deterministic numerical techniques, it is necessary to calculate both, the values of the probability constraint function and its gradient. Nevertheless, in some cases, the probability constraint function may be nonsmooth. This motivates us to propose a regularization employing the Moreau envelope of a scalar representation of the probability function $\varphi(x)$ given in (1.1). More precisely, to handle the random vector inequality we consider a compact convex $\mathcal{C} \subseteq \mathcal{Y}^*$, which generates the positive polar cone of $\mathcal{K}$, that is, $\mathrm{cl}^{w^*} \operatorname{cone} \mathcal{C} = \mathcal{K}^+$, where cl^{w^*} denotes the weak*-closure, and assume the existence of a continuously differentiable convex function $h : \mathcal{H} \to \mathbb{R}$ such that for all $v^* \in \mathcal{C}$, the function

$$\mathcal{H} \times \mathbb{R}^m \ni (x, z) \to \Phi^h_{v^*}(x, z) := \langle v^*, \Phi\rangle(x, z) + h(x) \tag{1.2}$$

is convex in both variables, where $\langle v^*, \Phi\rangle(x, z) := \langle v^*, \Phi(x, z)\rangle$. Now we are able to rewrite the probability function φ in (1.1) in its scalar representation as

$$\varphi(x) = \mathbb{P}\left(\omega \in \Omega : S^h_\Phi(x, \xi(\omega)) \le h(x)\right) \text{ for all } x \in \mathcal{H}.$$

where $S^h_\Phi(x, z) := \sup\left\{\langle v^*, \Phi\rangle(x, z) + h(x) : v^* \in \mathcal{C}\right\}$. Finally, the inner regularization of φ that we propose is

$$\varphi_\lambda(x) := \mathbb{P}\left(\omega \in \Omega : \mathsf{e}_\lambda\, \Phi^h(x, \xi(\omega)) \le h(x)\right), \tag{1.3}$$

where $\mathsf{e}_\lambda\, \Phi^h(x, \xi(\omega))$ stands for the Moreau envelope of the function $S^h_\Phi(x, z)$.

Example *Let $\xi \sim \mathcal{N}(0,1)$ and considering the probability function given by the nonsmooth single inequality*

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

where $\Phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is given by $\Phi(x,z) = 2f_1(x) + f_2(z) - 5$, where $f_1(x) = \max(|x| - 1, 0)$ and

$$f_2(z) = \begin{cases} z^2 & \text{if } z \geq 0 \\ -z & \text{otherwise.} \end{cases}$$

It is clear that in this case the cone $\mathcal{K}$ in consideration is given by the set of nonnegative real numbers, the generator of the positive polar cone is nothing more than the singleton $\mathcal{C} = \{1\}$. Also, notice that we can choose $h(x) = 0$ for all $x \in \mathcal{H}$. This probability function is not differentiable at $\bar{x} = 1, -1$ and, given $\lambda > 0$, its inner regularization is given by

$$\varphi_\lambda(x) := \mathbb{P}\left(\mathtt{e}_\lambda f_2(\xi) \leq -2\mathtt{e}_{2\lambda} f_1(x) + 5\right).$$

The details of the proof, the graphs, and the formulas for the Moreau envelope are given in Example 3.3.

Our inner regularization inherits variational properties of the Moreau envelope, for instance, its smoothness and variational convergence to the original function. More precisely, the variational convergence properties of the family φ_λ to the function φ is in terms of hypoconvergence:

Corollary *The sequence of regularizations φ_λ hypo-converges to the probability function φ. In addition, suppose that the function h in (1.2) is weakly continuous, then the sequence of regularizations φ_λ Mosco hypo-converges to the probability function φ.*

The gradient formula of the inner regularization φ_λ is obtained assuming that $\mathcal{H}$ is finite-dimensional, assuming the existence of a Slater point, and imposing a growth condition on f_ξ in order to apply the gradient formula using the so-called spherical radial decomposition.

Theorem *Let $\bar{x} \in \mathcal{H}$ be such that $S^h_\Phi(\bar{x}, 0) < h(\bar{x})$, and assume that f_ξ satisfies the following growth condition*

$$\lim_{\|z\| \to +\infty} \|z\|^{m+1} f_\xi(z) = 0. \tag{1.4}$$

Then, for any given $\lambda > 0$, the probability function φ_λ, defined in (1.3), is continuously differentiable on an appropriate neighbourhood U of $\bar{x}$ and it holds:

$$\nabla \varphi_\lambda(x) = \int_{\mathbb{S}^{m-1}} \mathcal{G}_\lambda(x,v) d\mu_\zeta(v), \quad \text{for all } x \in U,$$

where $\mathcal{G}_\lambda$ is as in (3.16). Moreover, the gradients of $\mathtt{e}_\lambda \Phi^h$ can be computed by the formula

$$\nabla \mathtt{e}_\lambda \Phi^h(x,z) = \frac{(x,z) - \mathtt{Prox}_{\lambda(\langle v^*, \Phi\rangle + h)}(x,z)}{\lambda}, \tag{1.5}$$

where v^* is any active vector at (x,z), that is, $v^* \in \mathcal{C}$ and $\mathtt{e}_\lambda \Phi^h_{v^*}(x,z) = \mathtt{e}_\lambda \Phi^h(x,z)$ in view *of Proposition 3.2.*

Let us consider a convex proper and lower semicontinuous function $\psi : \mathcal{H} \to \mathbb{R}_\infty$, a fixed reliability parameter $p \in [0, 1]$, and the optimization problem

$$\begin{aligned} \min \psi(x) \\ \text{s.t } \ x \in M(p), \end{aligned} \tag{P}$$

where $M(p) := \{x \in \mathcal{H} : \varphi(x) \geq p\}$ and φ is the probability function defined in (1.1). Related to problem (P) we consider the family of problems

$$\begin{aligned} \min \mathsf{e}_\lambda \psi(x) \\ \text{s.t } \ x \in M_\lambda(p), \end{aligned} \tag{P_λ}$$

where $M_\lambda(p) := \{x \in \mathcal{H} : \varphi_\lambda(x) \geq p\}$ for the regularized probability function φ_λ. Our main result states that the approximated probability functions, through their inner Moreau envelope, allow us to approximate the given optimization problem (P).

Theorem *Let $\psi : \mathcal{H} \to \mathbb{R}_\infty$ be a convex, coercive and lower semicontinuous function such that $M(p) \cap \operatorname{dom} \psi \neq \emptyset$. Then*

a) $v(P), v(P_\lambda) \in \mathbb{R}$ for all $\lambda > 0$ and $v(P_\lambda) \to v(P)$.

b) If (P) has a unique optimum x_0 and x_λ is any sequence of optimal solutions for (P_λ), then $x_\lambda \rightharpoonup x_0$, provided that the function h in (3.3) is sequentially weakly continuous. If, furthermore, $\operatorname{dom} \psi = \mathcal{H}$ and ψ^* is Fréchet differentiable on $\operatorname{dom} \partial\psi^*$, then $x_\lambda \to x_0$.

Chapter 3: Generalized differentiation of probability functions generated by set-valued mappings

In this chapter, we investigate a probability function formulated by

$$\varphi(x) := \mathbb{P}(\omega \in \Omega : \xi(\omega) \in \mathcal{S}_i(x) \text{ for all } i = 1, \dots, s), \tag{1.6}$$

where $\xi : \Omega \to \mathbb{R}^n$ is a random vector from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $\mathcal{X}$ is a separable reflexive Banach space, and $\mathcal{S}_i : \mathcal{X} \rightrightarrows \mathbb{R}^m$ with $i = 1, \dots, s$ is a family of set-valued mappings satisfying : there exists a neighborhood U of $\bar{x}$ such that

$$\begin{aligned} &a)\ 0 \in \mathcal{S}_i(x) \text{ for all } x \in U \\ &b)\ \mathcal{S}_i \text{ is locally Lipschitz-like at } (x, z) \in \operatorname{gph} \mathcal{S}_i \text{ and } x \in U \\ &c)\ \mathcal{S}_i \text{ has closed graph and convex values} \end{aligned} \tag{H}$$

In order to prove the local Lipschitz continuity of the function φ in (1.6) we consider the following enlargement: Given $\varepsilon > 0$,

$$\varphi_\varepsilon(x) := \mathbb{P}(\omega \in \Omega : \xi(\omega) \in \mathcal{S}_i(x) + \varepsilon\mathbb{B} \text{ for all } i = 1, \dots, s). \tag{1.7}$$

Now, since our intention is to use known results about the local Lipschitz continuity of a probability function (see Theorem 2.2) we reformulate the enlargement through the distance function as

$$\varphi_\varepsilon(x) = \mathbb{P}(\omega \in \Omega : \mathrm{d}(\xi(\omega), \mathcal{S}_i(x)) \leq \varepsilon \text{ for all } i = 1, \ldots, s).$$

Let us notice that due to the continuity of the probability measure we have that

$$\varphi(x) = \inf_{\varepsilon > 0} \varphi_\varepsilon(x).$$

By applying the results exposed in Theorem 2.2 to $g_i(x,z) := \frac{1}{2}\mathrm{d}^2(z, \mathcal{S}_i(x)) - \frac{\varepsilon^2}{2}$ we obtain

Theorem *Assume that the family of set-valued mappings $\mathcal{S}_i$ satisfy the η-growth condition given in Definition 4.12 at $\bar{x}$ and that each $\mathcal{S}_i$ satisfy (H) at $\bar{x}$.*

Then the probability function (1.7) is locally Lipschitz at $\bar{x}$ and on an appropriate neighborhood U' of $\bar{x}$ it holds:

$$\partial^{\mathsf{b}} \varphi_\varepsilon(x) \subseteq \mathrm{cl}^{w^*}\left(\int_{v \in \mathbb{S}^{m-1}} \mathrm{cl}\,\mathrm{co}\, \mathcal{M}_\varepsilon(x,v) d\mu_\zeta(v) \right), \text{ for all } x \in U', \tag{1.8}$$

where $\mathcal{M}_\varepsilon(x,v)$ are given in (4.14). In addition, if $\mathcal{H}$ is finite-dimensional the closure can be omitted.

We notice that the η-growth condition and the set $\mathcal{M}_\varepsilon(x,v)$ are given in terms of the coderivative of the set-valued mappings $\mathcal{S}_i$ due to Lemma 2.1.

In the sequel, we will require the following interior continuity property for set-valued mappings to ensure the continuity of the radial functions $\rho_\varepsilon(x,v)$ (Lemma 4.3) and boundedness of $\mathcal{M}_\varepsilon(x,v)$ (Lemma 4.6) on $(\varepsilon, x, v) \in (0, \varepsilon') \times U' \times \mathbb{S}^{m-1}$ for some $\varepsilon' > 0$ and some neighborhood U' of $\bar{x}$.

Definition 1.1 *We say that a set-valued mapping $\mathcal{S}$ has the interior continuity property on $U \subseteq \mathcal{H}$, if for every $x \in U$ and $z \in \mathrm{int}(\mathcal{S}(x))$ there exists $r > 0$ such that*

$$\mathbb{B}_r(z) \subseteq \mathcal{S}(x'), \text{ for all } x' \in \mathbb{B}_r(x).$$

Since the original probability function $\varphi(x)$ is the infimum of the arbitrary family $\{\varphi_\varepsilon(x)\}_{\varepsilon>0}$, the basic subdifferential of φ can be estimated in terms of the subgradients for the basic subdifferential of the members of the family (Lemma 4.2)

$$\partial^{\mathsf{b}} \varphi(x) \subseteq \left\{ x^* \in \mathcal{H} : \begin{array}{c} \text{There exist } \exists x_k \to x, \varepsilon_k \to 0^+ \\ \text{and } x_k^* \in \partial^{\mathsf{b}} \varphi_{\varepsilon_k}(x_k) \text{ s.t. } x_k^* \rightharpoonup x^* \end{array} \right\}.$$

This entails our main result

Theorem *Consider each $\mathcal{S}_i$ in the family of set-valued mappings satisfying Assumption (H) at $\bar{x} \in U$ with $0 \in \mathrm{int}(\mathcal{S}_i(x))$ for all $x \in U$ and having the interior continuity property on U. Moreover, assume that the family of set-valued mappings $\mathcal{S}_i$ satisfies the η-growth condition at $\bar{x}$ and that (2.10) holds true.*

Then the probability function (1.6) is locally Lipschitz at $\bar{x}$ and on an appropriate neigh-

borhood U' of $\bar{x}$ it holds:

$$\partial^{\mathsf{b}}\varphi(x) \subseteq \operatorname{cl}\left(\int_{v\in\mathcal{F}(x)} \mathcal{M}(x,v)d\mu_\zeta(v)\right), \text{ for all } x \in U'$$

where, $\mathcal{M}(x,v)$ is given by,

$$\mathcal{M}(x,v) = \left\{ \frac{\alpha}{\langle z^*, Lv\rangle}\cdot x^* : \begin{array}{c} \alpha \in I_\theta(\rho(x,v),v), z^* \in \mathrm{N}^{\mathsf{b}}_{\mathcal{S}(x)}(\rho(x,v)Lv)\cap \mathbb{S}^{m-1} \\ \\ i \in T_x(v), x^* \in D^*\mathcal{S}_i(x,\rho(x,v)Lv)(-z^*) \end{array} \right\}$$

for all $v \in \mathcal{F}(x)$ with

$$T_x(v) = \{i \in \{1,\dots,s\} : \rho_i(x,v) = \rho(x,v)\},$$

and by $\mathcal{M}(x,v) = \{0\}$ for all $v \in \mathcal{I}(x)$.

Finally, by considering $\mathcal{S}_i(x) := \{z : g_i(x,z) \le 0\}$ we prove the local Lipschitz continuity of Joint probability functions given by a system of inequality constraints with a relaxed convexity assumption.

Corollary *Consider the probability function*

$$\varphi(x) := \mathbb{P}(g_i(x,\xi) \le 0,\ \forall i = 1,\dots,s), \tag{1.9}$$

where $g_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are continuously differentiable, quasi-convex in z for all $i = 1,\dots,s$ and ξ has continuous density distribution. Suppose that a point of interest $\bar{x}$ is such that $g_i(\bar{x},0) < 0$ for all $i = 1,\dots,s$ and that the family g_i satisfies the η-growth condition given above at $\bar{x}$. Assume, moreover, that $\operatorname{int}\{z \in \mathbb{R}^m : g_i(x,z) \le 0\} = \{z \in \mathbb{R}^m : g_i(x,z) < 0\}$ for all $i = 1,\dots,s$. Then the probability function (1.9) is locally Lipschitz at $x \in U'$ and

$$\partial^{\mathsf{b}}\varphi(x) \subseteq -\int_{v\in F(x)} \bigcup_{i\in T_x(v)} \frac{\theta(\rho(x,v),v)}{\langle \nabla_z g_i(x,\rho(x,v)Lv), Lv\rangle}\nabla_x g_i(x,\rho(x,v)Lv)d\mu_\zeta(v), \text{ for all } x \in U'$$

Furthermore, if $\#T_x(v) = 1$ for all $x \in U'$, then the probability function (1.9) is continuously differentiable for all $x \in U'$ and

$$\nabla\varphi(x) = -\int_{v\in F(\bar{x})} \frac{\theta(\rho(x,v),v)}{\langle \nabla_z g_{T_x(v)}(x,\rho(x,v)Lv), Lv\rangle}\nabla_x g_{T_x(v)}(x,\rho(x,v)Lv)d\mu_\zeta(v), \text{ for all } x \in U'.$$

Chapter 4: Chebyshev sets: weak projection and nonconvexity estimates

The problem, in the context of Hilbert spaces, of whether a closed set C is convex when its associated projection mapping P_C is a single-valued mapping is still unanswered, and it is well known to be true for finite-dimensional Hilbert spaces and weakly closed sets in any Hilbert space. This is the famous Bunt-Klee theorem. In this chapter we give a partial positive answer to this problem, relaxing the concept of projection by the one of weak projection:

Given a nonempty set $C \subset \mathcal{H}$,

$$P_C^w(x) := \{y : \text{ there exists a net } y_i \in C \text{ such that } y_i \rightharpoonup y \text{ and } \|y_i - x\| \to d_C(x)\}.$$

A crucial observation is that the subdifferential of the weak closure of the Asplund function, defined as $\psi_C(x) := \frac{1}{2}\|x\|^2$ when $x \in C$ and $\psi_C(x) := +\infty$ otherwise, may be written in terms of the weak projection. Hence, using a special instance of a general characterization of the conjugate function given in [19] we obtain that

$$\partial \psi_C^*(x) = \overline{\text{co}}(P_C^w(x)).$$

Basing ourselves on this last formula, together with an integration criterion given in [18, Theorem 5.2] and a Hilbert version of the variational characterization of convexity given in [20, Corollary 11] we obtain our first main result in which we characterize the convexity of a set in terms of the weak projection:

Theorem *Let $C \subset \mathcal{H}$ be a proximinal set. Then the following are equivalent:*

(i) *C is convex.*

(ii) *$P_C^w(x)$ is a singleton for all $x \in \mathcal{H}$.*

(iii) *d_C^2 is Gâteaux differentiable on $\mathcal{H}$.*

(iv) *d_C^2 is Fréchet differentiable on $\mathcal{H}$.*

(v) *For all $x \in \mathcal{H}$, there exists a selection of P_C, norm-weak continuous at x.*

(vi) *For all $x \in \mathcal{H}$, there exists a selection of P_C, norm-norm continuous at x.*

There are many concepts used to measure the nonconvexity of a set, the most natural being the Hausdorff distance between the set and its closed convex hull. In order to quantify the Shapley-Folkman theorem some measures of nonconvexity have been given light, for instance, the effective standard deviation of a set $C \subset \mathcal{H}$, $v(C)$, due to Cassels [15], has a very special formulation in terms of the Asplund function:

$$v(C) = \sup\{v_C(x) : x \in \operatorname{co} C\} \text{ where } v_C^2 = \operatorname{co}\psi_C - \tfrac{1}{2}\|\cdot\|^2.$$

Due to this last formula, we are able to estimate the Hausdorff distance between a weakly closed set and its closed convex hull in terms of simultaneous projections onto the set and its closed convex hull with the condition $v(C) < +\infty$. More precisely, for a given weakly closed set C, we have that

$$d^2(C, \overline{\text{co}}C) \le \sup\left\{\|P_C(x) - P_{\overline{\text{co}}C}(x)\|^2 : x \in \Pi\right\}.$$

where $\Pi := \{x \in \mathcal{H} : P_{\overline{\text{co}}C}(x) \in \overline{\text{co}}(P_C(x))\}$. We call this last result a "dual estimate" of the Hausdorff distance since our estimate involves only the projections, whereas the "primal estimate," $d^2(C, \overline{\text{co}}C) \le \sup\{v_C^2(x) : x \in \overline{\text{co}}C\}$, concerns only the apparent shape of the set involved and clearly, harder to visualize:

Example *Consider $C \subset \mathbb{R}^2$ given by $C = \mathbb{B}_4(0,0) \setminus \mathbb{B}_2(0,0)$. To compute $v(C)$ we shall consider all the points in $\overline{\text{co}}C$ and all the possible ways those points may be written as a* *convex combination of elements in C. Instead, to compute our proposed "dual" we notice that the unique point in Π (defined above), who gives us information in order to obtain the estimate, is the center $x_0 = (0,0)$. Then we get*

$$d(\overline{\text{co}}C, C) \leq \|(2,0) - (0,0)\| = 2$$

where we used the fact that $P_C(0,0) = \mathbb{S}_2(0,0)$.

Finally, we observe that the effective standard deviation provides us with the following quantification of the Bunt-Klee theorem:

Corollary *Let $C \subset \mathcal{H}$ be weakly closed and $v(C) < +\infty$. Then*

$$d(C, \overline{\text{co}}C) \leq \sup\left\{\text{diam}(P_C(x)) : x \in \mathcal{H}\right\}.$$

Chapter 2

Preliminaries

2.1. Basic results and notation

Let $\mathcal{X}$ be a reflexive Banach space, $\mathcal{X}^*$ its topological dual and duality product $\langle x, x^* \rangle$ for $x \in \mathcal{X}$, $x^* \in \mathcal{X}^*$. We use $\to$ and cl to denote the convergence and closure with respect to the norm $\|\cdot\|$-topology on both $\mathcal{X}$ and $\mathcal{X}^*$. We also use cl^w (respectively cl^{w^*}) to mean the closure with respect to the weak topology (respectively weak* topology) and use $\rightharpoonup$ to mean the weak convergence in both topologies. $\mathbb{B}$ and $\mathbb{B}^*$ stand for the closed unit balls of the space and its dual, respectively. For $x \in \mathcal{X}$ and $r > 0$, $\mathbb{B}_r(x) := x + r\mathbb{B}$.

We denote by $\mathbb{N}$, $\mathbb{R}^m$, $\mathbb{S}^{m-1}$, $\mathbb{R}^m_+$, and $\mathbb{R}^m_{++}$ the set of natural numbers, the m-dimensional euclidean space, the unit sphere of $\mathbb{R}^m$, the non-negative orthant of $\mathbb{R}^m$, and the positive orthant of $\mathbb{R}^m$, respectively. We consider $\mathbb{R}_\infty := \mathbb{R} \cup \{+\infty\}$ adopting the conventions $(+\infty) + (-\infty) = (-\infty) + (+\infty) = 0 \cdot (+\infty) = +\infty$.

Given $C \subset \mathcal{X}$, we denote by $\operatorname{co} C$, $\operatorname{cone} C$, $\operatorname{aff} C$, $\operatorname{bd} C$, $\operatorname{int} C$, and $\operatorname{ri} C$, the convex hull, conic hull, affine hull, boundary, interior, and relative interior of the set C, respectively. More specifically,

$$\operatorname{co} C := \left\{ \sum_{i=1}^{k} \lambda_i x_i : x_i \in C, k \in \mathbb{N}, \lambda \in \Delta_k \right\},$$

where Δ_k denotes the *canonical simplex* in $\mathbb{R}^k$, i.e.,

$$\Delta_k := \left\{ \lambda \in \mathbb{R}^k_+ : \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$

The conic hull is $\operatorname{cone} C := \mathbb{R}_{++} C$ and the affine hull is

$$\operatorname{aff} C := \left\{ \sum_{i=1}^{k} \lambda_i x_i : x_i \in C, k \in \mathbb{N}, \lambda_i \in \mathbb{R} \text{ and } \sum_{i=1}^{k} \lambda_i = 1 \right\}.$$

Sometimes we write $\overline{\operatorname{co}} C$ to refer us to $\operatorname{cl} \operatorname{co} C$, the closed convex hull of C. The set $\operatorname{ri} C$ is the interior of C relative to $\operatorname{cl}(\operatorname{aff} C)$. When $\mathcal{X}$ is finite-dimensional $\operatorname{ri} C$ is the classical relative interior, that is, the interior of C relative to $\operatorname{aff} C$.

Given a cone $\mathcal{K} \subseteq \mathcal{X}$, we denote its positive and negative dual cone by

$$\mathcal{K}^+ := \{x^* \in \mathcal{X}^* : \langle x^*, x\rangle \geq 0 \text{ for all } x \in \mathcal{K}\},$$
$$\mathcal{K}^- := \{x^* \in \mathcal{X}^* : \langle x^*, x\rangle \leq 0 \text{ for all } x \in \mathcal{K}\},$$

respectively.

The epigraph, (effective) domain, lower α-level set and upper α-level set ($\alpha \in \mathbb{R}$) of an extended real-valued function $\psi : \mathcal{X} \to \mathbb{R}_\infty$ are defined and denoted by

$$\begin{aligned} \operatorname{epi}\psi &:= \{(x,\alpha) \in \mathcal{X} \times \mathbb{R} : \psi(x) \leq \alpha\}, \\ \operatorname{dom}\psi &:= \{x \in \mathcal{X} : \psi(x) < +\infty\}, \\ [\psi \leq \alpha] &:= \{x \in \mathcal{X} : \psi(x) \leq \alpha\}, \\ [\psi \geq \alpha] &:= \{x \in \mathcal{X} : \psi(x) \geq \alpha\}, \end{aligned}$$

respectively.

Given nonempty set $C \subseteq \mathcal{X}$, the indicator function $\delta_C : \mathcal{X} \to \mathbb{R}_\infty$ and the support function $\sigma_C : \mathcal{X}^* \to \mathbb{R}_\infty$ are given respectively by

$$\delta_C(x) := \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise} \end{cases} \quad \text{and} \quad \sigma_C(x^*) := \sup_{x \in C}\langle x^*, x\rangle.$$

We consider the distance function $d_C : \mathcal{X} \to \mathbb{R}_\infty$ (sometimes also denoted by $\mathrm{d}(\cdot, C)$) given by

$$d_C(x) := \inf_{y \in C} \|x - y\|$$

and define the (metric) projection of x to C as the set

$$P_C(x) := \{y \in C : \|x - y\| = d_C(x)\}.$$

A set $C \subseteq \mathcal{H}$ is called *proximinal* when $P_C(x) \neq \emptyset$, for all $x \in \mathcal{H}$. For instance, every weakly closed set is proximinal and, conversely, every proximinal set is closed (see, e.g.,[24, 28]). Moreover, a proximinal set C is called a *Chebyshev* set when $P_C(x)$ is a singleton, for all $x \in \mathcal{H}$. We say that an operator $T : \mathcal{X} \to \mathcal{X}$ is a selection of the projection P_C when $T(x) \in P_C(x)$ for all $x \in \mathcal{X}$. The projection P_C is said to be norm-norm (respectively norm-weak) continuous at $x \in \mathcal{H}$ if $P_C(x)$ is a singleton and $y_n \to P_C(x)$ (respectively $y_n \rightharpoonup P_C(x)$) whenever $y_n \in P_C(x_n)$ and $x_n \to x$ as $n \to \infty$.

Given two sets $C_1, C_2 \subseteq \mathcal{X}$ we denote by $C_1 + C_2$ the (Minkowski) sum of C_1 and C_2, that is, $C_1 + C_2 := \{x + y : x \in C_1, y \in C_2\}$, while for $\alpha \in \mathbb{R}$, $\alpha C_1 = \{\alpha x : x \in C_1\}$. We define and denote the Hausdorff distance between C_1 and C_2 as

$$d(C_1, C_2) = \max\{e(C_1, C_2), e(C_2, C_1)\},$$

where $e(C_1, C_2) := \sup\{d_{C_2}(x) : x \in C_1\}$ and $e(C_2, C_1)$ is defined symmetrically.

2.1.1. Generalized differentiation

For a given closed subset $C \subseteq \mathcal{X}$, the regular/Fréchet and the basic/limiting/Mordukhovich normal cones to C at x are denoted and defined respectively by

$$\mathrm{N}^{\mathsf{r}}_C(x) := \left\{ x^* \in \mathcal{X}^* \mid \limsup_{x' \xrightarrow{C} x} \frac{\langle x^*, x' - x\rangle}{\|x' - x\|} \leq 0 \right\}$$

and

$$\mathrm{N}^{\mathsf{b}}_C(x) := \left\{ x^* \in \mathcal{X}^* \mid \exists x_k \xrightarrow{C} x,\, \exists x_k^* \rightharpoonup x^* : \, x_k^* \in \mathrm{N}^{\mathsf{r}}_C(x_k) \right\},$$

where by $x' \xrightarrow{C} x$ we mean that $x' \to x$ with $x' \in C$. For a function ψ with closed epigraph (i.e., ψ is lower semi-continuous) its Fréchet/regular and Mordukhovich/basic/limiting subdifferentials at $x \in \mathcal{X}$ may be defined through the corresponding normal cones to its epigraph, or more explicitly, they can be represented as (see, e.g.,[39])

$$\partial^{\mathsf{r}}\psi(x) = \left\{ x^* \in \mathcal{X}^* \mid \liminf_{x' \to x} \frac{\psi(x') - \psi(x) - \langle x^*, x' - x\rangle}{\|x' - x\|} \geq 0 \right\}$$

and

$$\partial^{\mathsf{b}}\psi(x) := \{x^* \in \mathcal{X}^* \mid \exists x_k \to x, \;\text{ with } \psi(x_k) \to \psi(x) \text{ and } \exists x_k^* \rightharpoonup x^* : \, x_k^* \in \partial^{\mathsf{r}}\psi(x_k)\},$$

respectively.

We recall that a set-valued mapping $\mathcal{S} : \mathcal{X} \rightrightarrows \mathbb{R}^m$ is a mapping whose value at each $x \in \mathcal{X}$ is a subset $\mathcal{S}(x) \subset \mathbb{R}^m$ and is uniquely defined by its graph

$$\operatorname{gph} \mathcal{S} := \{(x, z) \in \mathcal{X} \times \mathbb{R}^m : z \in \mathcal{S}(x)\}.$$

For a set-valued mapping with closed graph we define its coderivative at $(x, z) \in \operatorname{gph} \mathcal{S}$ as the set-valued mapping $D^*\mathcal{S}(x, z) : \mathbb{R}^m \rightrightarrows \mathcal{X}^*$ such that

$$D^*\mathcal{S}(x, z)(z^*) := \left\{ x^* \in \mathcal{X}^* \mid (x^*, -z^*) \in \mathrm{N}^{\mathsf{b}}_{\operatorname{gph} \mathcal{S}}(x, z) \right\}.$$

The following property is an extension of Lipschitz continuity to set-valued mappings introduced by Aubin [6].

Definition 2.1 *Let $\mathcal{S} : \mathcal{X} \rightrightarrows \mathbb{R}^m$ be a set-valued mapping and $(x, z) \in \operatorname{gph} \mathcal{S}$. We say that $\mathcal{S}$ is locally Lipschitz-like at (x, z) if there exist $\kappa > 0$ and $\delta > 0$ such that*

$$\mathrm{d}(z', \mathcal{S}(x')) \leq \kappa \|x_1 - x_2\|, \, \forall x', x'' \in \mathbb{B}_\delta(x) \textit{ and } z \in \mathcal{S}(x'') \cap \mathbb{B}_\delta(z).$$

The following lemma that the local Lipschitz-like property of a set-valued mapping implies the local Lipschitz continuity of the distance function associated to its values.

Lemma 2.1 *Let $\mathcal{S} : \mathcal{X} \rightrightarrows \mathbb{R}^m$ be a set valued-mapping with closed graph and convex values. Assume that there exists a neighborhood U of $\bar{x}$ such that $0 \in S(x)$ and such that S has the local Lipschitz property at $(x, z) \in \operatorname{gph} \mathcal{S}$ with $x \in U$.*

Then, the function $u(x, z) = \frac{1}{2}\mathrm{d}^2(z, \mathcal{S}(x))$ is locally Lipschitz continuous at $(x, z) \in U \times \mathbb{R}^m$.

Moreover, for all $(x, z) \in U \times \mathbb{R}^m$

$$\partial^{\mathsf{b}} u(x, z) \subseteq D^*\mathcal{S}(x, P_{\mathcal{S}(x)}(z))(P_{\mathcal{S}(x)}(z) - z) \times \{z - P_{\mathcal{S}(x)}(z)\}. \tag{2.1}$$

Proof. First, let us check the local Lipschitz continuity of u at $(x, z) \in U \times \mathbb{R}^m$. Indeed, On the one hand if $(x, z) \in \operatorname{gph} \mathcal{S}$ then $\mathcal{S}$ has the locally Lipschitz-like property at (x, z), hence by [39, Theorem 1.41], the function u is locally Lipschitz at (x, z). On the other hand if $(x, z) \notin \operatorname{gph} \mathcal{S}$, we can apply [37, Corollary 5.4] to conclude that u is locally Lipschitz at (x, z). Now, let us verify (2.1). The function u can be rewritten as a marginal function in the following way:

$$u(x, z) = \inf\{\psi(x, z, y) : y \in \mathcal{T}(x, z)\},$$

where $\psi(x, z, y) := \frac{1}{2}\|z - y\|^2$, $\mathcal{T}(x, z) := \mathcal{S}(x)$ and where the argminimum for u is the single-valued mapping $M(x, z) := P_{\mathcal{S}(x)}(z)$. Since u is locally Lipschitz at $(x, z) \in U \times \mathbb{R}^m$ we may apply [39, Theorem 3.38 i)] to obtain

$$\partial^{\mathsf{b}} u(x, z) \subseteq \bigcup_{(x^*, z^*, y^*) \in \partial^{\mathsf{b}} \psi(x, z, P_{\mathcal{S}(x)}(z))} (x^*, z^*) + D^*\mathcal{T}(x, z, P_{\mathcal{S}(x)}(z))(y^*).$$

Finally, the definition of $\mathcal{T}$ together with the fact that

$$\partial^{\mathsf{b}} \psi(x, z, P_{\mathcal{S}(x)}(z)) = (0, z - P_{\mathcal{S}(x)}(z), P_{\mathcal{S}(x)}(z) - z)$$

allow us to rewrite the above inclusion as (2.1). $\square$

2.1.2. (Convex) Subdifferential

In this subsection, we condensed the majority of the results we will make use of in Chapter 5. For this reason, let us consider $\mathcal{X} = \mathcal{H}$ a Hilbert space.

By $\operatorname{cl} \psi$, $\operatorname{cl}^w \psi$, $\operatorname{co} \psi$ and $\overline{\operatorname{co}}\psi$ we refer to the closed hull (i.e., the largest lower semi-continuous function dominated by ψ), the weak convex hull (i.e., the largest weak lower semi-continuous function dominated by ψ), the convex hull (i.e., the largest convex function dominated by ψ) and the closed convex hull (i.e., the largest lower semi-continuous convex function dominated by ψ) of the function ψ, respectively. Furthermore,

$$\operatorname{co} \psi(x) = \inf\left\{\sum_{i=1}^{k} \lambda_i \psi(x_i) : x_i \in \mathcal{H},\, k \geq 1,\, \lambda \in \Delta_k,\, \sum_{i=1}^{k} \lambda_i x_i = x\right\}.$$

The set of all convex, proper, and lower semi-continuous functions is denoted by $\Gamma_0(\mathcal{H})$. We say that function $\psi : \mathcal{H} \to \mathbb{R}_\infty$ is coercive if the sets $[\psi \leq \alpha]$ are bounded for all $\alpha \in \mathbb{R}$. Moreover, for $\psi \in \Gamma_0(\mathcal{H})$, the above is equivalent to the condition $0 \in \operatorname{int}(\operatorname{dom} \psi^*)$ (see, e.g., [10, Proposition 14.16]).

The (Legendre-Fenchel) conjugate of an extended real-valued function $\psi : \mathcal{H} \to \mathbb{R}_\infty$ is the function $\psi^* : \mathcal{H} \to \mathbb{R}_\infty$ given by $\psi^*(x) = \sup\{\langle x, y\rangle - \psi(y) : y \in \mathcal{H}\}$. The notation ψ^{**} stands for $(\psi^*)^*$. The Fenchel's inequality $\langle x, y\rangle \leq \psi^*(x) + \psi(y)$ for all $x, y \in \mathcal{H}$ follows from the definition of conjugate. Moreover, it is easy to verify the equalities

$$\psi^* = (\mathrm{cl}\,\psi)^* = (\mathrm{cl}^w\,\psi)^* = (\overline{\mathrm{co}}\psi)^*.$$

Hence, by Fenchel-Moreau's Theorem, for every function ψ with proper conjugate we have that $\psi^{**} = \overline{\mathrm{co}}\psi$ and, consequently,

$$\Gamma_0(\mathcal{H}) = \{\psi : \mathcal{H} \to \mathbb{R}_\infty : \psi^{**} = \psi\}. \tag{2.2}$$

For example, $\delta_C^* = \sigma_C$ and as a consequence of the Fenchel-Moreau's theorem, $\sigma_C^* = \delta_{\overline{\mathrm{co}}C}$. For two functions $\psi_1, \psi_2 \in \Gamma_0(\mathcal{H})$ such that ψ_1 is continuous somewhere in $\mathrm{dom}\,\psi_2$, the Moreau-Rockafellar theorem entails that $(\psi_1 + \psi_2)^* = \psi_1^* + \psi_2^*$.

Next, we recall the so-called Toland's duality result. We present and prove here a slight extension to non-necessarily convex functions.

Proposition 2.1 *For functions $\psi_1, \psi_2 : \mathcal{H} \to \mathbb{R}_\infty$ such that $\overline{\mathrm{co}}\psi_2$ is proper, we have that*

$$\inf\{\psi_1(x) - \psi_2(x) : x \in \mathcal{H}\} = \inf\{\psi_2^*(x) - \psi_1^*(x) : x \in \mathcal{H}\},$$

and, consequently, provided that $\overline{\mathrm{co}}\psi_1$ is proper,

$$\sup\{\psi_1(x) - \psi_2(x) : x \in \mathcal{H}\} = \sup\{\psi_2^*(x) - \psi_1^*(x) : x \in \mathcal{H}\}.$$

Proof. First, we assume that $\psi_2 \in \Gamma_0(\mathcal{H})$. By (2.2), $\psi_2 = \psi_2^{**}$ and, so by definition of the conjugate we have that

$$\begin{aligned}
\inf\{\psi_1(x) - \psi_2(x) : x \in \mathcal{H}\} &= \inf\{\psi_1(x) - \sup\{\langle y, x\rangle - \psi_2^*(y) : y \in \mathcal{H}\} : x \in \mathcal{H}\} \\
&= \inf\{\psi_1(x) - \langle y, x\rangle + \psi_2^*(y) : x, y \in \mathcal{H}\} \\
&= \inf\{\psi_2^*(y) + \inf\{\psi_1(y) - \langle y, x\rangle : x \in \mathcal{H}\} : y \in \mathcal{H}\} \\
&= \inf\{\psi_2^*(x) - \psi_1^*(x) : x \in \mathcal{H}\}
\end{aligned}$$

yielding the first part in this case. Consequently, provided that also $\psi_1 \in \Gamma_0(\mathcal{H})$,

$$\begin{aligned}
\sup\{\psi_1(x) - \psi_2(x) : x \in \mathcal{H}\} &= -\inf\{\psi_2(x) - \psi_1(x) : x \in \mathcal{H}\} \\
&= -\inf\{\psi_1^*(x) - \psi_2^*(x) : x \in \mathcal{H}\} \\
&= \sup\{\psi_2^*(x) - \psi_1^*(x) : x \in \mathcal{H}\}.
\end{aligned}$$

Now, we assume ψ_2 is a general function such that $\overline{\mathrm{co}}\psi_2$ is proper, that is, $\overline{\mathrm{co}}\psi_2 \in \Gamma_0(\mathcal{H})$. Due to the first part and since $\overline{\mathrm{co}}\psi_2(x) = \sup\{\psi_3 : \psi_3 \in \Gamma_0(\mathcal{H}),\ \psi_3 \leq \psi_2\}$ we infer that

$$\inf\{\psi_1(x) - \psi_2(x) : x \in \mathcal{H}\} = \inf\{\inf\{\psi_1(x) - \psi_3(x) : x \in \mathcal{H}\} : \psi_3 \in \Gamma_0, \psi_3 \leq \psi_2\}. \tag{2.3}$$

Thus, because $\psi_3^* \geq \psi_2^*$ for all $\psi_3 \in \Gamma_0(\mathcal{H})$ satisfying $\psi_3 \leq \psi_2$, we deduce from (2.3) that

$$\inf\{\psi_1(x) - \psi_2(x) : x \in \mathcal{H}\} \geq \inf\{\psi_2^*(x) - \psi_1^*(x) : x \in \mathcal{H}\}.$$

Conversely, we have that $\psi_2 \geq \overline{\mathrm{co}}\psi_2$, and using again (2.3), we obtain

$$\inf\{\psi_1(x) - \psi_2(x) : x \in \mathcal{H}\} \leq \inf\{\psi_2^*(x) - \psi_1^*(x) : x \in \mathcal{H}\}.$$

concluding the first part. The second part follows similarly as in the first case. $\square$

For $\varepsilon \geq 0$, the ε-*subdifferential* of an extended real-valued function $\psi : \mathcal{H} \to \mathbb{R}_\infty$ is defined and denoted by $\partial_\varepsilon \psi(x) := \{z \in \mathcal{H} : \langle z, y - x\rangle \leq \psi(y) - \psi(x) + \varepsilon, \forall y \in \mathcal{H}\}$ when $x \in \operatorname{dom}\psi$. Otherwise, $\partial_\varepsilon\psi(x) = \emptyset$. We set $\partial\psi(x) := \partial_0\psi(x)$.

If $\psi \in \Gamma_0(\mathcal{H})$ then we have that $\partial\psi^* = (\partial\psi)^{-1}$ where the superscript -1 denotes the inverse image.

Proposition 2.2 *Let $\psi_1, \psi_2 \in \Gamma_0(\mathcal{H})$ and $\bar{x}$ be an ε-minimizer of the optimization problem*

$$\inf\{\psi_1(x) - \psi_2(x) : x \in \mathcal{H}\},$$

that is,

$$\inf\{\psi_1(x) - \psi_2(x) : x \in \mathcal{H}\} \geq \psi_1(\bar{x}) - \psi_2(\bar{x}) - \varepsilon.$$

Then

$$\partial\psi_2(x) \subset \partial_\varepsilon\psi_1(x).$$

The following characterization of the subdifferential of the conjugate function is a special instance of the general characterization given in [19].

Proposition 2.3 *Let $\psi : \mathcal{H} \to \mathbb{R}_\infty$ be such that $\operatorname{dom}\psi^* = \mathcal{H}$. Then, for every $x \in \mathcal{H}$,*

$$\partial\psi^*(x) = \overline{\mathrm{co}}((\partial \operatorname{cl}^w \psi)^{-1}(x)).$$

And provided that $\mathcal{H}$ is finite-dimensional,

$$\partial\psi^*(x) = \mathrm{co}((\partial \operatorname{cl} \psi)^{-1}(x)).$$

The following integration criterion given in [18, Theorem 5.2] extends the classical Moreau-Rockafellar integration of convex functions.

Proposition 2.4 *Let $\psi_1 \in \Gamma_0(\mathcal{H})$ be given. Then every lower semi-continuous function $\psi_2 : \mathcal{H} \to \mathbb{R}_\infty$ satisfying*

$$\partial\psi_1(x) \subset \partial\psi_2(x) \text{ for all } x \in \mathcal{H},$$

coincides with ψ_1 up to an additive constant, and is in particular convex.

The following result gives a Hilbert version of the variational characterization of convexity given in [20, Corollary 11]. For the sake of completeness, we give a short proof based on the previous propositions, taking advantage of the current Hilbert context.

Proposition 2.5 *Let $\psi : \mathcal{H} \to \mathbb{R}_\infty$ be a weakly lower semi-continuous function such that $\operatorname{dom}\psi^* = \mathcal{H}$. Then ψ is convex provided that $(\partial\psi)^{-1}(x)$ is convex for all x in some convex dense set $\mathcal{D} \subset \mathcal{H}$.*

Proof. By Proposition 2.3 we have that

$$\partial\psi^*(x) = (\partial\psi)^{-1}(x) \text{ for all } x \in \mathcal{D}.$$

Moreover, thanks to Lemma in [18], the last relation implies

$$\partial\psi^*(x) \subset (\partial\psi)^{-1}(x) \text{ for all } x \in \mathcal{H}.$$

Hence,

$$\partial\psi^{**}(x) = (\partial\psi^*)^{-1}(x) \subset \partial\psi(x) \text{ for all } x \in \mathcal{H}.$$

Therefore, according to Proposition 2.4, the (lower semicontinuous) function ψ is convex. □

2.1.3. Variational convergence and Moreau envelope

In this subsection, we provide variational concepts and results that play a crucial role in Chapter 3. Consider $\mathcal{X} = \mathcal{H}$ a Hilbert space.

Let us recall here the definition of epi/hypo-convergence. We refer to [3, 11, 40] for more details and properties and also to [46] for finite-dimensional spaces.

A sequence of sets $(C_k)_k \subset \mathcal{H}$ *Painlevé-Kuratowski* converges to a set C if the following conditions hold:

a) $C \subset \liminf\limits_{k\to\infty} C_k := \{x \in \mathcal{H} \colon \exists x_k \in C_k \text{ with } x_k \to x\}$, and

b) $\limsup\limits_{k\to\infty} C_k := \{x \in \mathcal{H} \colon \exists x_n \in C_{k_n} \text{ with } k_n \to \infty \text{ and } x_n \to x\} \subset C$.

The sequence $(C_k)_k$ is said to *Mosco converge* to C if condition *a*) is satisfied and *b*) is replaced by the following condition:

c) $w\text{-}\limsup\limits_{k\to\infty} C_k := \{x \in \mathcal{H} : \exists x_n \in C_{k_n} \text{ with } k_n \to \infty \text{ and } x_n \rightharpoonup x\} \subset C$.

Moreover, the limit set of a sequence of epigraphs is again an epigraph (in both of the above notions). Thus, we obtain two notions of convergence of functions which can be characterized as follows: A sequence of functions $\psi_k : \mathcal{H} \to \mathbb{R}_\infty$ *epi-converge* to $\psi : \mathcal{H} \to \mathbb{R}_\infty$ when the following two conditions hold:

a′) For all $x \in \mathcal{H}$, there exist $x_k \to x$ such that $\limsup\limits_{k\to\infty} \psi_k(x_k) \le \psi(x)$, and

b′) For all $x \in \mathcal{H}$ and for all $x_k \to x$, we have $\liminf\limits_{k\to\infty} \psi_k(x_k) \ge \psi(x)$.

The sequence ψ_k is said to *Mosco epi-converge* to ψ when condition *a′*) is satisfied and *b′*) is replaced by the following condition:

c′) For all $x \in \mathcal{H}$ and for all $x_k \rightharpoonup x$, we have $\liminf\limits_{k\to\infty} \psi_k(x_k) \ge \psi(x)$.

Hypo-convergences notions can be obtained by applying the above notions to the functions $-\psi, -\psi_k$. Moreover, a sequence of functions (ψ_k) *converges continuously* to ψ if $(\psi_k)_k$ epi-converges and hypo-converges to ψ, i.e., $\lim_{k\to\infty} \psi_k(x_k) = \psi(x)$ for all $x_k \to x$.

Given a function $\psi \in \Gamma_0(\mathcal{H})$ and $\lambda > 0$, the Moreau envelope of ψ of parameter λ is the function $\mathtt{e}_\lambda\psi : \mathcal{H} \to \mathbb{R}_\infty$ defined by

$$\mathtt{e}_\lambda\psi(x) := \inf\{\psi(z) + \tfrac{1}{2\lambda}\|x-z\|^2 : z \in \mathcal{H}\}.$$

In particular, when $\lambda = 1$ we have the following property, known as the Moreau decomposition,

$$\mathtt{e}_1\psi(x) + \mathtt{e}_1\psi^*(x) = \tfrac{1}{2}\|\cdot\|^2. \tag{2.4}$$

The infimum in the definition of the Moreau envelope is attained at a unique point, which is called the proximal point of ψ of index λ at x. It defines a non-expansive operator $\mathtt{Prox}_{\lambda\psi} : \mathcal{H} \to \mathcal{H}$ given by

$$\mathtt{Prox}_{\lambda\psi}(x) := \{z \in \mathcal{H} : \psi(z) + \tfrac{1}{2\lambda}\|x-z\|^2 = \mathtt{e}_\lambda\psi(x)\} = (I + \lambda\partial\psi)^{-1}(x).$$

Moreover, the Moreau envelope of $\psi \in \Gamma_0(\mathcal{H})$ is convex and continuously differentiable function with

$$\nabla\mathtt{e}_\lambda\psi(x) = \tfrac{1}{\lambda}(x - \mathtt{Prox}_{\lambda\psi}(x)) \text{ for all } x \in \mathcal{H}. \tag{2.5}$$

More specifically, $\nabla\mathtt{e}_\lambda\psi$ is $\frac{1}{\lambda}-$Lipschitz continuous on $\mathcal{H}$. It follows moreover from the above identification of the proximal operator with a resolvant that (see, e.g., [10, Proposition 16.44]):

$$\nabla\mathtt{e}_\lambda\psi(x) \in \partial\psi(\mathtt{Prox}_{\lambda\psi}(x)). \tag{2.6}$$

The following proposition summarizes some properties of the Moreau envelope in Hilbert spaces. We refer to [4, 5, 10] for more details.

Proposition 2.6 *Let $g\colon \mathcal{H} \to \mathbb{R}$ be a convex and lower semicontinuous function. Then the following hold.*

a) Monotone convergence: $\mathtt{e}_\lambda g(x) \nearrow g(x)$ as $\lambda \searrow 0$ for all $x \in \mathcal{H}$.

b) Convergence of resolvents: $\mathtt{Prox}_{\lambda g}(x) \to x$ as $\lambda \to 0$ for all $x \in \mathcal{H}$.

c) Lower epi-convergence: If $x_k \rightharpoonup x$ and $\lambda_k \searrow 0$, then

$$g(x) \leq \liminf_{k\to\infty} \mathtt{e}_{\lambda_k} g(x_k).$$

d) Continuous convergence: If $x_k \to x$ and $\lambda_k \searrow 0$, then

$$g(x) = \lim_{k\to\infty} \mathtt{e}_{\lambda_k} g(x_k).$$

Proof. Items a-c) can be found in [45, Proposition 2.2]. Let us focus on *d*). To this end, let a sequence $x_k \to x$ and $\lambda_k \searrow 0$ be given. Then, by virtue of the proximal operator being non-expansive we have

$$\begin{aligned}\|\mathtt{Prox}_{\lambda_k g}(x_k) - x\| &\leq \|\mathtt{Prox}_{\lambda_k g}(x_k) - \mathtt{Prox}_{\lambda_k g}(x)\| + \|\mathtt{Prox}_{\lambda_k g}(x) - x\| \\ &\leq \|x_k - x\| + \|\mathtt{Prox}_{\lambda_k g}(x) - x\|,\end{aligned} \tag{2.7}$$

which, by b), implies that $\mathtt{Prox}_{\lambda_k g}(x_k) \to x$, as $k \to +\infty$. Moreover, for all $k \in \mathbb{N}$

$$\begin{aligned} g(\mathtt{Prox}_{\lambda_k g}(x_k)) &\leq g(\mathtt{Prox}_{\lambda_k g}(x_k)) + \frac{1}{2\lambda_k}\|x_k - \mathtt{Prox}_{\lambda_k g}(x_k)\|^2 \\ &= \mathtt{e}_{\lambda_k} g(x_k) \leq g(x_k), \end{aligned}$$

where a) was used to derive the last inequality. Thus, by using (2.7), the continuity of g (g is lower semicontinuous with finite values) and taking the limit $k \to +\infty$ in the latter inequality, we obtain that $\lim_{k\to+\infty} \mathtt{e}_{\lambda_k} g(x_k) = g(x)$. $\square$

The following proposition gives a precise (uniform) bound on the distance between a function and its Moreau envelope in a finite dimensional setting.

Proposition 2.7 *Let S be a closed, convex and bounded subset of $\mathcal{H} = \mathbb{R}^n$ and let $g\colon \mathcal{H} \to \mathbb{R}$ be a convex function. Then, there exist $\ell \geq 0$, $\lambda_0 \in (0,1)$ and a constant $C > 0$ such that for all $\lambda \in (0, \lambda_0)$, the function $x \mapsto \mathtt{e}_\lambda g(x)$ is ℓ-Lipschitz on S and*

$$\sup_{x\in S} |\mathtt{e}_\lambda g(x) - g(x)| \leq \ell\sqrt{\lambda} C.$$

Proof. Let us consider the set $\tilde{S} := \{x \in \mathcal{H}\colon d_S(x) \leq 1\}$. Since $\tilde{S}$ is closed, convex and bounded, there exists $\ell \geq 0$ such that g is ℓ-Lipschitz on $\tilde{S}$. Moreover, for all $x \in S$

$$\begin{aligned} \frac{1}{2\lambda}\|x - \mathtt{Prox}_{\lambda g}(x)\|^2 \leq & g(x) - g(\mathtt{Prox}_{\lambda g}(x)) \\ \leq & g(x) - \langle x^*, \mathtt{Prox}_{\lambda g}(x)\rangle - \beta \\ \leq & g(x) + \|x^*\| \cdot \|x - \mathtt{Prox}_{\lambda g}(x)\| + \|x^*\| \cdot \|x\| - \beta \\ \leq & g(x) + \lambda\|x^*\|^2 + \frac{1}{4\lambda}\|x - \mathtt{Prox}_{\lambda g}(x)\|^2 \\ & + \|x^*\| \cdot \|x\| - \beta, \end{aligned}$$

where $x \mapsto \langle x^*, x\rangle + \beta$ is an arbitrary but fixed affine minorant of g. We have also used the inequality $ab \leq \frac{c^2}{2}a^2 + \frac{1}{2c^2}b^2$ for $c = \sqrt{2\lambda}$. Thus, for all $x \in S$, we have

$$\|x - \mathtt{Prox}_{\lambda g}(x)\|^2 \leq 4\lambda \left(g(x) + \lambda\|x^*\|^2 + \|x^*\| \cdot \|x\| - \beta\right).$$

Since the right-hand side of the latter inequality is uniformly bounded in S, $\lambda \leq 1$, it is possible to find a constant $C > 0$ such that

$$\|x - \mathtt{Prox}_{\lambda g}(x)\| \leq \sqrt{\lambda} C. \tag{2.8}$$

In particular, it is possible to find $\lambda_0 \in (0,1)$ such that for all $\lambda \in (0, \lambda_0)$

$$\|x - \mathtt{Prox}_{\lambda g}(x)\| \leq 1 \text{ for all } x \in S,$$

implying that $\mathtt{Prox}_{\lambda g}(x) \in \tilde{S}$. Hence, for all $\lambda \in (0, \lambda_0)$ and $x \in S$

$$\begin{aligned} 0 \leq g(x) - \mathtt{e}_\lambda g(x) &= g(x) - g(\mathtt{Prox}_{\lambda g}(x)) - \frac{1}{2\lambda}\|x - \mathtt{Prox}_{\lambda g}(x)\|^2 \\ &\leq \ell\|x - \mathtt{Prox}_{\lambda g}(x)\| \leq \ell\sqrt{\lambda}C, \end{aligned}$$

where we have used (2.8) and the fact that $\mathtt{Prox}_{\lambda g}(x) \in \tilde{S}$ for all $x \in S$ when $\lambda \in (0, \lambda_0)$. Finally, since $\mathtt{e}_\lambda g$ is convex and differentiable, for all $x \in S$

$$\mathtt{e}_\lambda g(y) \geq \mathtt{e}_\lambda g(x) + \langle \nabla \mathtt{e}_\lambda g(x), y - x\rangle \text{ for all } y \in \mathcal{H},$$

where $\nabla \mathtt{e}_\lambda g(x) \in \partial g(\mathtt{Prox}_{\lambda g}(x))$. Hence, since $\mathtt{Prox}_{\lambda g}(x) \in \tilde{S}$ and g is ℓ-Lipschitz on $\tilde{S}$, it follows that $\|\nabla \mathtt{e}_\lambda g(x)\| \leq \ell$ for all $x \in S$. Therefore, for all $x, y \in S$

$$\mathtt{e}_\lambda g(x) \leq \mathtt{e}_\lambda g(y) + \ell\|y - x\|,$$

which ends the proof, by showing that $x \mapsto \mathtt{e}_\lambda g(x)$ is ℓ-Lipschitz on S. □

2.2. Probability functions

Optimization problems have many real-life applications. They can be classified as deterministic or stochastic, depending on whether the input variables are known or random/uncertain. In the latter case, an essential topic is chance-constrained programming, finding applications in water management, telecommunications, electricity network expansion, mineral blending, chemical engineering, etc. (see, e.g., [44, 50]). Chance-constrained programming was initiated in 1958 by Charnes, Cooper and Symonds [16] who studied an optimization problem with a system of individual probabilistic constraints of the form

$$\varphi_i(x) = \mathbb{P}(\xi_i \leq A_i x) \geq p_i \text{ for all } i = 1, \ldots, s,$$

where $x \in \mathbb{R}^n$, A_i are matrices of dimension $m \times n$, ξ_i are m-dimensional random vectors, and $p_i \in [0, 1]$. Later, in 1965, Miller and Wagner [35] reformulated the problem with a single joint probabilistic constrained

$$\varphi(x) = \mathbb{P}(\xi_i \leq A_i x, \text{ for all } i = 1, \ldots, s) \geq p,$$

where $x \in \mathbb{R}^n$, A_i are matrices of dimension $m \times n$, ξ_i are m-dimensional independent random vectors, and $p \in [0, 1]$. In 1970, Prékopa [42] generalized the latter formulation to a more challenging one given by

$$\varphi(x) = \mathbb{P}(\xi \leq g_i(x), \text{ for all } i = 1, \ldots, s) \geq p$$

where $x \in \mathbb{R}^n$, $g_i : \mathbb{R}^n \to \mathbb{R}$, and ξ a random variable. In the mentioned article, Prékopa, gave the data in the problem standard optimality requirements.

More recently, and in this thesis, constraints are considered in the form

$$\varphi(x) = \mathbb{P}(g_i(x,\xi) \leq 0, \text{ for all } i = 1,\dots,s) \geq p, \tag{2.9}$$

where $g_i : \mathcal{X} \times \mathbb{R}^m \to \mathbb{R}$ are functions defined on a space $\mathcal{X}$, ξ is a m-dimensional random vector defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The function φ defined in (2.9) is called a *probability function* and the value $p \in [0,1]$ is called the *probability or safety level*. The meaning of the probabilistic constraint is that in order to declare a decision variable x as feasible it must satisfy the random inequality system $g_i(x,\xi) \leq 0$ with a probability of at least p.

In this section, we recall some classical and recent results about natural properties that probability functions may be asked to have as constituents of an optimization problem, such as continuity, differentiability, generalized differentiability, and generalized concavity. In the sequel, we assume $\mathcal{X}$ to be a reflexive separable Banach space and ξ admitting a density $f_\xi : \mathbb{R}^m \to \mathbb{R}_+$ with respect to the Lebesgue measure λ_m which is is bounded on compact sets, i.e.,

$$f_\xi \in L^\infty(K), \text{ for every compact set } K \subseteq \mathbb{R}^m. \tag{2.10}$$

Let us notice that the probability function in (2.9) can be rewritten as the Lebesgue integral

$$\varphi(x) = \int\limits_{\{z\in\mathbb{R}^m : g(x,z)\leq 0\}} f_\xi(z) d\lambda_m(z), \tag{2.11}$$

where

$$g(x,z) = \max_{i=1,\dots,s} g_i(x,z). \tag{2.12}$$

2.2.1. (Generalized) Differentiability of probability functions

In this subsection, we discuss and present results about some analytical properties of probability functions, such as continuity and differentiability. The topic of understanding the differentiability of probability functions has received great attention. Here we can indicate, e.g., [31, 48, 53] for recent contributions. For a recent introductory text to the topic with a perspective in variational analysis, we refer to [52].

The first natural question that arises is under which conditions are probability functions continuous. As the following example shows, continuity can not be assured even when the given data is continuous.

Example 2.1 *[52, Example 2.1] Consider $g_1, g_2 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by $g_1(x,z) = z - x$, $g_2(x,z) = -x$ and let ξ be a standard Gaussian random variable. Then for all $x < 0$, $\varphi(x) = 0$, whereas $\varphi(0) = \frac{1}{2}$. It thus follows that the probability function is not continuous.*

The following Lemma presents minimum requirements on the problem data ensuring continuity of the probability function, e.g., [27, eq. (3)].

Lemma 2.2 *Let us consider continuous functions $g_i : \mathcal{X} \times \mathbb{R}^m \to \mathbb{R}$ and assume that for all*

$x \in \mathcal{H}$,

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

where g is given by (2.12). Then φ in (2.9) is continuous for all $x \in \mathcal{H}$.

Proof. Consider $x_n \to x$. Due to the continuity of g we get that

$$\liminf_{k\to\infty} \mathbb{1}_{A_n}(z) = \mathbb{1}_A(z), \text{ for all } z \in \mathbb{R}^m,$$

where $A := \{z \in \mathbb{R}^m : g(x,z) > 0\}$ and $A_n := \{z \in \mathbb{R}^m : g(x_n,z) > 0\}$. By applying Fatou's Lemma we obtain

$$1 - \limsup_{n\to\infty} \varphi(x_n) = \liminf_{k\to\infty} \mathbb{P}(\xi \in A_n) \geq \mathbb{P}(\xi \in A) = 1 - \varphi(x),$$

which proves the upper semicontinuity of φ. Now, the assumption leads us to the following inequalities

$$\varphi(x) = \mathbb{P}(\xi \in \mathbb{R}^m \setminus A) \leq \liminf_{n\to\infty} \mathbb{P}(\xi \in \mathbb{R}^m \setminus A_n) \leq \liminf_{k\to\infty} \varphi(x_n),$$

concluding the proof. □

The condition of the set $\{z \in \mathbb{R}^m : g(x,z) = 0\}$ having Lebesgue measure zero in the last lemma is satisfied, for instance, if

$$\mathrm{bd}\{z \in \mathbb{R}^m : g(x,z) \leq 0\} = \{z \in \mathbb{R}^m : g(x,z) = 0\}.$$

This last equality is ensured, in terms of the data, when the functions g_i are convex in the second argument, admitting a common Slater point. Thus, continuity is present under not very restrictive conditions. Meanwhile, differentiability is not guaranteed even if, furthermore, we consider the data functions to be sufficiently smooth as the following example exposes:

Example 2.2 *[31, Example 1] Consider the probability function (2.9) given by a single inequality $g : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}$ defined by $g(x,z_1,z_2) = \alpha(x)h(z_1) + z_2 - 1$, where*

$$\alpha(x) := \begin{cases} x^2, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad \textit{and} \quad h(z_1) := \exp(-1 - 4\log(1 - \Phi(z_1)))$$

with Φ referring to the one-dimensional standard Gaussian distribution function. Also consider, ξ_1 and ξ_2 being standard Gaussian random variables. It follows that the probability function is not even locally Lipschitzian at $x = 0$, despite g being continuously differentiable, convex in the second argument, and such that $g(0,0,0) < 0$.

Figure 2.1 illustrates the nonsmoothness of φ at $x = 0$.

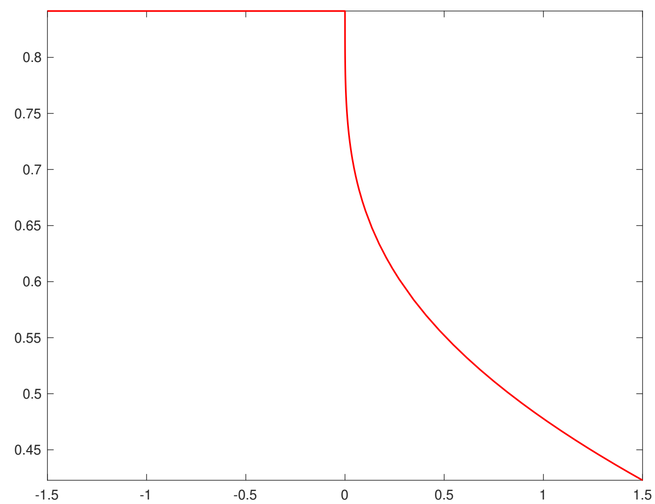

Figure 2.1: Graph of φ in Example 2.2

The failure of the compactness of the set $\{(z_1, z_2) \in \mathbb{R}^2 : g(0, z_1, z_2) \leq 0\}$ in Example 2.2 seems to be the cause of nonsmoothess behavior even though all input data are smooth. To assume compactness is not ideal since it rules out problems involving random vectors with unbounded support. So, to handle unboundedness, a growth condition on the functions g_i may be required instead. Nonetheless, with sufficiently nice data and even compactness, we can be led to nonsmoothness:

Example 2.3 *[54, Example 1.1] Consider* $g_1(x_1, x_2, x_3, z) = z - x_1$, $g_2(x_1, x_2, x_3, z) = z - x_2$, $g_3(x_1, x_2, x_3, z) = -z + x_3$ *and let* ξ *have a one-dimensional standard Gaussian distribution. Then, with* Φ *referring to the one-dimensional standard Gaussian distribution function, one has that*

$$\varphi(x_1, x_2, x_3) = \max\{\min\{\Phi(x_1), \Phi(x_2)\} - \Phi(x_3), 0\}.$$

Clearly φ *fails to be differentiable at* $x = (0, 0, -1)$*, while* $\{z \in \mathbb{R} : g(0, 0, -1, z)\} = [-1, 0]$ *is compact and satisfies Slater's condition.*

The growth condition that may be imposed will lead us to local Lipschitz continuity of φ. Thus, the motivation to obtain subdifferential formulas for φ. In turn, the differentiability of probability functions with single inequalities is guaranteed in the finite-dimensional setting by assuming that the density is continuous (see Corollary 2.1below).

As discussed in [52] differentiability has been studied through two paths. In this thesis we consider the path that focuses on a representation of the probability function in (2.11) relying on a parametrization via polar coordinates (see [29, Theorem 2.49 and Proposition 2.51]) of the random vector $\xi \in \mathbb{R}^m$.

Theorem 2.1 (spherical-radial decomposition [26, 31, 53, 55]) *The probability function in* (2.11) *can be rewritten as*

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

where μ_ζ is the uniform distribution on $\mathbb{S}^{m-1}$ and $e : \mathcal{X} \times \mathbb{S}^{m-1} \to \mathbb{R}_\infty$ is given by

$$e(x, v) = \frac{2\pi^{\frac{m}{2}} |\det(L)|}{\Gamma(\frac{m}{2})} \int\limits_{\{r \geq 0 : g(x, rLv) \leq 0\}} r^{m-1} f_\xi(rLv) dr. \tag{2.13}$$

where L is an arbitrary nonsingular matrix of dimension $m \times m$.

Following the terminology in [57] we refer to the function e in Theorem 2.1 as the *radial probability-like* function and, to simplify the notation, the *density-like* function to refer us to

$$\theta(r, v) := \frac{2\pi^{\frac{m}{2}} |\det(L)|}{\Gamma(\frac{m}{2})} r^{m-1} f_\xi(rLv). \tag{2.14}$$

The density-like function is independent of v when we consider the random vector $\xi \in \mathbb{R}^m$ to be elliptical symmetrically distributed. To see this, let us first define the family of such distributions and give some examples.

Definition 2.2 *We say that the random vector $\xi \in \mathbb{R}^m$ is elliptical symmetrically distributed with mean μ, positive definite covariance matrix Σ, and generator $\tilde{\theta} : \mathbb{R}_+ \to \mathbb{R}_+$, which is denoted by $\xi \sim \mathcal{E}(\mu, \Sigma, \tilde{\theta})$, if and only if its density is given by*

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

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

The family of elliptical random vectors includes the Gaussian random vectors and Student random vectors with the respective generators

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

For more examples, see [26] and [34].

Let us now notice that when $\xi \sim \mathcal{E}(\mu, \Sigma, \tilde{\theta})$ we can always assume it with $\mu = 0$ and Σ a correlation matrix by passing to the standardized vector $\tilde{\xi} := D(\xi - \mu)$ where D is the diagonal matrix with elements $D_{ii} := 1/\sqrt{\Sigma_{ii}}$, and consider $\tilde{g}(x, z) := g(x, D^{-1}z + \mu)$. Thus,

$$\tilde{\xi} \sim (0, R, \tilde{\theta}) \;\text{ and }\; \mathbb{P}(\tilde{g}(x, \tilde{\xi})) = \mathbb{P}(g(x, \xi)),$$

where R is the correlation matrix associated with Σ. Therefore, if ξ is elliptical symmetrically distributed and if the matrix L in Theorem 2.1 is the lower triangle matrix that satisfies $R = LL^T$ (the so-called Cholesky decomposition of R) we obtain that

$$f_\xi(rLv) = \det(L)^{-1} \tilde{\theta}(r^2).$$

For instance, when $\xi \sim \mathcal{N}(0,R)$, the radial probability-like function takes the form

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

where $\mu_{\mathcal{R}}$ is the one-dimensional Chi-distribution with m degrees of freedom.

Given a point of interest $\bar{x}$ we assume that on a neighborhood U of $\bar{x}$ the functions g_i are locally Lipschitz at any $(x,z) \in U \times \mathbb{R}^m$, convex in the second variable and satisfy

$$g_i(x,0) < 0, \qquad \text{for all } x \in U \text{ and all } i = 1, \ldots, s. \tag{2.15}$$

We now address the generalized differentiability of the probability function. To do this, let us give some preliminary definitions and results. For $x \in U$ the sets of finite and infinite directions with respect to g_i are the sets

$$\begin{aligned} F_i(x) &:= \{v \in \mathbb{S}^{m-1} \mid \exists r \geq 0 : g_i(x, rLv) = 0\}, && (2.16)\\ I_i(x) &:= \{v \in \mathbb{S}^{m-1} \mid \forall r \geq 0 : g_i(x, rLv) < 0\}, && (2.17) \end{aligned}$$

respectively. The finite and infinite directions with respect to g can be defined analogously and in fact they identify with $F(x) = \cup_{i=1}^s F_i(x)$ and $I(x) = \cap_{i=1}^s I_i(x)$. Now, let us define the radial functions

$$\rho_i\,(x,v) := \sup\left\{r > 0 : g_i(x, rLv) \leq 0\right\} \text{ for all } i = 1, \ldots, s \tag{2.18}$$

and the set of active indexes at (x,v), $T_x(v) := \{i = 1, \ldots, s : \rho_i(x,v) = \rho(x,v)\}$, where

$$\rho(x,v) = \min_{i=1,\ldots,s} \rho_i(x,v). \tag{2.19}$$

The following two lemmas collect some elementary properties of the objects defined above that can be found in, e.g., [31, 53, 54].

Lemma 2.3 *Let $x \in U$. Then the following hold:*

1. *$\{r \geq 0 \ : g_i(x, rLv) \leq 0\} = [0, \rho_i(x,v)]$ where $\rho_i(x,v) = +\infty$ is allowed, with the convention $[0,+\infty] = [0,+\infty)$.*
2. *$I_i(x) \cup F_i(x) = \mathbb{S}^{m-1}$.*
3. *$\rho_i(x,v) = \{r$ such that $g_i(x, rLv) = 0\}$ when $v \in F_i(x)$.*
4. *$\mathrm{dom}(\rho_i(x,\cdot)) = F_i(x)$.*
5. *For $z^* \in \partial_z g_i(x, \rho(x,v)Lv)$,*

$$\langle z^*, Lv\rangle \geq -\frac{g_i(x,0)}{\rho(x,v)} > 0.$$

Lemma 2.4 *The radial functions ρ_i are continuous at $(x,v) \in U \times \mathbb{S}^{m-1}$. In consequence, ρ is continuous at $(x,v) \in U \times \mathbb{S}^{m-1}$.*

The following lemma corresponds to a generalization of [57, Lemma 3.4].

Lemma 2.5 *For $x \in U$, $v \in F_i(x)$, we have*

$$\partial_x^{\mathsf{b}} \rho_i(x,v) \subseteq \operatorname{cl}\operatorname{co}\left\{ \frac{-1}{\langle z^*, Lv\rangle} x^* : (x^*, z^*) \in \partial^{\mathsf{b}} g_i(x, \rho_i(x,v)Lv) \right\}. \tag{2.20}$$

for all $i = 1, \ldots, s$. Moreover,

$$\partial_x^{\mathsf{b}} \rho(x,v) \subseteq \operatorname{cl}\operatorname{co}\left\{ \frac{-1}{\langle z^*, Lv\rangle} x^* : \begin{array}{c} (x^*, z^*) \in \partial^{\mathsf{b}} g_i(x, \rho(x,v)Lv) \\ \text{and } i \in T_x(v) \end{array} \right\}. \tag{2.21}$$

Proof. Fix $\bar{x} \in U$, $i \in \{1, \ldots, s\}$ and $\bar{v} \in F_i(\bar{x})$. To obtain (2.20) let us first prove that for every $y^* \in \partial_x^{\mathsf{b}} \rho_i(\bar{x}, \bar{v})$ and every $w \in \mathcal{X}$, there exists $(x^*, z^*) \in \partial^{\mathsf{b}} g_i(\bar{x}, \rho_i(\bar{x}, \bar{v})L\bar{v})$ such that

$$\langle y^*, w\rangle \leq \frac{-1}{\langle z^*, L\bar{v}\rangle} \langle x^*, w\rangle.$$

By continuity of ρ_i and since g_i is locally Lipschitz there exists $\varepsilon' > 0$ such that $\mathbb{B}(\bar{x}, \varepsilon') \subseteq U$ and for every $x \in \mathbb{B}(\bar{x}, \varepsilon')$ and $z \in \mathbb{B}(\rho_i(\bar{x}, \bar{v})L\bar{v}, \varepsilon')$ we have that $\bar{v} \in F_i(x)$, $g_i(x, 0) < 0$ and

$$\partial^{\mathsf{b}} g_i(x, z) \subseteq r\mathbb{B}^* \tag{2.22}$$

for some $r > 0$.

We claim that for every $y^* \in \partial_x^{\mathsf{r}} \rho_i(x, \bar{v})$ with $x \in \mathbb{B}(\bar{x}, \varepsilon'/2)$ and every $w \in \mathcal{X}$ there exists $(x^*, z^*) \in \partial^{\mathsf{b}} g_i(x, \rho_i(x, \bar{v})L\bar{v})$ such that

$$\langle y^*, w\rangle \leq \frac{-1}{\langle z^*, L\bar{v}\rangle} \langle x^*, w\rangle.$$

To see this, let $w \in \mathcal{X}$ and consider $t_k \to 0^+$ such that

$$x + t_k w \in \mathbb{B}(\bar{x}, \varepsilon'/2) \ \text{ and } \ \rho_i(x + t_k w, \bar{v})L\bar{v} \in \mathbb{B}(\rho_i(x, \bar{v})L\bar{v}, \varepsilon'/2), \ \text{ for all } k.$$

Applying the mean value inequality in [39, Corollary 3.51] we get

$$g_i(x+t_k w, \rho_i(x+t_k w, \bar{v})L\bar{v}) - g_i(x, \rho_i(x, \bar{v})L\bar{v}) \leq -t_k\langle x_k^*, w\rangle + [\rho_i(x, \bar{v}) - \rho_i(x+t_k w, \bar{v})]\langle z_k^*, L\bar{v}\rangle,$$

for some

$$(x_k^*, z_k^*) \in \partial^{\mathsf{b}} g_i(x_k, z_k) \tag{2.23}$$

with

$$x_k \in [x + t_k w, x) \text{ and } z_k \in [\rho_i(x + t_k w, \bar{v})L\bar{v}, \rho_i(x, \bar{v})L\bar{v}).$$

Hence, taking into account that

$$g_i(x, \rho(x, \bar{v})L\bar{v}) = 0 \ \text{ and } \ g_i(x + t_k w, \rho_i(x + t_k w, \bar{v})L\bar{v}) = 0,$$

it follows that

$$[\rho_i(x + t_k w, \bar{v}) - \rho_i(x, \bar{v})]\langle z_k^*, L\bar{v}\rangle \leq -t_k\langle x_k^*, w\rangle.$$

Now, by [38, Lemma 4.8], $z_k^* \in \partial_z g_i(x_k, \rho_i(x_k, \bar{v})L\bar{v})$ and hence $\langle z_k^*, L\bar{v}\rangle > 0$ (recall Lemma

2.3). Therefore,

$$\frac{\rho_i(x+t_kw,\bar v)-\rho_i(x,\bar v)}{t_k}\le\frac{-1}{\langle z_k^*,L\bar v\rangle}\langle x_k^*,w\rangle.$$

Considering $\varepsilon_k\to 0^+$ with $\varepsilon_k<\varepsilon'/2$ for all k, by definition of the basic subdifferential, we have that there exists $(\hat x_k^*,\hat z_k^*)\in\partial^{\mathsf r}g_i(\hat x_k,\hat z_k)$ with $\|\hat x_k-x_k\|\le\varepsilon_k$, $\|\hat z_k-z_k\|\le\varepsilon_k$, $\|\hat z_k^*-z_k^*\|\le\varepsilon_k$ and such that

$$\langle z_k^*,L\bar v\rangle\frac{\rho_i(x+t_kw,\bar v)-\rho_i(x,\bar v)}{t_k}\le-(\langle\hat x_k^*,w\rangle-\varepsilon_k). \tag{2.24}$$

Now using (2.22), we have that $\|\hat x_k^*\|\le r$ and $\|\hat z_k^*\|\le r$. Since $\mathcal X$ is reflexive there exists a subsequence $(\hat x_{n_k}^*,\hat z_{n_k}^*)$ and some $(x^*,z^*)\in\mathcal X\times\mathbb R^m$ such that $x_{n_k}^*\rightharpoonup_k x^*$ and $z_{n_k}^*\to z^*$. By definition of the basic subdifferential $(x^*,z^*)\in\partial^{\mathsf b}g_i(x,\rho_i(x,\bar v)L\bar v)$. Again by [38, Lemma 4.8], $z^*\in\partial_z g_i(x,\rho_i(x,\bar v)L\bar v))$ and thus $\langle z^*,L\bar v\rangle>0$ (recall Lemma 2.3). Thus, by applying inferior limit in (2.24), we conclude the proof of the claim by recalling the definition of the regular subdifferential.

Now, let $y^*\in\partial_x^{\mathsf b}\rho_i(\bar x,\bar v)$. Then there exist $y_l^*\to y^*$ and $x_l\to\bar x$ with $y_l^*\in\partial_x^{\mathsf r}\rho_i(x_l,\bar v)$. For l large enough such that

$$\|x_l-\bar x\|\le\varepsilon'/2 \text{ and } \|\rho_i(x_l,\bar v)L\bar v-\rho_i(\bar x,\bar v)L\bar v\|\le\varepsilon'/2,$$

we apply the claim proved above to obtain that there exists $(x_l^*,z_l^*)\in\partial^{\mathsf b}g_i(x_l,\rho_i(x_l,\bar v)L\bar v)$ such that

$$\langle y_l^*,w\rangle\le\frac{-1}{\langle z_l^*,L\bar v\rangle}\langle x_l^*,w\rangle.$$

By definition of the basic subdifferential and considering $\varepsilon_l\to 0^+$ with $\varepsilon_l<\varepsilon'/2$ for all l, there exists $(\hat x_l^*,\hat z_l^*)\in\partial^{\mathsf r}g_i(\hat x_l,\hat z_l)$ with $\|\hat x_l-x_l\|\le\varepsilon_l$, $\|\hat z_l-\rho_i(x_l,\bar v)L\bar v+\bar z\|\le\varepsilon_l$, $\|\hat z_l^*-z_l^*\|\le\varepsilon_l$ and such that

$$\langle y_l^*,w\rangle\le\frac{-1}{\langle z_l^*,L\bar v\rangle}(\langle\hat x_l^*,w\rangle-\varepsilon_l). \tag{2.25}$$

Again, using (2.22) we obtain that (under subsequence) that

$$(\hat x_l^*,\hat z_l^*)\rightharpoonup_l(x^*,z^*)\in\partial^{\mathsf b}g_i(\bar x,\rho_i(\bar x,\bar v)L\bar v).$$

Therefore, letting $l\to\infty$ in (2.25), we conclude that

$$\langle y^*,w\rangle\le\frac{-1}{\langle z^*,L\bar v\rangle}\langle x^*,w\rangle,$$

for some $(x^*,z^*)\in\partial^{\mathsf b}g_i(\bar x,\rho_i(\bar x,\bar v)L\bar v)$. Let us notice that this last result implies that

$$\langle y^*,w\rangle\le\sigma_{\mathcal A(\bar x,\bar v)}(w) \text{ for all } y^*\in\partial_x^{\mathsf b}\rho_i(\bar x,\bar v) \text{ and for all } w\in\mathcal X$$

where

$$\mathcal A(\bar x,\bar v):=\left\{\frac{-1}{\langle z^*,L\bar v\rangle}x^*:(x^*,z^*)\in\partial^{\mathsf b}g_i(\bar x,\rho_i(\bar x,\bar v)L\bar v)\right\}.$$

Therefore, $\sigma_{\partial_x^{\mathsf b}\rho(\bar x,\bar v)}(w)\le\sigma_{\mathcal A(\bar x,\bar v)}(w)$, for all $w\in\mathcal X$, which entails (2.20) (see, e.g. [60]). Finally, (2.21) follows from [39, Proposition 1.113]. $\square$

Following [57], we define the function $I_\theta : \mathbb{R}_+ \times \mathbb{S}^{m-1} \rightrightarrows \mathbb{R}_+$ by

$$I_\theta(r,v) := [\underline{\theta}(r,v), \overline{\theta}_+(r,v)] \cup [\underline{\theta}^-(r,v), \overline{\theta}(r,v)], \tag{2.26}$$

where

$$\begin{aligned} \overline{\theta}(r,v) &:= \inf\left\{k > 0 : \exists \varepsilon > 0 \text{ such that } \theta(u,v) \le k \text{ a.e. for } u \in [r-\varepsilon, r+\varepsilon]\right\}, \\ \overline{\theta}_+(r,v) &:= \inf\{k > 0 : \exists \varepsilon > 0 \text{ such that } \theta(u,v) \le k \text{ a.e. for } u \in [r, r+\varepsilon]\}, \\ \underline{\theta}(r,v) &:= \sup\{k > 0 : \exists \varepsilon > 0 \text{ such that } \theta(u,v) \ge k \text{ a.e. for } u \in [r-\varepsilon, r+\varepsilon]\}, \\ \underline{\theta}^-(r,v) &:= \sup\{k > 0 : \exists \varepsilon > 0 \text{ such that } \theta(u,v) \ge k \text{ a.e. for } u \in [r-\varepsilon, r]\}. \end{aligned} \tag{2.27}$$

As it was deduced in [57], property (2.10) implies that

$$I_\theta(r,v) \subseteq [0, \overline{\theta}(r,v)] \subseteq [0, +\infty). \tag{2.28}$$

Furthermore, from the definition of $\overline{\theta}(r,v)$, we have

$$\overline{\theta}(r,v) \le M_\theta, \ \forall r \le M, \text{ and } v \in \mathbb{S}^{m-1}, \tag{2.29}$$

where M_θ is the (finite) constant defined by

$$M_\theta := \operatorname*{ess\,sup}_{(r,v) \in [0,M+1] \times \mathbb{S}^{m-1}} \overline{\theta}(r,v).$$

To obtain a generalized differentiability result we require the following growth condition:

Definition 2.3 (η_θ-growth condition for nonsmooth functions) *Consider $\bar{x} \in U$ and $\bar{v} \in I(\bar{x})$. Let $\eta_\theta : \mathbb{R} \times \mathbb{S}^{m-1} \to [0, +\infty]$ be a mapping such that*

$$\lim_{\substack{x \to \bar{x} \\ v \to \bar{v}}} \rho(x,v) \bar{\theta}(\rho(x,v), v) \eta_\theta(\rho(x,v), v) = 0. \tag{2.30}$$

We say that the family of mappings $\{g_i\}_{i=1}^s$ satisfies the η_θ-growth condition at $(\bar{x}, \bar{v})$ if for some $l > 0$

$$\|\pi_x(\partial^{\mathsf{b}} g_i(x, \rho(x,v)Lv))\| \le l \eta_\theta(\rho(x,v), v), \ \forall (x,v) \in \mathbb{B}_{1/l}(\bar{x}) \times \mathbb{B}_{1/l}(\bar{v}), \ v \in F(x) \text{ and } i \in T_x(v); \tag{2.31}$$

where $\|\pi_x(\partial^{\mathsf{b}} g_i(x, \rho(x,v)Lv))\| := \sup\{\|x^*\| : (x^*, z^*) \in \partial^{\mathsf{b}} g_i(x, \rho(x,v)Lv) \text{ for all } z^*\}$.

The following theorem provides an extension of [57, Theorem 3.1] to the infinite-dimensional nonsmooth setting.

Theorem 2.2 *Let $\bar{x} \in U$ be given and assume that the family of mappings $\{g_i\}_{i=1}^s$ satisfies the η_θ-growth condition at $(\bar{x}, v)$ for all $v \in I(\bar{x})$ and that (2.15) holds true.*

Then the probability function (2.9) is locally Lipschitz at $\bar{x}$ and on an appropriate neigh-

borhood U' of $\bar{x}$ it holds:

$$\partial^{\mathsf{b}}\varphi(x) \subseteq \mathrm{cl}^{w^*}\left(\int\limits_{v\in\mathbb{S}^{m-1}} \partial_x^{\mathsf{b}} e(x,v) d\mu_\zeta(v)\right), \text{ for all } x \in U' \tag{2.32}$$

where, $\partial_x^{\mathsf{b}} e(x,v) \subseteq \{0\}$ if $v \in I(x)$ and

$$\partial_x^{\mathsf{b}} e(x,v) \subseteq \mathrm{cl\,co}\left\{ \frac{-\alpha}{\langle z^*, Lv\rangle} x^* : \begin{array}{c} (x^*, z^*) \in \partial^{\mathsf{b}} g_i(x, \rho(x,v) Lv) \\ i \in T_x(v),\ \alpha \in I_\theta(\rho(x,v), v) \end{array} \right\} \text{ if } v \in F(x) \tag{2.33}$$

where I_θ is given by (2.26). Moreover, the closure operator can be omitted in (2.32) if $\mathcal{X}$ is finite-dimensional.

Proof. First, let us show that for every fixed $\bar{v} \in \mathbb{S}^{m-1}$, there exist neighborhoods $U_{\bar{v}}$ of $\bar{x}$ and $V_{\bar{v}}$ of $\bar{v}$ and $K_{\bar{v}} > 0$ such that

$$\partial_x^{\mathsf{b}} e(x,v) \subset K_{\bar{v}} \mathbb{B}^* \text{ for all } (x,v) \in U_{\bar{v}} \times V_{\bar{v}}. \tag{2.34}$$

To this end we consider the following two cases:

1. Let $\bar{v} \in F(\bar{x})$. By continuity of ρ there exist neighborhoods $U_{\bar{v}}$ of $\bar{x}$, $V_{\bar{v}}$ of $\bar{v}$ and a constant $M > 0$ such that $\rho(x,v) \leq M$ and $g(x,0) < 0$ for all $(x,v) \in U_{\bar{v}} \times V_{\bar{v}}$. Hence, we may apply Lemma 2.5 to such neighborhoods and by using [57, Proposition 3.2] we obtain

   $$\partial_x^{\mathsf{b}} e(x,v) \subseteq \mathrm{cl\,co}\left\{ \frac{-\alpha}{\langle z^*, Lv\rangle} x^* : \begin{array}{c} (x^*, z^*) \in \partial^{\mathsf{b}} g_i(x, \rho(x,v) Lv) \\ i \in T_x(v),\ \alpha \in I_\theta(\rho(x,v), v) \end{array} \right\}, \text{ for all } (x,v) \in U_{\bar{v}} \times V_{\bar{v}}. \tag{2.35}$$

   Now, for each $(x^*, z^*) \in \partial^{\mathsf{b}} g_i(x, \rho(x,v) Lv)$ in (2.35) we have that $z^* \in \partial_z g_i(x, \rho(x,v) Lv)$ (see [38, Lemma 4.8]) and in consequence (recall Lemma 2.3)

   $$\langle z^*, Lv\rangle \geq -\frac{g_i(x,0)}{\rho(x,v)} > 0.$$

   Hence, by (2.35) and (2.28), for each $y^* \in \partial_x^{\mathsf{b}} e(x,v)$ and $i \in T_x(v)$ there exists $(x^*, z^*) \in \partial^{\mathsf{b}} g_i(x, \rho(x,v) Lv)$ such that

   $$\|y^*\| \leq \frac{-1}{|g_i(x,0)|} \rho(x,v) \bar{\theta}(\rho(x,v), v) \|x^*\|. \tag{2.36}$$

   Therefore, by the fact that g_i's are locally Lipschitz together with the continuity of ρ and (2.29) we may derive the estimate (2.34).

2. Let $\bar{v} \in I(\bar{x})$. Consider $l > 0$ given by the η_θ-growth condition in Definition 2.3 and let $\varepsilon > 0$ be arbitrary. By continuity of ρ, there exist neighborhoods $U_{\bar{v}}$ of $\bar{x}$ and $V_{\bar{v}}$ of $\bar{v}$, contained in $\mathbb{B}_{1/l}(\bar{x})$ and $\mathbb{B}_{1/l}(\bar{v})$ respectively, such that

   $$\rho(x,v) \geq l \text{ and } \rho(x,v)\bar{\theta}(\rho(x,v), v)\eta_\theta(\rho(x,v), v) \leq \varepsilon \text{ for all } (x,v) \in U_{\bar{v}} \times V_{\bar{v}}.$$

   We consider the following two further cases:

a) If $v \in F(x)$, we continue as in Item 1. until we obtain (2.36). Then from (2.31), for each for each $y^* \in \partial_x^{\mathsf{b}} e(x,v)$ we have that

$$\|y^*\| \le \frac{l}{|g(x,0)|}\bar{\theta}(\rho(x,v),v)\rho(x,v)\eta_\theta(\rho(x,v),v) \le \varepsilon l \sup_{x\in U}\frac{1}{|g(x,0)|} =: K_{\bar{v}}.$$

b) If $v \in I(x)$, similarly as in [57, Proposition 3.4 i)] we obtain that $\partial_x^{\mathsf{r}} e(x,v) \subseteq \{0\}$. Now let $y^* \in \partial_x^{\mathsf{b}} e(x,v)$ and choose $x_n \to x$ and $y_n^* \rightharpoonup y^*$ with $y_n^* \in \partial_x^{\mathsf{r}} e(x_n,v)$. If $v \in I(x_n)$ for all n then $y^* = 0$. Instead, if $v \in F(x_n)$ for all n, we consider a sequence $\varepsilon_n \to 0^+$, hence, from part a) with ε_n instead of ε we obtain that $y^* = 0$.

From a) and b) we obtain the estimate (2.34) in this case.

Now, since $\mathbb{S}^{m-1}$ is compact and the family of neighborhoods $V_{\bar{v}}$ covers $\mathbb{S}^{m-1}$, we can pick a finite subcover, that is, there exists $N \in \mathbb{N}$ and some $v_1,\dots,v_N \in \mathbb{S}^{m-1}$ such that

$$\mathbb{S}^{m-1} \subset \bigcup_{i=1}^{N} V_{v_i}.$$

Therefore, we choose a neighborhood U' of $\bar{x}$ such that

$$U' \subset \bigcap_{i=1}^{N} U_{v_i}$$

and define $\kappa := \max\{K_{v_i} :\ i = 1\dots,N\}$ to conclude that

$$\partial_x^{\mathsf{b}} e(x,v) \subset \kappa\mathbb{B}^* \ \text{ for all } x \in U' \text{ and } v \in \mathbb{S}^{m-1}. \tag{2.37}$$

Finally, (2.32) follows from [22, Corollary 4.4] □

By considering a single inequality in (2.9) we obtain a gradient formula for the probability function under stronger assumptions than in Theorem 2.2. We refer to [57] for similar results.

Corollary 2.1 (Corollary 3.2 [57]) *Consider $\mathcal{X} = \mathbb{R}^n$, $\bar{x} \in U$ and assume that g is continuously differentiable and satisfies the η_θ-growth condition at $(\bar{x},\bar{v})$ for all $\bar{v} \in I(\bar{x})$. Moreover, assume that f_ξ is continuous.*

Then the probability function φ defined in (2.9) is continuously differentiable on an appropriate neighbourhood U' of $\bar{x}$ with

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

where,

$$\nabla_x e(x,v) = \begin{cases} -\frac{\theta(\rho(x,v),v)}{\langle \nabla_z g(x,\rho(x,v)Lv), Lv\rangle}\nabla_x g(x,\rho(x,v)Lv) & \textit{if } v \in F(x), \\ 0 & \textit{if } v \in I(x). \end{cases}$$

Proof. Since f_ξ is continuous, the set $I_\theta(r, v) = \{\theta(r, v)\}$ and satisfies (2.10). Therefore, the result follows from the fact that $\mathcal{X} = \mathbb{R}^n$ together with [36, Theorem 4.17] upon noticing that the right-hand side of (2.33) is a singleton. ☐

The following example exposes that the η_θ-growth condition holds for a continuously differentiable function g when ξ is a standard Gaussian random variable and when the gradients $\nabla_x g$ satisfy an exponential growth condition.

Example 2.4 *Let $\xi \sim \mathcal{N}(0, R)$ and suppose that at x there exists $\varepsilon, C > 0$ such that*

$$\|\nabla_x g(x, z)\| \leq C\exp(\|z\|),\ \forall x \in \mathbb{B}_\varepsilon(\bar{x})\ \ \forall z : \|z\| \geq C.$$

Defining $\eta_\theta(r, v) := \exp(r\|Lv\|)$ we observe that

$$r\theta(r, v)\eta_\theta(r, v) = 2^{1-\frac{m}{2}} r^m \exp(-\tfrac{1}{2}r^2)\exp(r\|Lv\|) \to_{r\to+\infty} 0.$$

Furthermore, since L is nonsingular we may find $l \geq C$ such that $r \geq l$ implies that $r\|Lv\| \geq C$ for all $v \in \mathbb{S}^{m-1}$. Thus, when $r \geq l$ and $x \in \mathbb{B}_\varepsilon(\bar{x})$,

$$\|\nabla_x g(x, rLv)\| \leq C\exp(r\|Lv\|) \leq l\exp(r\|Lv\|) = l\eta_\theta(r, v).$$

That is, g satisfies the η_θ-growth condition above.

2.2.2. (Generalized) Concavity of probability functions

Along with continuity and differentiability, a fundamental question for an optimization problem is about the convexity of the feasible set $M(p) := [\varphi \geq p]$. It is well-known that $M(p)$ is convex if and only if φ is *quasi-concave* (see definition below). Investigations regarding the concavity of φ, based on the underlying probability distributions, start with classical works by Prékopa, e.g., [43]. To state the most general result we need the following generalized concavity definition.

Definition 2.4 (α-concavity) *A nonnegative function f defined on a convex set $D \subset \mathcal{X}$ is α-concave, where $\alpha \in [-\infty, \infty]$, if for all $x, y \in D$ and all $\lambda \in [0, 1]$ the following inequality holds:*

$$f(\lambda x + (1-\lambda)y) \geq m_\alpha(f(x), f(y), \lambda),$$

where $m_\alpha : \mathbb{R}_+ \times \mathbb{R}_+ \times [0, 1] \to \mathbb{R}$ is defined as follows:

$$m_\alpha(a, b, \lambda) = 0 \text{ if } ab = 0,\ \text{and } \alpha \leq 0$$

and for any other value of a and b,

$$m_\alpha(a, b, \lambda) = \begin{cases} a^\lambda b^{1-\lambda} & \text{if } \alpha = 0, \\ \max\{a, b\} & \text{if } \alpha = \infty, \\ \min\{a, b\} & \text{if } \alpha = -\infty, \\ (\lambda a^\alpha + (1-\lambda)b^\alpha)^{1/\alpha} & \text{otherwise.} \end{cases}$$

In the case $\alpha = 0$, the function is called log-concave, for $\alpha = 1$ concave, and $\alpha = -\infty$ quasi-concave.

Remark 2.1 *The mapping $\alpha \mapsto m_\alpha(a, b, \lambda)$, being nondecreasing (see [50, Lemma 4.8]), implies that all α-concave functions are quasi-concave.*

Example 2.5 *The density function f_ξ of $\xi \sim \mathcal{N}(0, R)$, with R a positive definite matrix, is a log-concave function. Indeed, the function*

$$\log f_\xi(z) = -(\tfrac{1}{2} z^T R z + \ln \sqrt{(2\pi)^m \det(R)})$$

is concave.

Definition 2.5 *A random vector ξ has α-concave probability distribution if the probability measure $\mathbb{P}_\xi(A) := \mathbb{P}(\xi \in A)$ induced by ξ on $\mathbb{R}^m$ satisfies that for any Borel measurable sets $A, B \subseteq \mathbb{R}^m$ and for all $\lambda \in [0, 1]$*

$$\mathbb{P}_\xi(\lambda A + (1 - \lambda) B) \geq m_\alpha(\mathbb{P}_\xi(A), \mathbb{P}_\xi(B), \lambda).$$

Example 2.6 *Random variables having log-concave density probability function have log-concave probability distribution (see [50, Theorem 4.15]). In particular, by Example 2.5, $\xi \sim \mathcal{N}(0, R)$ has log-concave probability distribution.*

One of the most general results is the following theorem.

Theorem 2.3 *[50, Theorem 4.39] Let $g : \mathcal{X} \times \mathbb{R}^m \to \mathbb{R}$ be quasi-convex[1] on both variables and $\xi \in \mathbb{R}^m$ be a random vector having an α-concave probability distribution. Then the function φ in (2.9) is α-concave on the set $\{x \in \mathcal{X} : \exists z \in \mathbb{R}^m$ s.t. $g(x, z) \leq 0\}$.*

[1] meaning $-g$ is quasi-concave.

Chapter 3

Inner Moreau envelope of probability functions

In convex analysis, the Moreau envelope (also called Moreau-Yosida regularization) is a useful regularization for general nonsmooth convex functions. The applications of such an envelope cover a variety of theoretical developments, and it is at the core of many numerical optimization methods. Nowadays, there are plenty of explicit formulations for the computation of the Moreau envelope of most common convex functions, and there are efficient algorithms to compute the envelope numerically for more complex data (see, e.g., [10] and the references therein).

In this chapter, we propose and investigate a general regularization of probabilistic functions, which employs the Moreau envelope of some functions. Formally, we consider a probability function $\varphi : \mathcal{H} \to [0, 1]$ given by

$$\varphi(x) := \mathbb{P}\left(\omega \in \Omega : \Phi(x, \xi(\omega)) \in -\mathcal{K}\right), \tag{3.1}$$

where $\mathcal{H}$ is a Hilbert space, $\xi : \Omega \to \mathbb{R}^m$ is an m-dimensional random vector, $\mathcal{K} \subset \mathcal{Y}$ is a (nonempty) convex cone of a Banach space $\mathcal{Y}$ and $\Phi : \mathcal{H} \times \mathbb{R}^m \to \mathcal{Y}$ is a function. The formulation of the probability function φ in (3.1) covers several of the most general classes of probabilistic constraints arising in chance constrained, joint-chance constrained, and even probabilistic/robust (probust) chance constrained optimization problems, as we will show in this work. Here, it is worth mentioning that the inclusion can be represented as an abstract inequality given by the cone order $x \preceq y$ if and only if $y - x \in \mathcal{K}$. A particular example covered is one wherein $\mathcal{K}$ is the cone of positive definite matrices, and thus $\Phi(x, \xi) \in \mathcal{K}$ represents that our (random) decision matrix $\Phi(x, \xi)$ should be positive semidefinite for most possible cases (see Section 3.5 for more details on such an application).

Since the random (possibly infinite-dimensional) constraint $\Phi(x, \xi(\omega)) \in -\mathcal{K}$ is challenging to handle, we propose a Moreau regularization of a (nonsmooth) scalarization of the function Φ. Then our regularization will be given by the probability function generated by the Moreau envelope of that regularization (see Section 3.1 for more details). Surprisingly, and under mild assumptions, such regularization inherits variational properties of the Moreau envelope, for instance, its smoothness and variational convergence to the original function. Those properties are used to provide a regularization of (general) chance constrained optimization problems and the convergence of the minimizers of the regularized problems to the minimizers of the original formulation. It is natural to understand such convergence as a naive form to propose a toolbox for solving general classes of nonsmooth chance constraints optimization problems. Consequently, our developments open a gate to study further improvements using the ideas exploited in deterministic optimization algorithms, which use Moreau envelops of functions in a future research project.

This chapter is organized as follows: section 3.1 suggests the setting of the work. Section 3.2 examines the convergence of the inner Moreau envelope of the probability function towards the nominal probability function (3.1). Differentiability of the approximating function is investigated in section 3.3. The manner in which the use of approximated probability functions, through their inner Moreau envelope, allows us to approximate a given optimization problem is investigated in section 3.4. Section 3.5 provides several examples and possible applications of the developed results.

3.1. Inner scalarization

In this section, we describe our inner regularization of the probability function (3.1). In order to set up a suitable framework to use the properties of the Moreau envelope, we need to impose that our nominal function Φ in (3.1) satisfies some convexity properties. A common assumption in the study of probability functions is that the inequality systems satisfy some property of convexity with respect to the random variable $\xi \in \mathbb{R}^m$, but not necessarily in the decision variable $x \in \mathcal{H}$. Since our function Φ is vector-valued, we propose an inner scalarization as the following: let us consider a (weak*-)compact convex set $\mathcal{C} \subseteq \mathcal{Y}^*$, which generates the positive polar cone of $\mathcal{K}$, that is,

$$\mathrm{cl}^{w^*} \operatorname{cone} \mathcal{C} = \mathcal{K}^+, \tag{3.2}$$

where cl^{w^*} denotes the weak*-closure. In what follows, we assume that there is a continuously differentiable convex function $h : \mathcal{H} \to \mathbb{R}$ such that for all $v^* \in \mathcal{C}$, the function

$$\mathcal{H} \times \mathbb{R}^m \ni (x, z) \to \Phi^h_{v^*}(x, z) := \langle v^*, \Phi\rangle(x, z) + h(x) \tag{3.3}$$

is convex in both variables, where $\langle v^*, \Phi\rangle(x, z) := \langle v^*, \Phi(x, z)\rangle$.

Example 3.1 (Separated variables in joint chance constrained optimization) *Let us consider the probability function $\varphi(x) = \mathbb{P}(\omega \in \Omega\colon g(x, \xi(\omega)) \leq 0)$, where $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^s$ is the function defined by $g(x, \xi) = \Psi(x) + A\xi$, where A is a matrix and $\Psi\colon \mathbb{R}^n \to \mathbb{R}^s$ a C^2 function. If we set $\Phi(x, z) = \Psi(x) + Az$, $\mathcal{K} := \mathbb{R}^s_+$, and $\mathcal{C}$ is any convex compact set with $\mathrm{cl}\operatorname{cone} \mathcal{C} = \mathbb{R}^s_+$, then Φ satisfies (3.3). Indeed, since $\Psi = (\Psi_1, \ldots, \Psi_s)$ is C^2, there are C^2 convex functions ψ^k_1 and ψ^k_2, for $k = 1, \ldots, s$, such that $\Psi_k = \psi^k_1 - \psi^k_2$ (see, e.g., [23, 32, 41]). Hence, since $\mathcal{C}$ is compact, there exists $C > 0$ such that Φ satisfies (3.3) with $h = C \sum_{k=1}^{s} \psi^k_2$.*

Next, let us introduce the supremum function $S^h_\Phi : \mathcal{H} \times \mathbb{R}^m \to \mathbb{R}$ given by

$$S^h_\Phi(x, z) := \sup \left\{\langle v^*, \Phi\rangle(x, z) + h(x) : v^* \in \mathcal{C}\right\}. \tag{3.4}$$

Moreover, for $h = 0$, we simply write $S_\Phi := S^0_\Phi$.

The next proposition enables us to rewrite the probability function (3.1) in terms of the supremum function (3.4).

Proposition 3.1 *Let $\mathcal{H}$ be a separable Hilbert space, $\xi : \Omega \to \mathbb{R}^m$ be a random vector, $\mathcal{K} \subset \mathcal{Y}$ be a (nonempty) convex cone of a (possibly nonseparable) Banach space and $\Phi : \mathcal{H} \times \mathbb{R}^m \to \mathcal{Y}$ be a function such that (3.3) holds. Then,*

$$\varphi(x) = \mathbb{P}\left(\omega \in \Omega : S^h_\Phi(x, \xi(\omega)) \le h(x)\right) \text{ for all } x \in \mathcal{H}. \tag{3.5}$$

Proof. Fix $x \in \mathcal{H}$ and $\omega \in \Omega$. Then, by the bipolar theorem (see, e.g., [25, Theorem 3.38 p.99]) we have that

$$\begin{aligned}
\Phi(x, \xi(\omega)) \in -\mathcal{K} &\Leftrightarrow -\Phi(x, \xi(\omega)) \in (\mathcal{K}^-)^- \\
&\Leftrightarrow \langle v^*, -\Phi\rangle(x, \xi(\omega)) \le 0, \forall v^* \in \mathcal{K}^- \\
&\Leftrightarrow \langle v^*, \Phi\rangle(x, \xi(\omega)) \le 0, \forall v^* \in \mathcal{K}^+ \\
&\Leftrightarrow \langle v^*, \Phi\rangle(x, \xi(\omega)) \le 0, \forall v^* \in \mathcal{C} \\
&\Leftrightarrow \langle v^*, \Phi\rangle(x, \xi(\omega)) + h(x) \le h(x), \forall v^* \in \mathcal{C} \\
&\Leftrightarrow S^h_\Phi(x, \xi(\omega)) \le h(x),
\end{aligned}$$

where we used the fact that $\mathcal{C}$ generates the positive polar cone of $\mathcal{K}$ (see (3.2)), which proves (3.5). □

The previous formula (3.5) for the probability function (3.1) allows us to propose an inner regularization based on the Moreau envelope. Given $\lambda > 0$, we define the inner regularization of φ as

$$\varphi_\lambda(x) := \mathbb{P}\left(\omega \in \Omega : \mathtt{e}_\lambda \Phi^h(x, \xi(\omega)) \le h(x)\right), \tag{3.6}$$

where $\mathtt{e}_\lambda \Phi^h := \mathtt{e}_\lambda S^h_\Phi$ is the Moreau envelope of the supremum function (3.4). It is worth to emphasize that the Moreau envelope of the supremum function (3.4) is the supremum of Moreau envelopes of the scalarizations (3.3), which is established in the next result.

Proposition 3.2 *Let $\Phi : \mathcal{H} \times \mathbb{R}^m \to \mathcal{Y}$ be a continuous functions satisfying (3.3) for some continuously differentiable convex function h. Then, for all $\lambda > 0$*

$$\mathtt{e}_\lambda \Phi^h(x, z) = \max_{v^* \in \mathcal{C}} \mathtt{e}_\lambda \Phi^h_{v^*}(x, z) \text{ for all } (x, z) \in \mathcal{H} \times \mathbb{R}^m.$$

Proof. By virtue of (3.3), it is clear that the function $(x, z, v^*) \to \Phi^h_{v^*}(x, z)$ is, by assumption, convex with respect to (x, z), and readily seen to be concave with respect to $v^* \in \mathcal{C}$. Moreover, the set $\mathcal{C}$ is (weak*-)compact and the function $v^* \to \Phi_{v^*}(x, z)$ is continuous for fixed (x, z). Thus, the result follows from [45, Theorem 3.1]. □

3.2. Variational convergence of φ_λ

In this section, we show that our inner regularization of the probability function (3.1) inherits similar variational properties from the Moreau envelope (see Proposition 2.6).

Theorem 3.1 *Let $\mathcal{H}$ be a separable Hilbert space, $\xi : \Omega \to \mathbb{R}^m$ be a random vector having density with respect to the Lebesgue measure, $\mathcal{K} \subset \mathcal{Y}$ be a (nonempty) convex cone of a (possible nonseparable) Banach space and $\Phi : \mathcal{H} \times \mathbb{R}^m \to \mathcal{Y}$ be a continuous function such that (3.3) holds. Then, the probability function φ given in (3.1) and the regularization φ_λ given in (3.6) satisfy the following properties:*

a) For all $\lambda_1 > \lambda_2 > 0$, $\varphi_{\lambda_1}(x) \geq \varphi_{\lambda_2}(x)$ and $\inf_{\lambda>0} \varphi_\lambda(x) = \varphi(x)$ for all $x \in \mathcal{H}$.

b) For any sequence $\lambda_k \to 0$ and $x_k \to x$ we have that

$$\limsup_{k\to\infty} \varphi_{\lambda_k}(x_k) \leq \varphi(x). \tag{3.7}$$

Furthermore, if the function h from (3.3) is sequentially weakly continuous on $\mathcal{H}$, then for any sequence $\lambda_k \to 0$ and $x_k \rightharpoonup x$ we have that (3.7) also holds.

c) For any sequence $\lambda_k \to 0$ and any sequence $x_k \to x \in \mathcal{D}$, we have

$$\lim_{k\to\infty} \varphi_{\lambda_k}(x_k) = \varphi(x). \tag{3.8}$$

where $\mathcal{D}$ is the open set $\mathcal{D} := \{x \in \mathcal{H} : \exists z \text{ s.t. } S^h_\Phi(x,z) < h(x)\}$.

d) The functions φ and φ_λ are sequentially weakly upper semicontinuous on $\mathcal{H}$ provided that the function h is sequentially weakly continuous on $\mathcal{H}$.

e) The functions φ and φ_λ are continuous on $\mathcal{D}$.

Proof. To prove a), let $\lambda_1 > \lambda_2 > 0$. Then, for any fixed (x,z) we have the inequalities $\mathsf{e}_{\lambda_1} \Phi^h(x,z) \leq \mathsf{e}_{\lambda_2} \Phi^h(x,z) \leq S^h_\Phi(x,z)$. Hence, $\varphi_{\lambda_1}(x) \geq \varphi_{\lambda_2}(x) \geq \varphi(x)$ and, by virtue of Proposition 2.6, item a), it follows that

$$\lim_{\lambda\searrow 0} \varphi_\lambda(x) = \inf_{\lambda>0} \varphi_\lambda(x) \geq \varphi(x).$$

Let $\lambda_k \to 0$ and fix $x \in \mathcal{H}$. By Proposition 2.6, item a) we have that

$$\lim_{k\to\infty} \mathsf{e}_{\lambda_k} \Phi^h(x,z) = S^h_\Phi(x,z), \text{ for all } z \in \mathbb{R}^m.$$

Thus, for all $z \in \mathbb{R}^m$, $\liminf_{k\to\infty} \mathbb{1}_{\hat{A}_k}(z) \geq \mathbb{1}_{\hat{A}}(z)$, where

$$\hat{A}_k := \{z \in \mathbb{R}^m : \mathsf{e}_{\lambda_k} \Phi^h(x,z) > h(x)\}, \ \hat{A} := \{z \in \mathbb{R}^m : S^h_\Phi(x,z) > h(x)\}.$$

Then, by using the fact that ξ has a density with respect to the Lebesgue measure and applying Fatou's Lemma, we get

$$1-\lim_{k\to\infty}\varphi_{\lambda_k}(x)=\lim_{k\to\infty}\mathbb{P}(\xi^{-1}(\hat{A}_k))=\liminf_{k\to\infty}\mathbb{P}(\xi^{-1}(\hat{A}_k))$$
$$\geq \mathbb{P}(\xi^{-1}(\hat{A}))=1-\varphi(x).$$

Therefore, $\lim_{k\to\infty}\varphi_{\lambda_k}(x)\leq\varphi(x)$, which concludes the proof of a).

To prove b), consider $\lambda_k\to 0$, $x_k\to x$ ($x_k\rightharpoonup x$, respectively) and the sets

$$A_k:=\{z\in\mathbb{R}^m : \mathsf{e}_{\lambda_k}\Phi^h(x_k,z)>h(x_k)\},\ A:=\{z\in\mathbb{R}^m : S_\Phi^h(x,z)>h(x)\}.$$

Now, due to Proposition 2.6, item c) and the continuity of h (sequentially weak continuity of h on $\mathcal{H}$, respectively) we get for any $z\in\mathbb{R}^m$ that

$$\liminf_{k\to\infty}\Big(\mathsf{e}_{\lambda_k}\Phi^h(x_k,z)-h(x_k)\Big)\geq S_\Phi^h(x,z)-h(x),$$

which implies

$$\liminf_{k\to\infty}\mathbb{1}_{A_k}(z)\geq\mathbb{1}_A(z),\ \text{ for all } z\in\mathbb{R}^m.$$

Then using again the fact that ξ has a density and applying Fatou's Lemma we get

$$1-\limsup_{k\to\infty}\varphi_{\lambda_k}(x_k)=\liminf_{k\to\infty}\mathbb{P}(\xi^{-1}(A_k))\geq\mathbb{P}(\xi^{-1}(A))=1-\varphi(x),$$

which proves (3.7).

Now, let us show c). Assume that $x_k\to x$, so by Proposition 2.6, item d) and the continuity of h we have that

$$\lim_{k\to\infty}(\mathsf{e}_{\lambda_k}\Phi^h(x_k,z)-h(x_k))=S_\Phi^h(x,z)-h(x) \text{ for all } z\in\mathbb{R}^m. \tag{3.9}$$

Hence, by using the sets A_k and A defined above and by similar arguments as before, we obtain

$$\limsup_{k\to\infty}\varphi_{\lambda_k}(x_k)\leq\varphi(x). \tag{3.10}$$

On the other hand, we consider the sets

$$B_k:=\{z\in\mathbb{R}^m : \mathsf{e}_{\lambda_k}\Phi(x_k,z)<h(x_k)\},\ B:=\{z\in\mathbb{R}^m : S_\Phi^h(x,z)<h(x)\}.$$

Then, mimicking the last proof, we obtain that

$$\liminf_{k\to\infty}\mathbb{P}(\xi^{-1}(B_k))\geq\mathbb{P}(\xi^{-1}(B)). \tag{3.11}$$

Since $x\in\mathcal{D}$ (and recalling that ξ has density), we have that $\mathbb{P}(S_\Phi^h(x,\xi)=h(x))=0$. Hence, by using (3.10) and (3.11), it follows that

$$\limsup_{k\to\infty}\varphi_{\lambda_k}(x_k)\leq\varphi(x)=\mathbb{P}(\xi^{-1}(B))\leq\liminf_{k\to\infty}\mathbb{P}(\xi^{-1}(B_k))\leq\liminf_{k\to\infty}\varphi(x_k),$$

which completes the proof of (3.8).

To prove d), we consider $x_n \rightharpoonup x$ and the sets

$$C_n := \{z \in \mathbb{R}^m : S_\Phi^h(x_n, z) > h(x_n)\},\ C := \{z \in \mathbb{R}^m : S_\Phi^h(x, z) > h(x)\}.$$

From the weak lower semicontinuity of S_Φ^h and the sequentially weak continuity of h,

$$\liminf_{n\to\infty} \left(S_\Phi^h(x_n, z) - h(x_n)\right) \geq S_\Phi^h(x, z) - h(x).$$

Hence, following an analogous argumentation, we can conclude that

$$\liminf_{n\to\infty} \mathbb{P}(\xi^{-1}(C_n)) \geq \mathbb{P}(\xi^{-1}(C)).$$

Thus, applying Fatou's Lemma, we get

$$1 - \limsup_{n\to\infty} \varphi(x_n) = \liminf_{n\to\infty} \mathbb{P}(\xi^{-1}(C_n)) \geq \mathbb{P}(\xi^{-1}(C)) = 1 - \varphi(x).$$

Therefore, $\limsup_{n\to\infty} \varphi(x_n) \leq \varphi(x)$. Now for a fixed $\lambda > 0$, the upper semicontinuity of φ_λ follows from similar arguments as before but upon considering the sets

$$\hat{C}_n := \{z \in \mathbb{R}^m : \mathsf{e}_\lambda\, \Phi^h(x_n, z) > h(x_n)\},\ \hat{C} := \{z \in \mathbb{R}^m : \mathsf{e}_\lambda\, \Phi^h(x, z) > h(x)\},$$

Finally, let us prove e). Assume that $x_n \to x$. By the continuity of S_Φ^h and continuity of h, we have (3.9) holds. Thus, by using, once again, similar arguments but with the sets

$$D_n := \{z \in \mathbb{R}^m : S_\Phi^h(x_n, z) < h(x_n)\},\ D := \{z \in \mathbb{R}^m : S_\Phi^h(x, z) < h(x)\},$$

we get that

$$\liminf_{n\to\infty} \mathbb{P}(\xi^{-1}(D_n)) \geq \mathbb{P}(\xi^{-1}(D)). \tag{3.12}$$

Now, since $x \in \mathcal{D}$ we have that $\mathbb{P}(S_\Phi^h(x, \xi) = h(x)) = 0$. Hence, by using part d) and (3.12), we have

$$\limsup_{n\to\infty} \varphi(x_n) \leq \varphi(x) = \mathbb{P}(\xi^{-1}(D)) \leq \liminf_{n\to\infty} \mathbb{P}(\xi^{-1}(D_n)) \leq \liminf_{n\to\infty} \varphi(x_n),$$

which yields the continuity of φ. The continuity of φ_λ follows from similar arguments. □

Remark 3.1 (Slater condition for S_Φ^h) *It is worth mentioning that, in order to have the existence of a point (x, z) such that $S_\Phi^h(x, z) < h(x)$, the set $\mathcal{C}$ cannot contain the zero vector. Indeed, if $0 \in \mathcal{C}$, then from (3.4) it follows that $S_\Phi^h(x, z) \geq h(x)$, for all $(x, z) \in \mathcal{H} \times \mathbb{R}^m$. On the other hand, if $\mathcal{C}$ is such that $\inf\{\|v^*\| : v^* \in \mathcal{C}\} > 0$ and (x, z) satisfy $\Phi(x, z) \in \operatorname{int}(-\mathcal{K})$, then $\langle v^*, \Phi(x, z)\rangle < -\eta\|v^*\|$ holds for all $v^* \in \mathcal{C}$ with some $\eta > 0$, and, thus, $S_\Phi^h(x, z) < h(x)$.*

Now, we formally describe the convergence properties of the family φ_λ to the function φ in terms of hypo-convergence.

Corollary 3.1 *Under the assumptions of Theorem 3.1, the sequence of regularizations φ_λ*

hypo-converges to the probability function φ. In addition, suppose that the function h in (3.3) is weakly continuous, then the sequence of regularizations φ_λ Mosco hypo-converges to the probability function φ.

Proof. The constant sequence $x_k = x$ together with the pointwise convergence in Theorem 3.1 Item a) gives us the existence of a sequence $x_k \to x$ such that

$$\liminf_{k\to\infty} \varphi_{\lambda_k}(x_k) \geq \varphi(x).$$

The remaining second condition to obtain hypo-convergence follows from (3.10) obtained in the proof of Theorem 3.1 Item c). If we suppose that the function h in (3.3) is weakly continuous, then the remaining second condition to obtain Mosco hypo-convergence is given by Theorem 3.1 Item b). $\square$

3.3. Differentiability and gradient formula for φ_λ

In this section, we assume that $\mathcal{H}$ is finite-dimensional. Here we apply the results of subsection 2.2.1 to give a formula for the gradients of our inner regularization of the probability function (3.1), and later we provide the consistency of the gradients of our inner regularization.

First, we provide the following lemma, which shows that the gradients of $\mathtt{e}_\lambda\,\Phi^h - h$ satisfies a growth condition.

Lemma 3.1 *Let $\lambda > 0$ be given but fixed. Let $\bar{x}$ be a point such that $S^h_\Phi(\bar{x}, 0) < h(\bar{x})$. Then, there exists $C_\lambda, \varepsilon > 0$ such that*

$$\|\nabla_x \mathtt{e}_\lambda\,\Phi^h(x,z) - \nabla h(x)\| \leq C_\lambda(\|z\| + 1), \quad \textit{for all } x \in \mathbb{B}_\varepsilon(\bar{x}) \textit{ and all } \; z \in \mathbb{R}^m. \tag{3.13}$$

Proof. We have, by (2.5) and the triangle inequality, that

$$\|\nabla_x \mathtt{e}_\lambda\,\Phi^h(x,z) - \nabla h(x)\| \leq \frac{1}{\lambda}\left(\|x\| + \|\mathtt{Prox}_{\lambda S^h_\Phi}(x,z)\|\right) + \|\nabla h(x)\|,$$

for all $x \in \mathcal{H}$ and all $z \in \mathbb{R}^m$. By the nonexpansiveness of the proximal mapping we get

$$\|\nabla_x \mathtt{e}_\lambda\,\Phi^h(x,z) - \nabla h(x)\| \leq \frac{1}{\lambda}\left(2\|x\| + \|z\| + \|\mathtt{Prox}_{\lambda S^h_\Phi}(0,0)\|\right) + \|\nabla h(x)\|,$$

for all $x \in \mathcal{H}$ and all $z \in \mathbb{R}^m$. Since ∇h is locally bounded at $\bar{x}$ (h is continuously differentiable), there exists $\varepsilon > 0$ and $M > 0$ such that

$$\|\nabla_x \mathtt{e}_\lambda\,\Phi^h(x,z) - \nabla h(x)\| \leq \frac{1}{\lambda}\left(2\varepsilon + 2\|\bar{x}\| + \|z\| + \|\mathtt{Prox}_{\lambda S^h_\Phi}(0,0)\|\right) + M,$$

for all $x \in \mathbb{B}_\varepsilon(\bar{x})$ and all $z \in \mathbb{R}^m$. By defining

$$C_\lambda := \max\{\frac{1}{\lambda}\left(2\varepsilon + 2\|\bar{x}\| + \|\mathtt{Prox}_{\lambda S^h_\Phi}(0,0)\|\right) + M, \frac{1}{\lambda}\},$$

we conclude the proof. $\square$

In order to apply the gradient formula given in Corollary 2.1 it will be convenient to introduce the following notation. Given a parameter $\lambda > 0$, let us assume x belonging to an appropriate neighbourhood of $\bar{x}$ such that $\mathtt{e}_\lambda \, \Phi^h(x, 0) < h(x)$. Then, we define the set of finite and infinite directions for the function $\mathtt{e}_\lambda \, \Phi^h$ by

$$F_\lambda(x) := \{v \in \mathbb{S}^{m-1} : \exists r > 0 : \mathtt{e}_\lambda \, \Phi^h(x, rLv) = h(x)\}, \quad I_\lambda(x) := \mathbb{S}^{m-1} \backslash F_\lambda(x), \tag{3.14}$$

respectively, and its associated radial function is given by

$$\rho_\lambda(x, v) := \sup \left\{ r > 0 : \mathtt{e}_\lambda \, \Phi^h(x, rLv) \leq h(x) \right\} \tag{3.15}$$

Remark 3.2 (Characterization of radial function) *It is important to recall that when we have* $\mathtt{e}_\lambda \, \Phi^h(x, 0) < h(x)$*, then the radial function* $\rho_\lambda(x, v)$ *can be characterized as the*

$$\rho_\lambda(x, v) = \inf\{r > 0 : \mathtt{e}_\lambda \, \Phi^h(x, rLv) > h(x)\},$$

with the convention $\inf \emptyset = +\infty$*. Furthermore, it also can be characterized as the unique solution of the equation*

$$\mathtt{e}_\lambda \, \Phi^h(x, rLv) = h(x)$$

for any finite direction $v \in F_\lambda(x)$*. We refer to [56, Proposition 2.6] for more details of the proof, which uses essentially the convexity and continuity. Nevertheless, it is clear that the continuity of the convex function is necessary, as was illustrated in [56, Example 2.7].*

Finally, let us introduce the gradient-like mapping $\mathcal{G}_\lambda : \mathcal{H} \times \mathbb{S}^{m-1} \to \mathcal{H}$ defined as

$$\mathcal{G}_\lambda(x, v) := \begin{cases} -\theta(\rho_\lambda(x, v), v) \left(\frac{\nabla_x \mathtt{e}_\lambda \, \Phi^h(x, \rho_\lambda(x,v)Lv) - \nabla h(x)}{\left\langle \nabla_z \mathtt{e}_\lambda \, \Phi^h(x, \rho_\lambda(x,v)Lv, Lv \right\rangle} \right) & \text{if } v \in F_\lambda(x) \\ 0 & \text{if } v \in I_\lambda(x) \end{cases} \tag{3.16}$$

and the factor θ is defined in (2.14). Using the above notation we are able to provide a gradient formula for the probability function φ_λ

Theorem 3.2 *Let* $\bar{x} \in \mathcal{H}$ *be such that* $S^h_\Phi(\bar{x}, 0) < h(\bar{x})$*, and assume that* f_ξ *satisfies the following growth condition*

$$\lim_{\|z\| \to +\infty} \|z\|^{m+1} f_\xi(z) = 0. \tag{3.17}$$

Then, for any given $\lambda > 0$*, the probability function* φ_λ*, defined in (3.6), is continuously differentiable on an appropriate neighbourhood* U *of* $\bar{x}$ *and it holds:*

$$\nabla \varphi_\lambda(x) = \int_{\mathbb{S}^{m-1}} \mathcal{G}_\lambda(x, v) d\mu_\zeta(v), \ \ \textit{for all } x \in U,$$

where $\mathcal{G}_\lambda$ is as in (3.16). Moreover, the gradients of $\mathtt{e}_\lambda \Phi^h$ can be computed by the formula

$$\nabla \mathtt{e}_\lambda \Phi^h(x,z) = \frac{(x,z) - \mathtt{Prox}_{\lambda(\langle v^*, \Phi\rangle + h)}(x,z)}{\lambda}, \tag{3.18}$$

where v^* is any active vector at (x,z), that is, $v^* \in \mathcal{C}$ and $\mathtt{e}_\lambda \Phi^h_{v^*}(x,z) = \mathtt{e}_\lambda \Phi^h(x,z)$ in view of Proposition 3.2.

Proof. Let $\lambda > 0$ be given but fixed. Due to Proposition 2.6, $\mathtt{e}_\lambda \Phi^h \leq S^h_\Phi$. Thus, $\mathtt{e}_\lambda \Phi^h(\bar{x}, 0) < h(\bar{x})$. Therefore, due to continuity, we can set aside an appropriate neighbourhood U of $\bar{x}$ on which this continues to hold and on which the objects in equations (3.14), (3.15) and (3.16) are well defined. This neighborhood can be taken independently of $\lambda > 0$.

Thus to prove the first part, by Corollary 2.1, it is enough to prove the η_θ-growth condition. To this end let us pick an arbitrary $\bar{v} \in I_\lambda(\bar{x})$. In view of (3.13), we define

$$\eta_\theta(r, v) := C_\lambda \left(r \|Lv\| + 1 \right),$$

choose $l \geq 1/\varepsilon$ and notice, by (2.14), that

$$\begin{aligned} r\theta(r,v)\eta_\theta(r,v) &= \frac{2\pi^{\frac{m}{2}} |\det(L)|}{\Gamma(\frac{m}{2})} C_\lambda (r^{m+1} \|Lv\| f_\xi(rLv) + r^m f_\xi(rLv)) \\ &= \frac{2\pi^{\frac{m}{2}} |\det(L)|}{\Gamma(\frac{m}{2})} C_\lambda \|rLv\|^{m+1} f_\xi(rLv) \left(\frac{1}{\|Lv\|^m} + \frac{1}{r\|Lv\|^{m+1}} \right) \xrightarrow[\substack{r \to +\infty \\ v \to \bar{v}}]{} 0, \end{aligned}$$

where the last limit follows from assumption (3.17). Therefore, as a result of (2.30), the η_θ-growth condition is satisfied. The computation for the gradient (3.18) follows from [45, Theorem 3.5]. ☐

The so-called radial function ρ_λ is used in the last gradient formula. The following proposition shows that this mapping is continuous on the three parameters (λ, x, v), which is a key property for numerical computations, and provides the asymptotic behavior of the gradients of the probability function φ_λ to the (sub-)gradients to the nominal function φ.

Proposition 3.3 *Let us consider the radial function in (3.15) and the open set defined by $U := \{x \in \mathcal{H} : S^h_\Phi(x, 0) < h(x)\}$. Then, for every sequence $(\lambda_k, x_k, v_k) \to (\lambda, x, v)$ with $(\lambda, x, v) \in [0, +\infty) \times U \times \mathbb{S}^{m-1}$ we have that $\rho_{\lambda_k}(x_k, v_k) \to \rho_\lambda(x, v)$, where ρ_0 is defined by $\rho_0(x, v) := \sup\{r > 0 : S_\Phi(x, rLv) \leq h(x)\}$.*

Proof. Let us focus on the case $(\lambda_k, x_k, v_k) \to (0, x, v)$ since the proof for $\lambda > 0$ is analogous. Let us first assume that the sequence $\{\rho_{\lambda_k}(x_k, v_k)\}$ admits a cluster point called r. Then for some subsequence of $\{\rho_{\lambda_k}(x_k, v_k)\}$ we have $\rho_{\lambda_k}(x_{k_l}, v_{k_l}) \to_l r$. By Proposition 2.6 Item d), continuity of h and again by the characterization of the radial function as the unique solution (see Remark 3.2) we have that

$$0 = \mathtt{e}_{\lambda_{k_l}} \Phi^h(x_{k_l}, \rho_{\lambda_k}(x_{k_l}, v_{k_l}) L v_{k_l}) - h(x_{k_l}) \to_l S^h_\Phi(x, rLv) - h(x),$$

then $r = \rho_0(x, v)$. Since this holds true for all possible cluster points, we have in fact that $\rho_{\lambda_k}(x_k, v_k)$ converges to $\rho(x, v)$, whenever the sequence $(\rho_{\lambda_k}(x_k, v_k))_{k \in \mathbb{N}}$ has a cluster point.

Next let us assume that, $\rho_{\lambda_k}(x_k, v_k) \to +\infty$, and by contradiction suppose that $r := \rho(x, v) < +\infty$.

Then, by Proposition 2.6 Item d), we have that for all large enough k

$$0 < \mathsf{e}_{\lambda_k} \Phi^h(x_k, (r+1)Lv_k) - h(x_k),$$

which implies that $\rho_{\lambda_k}(x_k, v_k) < (r+1)$ for all large enough k (see Remark 3.2), which contradicts our assumption, and concludes the proof. □

The next proposition shows that the radial function ρ_λ, given in (3.15), can be computed using the associated radial function to the function $\mathsf{e}_\lambda \Phi^h_{v^*}$, defined in (3.3), that is, for a given $v^* \in \mathcal{C}$, and $\lambda > 0$ we set

$$\rho_\lambda^{v^*}(x, v) := \sup\{r > 0 : \mathsf{e}_\lambda \Phi^h_{v^*}(x, rLv) \leq h(x)\}.$$

Proposition 3.4 *In the setting of Proposition 3.3, we have that*

$$\rho_\lambda(x, v) := \min\{\rho_\lambda^{v^*}(x, v) : v^* \in \mathcal{C}\}.$$

Proof. The proof follows the same lines of arguments that [56, Proposition 2.6], which only uses the supremum structure of the function. □

Now, we focus on well-posedness of the gradient approximation, that is, the study of convergence properties of the gradients of the regularized probability functions φ_λ. The following proposition provides a (sub-)differential variational principle for the (not necessarily smooth) probability function φ using the inner regularized functions φ_λ.

Proposition 3.5 *Under the assumption of Theorem 3.2 we have that for every $x^* \in \partial^{\mathsf{r}}\varphi(\bar{x})$ and every $\varepsilon > 0$ there exists $\lambda > 0$, $x_\lambda \in \mathcal{H}$ such that*

$$\|\bar{x} - x_\lambda\| + \|x^* - \nabla\varphi_\lambda(x_\lambda)\| + |\varphi(\bar{x}) - \varphi(x_\lambda)| \leq \varepsilon.$$

Particularly, we have that $\partial^{\mathsf{b}}\varphi(\bar{x}) \subseteq \limsup_{x\to\bar{x},\, \lambda\to 0^+}\{\nabla\varphi_\lambda(x)\}$.

Proof. The first part follows from a direct application of [56, Lemma 2.1]. For the second part, consider a point $x^* \in \partial^{\mathsf{b}}\varphi(\bar{x})$, by definition there are $x_k^* \in \partial^{\mathsf{r}}\varphi(x_k)$ with $x_k \to \bar{x}$, $\varphi(x_k) \to \varphi(\bar{x})$ and $x_k^* \to x^*$. By the last part applied to x_k (for large enough k) we have that each x_k^* can be approximated by gradients of the probability functions φ_λ, which by a classical diagonal argument shows the desire inclusion. □

The last result shows that the basic subdifferential of the probability function φ can be upper-estimated by using the gradients of the probability function φ_λ. In the rest of this subsection, we will focus on providing the opposite inclusion, that is to say, the accumulations points of gradients are points in the basic subdifferential.

Lemma 3.2 *Let us suppose the mapping S^h_Φ defined in (3.4) is bounded from below by an affine linear function $\bar{h} : \mathcal{H} \to \mathbb{R}$, let $\bar{x} \in \mathcal{H}$ such that $S^h_\Phi(\bar{x}, 0) < h(\bar{x})$. Given $\varepsilon > 0$, there exists $\lambda_0, \varepsilon_0 > 0$ such that for all $(\lambda, x, v) \in (0, \lambda_0) \times \mathbb{B}_{\varepsilon_0}(\bar{x}) \times \mathbb{S}^{m-1}$ with $v \in F_\lambda(x)$*

$$\|x - \hat{x}\| + \|\rho_\lambda(x, v)Lv - \hat{z}\| \leq \varepsilon, \tag{3.19}$$

where $(\hat{x}, \hat{z}) := \mathtt{Prox}_{\lambda S^h_\Phi}(x, \rho_\lambda(x, v)Lv)$.

Proof. Let $\varepsilon_0 \in (0, \varepsilon)$ such that $\mathbb{B}_{\varepsilon_0}(\bar{x}) \subset U$, and pick (λ, x, v) with $x \in \mathbb{B}_{\varepsilon_0}(\bar{x})$, $\lambda \in (0,1)$ and $v \in F_\lambda(x)$. We first notice that

$$S^h_\Phi(\hat{x}, \hat{z}) + \tfrac{1}{2\lambda}\|\hat{x} - x\|^2 + \tfrac{1}{2\lambda}\|\hat{z} - z\|^2 = \mathtt{e}_\lambda\, \Phi^h(x, z) = h(x)$$

where $z := \rho_\lambda(x, v)Lv$ and the last equality follows from the definition of the latter term. Now, let us suppose that $\bar{h} = \langle x^*, \cdot\rangle + \beta$, so

$$\langle x^*, \hat{x}\rangle + \beta + \tfrac{1}{2\lambda}\|\hat{x} - x\|^2 + \tfrac{1}{2\lambda}\|\hat{z} - z\|^2 \le h(x). \tag{3.20}$$

On the other hand, the inequality

$$|\langle x^*, \hat{x} - x\rangle| \le \tfrac{1}{2}\|x^*\|^2 + \tfrac{1}{2}\|\hat{x} - x\|^2$$

implies

$$|\langle x^*, \hat{x}\rangle| \le \tfrac{1}{2}\|x^*\|^2 + \tfrac{1}{2}\|\hat{x} - x\|^2 + \tfrac{1}{2}\|x^*\|^2 + \tfrac{1}{2}\|x\|^2. \tag{3.21}$$

From (3.20) and (3.21) we have that

$$\left(\tfrac{1}{2\lambda} - \tfrac{1}{2}\right)\|\hat{x} - x\|^2 + \tfrac{1}{2\lambda}\|\hat{z} - z\|^2 \le h(x) - \beta + \|x^*\|^2 + \tfrac{1}{2}\|x\|^2.$$

Since $\frac{1}{1-\lambda} > 1$, and due to continuity of h, a constant $M > 0$ such that

$$\|\hat{x} - x\|^2 + \|\hat{z} - z\|^2 \le \left(\frac{2\lambda}{1 - \lambda}\right) M.$$

Now, considering $\lambda_0 > 0$ small enough, we can conclude that

$$\|\hat{x} - x\|^2 + \|\hat{z} - z\|^2 \le \varepsilon^2,$$

for all $\lambda \in (0, \lambda_0)$ and $x \in \mathbb{B}_{\varepsilon_0}(\bar{x})$, which shows that (3.19) holds. ☐

In the following lemma we will require that the mapping S^h_Φ defined in (3.4) satisfies the following growth condition at $\bar{x}$: there exist constants $\varepsilon, \ell > 0$ such that

$$\|\partial_x S^h_\Phi(x, z)\| \le \eta(\|z\|),\ \forall x \in \mathbb{B}_\varepsilon(\bar{x}),\ \forall \|z\| \ge \ell, \tag{3.22}$$

for some nondecreasing function η satisfying

$$\lim_{\|z\|\to\infty} \|z\|^m f_\xi(z)\eta(\|z\| + \alpha) = 0$$

for some $\alpha > 0$. Here, the norm of a sub-differential set is defined as follows:

$$\|\partial_x S^h_\Phi(x, z)\| := \sup\{\|x^*\| : x^* \in \partial_x S^h_\Phi(x, z)\}$$

Lemma 3.3 *Let us suppose the mapping S^h_Φ defined in (3.4) is bounded from below by an affine linear function $\bar{h} : \mathcal{H} \to \mathbb{R}$ and satisfies the growth condition (3.22) at $\bar{x}$, where*

$S_\Phi^h(\bar{x},0) < h(\bar{x})$. *Then there exists* $\gamma > 0$ *and* $\kappa > 0$ *such that*

$$\|\mathcal{G}_\lambda(x,v)\| \leq \kappa, \quad \forall(\lambda,x,v) \in (0,\gamma) \times \mathbb{B}_\gamma(\bar{x}) \times \mathbb{S}^{m-1} \tag{3.23}$$

where $\mathcal{G}_\lambda$ *is defined in* (3.16). *Moreover, for all* $v \in I(\bar{x})$ *and all* $\varepsilon > 0$ *there exists* $\gamma > 0$ *such that*

$$\|\mathcal{G}_\lambda(x,v)\| \leq \varepsilon, \quad \forall(\lambda,x,v) \in (0,\gamma) \times \mathbb{B}_\gamma(\bar{x}) \times \mathbb{B}_\gamma(\bar{v}). \tag{3.24}$$

Proof. First, let us show that for every $\bar{v} \in \mathbb{S}^{m-1}$ there exist $\varepsilon_{\bar{v}} > 0$ and $M_{\bar{v}} > 0$ such that

$$\|\mathcal{G}_\lambda(x,v)\| \leq M_{\bar{v}}, \text{ for all } (\lambda,x,v) \in (0,\varepsilon_{\bar{v}}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{x}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{v}).$$

First we notice that there exist $\varepsilon_1 > 0$ and $\beta_1 > 0$ such that we have $S_\Phi^h(x,0) - h(x) \leq -\beta_1$ for all $x \in \mathbb{B}_{\varepsilon_1}(\bar{x})$. Then, for all $v \in F_\lambda(x)$, (see, e.g., [54, Lemma 2.1 item 2])

$$\begin{aligned}
&\tfrac{-\rho_\lambda(x,v)}{2}\langle \nabla_z \mathsf{e}_\lambda \Phi^h(x,\rho_\lambda(x,v)Lv), Lv\rangle \\
&\quad = \langle \nabla_z \mathsf{e}_\lambda \Phi^h(x,\rho_\lambda(x,v)Lv), \tfrac{\rho_\lambda(x,v)}{2}Lv - \rho_\lambda(x,v)Lv\rangle \\
&\quad \leq \mathsf{e}_\lambda \Phi^h(x, \tfrac{\rho_\lambda(x,v)}{2}Lv) - \mathsf{e}_\lambda \Phi^h(x,\rho_\lambda(x,v)Lv) \\
&\quad = \mathsf{e}_\lambda \Phi^h(x, \tfrac{\rho_\lambda(x,v)}{2}Lv) - h(x) \\
&\quad \leq \tfrac{1}{2}\mathsf{e}_\lambda \Phi^h(x,0) + \tfrac{1}{2}\mathsf{e}_\lambda \Phi^h(x,\rho_\lambda(x,v)Lv) - h(x) \\
&\quad = \tfrac{1}{2}\mathsf{e}_\lambda \Phi^h(x,0) - \tfrac{1}{2}h(x) \\
&\quad \leq \tfrac{1}{2}S_\Phi^h(x,0) - \tfrac{1}{2}h(x) \\
&\quad \leq -\tfrac{\beta_1}{2}.
\end{aligned}$$

Since ∇h is locally bounded we have

$$\begin{aligned}
\|\mathcal{G}_\lambda(x,v)\| \leq &\tfrac{2\pi^{m/2}\det(L)}{\Gamma(m/2)\beta_1}\rho_\lambda(x,v)^m f_\xi(\rho_\lambda(x,v)Lv) \\
&\times \left(\|\nabla_x \mathsf{e}_\lambda \Phi^h(x,\rho_\lambda(x,v)Lv)\| + \beta_2\right)
\end{aligned} \tag{3.25}$$

for some $\beta_2 > 0$ and for all $x \in \mathbb{B}_{\varepsilon_1}(\bar{x})$ and $v \in F_\lambda(x)$.

Now let $\bar{v} \in \mathbb{S}^{m-1}$ be fixed. If $\bar{v} \notin I(\bar{x})$ then there exist $\varepsilon_{\bar{v}} > 0$ such that $v \notin I_\lambda(x)$ for all $(\lambda,x,v) \in (0,\varepsilon_{\bar{v}}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{x}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{v})$ where $I_\lambda(x)$ was defined in (3.14). Indeed, if it is not true, then there exists a sequence $(\lambda_k,x_k,v_k) \to (0,\bar{x},\bar{v})$ with $v_k \in I_{\lambda_k}(x_k)$. Hence, $\rho_{\lambda_k}(x_k,v_k) = \infty$ and so $\rho_0(\bar{x},\bar{v}) = \infty$ by Proposition 3.3. This yields a contradiction with $\bar{v} \notin I(\bar{x})$.

Since S_Φ^h is continuous at $(\bar{x},\bar{z})$, where $\bar{z} := \rho_0(\bar{x},\bar{v})L\bar{v}$, there exist $\varepsilon_2 > 0$ and $\beta_3 > 0$ such that for all $(x,z) \in \mathbb{B}_{\varepsilon_2}(\bar{x},\bar{z})$

$$\|(u^*,v^*)\| \leq \beta_3, \text{ for all } (u^*,v^*) \in \partial S_\Phi^h(x,z). \tag{3.26}$$

Now, by Proposition 3.3 and Lemma 3.2, and considering $\varepsilon_{\bar{v}}$ small enough, we get that

$$\|(\bar{x},\rho_0(\bar{x},\bar{v})L\bar{v}) - \mathtt{Prox}_{\lambda S_\Phi^h}(\bar{x},\rho_0(\bar{x},\bar{v})L\bar{v})\| \leq \tfrac{\varepsilon_2}{2}$$

and

$$\|(\bar{x},\rho_0(\bar{x},\bar{v})L\bar{v}) - (x,\rho_\lambda(x,v)Lv)\| \leq \tfrac{\varepsilon_2}{2}$$

for all $(\lambda, x, v) \in (0, \varepsilon_{\bar{v}}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{x}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{v})$. Thus, since the proximal mapping is nonexpansive we have that

$$\begin{aligned}\|(\bar{x}, \rho_0(\bar{x}, \bar{v})L\bar{v}) - &\mathtt{Prox}_{\lambda S_\Phi^h}(x, \rho_\lambda(x, v)Lv)\| \\ &\leq \|(\bar{x}, \rho_0(\bar{x}, \bar{v})L\bar{v}) - \mathtt{Prox}_{\lambda S_\Phi^h}(\bar{x}, \rho_0(\bar{x}, \bar{v})L\bar{v})\| \\ &\quad + \|(\bar{x}, \rho_0(\bar{x}, \bar{v})L\bar{v}) - (x, \rho_\lambda(x, v)Lv)\| \leq \varepsilon_2 \end{aligned} \tag{3.27}$$

for all $(\lambda, x, v) \in (0, \varepsilon_{\bar{v}}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{x}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{v})$. Now, by (3.26), (3.27) and since due to (2.6)

$$\nabla \mathtt{e}_\lambda \Phi^h(x, \rho_\lambda(x, v)Lv) \in \partial S_\Phi^h(\mathtt{Prox}_{\lambda S_\Phi^h}(x, \rho_\lambda(x, v)Lv))$$

we have that

$$\|\nabla_x \mathtt{e}_\lambda \Phi^h(x, \rho_\lambda(x, v)Lv)\| \leq \|\nabla \mathtt{e}_\lambda \Phi^h(x, \rho_\lambda(x, v)Lv)\| \leq \beta_3 \tag{3.28}$$

for all $(\lambda, x, v) \in (0, \varepsilon_{\bar{v}}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{x}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{v})$. By (3.28), Proposition 3.3, (3.25) and considering $\varepsilon_{\bar{v}} < \varepsilon_1$ smaller if needed, we get that

$$\|\mathcal{G}_\lambda(x, v)\| \leq M_1$$

for all $(\lambda, x, v) \in (0, \varepsilon_{\bar{v}}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{x}) \times \mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{v})$.

Now, let us assume $\bar{v} \in I(\bar{x})$ and consider $\gamma > 0$. By the growth condition, we have that there exists $\ell, \varepsilon > 0$ such that

$$\rho_\lambda(x, v)^m f_\xi(\rho_\lambda(x, v)Lv)\eta(\rho_\lambda(x, v)\|Lv\| + \alpha) \leq \gamma \tag{3.29}$$

whenever $\rho_\lambda(x, v)\|Lv\| \geq \ell$, and

$$\|\partial_x S_\Phi^h(x, z)\| \leq \eta(\|z\|), \ \forall x \in \mathbb{B}_\varepsilon(\bar{x}), \ \forall \|z\| \geq \ell; \tag{3.30}$$

Now, by Lemma 3.2, we can consider $\varepsilon_0, \lambda_0 > 0$ such that $\hat{x} \in \mathbb{B}_\varepsilon(\bar{x})$ and

$$\rho_\lambda(x, v)\|Lv\| + \alpha \geq \|\hat{z}\| \geq \rho_\lambda(x, v)\|Lv\| - \alpha, \tag{3.31}$$

for all $(\lambda, x, v) \in (0, \lambda_0) \times \mathbb{B}_{\varepsilon_0}(\bar{x}) \times \mathbb{S}^{m-1}$ with $v \in F_\lambda(x)$, where $(\hat{x}, \hat{z}) := \mathtt{Prox}_{\lambda S_\Phi^h}(x, \rho_\lambda(x, v)Lv)$. Moreover, using Proposition 3.3, when considering a small enough $\varepsilon_3 \in (0, \min\{\varepsilon_0, \lambda_0\})$ it follows that:

$$\rho_\lambda(x, v) \geq \frac{\ell + \alpha}{\|Lv\|}, \text{ for all } (\lambda, x, v) \in (0, \varepsilon_3) \times \mathbb{B}_{\varepsilon_3}(\bar{x}) \times \mathbb{B}_{\varepsilon_3}(\bar{v}). \tag{3.32}$$

Now, mixing equations (3.30), (3.31) and (3.32), we conclude that for all

$$(\lambda, x, v) \in (0, \varepsilon_3) \times \mathbb{B}_{\varepsilon_3}(\bar{x}) \times \mathbb{B}_{\varepsilon_3}(\bar{v}),$$

we have

$$\|\nabla_x \mathtt{e}_\lambda \Phi^h(x, \rho_\lambda(x, v)Lv)\| \leq \eta(\|\hat{z}\|) \leq \eta(\rho_\lambda(x, v)\|Lv\| + \alpha),$$

where we have used the fact that η is non-decreasing and $\nabla \mathtt{e}_\lambda \Phi^h(x, \rho_\lambda(x, v)Lv)$ belongs to the set $\partial S_\Phi^h(\mathtt{Prox}_{\lambda S_\Phi^h}(x, \rho_\lambda(x, v)Lv))$. Then, replacing this into (3.25), and using (3.29), we get that

$$\|\mathcal{G}_\lambda(x,v)\| \le \gamma, \text{ for all } (\lambda,x,v) \in (0,\varepsilon_3)\times \mathbb{B}_{\varepsilon_3}(\bar{x})\times \mathbb{B}_{\varepsilon_3}(\bar{v}). \tag{3.33}$$

Since $\mathbb{S}^{m-1}$ is compact and the family of neighborhoods $\mathbb{B}_{\varepsilon_{\bar{v}}}(\bar{v})$ covers $\mathbb{S}^{m-1}$, we can pick a finite subcover, that is, there exists $N \in \mathbb{N}$ and some $v_1,\dots,v_N \in \mathbb{S}^{m-1}$ such that

$$\mathbb{S}^{m-1} \subset \bigcup_{i=1}^{N} \mathbb{B}_{\varepsilon_{v_i}}(v_i).$$

Therefore, we choose $\gamma > 0$ such that

$$(0,\gamma) \subset \min\{\varepsilon_{v_i} : \ i = 1\dots,N\} \ \text{ and } \ \mathbb{B}_\gamma(\bar{x}) \subset \bigcap_{i=1}^{N} \mathbb{B}_{\varepsilon_{v_i}}(\bar{x})$$

and define $\kappa := \max\{M_{v_i} : \ i = 1\dots,N\}$ to conclude the proof of (3.23). Finally, the proof of (3.24) follows from the more precise estimation (3.33). □

Theorem 3.3 (Gradient Consistency) *Let us suppose the mapping S^h_Φ defined in (3.4) satisfies the η_θ-growth condition at $(\bar{x},\bar{v})$ for all $\bar{v} \in I(x)$, and assume that*

$$\partial S^h_\Phi(\bar{x},\rho(\bar{x},v)Lv) \textit{ is single valued for almost all } v \in \mathbb{S}^{m-1}.$$

Then, the probability function φ, given in (3.1) is Locally Lipschitzian at $\bar{x}$ and in fact Fréchet différentiable at $\bar{x}$. Moreover, any accumulation point of sequences $\{\nabla\varphi_{\lambda_k}(x_k)\}_{k\ge 0}$ with $\lambda_k \to 0^+$ and $x_k \to \bar{x}$ are equal to $\nabla\varphi(\bar{x})$.

Proof. First, let us notice that by Proposition 3.5 we have that for all x close enough to $\bar{x}$

$$\partial^{\mathsf{b}}\varphi(x) \subset \limsup_{\lambda\to 0^+, x'\to x} \nabla\varphi_\lambda(x').$$

Now, by Lemma 3.3 we have that the right-hand side set of the above inclusion is bounded, and consequently, the function φ is locally Lipschitz at $\bar{x}$ (see, [36, Theorem 4.15]). Then, due to [36, Theorem 4.17] it is enough to show that $\limsup_{\lambda\to 0^+, x'\to\bar{x}} \nabla\varphi_\lambda(x')$ is single valued. Indeed, by Lemma 3.3 we can apply Fatou's type theorem (see, e.g., [9, Corollary 4.1]) and obtain that

$$\limsup_{\lambda\to 0^+, x'\to\bar{x}} \nabla\varphi_\lambda(x') \subset \int_{\mathbb{S}^{m-1}} \limsup_{\lambda\to 0^+, x'\to\bar{x}} \mathcal{G}_\lambda(x',v)d\mu_\zeta(v). \tag{3.34}$$

Now, let $v \in F(\bar{x})$ and consider

$$w \in \limsup_{\lambda\to 0^+, x'\to\bar{x}} \mathcal{G}_\lambda(x',v).$$

Then there exist $x_k \to \bar{x}$ and $\lambda_k \to 0^+$ such that $\mathcal{G}_{\lambda_k}(x_k,v) \to w$. By Proposition 3.3, (3.16) and since

$$\limsup_{\lambda\to 0^+, x'\to\bar{x}} \nabla \mathsf{e}_\lambda\, \Phi^h(x',\rho_\lambda(x',v)Lv) = \partial S^h_\Phi(\bar{x},\rho(\bar{x},v)Lv)$$

(see, e.g., [3, Theorem 3.66, p. 373]), we have that

$$w = -\theta(\rho(\bar{x}, v), v) \begin{pmatrix} \dfrac{x^* - \nabla h(x)}{\langle z^*, Lv\rangle} \end{pmatrix} \text{ for some } (x^*, z^*) \in \partial S^h_\Phi(\bar{x}, \rho(\bar{x}, v)Lv).$$

On the other hand, if $v \in I(\bar{x})$ and

$$w \in \limsup_{\lambda\to 0^+, x'\to\bar{x}} \mathcal{G}_\lambda(x', v),$$

we can conclude from (3.24) that $w = 0$. Therefore,

$$\begin{aligned}
&\limsup_{\lambda\to 0^+, x'\to\bar{x}} \mathcal{G}_\lambda(x', v) \\
&\subset \begin{cases} \left\{-\theta(\rho(\bar{x}, v), v) \left(\frac{x^*-\nabla h(\bar{x})}{\langle z^*, Lv\rangle}\right) \text{ s.t } (x^*, z^*) \in \partial S^h_\Phi(\bar{x}, \rho(\bar{x}, v)Lv)\right\} & \text{if } v \in F(\bar{x}) \\ \{0\} & \text{if } v \in I(\bar{x}), \end{cases}
\end{aligned}$$

and since $\partial S^h_\Phi(\bar{x}, \rho(\bar{x}, v)Lv)$ is single valued, we conclude the proof from (3.34). □

3.4. Consistency in nonsmooth conic chance constrained optimization problems

In this section, we study the convergence of the solutions of optimization problems generated by replacing the probability function with our Moreau regularized versions. Formally, for a fixed reliability parameter $p \in [0, 1]$, let us consider a convex proper and lsc function $\psi : \mathcal{H} \to \mathbb{R}_\infty$ and the optimization problem

$$\begin{aligned} &\min \psi(x) \\ &\text{s.t } \ x \in M(p) \end{aligned} \tag{P}$$

where $M(p) := [\varphi \geq p] = \{x \in \mathcal{H} : \varphi(x) \geq p\}$ and φ is the probability function defined in (3.1). Furthermore, we consider the family of problems

$$\begin{aligned} &\min \mathsf{e}_\lambda\psi(x) \\ &\text{s.t } \ x \in M_\lambda(p), \end{aligned} \tag{P_λ}$$

where $M_\lambda(p) := \{x \in \mathcal{H} : \varphi_\lambda(x) \geq p\}$ for the regularized probability function φ_λ given in (3.6). In the same spirit, the objective function of problem (P) is replaced by its Moreau regularization to have that the optimization problems (P_λ) have smooth data. Let us denote by $v(P)$ and $v(P_\lambda)$ the values of the problems (P) and (P_λ), respectively.

The first result provides the Painlevé-Kuratowski and Mosco convergence of the feasible sets of problem (P_λ) to the feasible set given in the original optimization problem (P).

Proposition 3.6 *Consider $p \in [0, 1]$. Then,*

a) The sets $M_\lambda(p)$ Painlevé-Kuratowski converge to $M(p)$.

b) The sets $M_\lambda(p)$ Mosco converge to $M(p)$, provided that the function h in (3.3) is sequentially weakly continuous.

Proof. Let us consider a sequence $x_k \in M_{\lambda_k}(p)$ with $\lambda_k \to 0$ and $x_k \to x$ ($x_k \rightharpoonup x$, respectively). Then by Item b) of Theorem 3.1 we have that

$$p \leq \limsup_{k\to\infty} \varphi_{\lambda_k}(x_k) \leq \varphi(x),$$

which shows that $x \in M(p)$. Now, by Item a) of Theorem 3.1 we have that $M(p) \subset M_\lambda(p)$, which particularly implies that item a) holds (item b) holds, respectively). $\square$

It is worth mentioning that Mosco convergence is commonly related to convex sets because of the weak convergence needed in the definition. Nevertheless, a probability function cannot be convex (unless it is a constant mapping) because it takes values on $[0, 1]$. Furthermore, the sets $M_\lambda(p)$ are not necessarily convex even in finite dimension, as the following example shows.

Example 3.2 *Let $\xi \sim \mathcal{N}(0, 1)$ and consider the probability function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$\varphi(x_1, x_2) = \mathbb{P}(g(x_1, x_2, \xi) \leq 0)$$

where $g : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ is the nonconvex function $g(x_1, x_2, z) = z + f(x_1) + \frac{1}{2}x_2^2$ and

$$f(x_1) = \begin{cases} -\frac{1}{2}x_1^2 & \text{if } x_1 \leq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Consider the convex and differentiable function $h(x_1, x_2) = \frac{1}{2}x_1^2$ and notice that

$$g(x_1, x_2, z) + h(x_1, x_2) = z + \hat{f}(x_1) + \tfrac{1}{2}x_2^2$$

is convex, where

$$\hat{f}(x_1) = \begin{cases} \frac{1}{2}x_1^2 & \text{if } x_1 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The regularized probability function is

$$\varphi_\lambda(x) = \mathbb{P}\left(\xi - \tfrac{1}{2}\lambda + \mathsf{e}_\lambda \hat{f}(x_1) + \tfrac{1}{2(\lambda+1)}x_2^2 \leq \tfrac{1}{2}x_1^2\right)$$

where

$$\mathsf{e}_\lambda \hat{f}(x_1) = \begin{cases} \frac{1}{2(\lambda+1)}x_1^2 & \text{if } x_1 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The upper level sets $M_\lambda(p)$ of $\varphi_\lambda(x)$ with $\lambda = 0.5$ are not all convex: Figure 3.1 illustrates the graph of φ_λ for $\lambda = 0.5$ and Figure 3.2 illustrates its contour plot.

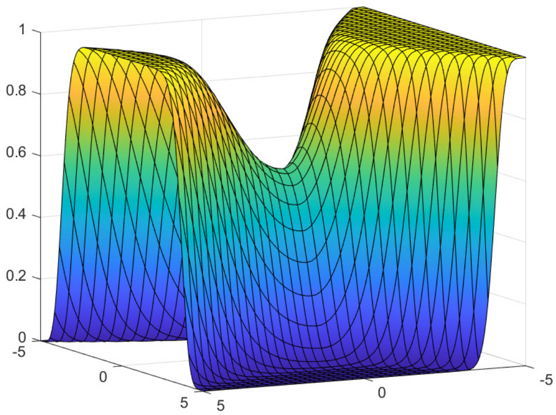

Figure 3.1: Graph of φ_λ for $\lambda = 0.5$ in Example 3.2

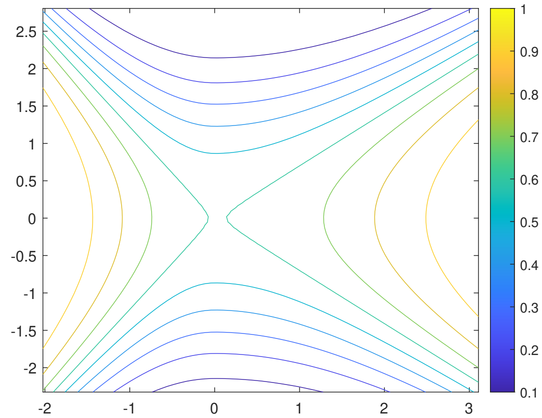

Figure 3.2: Contour plot of φ_λ for $\lambda = 0.5$ in Example 3.2

Here, it is important to notice that our regularized feasible set contains the initial one. Nevertheless, the next proposition shows that under a small enlargement it is possible to show a partial appositive inclusion, which measures how far we are from our initial feasible set in terms of the random inequality.

Proposition 3.7 *Let $\mathcal{H}$ be finite-dimensional space, and $\varepsilon, \eta > 0$ be given and $C \subseteq H$ be a bounded closed convex set. Let us define the following enlargements:*

$$M^\varepsilon(p) := \{x \in \mathcal{H} : \varphi^\varepsilon(x) \geq p\} \text{ and } \varphi^\varepsilon(x) = \mathbb{P}(S_\Phi(x, \xi) \leq h(x) + \varepsilon).$$

Then there exists $\lambda_0 > 0$ such that for all $p \in \mathbb{R}$ and all $\lambda \in (0, \lambda_0)$

$$M^\varepsilon(p - \eta) \supseteq M_\lambda(p) \cap C, \tag{3.35}$$

In addition, if f_ξ has bounded support we have

$$M^\varepsilon(p) \supseteq M_\lambda(p) \cap C.$$

Proof. We recall that the probability measure induced by ξ is Borellian and hence tight. Therefore, for any $\eta > 0$, we can find $r > 0$ such that $\mathbb{P}(\|\xi\| > r) \leq \eta$. Let us define the set $K = C \times \mathbb{B}_r$.

Then, by Proposition 2.7, we can find $\ell > 0$ and λ_1 such that for all $\lambda \in (0, \lambda_1)$

$$\sup_{(x,z)\in K} |\mathsf{e}_\lambda \Phi(x, z) - S_\Phi(x, z)| \leq \ell\sqrt{\lambda}.$$

Now, consider $\lambda_0 < \lambda_1$ such that $\ell\sqrt{\lambda_0} < \varepsilon$. So, for every $x \in C$, the following inclusion is valid:

$$\{z \in \mathbb{B}_r : \mathsf{e}_\lambda \Phi(x, z) \leq h(x)\} \subset \{z \in \mathbb{B}_r : S_\Phi(x, z) \leq h(x) + \varepsilon\}.$$

As a result, for $x \in M_\lambda(p) \cap C$, we have

$$\begin{aligned} p \le \varphi_\lambda(x) &= \mathbb{P}(\mathsf{e}_\lambda \Phi(x,\xi) \le h(x)) \\ &= \mathbb{P}(\|\xi\| \le r, \mathsf{e}_\lambda \Phi(x,\xi) \le h(x)) + \mathbb{P}(\|\xi\| > r, \mathsf{e}_\lambda \Phi(x,\xi) \le h(x)) \\ &\le \mathbb{P}(\|\xi\| \le r, \mathsf{e}_\lambda \Phi(x,\xi) \le h(x)) + \mathbb{P}(\|\xi\| > r) \le \varphi^\varepsilon(x) + \mathbb{P}(\|\xi\| > r). \end{aligned}$$

From this we can deduce $\varphi^\varepsilon(x) \ge p - \eta$, i.e., $x \in M^\varepsilon(p-\eta)$. If f_ξ has bounded support, we may in particular find an appropriate r when $\eta = 0$ is chosen, since then there is $r > 0$ such that $\mathbb{P}(\|\xi\| > r) = 0$. Then (3.35) allows us to conclude. $\square$

Now, we provide the main result of this section which establishes a relation between problems (P) and (P_λ).

Theorem 3.4 *Let $\psi : \mathcal{H} \to \mathbb{R}_\infty$ be a convex, coercive and lower semicontinuous function such that $M(p) \cap \operatorname{dom} \psi \ne \emptyset$. Then*

a) $v(P), v(P_\lambda) \in \mathbb{R}$ for all $\lambda > 0$ and $v(P_\lambda) \to v(P)$.

b) If (P) has a unique optimum x_0 and x_λ is any sequence of optimal solutions for (P_λ), then $x_\lambda \rightharpoonup x_0$, provided that the function h in (3.3) is sequentially weakly continuous. If, furthermore, $\operatorname{dom} \psi = \mathcal{H}$ and ψ^* is Fréchet differentiable on $\operatorname{dom} \partial\psi^*$, then $x_\lambda \to x_0$.

Proof. a) Let $\bar{x} \in M(p) \cap \operatorname{dom} \psi$ be given. By Item d) of Theorem 3.1, the set $M(p)$ is weakly closed, then the nonempty set $M := M(p) \cap \{x \in \mathcal{H} : \psi(x) \le \psi(\bar{x})\}$ is weakly compact. Since ψ is weakly lower semicontinuous, by Weierstrass' theorem, ψ has a minimizer in M. Therefore, $v(P) \in \mathbb{R}$. Similarly, since the Moreau envelope $\mathsf{e}_\lambda \psi$ is convex, coercive ($0 \in \operatorname{int} \operatorname{dom} \psi^* = \operatorname{int} \operatorname{dom} (\mathsf{e}_\lambda \psi)^*$), lower semicontinuous and

$$M(p) \cap \operatorname{dom} \psi \subset M_\lambda(p) \cap \operatorname{dom} \mathsf{e}_\lambda \psi$$

for all λ, we have, similarly, that $v(P_\lambda) \in \mathbb{R}$ for all λ.

To prove

$$\liminf_{\lambda \to 0} v(P_\lambda) \ge v(P),$$

let us proceed by contradiction. That is, for some $\alpha < v(P)$, there is a subsequence $x_{\lambda_k} \in M_{\lambda_k}(p)$ with

$$\mathsf{e}_{\lambda_k} \psi(x_{\lambda_k}) \le \alpha \tag{3.36}$$

for all k. Since by Item a) of Proposition 2.6, $\mathsf{e}_{\lambda_1} \psi(x_{\lambda_k}) \le \mathsf{e}_{\lambda_k} \psi(x_{\lambda_k}) \le \alpha$ for all k and $\mathsf{e}_{\lambda_1} \psi$ is coercive, convex and lower semicontinuous, there is a subsequence $x_{\lambda_{k_i}} \rightharpoonup x$. Indeed the level set of $\mathsf{e}_{\lambda_1} \psi$ is bounded. By (3.36) and Item c) of Proposition 2.6 we get $\psi(x) \le \alpha$, and on the other hand, by Proposition 3.6, $x \in M(p)$. Thus, $v(P) \le \alpha$, which is a contradiction.

Now, to prove

$$\limsup_{\lambda \to 0} v(P_\lambda) \le v(P),$$

notice that, by Item a) of Theorem 3.1, $M(p) \subset M_\lambda(p)$ for all λ and since $\mathsf{e}_\lambda \psi \le \psi$ for all λ we get $v(P_\lambda) \le v(P)$ for all λ.

b) Suppose x_0 is the unique optimum of (P), but that the statement of b) is false. Then for some, $\varepsilon > 0$, $u \in \mathcal{H}$ and some subsequence $\langle x_{\lambda_k} - x_0, u\rangle \geq \varepsilon$ for all k. Notice that, by Proposition 2.6 item a) and optimality of x_{λ_k} we have

$$\mathsf{e}_{\lambda_k}\psi(x_{\lambda_k}) \leq \mathsf{e}_{\lambda_k}\psi(x_0) \leq \psi(x_0) \tag{3.37}$$

for all k. Then, reproducing some of the above arguments with $\alpha = \psi(x_0)$, we deduce that $\mathsf{e}_{\lambda_1}\psi(x_{\lambda_k}) \leq \psi(x_0)$ for all k and since $\mathsf{e}_{\lambda_1}\psi$ is coercive, convex and lower semicontinuous, there is a subsequence $x_{\lambda_k} \rightharpoonup x$. By (3.37) and Item c) of Proposition 2.6 we get $\psi(x) \leq \psi(x_0)$, and on the other hand, by Proposition 3.6 $x \in M(p)$. By uniqueness $x = x_0$, which contradicts $\langle x_{\lambda_k} - x_0, u\rangle \geq \varepsilon$ for all k. Therefore, $x_\lambda \rightharpoonup x_0$.

Now suppose, furthermore, that $\operatorname{dom}\psi = \mathcal{H}$ and ψ^* is Fréchet differentiable on $\operatorname{dom}\partial\psi^*$. It follows that $\operatorname{dom}\partial\psi = \mathcal{H}$. Hence there exists $u \in \partial\psi(x_0)$, which implies $x_0 = \nabla\psi^*(u)$. Particularly, due to [13, Theorem 5.2.3], the function $\psi(\cdot) - \langle u, \cdot\rangle$ attains a strong minimum at x_0. We claim that

$$\psi(\hat{x}_\lambda) - \langle u, \hat{x}_\lambda\rangle \to \psi(x_0) - \langle u, x_0\rangle,$$

where $\hat{x}_\lambda = \operatorname{Prox}_\lambda\psi(x_\lambda)$. Indeed, since $\inf_{z\in\mathcal{H}}\psi(z) > -\infty$ as a result of ψ being convex, coercive and l.s.c., and

$$\inf_{z\in\mathcal{H}}\psi(z) + \frac{1}{2\lambda}\|x_\lambda - \hat{x}_\lambda\|^2 \leq \psi(\hat{x}_\lambda) + \frac{1}{2\lambda}\|x_\lambda - \hat{x}_\lambda\|^2 = \mathsf{e}_\lambda\psi(x_\lambda) \leq \psi(x_0),$$

we have that

$$\|x_\lambda - \hat{x}_\lambda\| \leq \sqrt{\lambda}C, \tag{3.38}$$

for $C \geq \sqrt{(\psi(x_0) - \inf_{z\in\mathcal{H}}\psi(z))} \in \mathbb{R}$. Thus also $\hat{x}_\lambda \rightharpoonup x_0$ and

$$\begin{aligned}\psi(x_0) \leq& \liminf_{\lambda\to 0}\psi(\hat{x}_\lambda) \leq \limsup_{\lambda\to 0}\psi(\hat{x}_\lambda)\\ &\leq \limsup_{\lambda\to 0}\psi(\hat{x}_\lambda) + \frac{1}{2\lambda}\|x_\lambda - \hat{x}_\lambda\|^2 = \limsup_{\lambda\to 0}\mathsf{e}_\lambda\psi(x_\lambda) = \psi(x_0),\end{aligned}$$

and together yields $\psi(\hat{x}_\lambda) - \langle u, \hat{x}_\lambda\rangle \to \psi(x_0) - \langle u, x_0\rangle$. Therefore, $\|\hat{x}_\lambda - x_0\| \to 0$ because x_0 is a strong minimum of $\psi(\cdot) - \langle u, \cdot\rangle$, so by (3.38), we can then conclude $\|x_\lambda - x_0\| \to 0$. ☐

The uniqueness of the minimizer is intrinsically related to the convexity of the optimization problems. The following result provides conditions under the problems optimization problems (P) and (P_λ) are convex and consequently all the assumptions of Theorem 3.4 hold.

Corollary 3.2 *Let us suppose that ξ has an α-concave probability distribution and Φ satisfies (3.3) with $h = 0$. Then, for every $\lambda > 0$, and any $p \in (0, 1)$ the functions φ_λ and φ are α-concave on the sets*

$$\{x \in \mathcal{H} : \exists z \in \mathbb{R}^m \ s.t\ \mathsf{e}_\lambda\Phi(x, z) \leq 0\} \ \text{ and } \ \{x \in \mathcal{H} : \exists z \in \mathbb{R}^m \ s.t\ \Phi(x, z) \leq 0\},$$

respectively. Consequently, for any $p \in (0, 1]$ the sets $M_\lambda(p)$ and $M(p)$ are convex. Moreover, suppose that the objective function ψ in the optimization problem (P) is convex, coercive, lower semicontinuous, $M(p) \cap \operatorname{int}(\operatorname{dom}\psi) \neq \emptyset$ and ψ^* is Fréchet differentiable on $\operatorname{dom}\partial\psi^*$. Then, the sequence of unique solutions of problems (P_λ) converges to the unique minimizer *of* (P).

Proof. The α-concavity of the functions φ and φ_λ follows from a direct application of [50, Theorem 4.39, p. 108]. In particular, φ and φ_λ are quasi-concave, hence the sets $M(p)$ and $M_\lambda(p)$, being upper level sets of these functions, are convex. Now suppose that the objective function ψ in the optimization problem (P) is convex, coercive, lower semicontinuous and $M(p) \cap \operatorname{int}(\operatorname{dom}\psi) \neq \emptyset$. Then Item a) of Theorem 3.4 follows and by [10, Corollary 16.38] we have

$$0 \in \partial(\psi + \delta_{M(p)})(x_0) = \partial\psi(x_0) + \partial\delta_{M(p)}(x_0),$$

where $\delta_{M(p)}$ is the indicator function of $M(p)$ and x_0 is an optimal solution of (P). Thus $\partial\psi(x_0) \neq \emptyset$, so the set of optimal solutions of (P) is a convex subset of $\operatorname{dom}\partial\psi$. By the differentiability assumption over ψ^*, the function ψ must be strictly convex on this set (see, e.g., [13, section 7.3]). Then (P) has a unique optimal solution. Similarly, since the objective functions $\mathtt{e}_\lambda\psi$ satisfy the same hypothesis as ψ, the problems (P_λ) also have unique optimal solutions. Then, the convergence of optimal solutions follows from Theorem 3.4. ☐

3.5. Examples and applications

In this section, we review some examples of the potential applications of our results. Formally, we discuss how our approach can be used to rewrite several classes of probability functions arising in (nonsmooth) optimizing models, and consequently, it illustrates the versatility of our research. Our first examples will demonstrate the smoothing effect of the suggested regularization. Then we will examine the situation of a so-called "joint chance constraint". The section will end with the investigation of a situation wherein $\mathcal{K}$ is the cone of positive definite matrices as well as the case wherein $\mathcal{K}$ describes infinitely many inequalities. In each situation, we will carefully investigate how (3.3) can be concretely shown to hold true.

3.5.1. Nonsmooth inequality constraint

First, we start our analysis considering a probability function given by a nonsmooth single inequality, that is,

$$\varphi(x) := \mathbb{P}\left(g(x,\xi) \leq 0\right).$$

where $g : \mathcal{H} \times \mathbb{R}^m \to \mathbb{R}$ is a (possible nonsmooth) function. It is clear that in that case the cone $\mathcal{K}$ in consideration is given by the set of nonnegative real numbers, the generator of the positive polar cone is nothing more than the singleton $\mathcal{C} = \{1\}$, and our function Φ is nothing more than the same function g. Moreover, in this setting assumption (3.3) is equivalent to the existence of a continuously differentiable function h such that $(x,z) \to g(x,z) + h(x)$ is convex. For simplicity, in the following two examples we chose $h(x) = 0$ for all $x \in \mathcal{H}$.

Example 3.3 *Let $\xi \sim \mathcal{N}(0,1)$ and consider the nonsmooth function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by*

$$g(x,z) = 2f_1(x) + f_2(z) - 5,$$

where $f_1(x) = \max(|x| - 1, 0)$ *and*

$$f_2(z) = \begin{cases} z^2 & \textit{if } z \geq 0 \\ -z & \textit{otherwise.} \end{cases}$$

The probability function

$$\varphi(x) = \mathbb{P}\left(2f_1(x) - 5 \leq \xi \leq \sqrt{-2f_1(x) + 5}\right)$$

is not differentiable at $\bar{x} = 1, -1$*. Indeed, the left derivative of* φ *at* $\bar{x} = 1$ *is* $\varphi'_-(1) = 0$ *and the right derivative of* φ *at* $\bar{x} = 1$ *is*

$$\varphi'_+(1) = -\tfrac{1}{\sqrt{2\pi}}\left[\tfrac{1}{\sqrt{5}}\exp(-5/2) + 2\exp(-25/2)\right] < 0.$$

Similarly, φ *is not differentiable at* $\bar{x} = -1$*. Given* $\lambda > 0$*, we have*

$$\varphi_\lambda(x) := \mathbb{P}\left(\mathsf{e}_\lambda f_2(\xi) \leq -2\mathsf{e}_{2\lambda} f_1(x) + 5\right),$$

where

$$\mathsf{e}_\lambda f_1(x) = \begin{cases} f_1(x) & \textit{if } |x| \leq 1 \\ |x| - \frac{\lambda}{2} - 1 & \textit{if } |x| \geq \lambda + 1 \\ \frac{1}{2\lambda}(|x| - 1)^2 & \textit{otherwise} \end{cases}$$

and

$$\mathsf{e}_\lambda f_2(\xi) = \begin{cases} -\xi - \frac{\lambda}{2} & \textit{if } \xi \leq \lambda \\ \frac{1}{2\lambda+1}\xi^2 & \textit{if } \xi \geq 0 \\ \frac{1}{2\lambda}\xi^2 & \textit{otherwise.} \end{cases}$$

Figure 3.3 illustrates the graph of the functions φ_λ *for* $\lambda \in \{0, 0.03, 0.1, 0.3\}$ *where* $\varphi_0 := \varphi$ *and Figure 3.4 illustrates a zoomed version for* $\lambda \in \{0, 0.0001, 0.0005\}$ *where we can clearly see the smoothness of the regularized probability function* φ_λ *at* $\bar{x} = -1$*.*

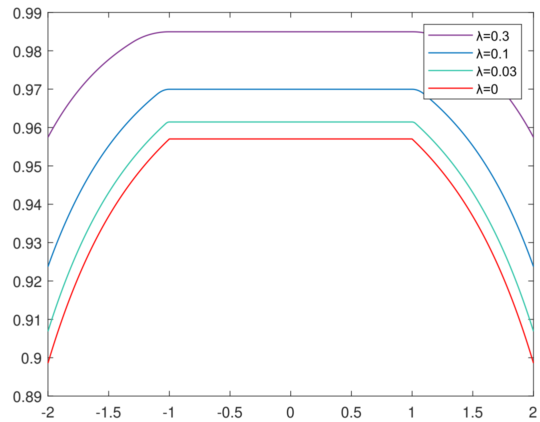

Figure 3.3: Graph of φ_λ for $\lambda \in \{0, 0.03, 0.1, 0.3\}$ in Example 3.3

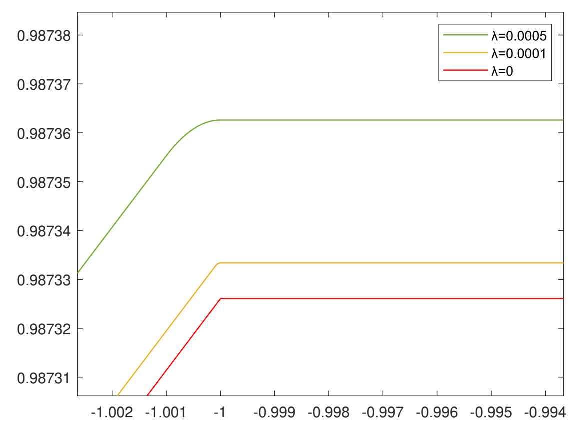

Figure 3.4: Graph of φ_λ for $\lambda \in \{0, 0.0001, 0.0005\}$ in Example 3.3

Example 3.4 *Let $\xi_1, \xi_2 \sim \mathcal{N}(0,1)$ and consider the nonsmooth function $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ given by $g(x_1, x_2, z_1, z_2) = f(x_1, x_2) + |z_1| + z_2 - 3$ where $f(x_1, x_2) = \max(\sqrt{x_1^2 + x_2^2} - 2, 0)$. The function*

$$\varphi(x_1, x_2) = \mathbb{P}\left(\xi_2 \leq -f(x_1, x_2) - |\xi_1| + 3\right)$$

does not have a directional derivative at $(2,0)$ in the direction $(x_1, x_2) = (1,0)$ since the left derivative is

$$\varphi'_-(2,0) := \lim_{t \to 0-} \frac{\varphi(2+t, 0) - \varphi(2,0)}{t} = 0$$

and the right derivative is

$$\varphi'_+(2,0) := \lim_{t \to 0+} \frac{\varphi(2+t, 0) - \varphi(2,0)}{t} = -\frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-z_1^2 + 3|z_1| - \tfrac{9}{2}) dz_1 < 0.$$

Given $\lambda > 0$, we have

$$\varphi_\lambda(x_1, x_2) = \mathbb{P}\left(\xi_2 \leq -\mathsf{e}_\lambda f(x_1, x_2) - \mathsf{e}_\lambda |\xi_1| + 3\right)$$

where

$$\mathsf{e}_\lambda f(x_1, x_2) = \begin{cases} f(x_1, x_2) & \text{if } \sqrt{x_1^2 + x_2^2} \leq 2 \\ \sqrt{x_1^2 + x_2^2} - \frac{\lambda}{2} - 2 & \text{if } \sqrt{x_1^2 + x_2^2} \geq \lambda + 2 \\ \frac{1}{2\lambda}(\sqrt{x_1^2 + x_2^2} - 2)^2 & \text{otherwise} \end{cases}$$

and

$$\mathsf{e}_\lambda |\xi_1| = \begin{cases} \frac{1}{2\lambda}|\xi_1|^2 & \text{if } |\xi_1| \leq \lambda \\ |\xi_1| - \frac{\lambda}{2} & \text{otherwise.} \end{cases}.$$

Figure 3.5 illustrates the nonsmoothness of φ on $\{(x_1, x_2) : x_1^2 + x_2^2 = 2\}$.

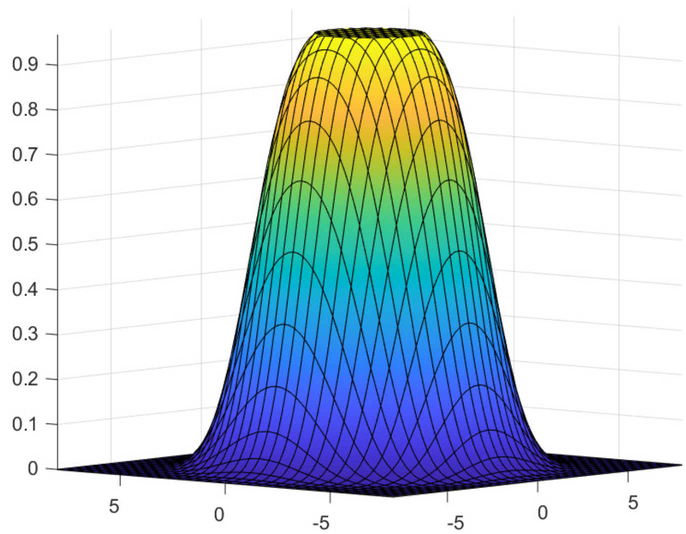

Figure 3.5: Graph of φ in Example 3.4

3.5.2. Joint Chance constraint

Let us consider a family of functions $g_i : \mathcal{H} \times \mathbb{R}^m \to \mathbb{R}$ with $i = 1, \ldots, s$ and the probability function

$$\varphi(x) := \mathbb{P}\left(g_i(x, \xi) \leq 0, \text{ for all } i = 1, \ldots, s\right). \tag{3.39}$$

Then, considering $\Phi : \mathcal{H} \times \mathbb{R}^m \to \mathbb{R}^s$ given by

$$\Phi(x, z) := \begin{pmatrix} g_1(x, z) \\ \vdots \\ g_s(x, z) \end{pmatrix} \tag{3.40}$$

and the cone $\mathcal{K} := \mathbb{R}^s_+$, the probability function in (3.39) can be rewritten in the form $\varphi(x) = \mathbb{P}(\Phi(x, \xi) \in -\mathcal{K})$, which places us in the framework of (3.1). It is easy to see that for a given function $h : \mathcal{H} \to \mathbb{R}$, and considering the unit simplex $\mathcal{C} := \Delta_s$, effectively "generating" the positive polar cone of $\mathcal{K}$, we have that $S^h_\Phi(x, z) = \max_{i=1,\ldots,s} g_i(x, z) + h(x)$. Furthermore, the next proposition gives us a simple characterization of the condition (3.3) in terms of the nominal data g_i.

Proposition 3.8 *Let $g_i : \mathcal{H} \times \mathbb{R}^m \to \mathbb{R}$ be a family of functions for $i = 1, \ldots, s$ and consider the vector-valued function Φ given in (3.40). Then the following are equivalent*

a) There exists a continuously differentiable convex function $h : \mathcal{H} \to \mathbb{R}$ such that Φ satisfies (3.3).

b) For every $i = 1, \ldots, s$ there exists a continuously differentiable convex function $h_i : \mathcal{H} \to \mathbb{R}$ such that $(x, z) \to g_i(x, z) + h_i(x)$ is convex.

Proof. To prove a) implies b) consider $w^* = e_i$ in (3.3) where e_i is the i-th standard basic vector of $\mathbb{R}^s$. To prove the converse, let $w^* \in \mathcal{C}$ and set $h(x) := \sum_{i=1}^s h_i(x)$. Then since

$0 \leq w_i^* \leq 1$ and the functions $g_i(x,z) + h_i(x)$ and $h_i(x)$ are convex we have that

$$\langle w^*, \Phi\rangle(x,z) + h(x) = \sum_{i=1}^{s} w_i^*(g_i(x,z) + h_i(x)) + (1 - w_i^*)h_i(x)$$

is convex. □

The final example in this subsection illustrates the convergence of the solution and minimizers established in Corollary 3.2.

Example 3.5 (Illustrative example) *Let $\xi_1, \xi_2 \sim \mathcal{N}(0,1)$ and consider problem (P) of section 3.4 with $p = 0.95$ and the nonsmooth functions $\psi, \varphi : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$\begin{aligned}\psi(x_1,x_2) &= |x_1 - 5| + \tfrac{1}{2}x_2^2 + x_2 + 8\\ \varphi(x_1,x_2) &= \mathbb{P}(\sqrt{x_1^2 + x_2^2} + |\xi_1| + \xi_2 \leq 5, \ \textit{and}\ |\xi_1| + \xi_2 \leq 3)\end{aligned}$$

In this case, we can consider the vector valued function $\Phi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\Phi(x_1,x_2,z_1,z_2) = \begin{pmatrix} \sqrt{x_1^2 + x_2^2} + |\xi_1| + \xi_2 - 5 \\ |\xi_1| + \xi_2 - 3 \end{pmatrix}$$

Then, the probability function can be recast as

$$\varphi(x_1,x_2) = \mathbb{P}\left(\Phi(x_1,x_2,\xi_1,\xi_2) \in -\mathbb{R}^2_+\right).$$

In Table 3.5 we give the optimal values and the minimizers of (P_λ) associated with problem (P).

Table 3.1: Results obtained by MatLab's optimization algorithm `fmincon`

λ	$v(P_\lambda)$	x_λ
1	8.19472	(2.96739,-1.19475)
0.1	10.19347	(2.21702,-0.73521)
0.01	10.39840	(2.13857,-0.68601)
0.001	10.41892	(2.13071,-0.68105)
0.0001	10.42098	(2.12992,-0.68055)
0.00001	10.42118	(2.12985,-0.68050)
	$v(P)$ =10.42121	x_0 =(2.12984,-0.68049)

3.5.3. Semidefinite chance constraint

In this section, we consider the following probability function

$$\varphi(x) := \mathbb{P}\left(\Phi(x,\xi) \preceq 0\right), \tag{3.41}$$

where $\Phi : \mathcal{H} \times \mathbb{R}^s \to \mathcal{S}^p$ is a function with $\mathcal{S}^p$ the set of $p \times p$ symmetric matrices, and the symbol $A \preceq 0$ means that the matrix A is negative semidefinite. It is important to notice that the probability function (3.41) appears as a natural alternative to deal with semidefinite mathematical programs where there exists a random inflow in the model.

It is well known that the partial order $\preceq$ can be characterized by the cone of negative definite matrices $\mathcal{S}^p_-$. Let us recall that the space $\mathcal{S}^p$ is a Hilbert space endowed with the inner product $\langle A, B \rangle := \mathrm{Tr}(AB)$, where Tr represents the trace operator (see, e.g., [12]). Using this topological structure, the positive polar cone of $\mathcal{S}^p_-$ is given by the set of positive definite symmetric matrices $\mathcal{S}^p_+$. It is straightforward to see that the set $\mathcal{C} := \{A \in \mathcal{S}^p_+ : \mathrm{Tr}(A) = 1\}$ generates the cone $\mathcal{S}^p_+$. Furthermore, in order to fulfill (3.3), we need to assume an appropriate notion of convexity for this precise setting. The following result establishes an equivalent characterization of (3.3) through simpler quadratic scalarizations.

Proposition 3.9 *Let $\Phi : \mathcal{H} \times \mathbb{R}^m \to \mathcal{S}^p$ be a function and $h : \mathcal{H} \times \mathbb{R}$ be a convex and continuously differentiable function. Then, the following are equivalent:*

a) For every $A \in \mathcal{C} = \{A \in \mathcal{S}^p_+ : \mathrm{Tr}(A) = 1\}$ the function $(x, z) \to \langle A, \Phi(x, z) \rangle + h(x)$ is convex.

b) For every $v \in \mathbb{S}^{p-1}$ the function $(x, z) \to v^\top \Phi(x, z) v + h(x)$ is convex.

Proof. On the one hand, let us suppose that $a)$ holds, and consider a vector $v \in \mathbb{S}^{p-1}$, that is $v \in \mathbb{R}^p$ with $\|v\| = 1$, then let us define the symmetric matrix $A := vv^\top$, which has $\mathrm{Tr}(A) = \|v\|^2 = 1$. Moreover, the matrix A is positive semidefinite as is clear. Finally, $\langle A, \Phi(x, z) \rangle = v^\top \Phi(x, z) v$, which shows that the function $x \to v^\top \Phi(x, z) v + h(x)$ is convex and that hence b) holds true.

On the other hand, let us assume that $b)$ holds, and consider $A \in \mathcal{S}^p_+$ with $\mathrm{Tr}(A) = 1$. Using the spectral decomposition we have that the matrix A can be decomposed into $A = PDP^\top = \sum_{i=1}^p \lambda_i(A) v_i v_i^\top$, where P is a $p \times p$ orthogonal matrix, and its columns are the vector $v_i \in \mathbb{R}^p$ with $\|v_i\| = 1$, and D is a diagonal given by the eigenvalues of the matrix A, denoted by $\lambda_1(A), \ldots, \lambda_p(A)$, allowing for multiplicity. Then, we can compute the inner product of this matrix and $\Phi(x, z)$ by

$$\langle A, \Phi(x, z) \rangle = \sum_{i=1}^p \lambda_i(A) \langle v_i v_i^\top, \Phi(x, z) \rangle = \sum_{i=1}^p \lambda_i(A) v_i^\top \Phi(x, z) v_i.$$

Finally, since $\sum_{i=1}^p \lambda_i(A) = \mathrm{Tr}(A) = 1$ and $\lambda_i(A) \geq 0$, we get that

$$\langle A, \Phi(x) \rangle + h(x) = \sum_{i=1}^p \lambda_i(A) \left(v_i^\top \Phi(x) v_i + h(x) \right),$$

consequently, the above function is convex, and that concludes the proof. $\square$

Remark 3.3 (Matrix convexity) *It is important to mention that Proposition 3.9 establishes that our desired assumptions hold under the so-called matrix convexity, that is, the assumption that for every $v \in \mathbb{S}^{p-1}$ the function $(x, z) \to v^\top \Phi(x, z) v$ is convex. We refer to [12, Section 5.3.2] for more details and references to these properties.*

Example 3.6 *Let us consider a family of matrices $B_j \in \mathcal{S}^p$ for $j = 0, \dots, m$ and C^2 functions $g_i : \mathbb{R}^s \to \mathbb{R}$ for $i = 1, \dots, s$. Define the mapping $\Phi : \mathbb{R}^s \times \mathbb{R}^m \to \mathcal{S}^p$ given by*

$$\Phi(x, z) := \sum_{i=1}^{s} g_i(x) A_i + \sum_{j=1}^{m} z_i B_i + B_0.$$

For $i = 1, \dots, s$, consider a convex and continuously differentiable function h_i such that $\pm g_i(x) + h_i(x)$ are convex.

Let $C > 0$ be a constant greater than any of the absolute values of the eigenvalues of the matrices A_i. Then, defining $h := C \sum_{i=1}^{s} h_i$, we have that for any $v \in \mathbb{S}^{p-1}$, we have that

$$v^\top \Phi(x, z) v + h(x) = \sum_{i=1}^{s} \left(v^\top A_i v g_i(x) + |v^\top A_i v| h(x)) + (C - |v^\top A_i v|) h_i(x) \right),$$

is a convex function, which due to Proposition 3.9 shows that the mapping Φ satisfies (3.3).

3.5.4. Probabilistic/Robust (Probust) Chance Constraint

Let us consider a compact Hausdorff space T and a function $g : T \times \mathcal{H} \times \mathbb{R}^m \to \mathbb{R}$ such that $t \to g(t, x, z)$ is continuous for all $(x, z) \in \mathcal{H} \times \mathbb{R}^m$. Consider the probability function

$$\varphi(x) := \mathbb{P}\left(g_t(x, \xi) \le 0, \text{ for all } t \in T\right). \tag{3.42}$$

Then, let us define $\Phi \colon \mathcal{H} \times \mathbb{R}^m \to C(T)$ given by

$$(x, z) \to \Phi(x, z) \in C(T) \text{ defined by } t \to \Phi(x, z)(t) := g(t, x, z), \tag{3.43}$$

where $C(T)$ is the space of continuous functions from T to $\mathbb{R}$ and considering the closed convex cone $\mathcal{K} := \{f \in C(T) : f(t) \ge 0 \text{ for all } t \in T\}$. Using this setting, we have that the probability function (3.42) can be expressed as (3.1), that is, $\varphi(x) = \mathbb{P}(\Phi(x, \xi) \in -\mathcal{K})$.

Now, we are going to write the probability function (3.42) using a suitable cone $\mathcal{C}$, which generates the positive polar cone of $\mathcal{K}$. In order to do that let us recall some concepts of measure theory. Let us denote by $\mathcal{B}(T)$ the Borel σ-algebra, which is the smallest σ-algebra generated by open sets, a signed measure $\mu : \mathcal{B}(T) \to \mathbb{R}$ is called regular if for every $A \in \mathcal{B}(T)$

$$\mu(A) = \inf\left\{\mu(U) : U \text{ is open and } A \subset U\right\} = \sup\left\{\mu(F) : F \text{ is closed and } F \subset A\right\}.$$

By Riesz representation theorem (see, e.g., [1, Theorem 14.14]) the dual space of $C(T)$ can be identified as the linear space of regular signed measures. Moreover, in this framework the positive polar cone of the set of positive functions is given by the set of (positive) regular measures $\mu : \mathcal{B}(T) \to \mathbb{R}$ (see, e.g., [1, Theorem 14.12]). Consequently, a suitable generator of that cone corresponds to the set $\mathcal{C}$ of probability measures on $(T, \mathcal{B}(T))$, which means that our supremum function is given by

$$S^h_\Phi(x, z) = \sup\left\{ \int_T g(t, x, z) d\mu(t) + h(x) : \mu \in \mathcal{C} \right\}. \tag{3.44}$$

The next proposition establishes formally that the general supremum function provided in (3.44) of the vector function (3.43) with scalarization over the set of probability measures is indeed nothing more than the pointwise supremum of the function g with respect to the parameter $t \in T$ plus the function h.

Proposition 3.10 *Let T be a compact Hausdorff space and $g : T \times \mathcal{H} \times \mathbb{R}^m \to \mathbb{R}$ be such that $t \to g(t, x, z)$ is continuous for all $(x, z) \in \mathcal{H} \times \mathbb{R}^m$. Then, for a given function $h : \mathcal{H} \to \mathbb{R}$ the following holds true:*

$$S_\Phi^h(x, z) = \sup_{t \in T} g(t, x, z) + h(x) \text{ for all } (x, z) \in \mathcal{H} \times \mathbb{R}^m$$

Proof. Defining $\mathcal{C}$ of probability measures on $(T, \mathcal{B}(T))$. First, we have that

$$S_\Phi^h(x, z) \leq \sup_{t \in T} g(t, x, z) + h(x) \tag{3.45}$$

for all $(x, z) \in \mathcal{H} \times \mathbb{R}^m$. Moreover, given a point $(x, z) \in \mathcal{H} \times \mathbb{R}^m$, we can take $\bar{t} \in T$ (since T is a compact Hausdorff space) such that $g(\bar{t}, x, z) = \sup_{t \in T} g(t, x, z)$, then if we considering the Dirac measure over $\bar{t}$, that is,

$$\mu_{\{\bar{t}\}}(A) = \begin{cases} 1 & \text{if } \bar{t} \in A \\ 0 & \text{otherwise.} \end{cases}$$

we obtain the equality in (3.45), which ends the proof. □

The final result of this section shows a sufficient condition to ensure condition (3.3) in the setting of probust chance constrained optimization.

Proposition 3.11 *Let T be a compact Hausdorff space and $g : T \times \mathcal{H} \times \mathbb{R}^m \to \mathbb{R}$ be such that $t \to g(t, x, z)$ is continuous for all $(x, z) \in \mathcal{H} \times \mathbb{R}^m$. Suppose the existence of a convex continuously differentiable function $h : \mathcal{H} \to \mathbb{R}$, such that for all $t \in T$, the function $(x, z) \to g(t, x, z) + h(x)$ is convex. Then, the function Φ defined in (3.43) satisfies condition (3.3).*

Proof. For all positive regular measures $\mu : \mathcal{B}(T) \to \mathbb{R}$ we have that (3.3) is given by

$$\langle \mu, \Phi \rangle(x, z) + h(x) = \int_T (g(t, x, z) + h(x)) d\mu(t) \text{ for all } (x, z) \in \mathcal{H} \times \mathbb{R}^m,$$

thus its convexity follows from the convexity of the function $g(t, x, z) + h(x)$ for all $(x, z) \in \mathcal{H} \times \mathbb{R}^m$, which is preserved under the integral sign (see, e.g., [17, 21] and the references therin for more details). □

Chapter 4

Generalized differentiation of probability functions generated by set-valued mappings

In this chapter, we investigate a probability function formulated by

$$\varphi(x) := \mathbb{P}(\omega \in \Omega : \xi(\omega) \in \mathcal{S}_i(x) \text{ for all } i = 1, \dots, s), \tag{4.1}$$

where $\xi : \Omega \to \mathbb{R}^n$ is a random vector from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{S}_i : \mathcal{X} \rightrightarrows \mathbb{R}^m$ with $i = 1, \dots, s$ is a family of set-valued mappings. It is clear that φ can also be represented as follows:

$$\varphi(x) = \int\limits_{\{z \in \mathbb{R}^m : z \in \mathcal{S}_i(x) \forall i = 1, \dots, s\}} f_\xi(z) d\lambda_m(z). \tag{4.2}$$

Throughout this work we will make the assumption that f_ξ satisfies (2.10).

In order to derive analytical properties of φ we will assume that given a point of interest $\bar{x} \in \mathcal{X}$ the following basic assumptions for each set-valued mapping $\mathcal{S}_i$ hold: There exists a neighborhood U of $\bar{x}$ such that

$$\begin{aligned}
&a)\ 0 \in \mathcal{S}_i(x) \text{ for all } x \in U\\
&b)\ \mathcal{S}_i \text{ is locally Lipschitz-like at } (x, z) \in \operatorname{gph} \mathcal{S}_i \text{ and } x \in U\\
&c)\ \mathcal{S}_i \text{ has closed graph and convex values}
\end{aligned} \tag{H}$$

We notice that when $\bar{z} \in \mathcal{S}(x)$ for all $x \in U$ and $\mathcal{S}$ satisifes conditions $b)$ and $c)$ on (H), we may consider the set-valued mapping $\tilde{\mathcal{S}}(x) = \mathcal{S}(x) - \bar{z}$ satisfying (H).

This chapter is organized as follows: Section 4.1 presents and investigates properties of the inner enlargement of the probability function (4.1). The local Lipschitz continuity of the probability function (4.1) is exposed in Section 4.2.

4.1. Inner enlargement of probability function (4.1)

In this section, we present the inner enlargement of the probability function (4.1): Given $\varepsilon > 0$ we define the probability function

$$\varphi_\varepsilon(x) := \mathbb{P}(\omega \in \Omega : \xi(\omega) \in \mathcal{S}_i(x) + \varepsilon\mathbb{B} \text{ for all } i = 1, \dots, s). \tag{4.3}$$

The aim of this section is to prove the local Lipschitz continuity of the enlargement by applying Theorem 2.2. To that end, we rewrite as a joint probability function given by an inequality system, this is done through the distance function. We notice that $z \in \mathcal{S}_i(x) + \varepsilon\mathbb{B}$ if and only if $\mathrm{d}(z, \mathcal{S}_i(x)) \leq \varepsilon$. Therefore, the probability function (4.3) can be reformulated as

$$\varphi_\varepsilon(x) = \mathbb{P}(\omega \in \Omega : \mathrm{d}(\xi(\omega), \mathcal{S}_i(x)) \leq \varepsilon \text{ for all } i = 1, \dots, s).$$

According to this reformulation, we define on U the finite and infinite directions with respect to $\mathcal{S}_i$ as the sets defined by

$$\mathcal{F}_i(x) := \{v \in \mathbb{S}^{m-1} \mid \exists r > 0 : \mathrm{d}(rLv, \mathcal{S}_i(x)) > 0\}, \tag{4.4}$$
$$\mathcal{I}_i(x) := \{v \in \mathbb{S}^{m-1} \mid \forall r \geq 0 : \mathrm{d}(rLv, \mathcal{S}_i(x)) = 0\}, \tag{4.5}$$

respectively. We also define

$$\mathcal{F}(x) := \{v \in \mathbb{S}^{m-1} \mid \exists r > 0 : \max_{i=1,\dots,s} \mathrm{d}(rLv, \mathcal{S}_i(x)) > 0\}, \tag{4.6}$$
$$\mathcal{I}(x) := \{v \in \mathbb{S}^{m-1} \mid \forall r \geq 0 : \max_{i=1,\dots,s} \mathrm{d}(rLv, \mathcal{S}_i(x)) = 0\}. \tag{4.7}$$

Now, let us define the radial functions associated with the spherical radial decomposition of our enlargement. Given $\varepsilon \geq 0$ we define $\rho_i^\varepsilon : U \times \mathbb{S}^{m-1} \to \mathbb{R} \cup \{+\infty\}$ by

$$\rho_i^\varepsilon(x, v) := \sup\{r > 0 : \mathrm{d}(rLv, \mathcal{S}_i(x)) \leq \varepsilon\} \tag{4.8}$$

and set

$$\rho_\varepsilon(x, v) := \min_{1 \leq i \leq s} \rho_i^\varepsilon(x, v). \tag{4.9}$$

Particularly, we simply denote $\rho_i(x, v) := \rho_i^\varepsilon(x, v)$ and $\rho(x, v) := \rho_\varepsilon(x, v)$, for $\varepsilon = 0$.

The next lemma establishes some basic properties of the radial functions defined above, which allows us to understand better the behavior of these functions.

Lemma 4.1 *Let each $\mathcal{S}_i$ of the of the family of set-valued mappings satisfy (H) at $\bar{x}$. Then, we have that:*

a) For all $\varepsilon \geq 0$ and all $x \in U$, $\mathcal{F}_i(x) = \operatorname{dom} \rho_i^\varepsilon(x, \cdot)$ and $\mathcal{F}(x) = \operatorname{dom} \rho_\varepsilon(x, \cdot)$.

b) For all $v \in \mathcal{F}_i(x)$,

$$\mathrm{d}(r_1 Lv, \mathcal{S}_i(x)) < \mathrm{d}(r_2 Lv, \mathcal{S}_i(x)), \text{ for all } r_2 > r_1 > \rho_i(x, v). \tag{4.10}$$

c) For all $v \in \mathcal{F}_i(x)$, $\lim_{r\to\infty} \mathrm{d}(rLv, \mathcal{S}_i(x)) = +\infty$.

d) For all $v \in \mathcal{F}_i(x)$ and all $r > \rho_i^{\varepsilon}(x, v)$ we have $\mathrm{d}(rLv, \mathcal{S}_i(x)) > \varepsilon$.

e) For all $\varepsilon \geq 0$, we have $\rho_i^{\varepsilon}(x, v) = \inf\{r > 0 : \mathrm{d}(rLv, \mathcal{S}_i(x)) > \varepsilon\}$ with the convention $\inf \emptyset = +\infty$.

f) For all $v \in \mathcal{F}_i(x)$, and all $\varepsilon > 0$, $\rho_i^{\varepsilon}(x, v)$ is the unique $r > 0$ such that $\mathrm{d}(rLv, \mathcal{S}_i(x)) = \varepsilon$.

Proof. Let us first prove b). Fix i, let $r, \beta \in \mathbb{R}$ and consider the function $\gamma_{r,\beta}(t) := \mathrm{d}((t + r)Lv, \mathcal{S}_i(x)) - \beta$ for $t \geq 0$. This function is convex, and so, whenever $\gamma_{r,\beta}(0) < 0$ and $\gamma_{r,\beta}(t_2) \geq 0$, we have that

$$\gamma_{r,\beta}(t_1) < \frac{t_1}{t_2}\gamma_{r,\beta}(t_2), \text{ for all } 0 < t_1 < t_2. \tag{4.11}$$

Now, if $v \in \mathcal{F}_i(x)$, then, for some $r' > 0$, we have that $\mathrm{d}(r'Lv, \mathcal{S}_i(x)) > 0$. Hence, by convexity of the distance function we have that $r_0 := \rho_i(x, v) < r' < +\infty$. Now, consider $r_2 > r_1 > r_0$ and fix $\beta \in (0, \mathrm{d}(r_2Lv, \mathcal{S}_i(x)))$. Then, using inequality (4.11) with $t_1 = r_1 - r_0$, $t_2 = r_2 - r_0$ and $r = r_0$ we have that (4.10) holds proving part b) and in consequence part c) follows. Now, let us prove a). If $v \in \mathcal{F}(x)$ then $v \in \mathcal{F}_i(x)$ for some $i \in \{1, \ldots, s\}$ and by Item b) the set $\{r \geq 0 : \mathrm{d}(rLv, \mathcal{S}_i(x)) \leq \varepsilon\}$ must be bounded yielding $\rho_i^{\varepsilon}(x, v) < +\infty$ and in consequence $\rho_{\varepsilon}(x, v) < +\infty$. On the other hand if $v \in \mathcal{I}(x)$ we have that $\rho_i^{\varepsilon}(x, v) = +\infty$ for all i and so $\rho_{\varepsilon}(x, v) = +\infty$ concluding the proof of Item a). Item d) follows by using (4.10) with $r_1 = \rho_i^{\varepsilon}(x, v)$ and $r_2 = r$. Item e) follows from Item d) and the continuity of the distance function. Finally, Item f) follows from Items d) and e). ☐

Definition 4.1 (η-growth condition for a family of set-valued mappings) *Consider $\bar{x} \in U$. Let $\eta : \mathbb{R} \to [0, +\infty]$ be a non-decreasing mapping such that*

$$\lim_{\|z\|\to+\infty} \|z\|^m \bar{f}_{\xi}(z)\eta(\|z\|) = 0. \tag{4.12}$$

We say that the family of set-valued mappings $\mathcal{S}_i$ satisfies the η-growth condition at $\bar{x}$ if for some $l > 0$

$$\|D^*\mathcal{S}_i(x, z)\| \leq l\eta(\|z\|), \ \forall x \in \mathbb{B}_{1/l}(\bar{x}), \ \ \forall z \in \mathbb{R}^m, \tag{4.13}$$

where $\|D^*\mathcal{S}_i(x, z)\| := \sup\{\|x^*\| : x^* \in D^*\mathcal{S}_i(x, z)(z^*)$ and $\|z^*\| \leq 1\}$.

Let us define, for $x \in U$, $\varepsilon > 0$ and $v \in \mathcal{F}(x)$

$$\mathcal{M}_{\varepsilon}(x, v) := \left\{ \frac{-\alpha}{\left\langle z_{\varepsilon}^{x,v} - P_{\mathcal{S}_i(x)}(z_{\varepsilon}^{x,v}), Lv \right\rangle} \cdot x^* : \begin{array}{c} i \in T_x^{\varepsilon}(v), \alpha \in I_{\theta}(\rho_{\varepsilon}(x, v), v) \\ x^* \in D^*\mathcal{S}_i(x, P_{\mathcal{S}_i(x)}(z_{\varepsilon}^{x,v}))(P_{\mathcal{S}_i(x)}(z_{\varepsilon}^{x,v}) - z_{\varepsilon}^{x,v}) \end{array} \right\} \tag{4.14}$$

where $z_{\varepsilon}^{x,v} := \rho_{\varepsilon}(x, v)Lv$ and $T_x^{\varepsilon}(v) = \{i \in \{1, \ldots, s\} : \rho_i^{\varepsilon}(x, v) = \rho_{\varepsilon}(x, v)\}$. For convenience we define $\mathcal{M}_{\varepsilon}(x, v) = \{0\}$ for all $v \in \mathcal{I}(x)$.

Theorem 4.1 *Let $\bar{x} \in U$ be given and assume that (2.10) holds true. Moreover, assume that the family of set-valued mappings $\mathcal{S}_i$ satisfy the η-growth condition for set-valued mappings at $\bar{x}$ and that each $\mathcal{S}_i$ satisfy (H) at $\bar{x}$.*

Then the probability function (4.3) is locally Lipschitz at $\bar{x}$ and on an appropriate neighborhood U' of $\bar{x}$ it holds:

$$\partial^{\mathsf{b}}\varphi_\varepsilon(x) \subseteq \operatorname{cl}^{w^*}\left(\int\limits_{v\in\mathcal{F}(x)} \partial^{\mathsf{b}}_x e_\varepsilon(x,v)d\mu_\zeta(v)\right), \text{ for all } x \in U' \tag{4.15}$$

where, e_ε refers to the radial probability-like function defined in (2.13) associated to the sub-level $\{z \in \mathbb{R}^m : \mathrm{d}(z, \mathcal{S}_i(x)) \le \varepsilon$ for all $i = 1, \dots, s\}$. Moreover, we have that

$$\partial^{\mathsf{b}}_x e_\varepsilon(x,v) \subseteq \operatorname{cl}\operatorname{co}\mathcal{M}_\varepsilon(x,v) \tag{4.16}$$

for all $v \in \mathbb{S}^{m-1}$. In addition, if $\mathcal{X}$ is finite-dimensional the closure can be omitted.

Proof. Let us notice that the probability function (4.3) can be written as

$$\varphi_\varepsilon(x) := \mathbb{P}(\omega \in \Omega : g(x,\xi(\omega)) \le 0) \tag{4.17}$$

where $g(x,z) := \max\{g_i(x,z) : i = 1,\dots,s\}$ and

$$g_i(x,z) = \frac{1}{2}\mathrm{d}^2(z,\mathcal{S}_i(x)) - \frac{\varepsilon^2}{2}.$$

Let us prove that the assumptions of Theorem 2.2 in Appendix are satisfied. Indeed, by Assumption (H) we have that $g_i(x,0) = -\frac{\varepsilon^2}{2} < 0$ and that $g_i's$ are convex on the second variable. Due to Lemma 2.1 the functions $g_i's$ are locally Lipschitz at $(x,z) \in U \times \mathbb{R}^m$. Also, by Lemma 4.1 Item f), the sets of finite direction $\mathcal{F}_i$ defined in (4.4) coincide with the sets of finite directions $F_i(x)$ defined in Appendix (Eq. (2.16)). That is, $\mathcal{F}_i(x) = F_i(x)$, and so, by taking complements we also have that $\mathcal{I}_i(x) = I_i(x)$ where $I_i(x)$ is the set of infinite directions defined in Appendix (Eq. (2.17)). In consequence $\mathcal{F}(x) = F(x)$ and $\mathcal{I}(x) = I(x)$. The radial functions $\rho_i^\varepsilon(x,v)$ are equal to the radial functions defined in Appendix Eq. (2.18), and so, $\rho_\varepsilon(x,v)$ is equal to the radial function defined in Appendix Eq. (2.19). Now, let us prove that the family of functions $g_i's$ that we defined above satisfy the η_θ-growth condition given in Appendix (Definition 2.3) for all $v \in \mathcal{I}(\bar{x})$. Consider $\bar{v} \in \mathcal{I}(\bar{x})$. Since, the family of set-valued mappings $\mathcal{S}_i$ satisfy the η-growth condition for set-valued mappings at $\bar{x}$, we have that there exists $\hat{l} > 0$ such that the family $\mathcal{S}_i$ satisfy (4.13). Let us set $l := \hat{l}\varepsilon$ and consider $(x,v) \in \mathbb{B}_{1/l}(\bar{x}) \times \mathbb{B}_{1/l}(\bar{v})$ with $v \in \mathcal{F}(x)$ and $\rho_\varepsilon^i(x,v) \ge l$ with $i \in T_x^\varepsilon(v)$. By Lemma 2.1, the fact that η is non-decreasing and by the nonexpansiveness of the projection mapping we obtain the following sequence of inequalities

$$\|\pi_x(\partial^{\mathsf{b}}_x g_i(x,z_\varepsilon^{x,v}))\| \le \|D^*\mathcal{S}_i(x, P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v}))\|\varepsilon \le \hat{l}\varepsilon\eta(\|P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v})\|) \le l\eta(\|z_\varepsilon^{x,v}\|).$$

Hence it is enough to consider $\eta_\theta(\rho_\varepsilon(x,v),v) = \eta(\|z_\varepsilon^{x,v}\|)$. Therefore, the proof follows from Theorem 2.2 in Appendix, upon noticing that the inclusion (4.16) is obtained using Appendix Eq. (2.33) together with Lemma 2.1 when $v \in \mathcal{F}(x)$ and that $\partial^{\mathsf{b}}_x e_\varepsilon(x,v) \subseteq \{0\} = \mathcal{M}_\varepsilon(x,v)$ when $v \in \mathcal{I}(x)$. $\square$

Remark 4.1 *In Theorem 4.1 we have, furthermore, the inclusion*

$$\partial^{\mathsf{b}}\varphi_\varepsilon(x) \subseteq \mathrm{cl}^{w^*}\left(\int\limits_{v\in\mathbb{S}^{m-1}} \mathcal{M}_\varepsilon(x,v)d\mu_\zeta(v)\right) \text{ for all } x \in U',$$

as a consequence of [8, Theorem 8.6.4] since $\mathcal{X}$ is a separable reflexive Banach space, $\mathcal{M}_\varepsilon(x,v)$ is integrably bounded (see Lemma 4.6 below) and μ_ζ is nonatomic.

4.2. Lipschitz continuity of probability function (4.1)

In this section, we show that the probability function (4.1) is locally Lipschitz continuous. First, we provide the following lemma which establishes a variational upper-estimate for the basic subdifferential of the probability function (4.1) in terms of the subgradients for the basic subdifferential of the enlargements.

Lemma 4.2 (Approximation of subgradients) *Consider the probability function φ defined in (4.1), and the family of probability functions φ_ε given by (4.3). Then, for all $x \in U$*

$$\varphi(x) = \inf_{\varepsilon>0} \varphi_\varepsilon(x). \tag{4.18}$$

$$\partial^{\mathsf{b}}\varphi(x) \subseteq \left\{ x^* \in \mathcal{X}^* : \begin{array}{c} \text{There exist } \exists x_k \to x, \varepsilon_k \to 0^+ \\ \text{and } x_k^* \in \partial^{\mathsf{b}}\varphi_{\varepsilon_k}(x_k) \text{ s.t. } x_k^* \rightharpoonup x^* \end{array} \right\}. \tag{4.19}$$

Proof. A direct application of the continuity of the probability measure shows (4.18). Now, consider a point $x^* \in \partial^{\mathsf{b}}\varphi(x)$, it follows from definition that there are sequences $x_k \to x$ with $\varphi(x_k) \to \varphi(x)$ and $x_k^* \rightharpoonup x^*$ such that $x_k^* \in \partial^{\mathsf{r}}\varphi(x_k)$. Hence, using [56, Lemma 2.1] for each point x_k^* we can get sequences $x_{k,j} \to x_k$ with $\varphi(x_{k,j}) \to \varphi(x_k)$ and $x_{k,j}^* \to x_k^*$ such that $x_{k,j}^* \in \partial^{\mathsf{r}}\varphi_{\varepsilon_{k,j}}(x_{k,j})$. Now, since $\mathcal{X}$ is reflexive and separable we have that the weak*-topology is metrizable on bounded sets (see, e.g., [25]), and it allows us to use a diagonal argument to conclude the result. ☐

In order to establish continuity of the radial functions (4.8) and (4.9) and consequently, boundedness of the sets $\mathcal{M}_\varepsilon(x,v)$ we define the following continuity property for set-valued mappings.

Definition 4.2 *We say that a set-valued mapping $\mathcal{S}$ has the interior continuity property on $U \subseteq \mathcal{X}$, if for every $x \in U$ and $z \in \mathrm{int}(\mathcal{S}(x))$ there exists $r > 0$ such that*

$$\mathbb{B}_r(z) \subseteq \mathcal{S}(x'), \text{ for all } x' \in \mathbb{B}_r(x).$$

Remark 4.2 *We notice that when $\mathcal{S}$ has the interior continuity property then*

- *we can replace Assumption (H) Item b) by $\mathcal{S}$ being locally Lipschitz-like at $z \in \mathrm{bd}\,\mathcal{S}(x)$ while maintaining Lemma 2.1 and the subsequent results. Indeed, if $z \in \mathrm{int}\,\mathcal{S}(x)$ we have thus that $d(z', \mathcal{S}(x')) = 0$ for all (x', z') close enough to (x, z), and consequently the function u defined in Lemma 2.1 is trivially locally Lipschitz.*
- *when computing $\|D^*\mathcal{S}(x,z)\|$ in the η-growth condition in Definition 4.1, the nontrivial*

points are when $z \in \operatorname{bd} \mathcal{S}(x)$ since if $z \in \operatorname{int} \mathcal{S}(x)$, due to the interior continuity property of $\mathcal{S}$, we have that $(x, z) \in \operatorname{int}(\operatorname{gph} \mathcal{S})$ and in consequence $D^*\mathcal{S}(x, z)(z^*)$ is either empty or $0 \in \mathcal{X}$.

Lemma 4.3 *Let each $\mathcal{S}_i$ in the family of set-valued mappings satisfy Assumption (H) and have the interior continuity property on U. Then, there is an open neighborhood $U' \subseteq U$, such that for every sequence $[0, +\infty) \times U' \times \mathbb{S}^{m-1} \ni (\varepsilon_k, x_k, v_k) \to (\varepsilon, x, v) \in [0, +\infty) \times U' \times \mathbb{S}^{m-1}$ we have that*

$$\rho_i^\varepsilon(x, v) = \lim_{k\to+\infty} \rho_i^{\varepsilon_k}(x_k, v_k) \text{ for each } i = 1, \ldots, s. \tag{4.20}$$

Furthermore,

$$\rho_\varepsilon(x, v) = \lim_{k\to+\infty} \rho_{\varepsilon_k}(x_k, v_k). \tag{4.21}$$

Proof. Fix i and consider $(\varepsilon_k, x_k, v_k) \to (\varepsilon, x, v)$. Let us first prove equality (4.20) by assuming that the sequence $\rho_i^{\varepsilon_k}(x_k, v_k)$ diverges. Suppose by contradiction that $\rho_i^\varepsilon(x, v) < +\infty$ and consider $r > \rho_i^\varepsilon(x, v)$. By Lemma 4.1 we have that $\mathrm{d}(rLv, \mathcal{S}_i(x)) > \varepsilon$, so by continuity of the distance function (recall Lemma 2.1) we have that $\mathrm{d}(rLv_k, \mathcal{S}_i(x_k)) > \varepsilon$ for k large enough, which, again by Lemma 4.1 means that $r \geq \rho_i^{\varepsilon_k}(x_k, v_k)$ for k large enough. A contradiction.

Thus, we assume that the sequence $\rho_i^{\varepsilon_k}(x_k, v_k)$ admits a cluster point r'. Then for some subsequence we have that $\rho_i^{\varepsilon_{k_l}}(x_{k_l}, v_{k_l}) \to_l r'$. Let us prove that $r' = \rho_i^\varepsilon(x, v)$. By Lemma 4.1 we have the equality $\mathrm{d}(\rho_i^{\varepsilon_{k_l}}(x_{k_l}, v_{k_l})Lv_{k_l}, \mathcal{S}_i(x_{k_l})) = \varepsilon_{k_l}$ which by continuity of the distance function yield us to the equality $\mathrm{d}(r'Lv, \mathcal{S}_i(x)) = \varepsilon$. From the uniqueness shown in Lemma 4.1 the result for the case $\varepsilon > 0$ follows. Thus it remains to prove it for $\varepsilon = 0$. In this case, by definition of the radial function we have that $r' \leq \rho_i(x, v)$. Suppose by contradiction that $r' < \rho_i(x, v)$. Therefore, $r'Lv \in \operatorname{int}(\mathcal{S}_i(x))$ and hence $\rho_i^{\varepsilon_{k_l}}(x_{k_l}, v_{k_l})Lv \in \operatorname{int}(\mathcal{S}_i(x))$ for l large enough. By the interior continuity property of $\mathcal{S}_i$ (recall Definition 4.2) we have that there exists $\gamma > 0$ such that $(\rho_i^{\varepsilon_{k_l}}(x_{k_l}, v_{k_l}) + \gamma)Lv \in \operatorname{int}(\mathcal{S}_i(x_k))$, which particularly, by definition of the radial function, implies that $\rho_i^{\varepsilon_{k_l}}(x_{k_l}, v_{k_l}) + \gamma \leq \rho_i(x_{k_l}, v_{k_l})$ for l large enough, which is a contradiction. Therefore, we have that $r' = \rho_i(x, v)$ and since this holds true for all possible cluster points we conclude (4.20).

Now let us prove (4.21). On the one hand, there is some $i \in \{1, \ldots, s\}$ such that $\rho_\varepsilon(x, v) = \rho_i^\varepsilon(x, v)$ which together with (4.20) lead us to $\rho_\varepsilon(x, v) = \lim_k \rho_i^{\varepsilon_k}(x_k, v_k)$. Since, by definition, $\rho_i^{\varepsilon_k}(x_k, v_k) \geq \rho_{\varepsilon_k}(x_k, v_k)$ for all k, we conclude that $\rho_\varepsilon(x, v) \geq \lim_k \rho_i^{\varepsilon_k}(x_k, v_k)$. On the other hand, for each k there exists $i_k \in \{1, \ldots, s\}$ such that $\rho_{\varepsilon_k}(x_k, v_k) = \rho_{i_k}^{\varepsilon_k}(x_k, v_k)$. Under subsequence, we may assume that $\rho_{\varepsilon_k}(x_k, v_k) = \rho_i^{\varepsilon_k}(x_k, v_k)$ for some fixed i. By taking limits on this last equality and thus by (4.20) we obtain that $\lim_k \rho_{\varepsilon_k}(x_k, v_k) = \rho_i^\varepsilon(x, v)$ which by definition of $\rho_\varepsilon(x, v)$ lead us to $\lim_k \rho_{\varepsilon_k}(x_k, v_k) \geq \rho_\varepsilon(x, v)$, concluding the proof of (4.21) and thus of the lemma. □

Before addressing the boundedness of $\mathcal{M}_\varepsilon(x, v)$ we need the following two lemmas concerning the projection mapping.

Lemma 4.4 *Under assumptions (H) the mapping $(x, z) \to P_{\mathcal{S}(x)}(z)$ is continuous on $U \times \mathbb{R}^m$.*

Proof. Consider a sequence $(x_k, z_k) \to (x, z) \in U \times \mathbb{R}^m$. We have that

$$\begin{aligned}\|P_{\mathcal{S}(x_k)}(z_k) - P_{\mathcal{S}(x)}(z)\| &\leq \|P_{\mathcal{S}(x_k)}(z_k) - P_{\mathcal{S}(x_k)}(z)\| + \|P_{\mathcal{S}(x_k)}(z) - P_{\mathcal{S}(x)}(z)\| \\ &\leq \|z_k - z\| + \|P_{\mathcal{S}(x_k)}(z) - P_{\mathcal{S}(x)}(z)\|,\end{aligned}$$

where in the second inequality we used the nonexpansiveness of the projection mapping (see, e.g., []). Now, define $y_k := P_{\mathcal{S}(x_k)}(z)$. By continuity of the distance function (see Lemma 2.1) we can assume that the sequence (y_k) is bounded. Hence, it is enough to show that each cluster point of (y_k) is equal to $P_{\mathcal{S}(x)}(z)$. Indeed, let $y_{k_l} \to y$. By closedness of the graph of $\mathcal{S}$ we have that $y \in \mathcal{S}(x)$. Furthermore, by definition of projection we have that $\|y_{k_l} - z\| = \mathrm{d}(z, \mathcal{S}(x_{k_l}))$ which by continuity of the distance function (Lemma 2.1) yield us to $\|y - z\| = d(z, \mathcal{S}(x))$. Finally, from the uniqueness of the projection onto convex sets we conclude that $y = P_{\mathcal{S}(x)}(z)$, and that ends the proof. □

Lemma 4.5 *Consider $\mathcal{S}$ satisfying Assumption (H) with $0 \in \operatorname{int}\mathcal{S}(x)$ for all $x \in U$. Then there exists a neighborhood of U' of $\bar{x}$ and $r > 0$ such that for all $x \in U'$ we have that*

$$\langle z - P_{\mathcal{S}(x)}(z), z\rangle \geq r\,\mathrm{d}(z, \mathcal{S}(x)), \text{ for all } (x, z) \in U' \times \mathbb{R}^m \tag{4.22}$$

Proof. Let U' and $r > 0$ be such that $r\mathbb{B} \subseteq \mathcal{S}(x)$ for all $x \in U'$. On the one hand, if $z \in \mathcal{S}(x)$ the inequality holds trivially. On the other hand, take $y := r\frac{z - P_{\mathcal{S}(x)}(z)}{\|z - P_{\mathcal{S}(x)}(z)\|}$, so

$$\begin{aligned}\langle z - P_{\mathcal{S}(x)}(z), z\rangle &= \langle z - P_{\mathcal{S}(x)}(z), z - P_{\mathcal{S}(x)}(z)\rangle + \langle z - P_{\mathcal{S}(x)}(z), P_{\mathcal{S}(x)}(z)\rangle \\ &= \|z - P_{\mathcal{S}(x)}(z)\|^2 + \langle z - P_{\mathcal{S}(x)}(z), P_{\mathcal{S}(x)}(z) - y\rangle + \langle z - P_{\mathcal{S}(x)}(z), y\rangle \\ &= \|z - P_{\mathcal{S}(x)}(z)\|^2 + \langle z - P_{\mathcal{S}(x)}(z), P_{\mathcal{S}(x)}(z) - y\rangle + r\|z - P_{\mathcal{S}(x)}(z)\|\end{aligned}$$

Now, we notice that $\langle z - P_{\mathcal{S}(x)}(z), P_{\mathcal{S}(x)}(z) - w\rangle \geq 0$ since $w \in \mathcal{S}(x)$ and by definition of the projection onto convex sets. Therefore,

$$\langle z - P_{\mathcal{S}(x)}(z), z\rangle \geq r\|z - P_{\mathcal{S}(x)}(z)\| = r\,\mathrm{d}(z, \mathcal{S}(x)),$$

concluding the proof. □

Lemma 4.6 *Let each $\mathcal{S}_i$ of the family of set-valued mappings satisfy Assumption (H) at $\bar{x} \in U$ with $0 \in \operatorname{int}(\mathcal{S}_i(x))$ for all $x \in U$, and have the interior continuity property on U. Moreover, assume that the family of set-valued mappings satisfy the η-growth condition at $\bar{x}$ and that (2.10) holds true.*

Then, there is a neighborhood U' of $\bar{x}$ and $\varepsilon' > 0$ such that

$$\sup\{\|x^*\| : x^* \in \mathcal{M}_\varepsilon(x, v),\ v \in \mathbb{S}^{m-1},\ x \in U',\ \varepsilon \in (0, \varepsilon')\} < \infty.$$

Moreover, $\limsup_{(\varepsilon,x,v)\to(0,\bar{x},\bar{v})} \mathcal{M}_\varepsilon(x, v) = \{0\}$ *for all $\bar{v} \in \mathcal{I}(\bar{x})$.*

Proof. Due to the compactness of $\mathbb{S}^{m-1}$ it is enough to show that for all $v \in \mathbb{S}^{m-1}$ there are neighbourhoods U_v of $\bar{x}$, V_v of v, and $\varepsilon_v > 0$ such that

$$\sup\{\|x^*\| : x^* \in \mathcal{M}_\varepsilon(x,v),\ v \in V_v,\ x \in U_v,\ \varepsilon \in (0, \varepsilon_v')\} < \infty. \tag{4.23}$$

Fix $\bar{v} \in \mathbb{S}^{m-1}$ and let us suppose first that $\bar{v} \in \mathcal{F}(\bar{x})$. By Lemma 4.3, there are neighbourhoods $U_{\bar{v}}$ of $\bar{x}$, $V_{\bar{v}}$ of $\bar{v}$, and $\varepsilon_{\bar{v}} > 0$ such that

$$\sup\{\rho_\varepsilon(x,v) : (\varepsilon, x, v) \in W := [0, \varepsilon_{\bar{v}}] \times U_{\bar{v}} \times V_{\bar{v}}\} < +\infty.$$

Particularly, $v \in \mathcal{F}(x)$ for all $(x,v) \in U_{\bar{v}} \times V_{\bar{v}}$. Now, fix $(\varepsilon, x, v) \in W$ and consider a point $w^* \in \mathcal{M}_\varepsilon(x,v)$ of the form:

$$\frac{-\alpha}{\left\langle z_\varepsilon^{x,v} - P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v}), Lv \right\rangle} \cdot x^*, \tag{4.24}$$

for some $i \in T_x^\varepsilon(v)$, $\alpha \in I_\theta(\rho_\varepsilon(x,v), v)$, $x^* \in D^*\mathcal{S}_i(x, P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v}))(P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v}) - z_\varepsilon^{x,v})$ where $z_\varepsilon^{x,v} := \rho_\varepsilon(x,v)Lv$. Since the set $I_\theta(\rho_\varepsilon(x,v), v)$ remains bounded on W, we have that α is uniformly bounded on W, let us say by $\bar{\alpha}$. Now, since $\mathcal{S}_i$ has the local Lipschitz-Like property at $(x, P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v}))$ we have that there exists $\kappa_i \geq 0$ such that

$$\|x^*\| \leq \kappa_i \|z_\varepsilon^{x,v} - P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v})\|.$$

On the other hand, by Lemma 4.5 (shrinking enough the neighborhood $U_{\bar{v}}$) there exists some $r_i > 0$ such that

$$\left\langle z_\varepsilon^{x,v} - P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v}), Lv \right\rangle \geq \frac{r_i}{\rho_\varepsilon(x,v)} \|z_\varepsilon^{x,v} - P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v})\|. \tag{4.25}$$

Therefore, we obtain that $\|w^*\| \leq \frac{\bar{\alpha}\kappa\bar{\rho}}{r}$, where $\kappa := \max \kappa_i$, $r := \min r_i$ and $\bar{\rho} := \sup\{\rho_\varepsilon(x,v) : (\varepsilon, x, v) \in W\}$ which means that (4.23) holds for $\bar{v} \in \mathcal{F}(\bar{x})$.

Now, consider the case when $\bar{v} \in \mathcal{I}(\bar{x})$. Let $\gamma > 0$ and let $l > 0$ be such that the family of $\mathcal{S}_i$ satisfy the η-growth condition at $\bar{x}$ (see Definition 4.1). By Lemma 4.3 we have that there are neighborhoods $U_{\bar{v}}$ of $\bar{x}$, $V_{\bar{v}}$ of $\bar{v}$ and $\varepsilon_{\bar{v}} > 0$ such that $\rho_\varepsilon(x,v) > l$ for all $(\varepsilon, x, v) \in W$. Moreover, by Lemma 4.5, there exists $r > 0$ such that (4.25) holds. Therefore, $w^* \in \mathcal{M}_\varepsilon(x,v)$ in the form of (4.24) satisfies

$$\|w^*\| \leq \bar{\theta}(\rho_\varepsilon(x,v), v) \frac{\rho_\varepsilon(x,v)}{r\|z_\varepsilon^{x,v} - P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v})\|} l\eta(\|z_\varepsilon^{x,v}\|)\|z_\varepsilon^{x,v} - P_{\mathcal{S}_i(x)}(z_\varepsilon^{x,v})\|.$$

Furthermore, for some constant $C > 0$ we have that $\bar{\theta}(\rho_\varepsilon(x,v), v) \leq C(2\rho_\varepsilon(x,v))^{m-1}\bar{f}_\xi(z_\varepsilon^{x,v})$ and thus

$$\|w^*\| \leq \frac{Cl2^{m-1}}{r\|Lv\|^{m-1}} \|z_\varepsilon^{x,v}\|^m \bar{f}_\xi(z_\varepsilon^{x,v})\eta(\|z_\varepsilon^{x,v}\|).$$

Since, $\rho_\varepsilon(x,v)$ can be chosen arbitrarily large (shrinking W if it is necessary) we can assume that $\|w^*\| \leq \gamma$, and that ends the proof. $\square$

Now, we provide the main result of this chapter.

Theorem 4.2 *Consider each $\mathcal{S}_i$ in the family of set-valued mappings satisfying Assumption*

(H) *at* $\bar{x} \in U$ *with* $0 \in \operatorname{int}(\mathcal{S}_i(x))$ *for all* $x \in U$ *and having the interior continuity property on* U*. Moreover, assume that the family of set-valued mappings* $\mathcal{S}_i$ *satisfies the* η*-growth condition at* $\bar{x}$ *and that* (2.10) *holds true.*

Then the probability function (4.1) *is locally Lipschitz at* $\bar{x}$ *and on an appropriate neighborhood* U' *of* $\bar{x}$ *it holds:*

$$\partial^{\mathsf{b}}\varphi(x) \subseteq \operatorname{cl}\left(\int\limits_{v\in\mathcal{F}(x)} \mathcal{M}(x,v)d\mu_\zeta(v)\right), \text{ for all } x \in U' \tag{4.26}$$

where, $\mathcal{M}(x,v)$ *is given by,*

$$\mathcal{M}(x,v) = \left\{ \frac{-\alpha}{\langle z^*, Lv\rangle}\cdot x^* : \begin{array}{c} \alpha \in I_\theta(\rho(x,v),v), z^* \in \mathrm{N}^{\mathsf{b}}_{\mathcal{S}(x)}(\rho(x,v)Lv)\cap\mathbb{S}^{m-1} \\ \\ i \in T_x(v), x^* \in D^*\mathcal{S}_i(x,\rho(x,v)Lv)(-z^*) \end{array} \right\}$$

for all $v \in \mathcal{F}(x)$ *with*

$$T_x(v) = \{i \in \{1,\dots,s\} : \rho_i(x,v) = \rho(x,v)\},$$

and by $\mathcal{M}(x,v) = \{0\}$ *for all* $v \in \mathcal{I}(x)$*.*

Proof. Consider $x^* \in \partial^{\mathsf{b}}\varphi(x)$. We divide the proof into four claims.

Claim 1: There exist sequences $x_k \to x$, $\varepsilon_k \to 0^+$ and $x_k^* \rightharpoonup x^*$ with

$$x_k^* \in \int\limits_{v\in\mathbb{S}^{m-1}} \mathcal{M}_{\varepsilon_k}(x_k,v)d\mu_\zeta(v)$$

Indeed, by Lemma 4.2 there exists $x_k^* \in \partial^{\mathsf{b}}\varphi_{\varepsilon_k}(x_k)$ with $x_k \to x$, $\varepsilon_k \to 0^+$ and $x_k^* \rightharpoonup x^*$. Furthermore, by Remark 4.1, we have that (for large enough k)

$$x_k^* \in C_k := \operatorname{cl}^{w^*}\left(\int\limits_{v\in\mathbb{S}^{m-1}} \mathcal{M}_{\varepsilon_k}(x_k,v)d\mu_\zeta(v)\right).$$

Let us notice that by Lemma 4.6 there is some $k_0 \in \mathbb{N}$ such that the set $\cup_{k\geq k_0} C_k$ is bounded. Since $\mathcal{X}$ is separable and reflexive, the weak*-topology is metrizable on bounded sets allowing us to take sequences $x_{j,k}^* \rightharpoonup x_k^*$ with $x_{j,k}^* \in \int_{v\in\mathbb{S}^{m-1}} \mathcal{M}_{\varepsilon_k}(x_k,v)d\mu_\zeta(v)$ as well as to use a diagonal argument so we can construct the desired sequence.

Claim 2: There exists a sequence of (Bochner) integrable functions $y_k : \mathbb{S}^{m-1} \to \mathcal{X}$ such that

$$y_k^*(v) \in \mathcal{M}_{\varepsilon_k}(x_k,v) \quad \mu_\zeta\text{-a.e. and } x_k^* = \int_{\mathbb{S}^{m-1}} y_k^*(v)d\mu_\zeta(v).$$

Using the definition of the integral of a set-valued mapping we get the existence of such a sequence.

Claim 3: We have that $x^* \in \operatorname{cl}\left(\int_{\mathbb{S}^{m-1}} F(v)d\mu_\zeta(v)\right)$, where $F(v)$ is the set of sequential weak* limits of $\{y_k^*(v)\}$. Moreover, the closure can be omitted when $\mathcal{X}$ is finite-dimensional.

Indeed, by Lemma 4.6, we have that there exists $M > 0$ such that $\|y_k^*(v)\| \leq M$ for almost all $v \in \mathbb{S}^{m-1}$. Now, by [9, Corollary 4.1], we have that $x^* \in \mathrm{cl}^{w^*}\left(\int_{\mathbb{S}^{m-1}} F(v) d\mu_\zeta(v)\right)$, where the closure operation can be omitted if the space $\mathcal{X}$ is finite-dimensional. Furthermore, by Lyapunov convexity theorem, we have that $\mathrm{cl}\left(\int_{\mathbb{S}^{m-1}} F(v)\mu_\zeta\right)$ is convex, so

$$\mathrm{cl}^{w^*}\left(\int_{\mathbb{S}^{m-1}} F(v) d\mu_\zeta\right) = \mathrm{cl}\left(\int_{\mathbb{S}^{m-1}} F(v) d\mu_\zeta\right),$$

and that ends the proof of our claim.

Claim 4: We have that $F(v) \subseteq \mathcal{M}(x,v)$ for almost all $v \in \mathbb{S}^{m-1}$.

Indeed, consider a set of full measure $S \subseteq \mathbb{S}^{m-1}$ such that $y_k^*(v) \in \mathcal{M}_{\varepsilon_k}(x_k, v)$ for all $k \in \mathbb{N}$ and $v \in S$. Then, fix $v \in S$. First, if $v \in \mathcal{I}(x)$, we have by Lemma 4.6 that $F(v) \subseteq \limsup \mathcal{M}_{\varepsilon_k}(x_k, v) \subseteq \{0\}$. Now, assume that $v \in \mathcal{F}(x)$ and consider $y_v^* \in F(v)$. Then there exists a sequence $y_{k_j}^*(v)$ such that $y_{k_j}^*(v) \rightharpoonup y_v^*$ and

$$y_{k_j}^*(v) = \frac{\alpha_j}{\left\langle z_j - P_{\mathcal{S}_{i_j}(x_j)}(z_j), Lv \right\rangle} \cdot x_j^*,$$

for some $i_j \in T_{x_j}^{\varepsilon_j}(v)$, $\alpha_j \in I_\theta(\rho_{\varepsilon_j(x_j,v)}, v)$ and $x_j^* \in D^*\mathcal{S}_{i_j}(x_j, P_{\mathcal{S}_{i_j}(x_j)}(z_j))(P_{\mathcal{S}_{i_j}(x_j)}(z_j) - z_j)$, where $z_j := \rho_{\varepsilon_j}(x_j, v)Lv$. Now, we may assume (by passing to a subsequence), that for some fixed $i \in T_{x_j}^{\varepsilon_j}(v)$,

$$y_{k_j}^*(v) = \cdot \frac{\alpha_j}{\left\langle z_j - P_{\mathcal{S}_i(x_j)}(z_j), Lv \right\rangle} \cdot x_j^*, \tag{4.27}$$

with $\alpha_j \in I_\theta(\rho_{\varepsilon_j}(x_j, v), v)$ and $x_j^* \in D^*\mathcal{S}_i(x_j, P_{\mathcal{S}_i(x_j)}(z_j))(P_{\mathcal{S}_i(x_j)}(z_j) - z_j)$, where $z_j := \rho_{\varepsilon_j}(x_j, v)Lv$.

First we notice that, by Lemma 4.3, we have $\rho_i(x,v) = \rho(x,v)$, that is, $i \in T_x(v)$. Moreover, the functions $\bar{\theta}$, $\underline{\theta}$, defined in (2.27), are upper semicontinuous and lower semicontinuous, respectively. Therefore, we can assume (by passing to a subsequence) that $\alpha_j \to \alpha \in I(\rho(x,v), v)$. Now, define $w_j^* := \frac{z_j - P_{\mathcal{S}_i(x_j)}(z_j)}{\|z_j - P_{\mathcal{S}_i(x_j)}(z_j)\|}$, $v_j^* := \frac{x_j^*}{\|z_j - P_{\mathcal{S}_i(x_j)}(z_j)\|}$. Since, $w_j^* \in \mathbb{R}^m$ and it has unit norm, we can assume that $w_j^* \to z^*$. Now, using the fact that $\mathcal{S}_i$ is locally Lipschitz-like at $(x_j, P_{\mathcal{S}_i(x_j)}(z_j))$, we have that the sequence v_j^* is bounded, and from the fact that $\mathcal{X}$ is reflexive, we can assume that $v_j^* \rightharpoonup v^*$ for some $v^* \in \mathcal{X}^*$. Let us now prove that $z^* \in N_{\mathcal{S}_i(x)}(\rho(x,v)Lv)$ and $x^* \in D^*\mathcal{S}_i(x, \rho(x,v)Lv)(-z^*)$. Indeed,

i) $z^* \in N_{\mathcal{S}_i(x)}(\rho(x,v)Lv)$: Since $\mathcal{S}_i(x_j)$ is closed and convex, we can apply [39, Corollary 1.96] to get that $w_j^* \in \partial_z d(z_j, \mathcal{S}_i(x_j))$ for all $j \in \mathbb{N}$. Now, we have that for all $j \in \mathbb{N}$ and all $w \in \mathbb{R}^m$

$$\langle w_j^*, w - x_j \rangle \leq d(w, \mathcal{S}_i(x_j)) - d(z_j, \mathcal{S}_i(x_j)).$$

Hence, taking limits in the above inequality and recalling that the distance function is Lipschitz continuous (see Lemma 2.1) we can conclude that $z^* \in \partial_z d(\rho(x,v)Lv, \mathcal{S}_i(x))$. Finally, using again [39, Corollary 1.96], we get that $z^* \in N_{\mathcal{S}_i(x)}(\rho(x,v)Lv)$.

ii) $v^* \in D^*\mathcal{S}_i(x, \rho(x,v)Lv)(-z^*)$: First, by definition of coderivative, we have that the sequence $(v_j^*, w_j^*) \in \mathrm{N}^{\mathbf{b}}_{\mathrm{gph}\, S_i}(x_j, z_j)$. Moreover, by [39, Theorem 3.60] the graph of the mapping $(u,v) \to \mathrm{N}^{\mathbf{b}}_{\mathrm{gph}\, \mathcal{S}_i}(u,v)$ is locally closed with respect to the $\|\cdot\| \times w^*$-topology at $(x, \rho(x,v)Lv)$ provided that the SNC property holds at $(x, \rho(x,v)Lv)$. Since, $\mathcal{S}_i$

is Lipschitz-like at $(x, \rho(x,v)Lv)$ we can apply [39, Proposition 1.68] to get that the mapping $\mathcal{S}_i$ is SNC at $(x, \rho(x,v)Lv)$ (see also [39, Definition 1.67]). Therefore, we have that $(v^*, z^*) \in \mathrm{N}^{\mathsf{b}}_{\mathrm{gph}\, \mathcal{S}_i}(x, \rho(x,v)Lv)$, and we can conclude by definition of coderivative.

Finally, by taking limits on (4.27) we get that $y_v^* \in \mathcal{M}(x,v)$. $\square$

For the last result of the section, we apply Theorem 4.2 to prove the local Lipschitz continuity of the following joint probability function

$$\varphi(x) := \mathbb{P}(g_i(x,\xi) \leq 0,\ \forall i = 1, \dots, s), \tag{4.28}$$

with continuously differentiable functions g_i which are quasi-convex on the second argument. To that end, it will be convenient to recall the following: associated with a point of interest $\bar{x}$ such that $g_i(\bar{x}, 0) < 0$, we define the finite and infinite directions with respect to g_i by

$$\begin{aligned} F_i(\bar{x}) &:= \{v \in \mathbb{S}^{m-1} \mid \exists r \geq 0 : g_i(\bar{x}, rLv) = 0\}, \\ I_i(\bar{x}) &:= \{v \in \mathbb{S}^{m-1} \mid \forall r \geq 0 : g_i(\bar{x}, rLv) < 0\}, \end{aligned}$$

respectively. The finite and infinite directions with respect to g can be defined analogously. The so-called radial functions are defined by

$$\rho_i\,(x,v) := \sup\{r > 0 : g_i(x, rLv) \leq 0\} \text{ for all } i = 1, \dots, s$$

and $\rho(x,v) = \min_{i=1,\dots,s} \rho_i(x,v)$. Finally, $T_x(v) := \{i = 1, \dots, s : \rho_i(x,v) = \rho(x,v)\}$ is called the set of active indexes at (x,v).

Definition 4.3 (η-growth condition for smooth functions) *Let $\eta : \mathbb{R} \to [0, +\infty)$ be a non-decreasing mapping such that*

$$\lim_{\|z\| \to +\infty} \|z\|^m \bar{f}_\xi(z)\eta(\|z\|) = 0.$$

We say that a family of continuously differentiable mappings $\{g_i\}_{i=1}^s$ satisfies the η-growth condition for smooth functions at $\bar{x}$ if for some $l > 0$

$$\nabla_z g_i(x,z) \neq 0 \textit{ and } \|\nabla_x g_i(x,z)\| \leq l\eta(\|z\|)\|\nabla_z g_i(x,z)\|,\ \forall x \in \mathbb{B}_{1/l}(\bar{x}),\ \forall z \in \{z : g_i(x,z) = 0\}. \tag{4.29}$$

Corollary 4.1 *Consider the probability function (4.28) where $g_i : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are continuously differentiable, quasi-convex in z for all $i = 1, \dots, s$ and ξ has continuous density distribution. Suppose that a point of interest $\bar{x}$ is such that $g_i(\bar{x}, 0) < 0$ for all $i = 1, \dots, s$ and that the family g_i satisfies the η-growth condition given above at $\bar{x}$. Assume, moreover, that $\operatorname{int}\{z \in \mathbb{R}^m : g_i(x,z) \leq 0\} = \{z \in \mathbb{R}^m : g_i(x,z) < 0\}$ for all $i = 1, \dots, s$. Then the probability function (4.28) is locally Lipschitz at $x \in U'$ and*

$$\partial^{\mathsf{b}}\varphi(x) \subseteq -\int_{v \in F(x)} \bigcup_{i \in T_x(v)} \frac{\theta(\rho(x,v), v)}{\langle \nabla_z g_i(x, \rho(x,v)Lv), Lv\rangle} \nabla_x g_i(x, \rho(x,v)Lv) d\mu_\zeta(v), \textit{ for all } x \in U' \tag{4.30}$$

Furthermore, if $\#T_x(v) = 1$ μ_ζ-almost all $v \in \mathbb{S}^{m-1}$ for all $x \in U'$, then the probability function (4.28) is continuously differentiable for all $x \in U'$ and

$$\nabla\varphi(x) = -\int_{v\in F(\bar{x})} \frac{\theta(\rho(x,v),v)}{\langle \nabla_z g_{T_x(v)}(x,\rho(x,v)Lv), Lv\rangle} \nabla_x g_{T_x(v)}(x,\rho(x,v)Lv) d\mu_\zeta(v), \text{ for all } x \in U'. \tag{4.31}$$

Proof. Let us consider $\mathcal{S}_i(x) := \{z : g_i(x,z) \le 0\}$. By continuity of g_i, quasi-convexity of g_i in z and since $g_i(\bar{x},0) < 0$ there exists a neighborhood U of $\bar{x}$ such that $\mathcal{S}_i$ satisfies Item c) in Assumption (H), $0 \in \operatorname{int}(\mathcal{S}_i(x))$ for all $x \in U$ and such that $\mathcal{S}_i$ has the interior continuity property on U. Let us prove that $\mathcal{S}_i$ is locally Lipschitz-like at $(x,z) \in \operatorname{gph} \mathcal{S}_i$ with $x \in U$. Since $\mathcal{S}_i$ has the interior continuity property on U it is enough to prove it for (x,z) such that $x \in U$ and $z \in \operatorname{bd} \mathcal{S}_i(x) = \{z \in \mathbb{R}^m : g_i(x,z) = 0\}$ (Recall Remark 4.2) then the result follows from [39, Corollary 4.39] upon noting that the assumptions therein are satisfied since in our case the constraint system on $\mathcal{S}_i$ is given by a unique function g_i satisfying (4.29). Now, let us prove that $\mathcal{S}_i$ satisfies the η-growth condition for set-valued mappings at $\bar{x}$. Since g_i satisfies the growth condition in Definition 4.3 at $\bar{x}$, we have that there exists $l > 0$ such that g_i satisfies (4.29). By [39, Corollary 4.35] we have that

$$D^*\mathcal{S}_i(x,z)(-z^*) = \{x^* : x^* = \lambda\nabla_x g_i(x,z), z^* = \lambda\nabla_z g_i(x,z), \lambda \ge 0\},$$

and thus when $z \in \operatorname{bd} \mathcal{S}_i(x)$ (recall Remark 4.2) we have that

$$\|D^*\mathcal{S}_i(x,z)\| \le \frac{\|\nabla_x g_i(x,z)\|}{\|\nabla_z g_i(x,z)\|} \le l\eta(\|z\|).$$

Hence the η-growth condition for set-valued mappings follows by considering the same non-decreasing mapping η. Therefore, since the family $\mathcal{S}_i$ satisfies the assumptions of Theorem 4.2, we have that φ is locally Lipschitz at $x \in U$. Now let us verify (4.30). We have that

$$N_{\mathcal{S}_i(x)}(\rho(x,v)Lv) = \{\lambda\nabla_z g_i(x,\rho(x,v)Lv) : \lambda \ge 0\}$$

and, again by [39, Corollary 4.35],

$$\begin{aligned} &D^*\mathcal{S}_i(x,\rho(x,v)Lv)(-z^*) \\ &\qquad = \{x^* : x^* = \lambda\nabla_x g_i(x,\rho(x,v)Lv), z^* = \lambda\nabla_z g_i(x,\rho(x,v)Lv) \text{ and } \lambda \ge 0\}. \end{aligned}$$

Hence, when $\|z^*\| = 1$ and $z^* \in N_{\mathcal{S}(x)}(\rho(x,v)Lv)$, we obtain that

$$D^*\mathcal{S}(x,\rho(x,v)Lv)(-z^*) = \|\nabla_z g(x,\rho(x,v)Lv)\|^{-1}\nabla_x g(x,\rho(x,v)Lv).$$

Clearly $I_i(x) = \mathcal{I}_i(x)$ entailing $F_i(x) = \mathcal{F}(x)$. And as was stated in [57] when f_ξ is continuous, $\bar{\theta} = \theta$ and $I_\theta(\rho(x,v),v) = \{\theta(\rho(x,v),v)\}$. Hence (4.30) follows from (4.26) where we omitted the closure operator due to $\mathcal{X} = \mathbb{R}^n$. When $\#T_x(v) = 1$ then (4.31) follows due to $\mathcal{X} = \mathbb{R}^n$ and [36, Theorem 4.17]. ∎

Chapter 5

Chebyshev sets: weak projection and nonconvexity estimates

It is well-known, in the context of Hilbert spaces, that nonempty closed convex sets are Chebyshev sets (see, e.g., [10]). What about the converse? It is known that every Chebyshev set is convex provided that the underlying space is finite-dimensional (Bunt's Theorem [14]).

In 1966, Klee [33] proposed the conjecture that non-convex Chebyshev sets exist in (some) infinite-dimensional Hilbert space. In the same work, Klee gave the following partial positive answer.

Theorem 5.1 (Klee (1966)) *If C is Chebyshev and weakly closed, then C is convex.*

In 1969, Asplund [2] gave the following criterion for the convexity of a Chebyshev set in terms of the continuity of the projection.

Theorem 5.2 (Asplund (1969)) *If C is Chebyshev and P_C is norm-weak continuous, then C is convex.*

For more results of this kind and historical details we recommend the book [24].

In the first section of this chapter, we give a partial positive answer to Klee's problem in Theorem 5.7 where we relaxed the concept of projection to the one of weak projection. More specifically, we prove that a closed subset of a Hilbert space is convex if and only if the associated weak projection mapping is single-valued. To this end, we based ourselves on the variational characterization of the convexity given in propositions 2.3, 2.4, and 2.5 exposed in Chapter 2.

In relation to the non-convexity of sets, there are many concepts used to measure the non-convexity of sets, including naturally the Hausdorff distance between the set and its closed convex hull. The effective standard deviation (see Definition 5.1) regards the deviation of the elements of the convex combinations of every given point in the convex hull from the point itself. All these concepts and other classical ones (see [30] and references therein) are then based on the shape of the convex hull of the set. On the side of convex analysis and approximation theory, Klee's Theorem provides criteria for convexity of sets based on the size (indeed, uniqueness) of the associated projections onto the involved set.

In the second section of this chapter, the main theorem is Theorem 5.8 where we show that the Hausdorff distance between the set and its closed convex hull can be fully characterized by means of the simultaneous projection onto the set and its closed convex envelope, by appealing exclusively to convex combinations composed of projections onto the set. The most explicit form of this result is Corollary 5.1 which confirms that the Hausdorff distance between the set and its closed convex hull is proportional to the size (diameter) of the projections onto the set, establishing a quantified version of the Bunt-Klee Theorem.

In this line, the standard deviation function provides a primal estimate of the non-convexity, since it concerns only the apparent shape of the set involved, whereas our estimates are rather of dual type because only the projections are solicited, so it can be calculated relatively easily.

5.1. Nonconvexity estimators

The Shapley-Folkman Theorem was derived by Shapley and Folkman in private communications and was introduced by Starr [51]. In [49], Schneider interprets the theorem by saying that the Minkowski sum is, in some sense, a convexifying operation. A simple proof of this Theorem, involving the conic version of Carathéodory's theorem (see, e.g. [47]), can be found in [59].

Theorem 5.3 (Shapley-Folkman) *Let $C_i \subset \mathbb{R}^m$, $i = 1, \ldots, k$. If*

$$x \in \operatorname{co}\left(\sum_{i=1}^{k} C_i\right) = \sum_{i=1}^{k} \operatorname{co}(C_i)$$

then

$$x \in \sum_{[1,m]\setminus\mathcal{I}_x} C_i + \sum_{\mathcal{I}_x} \operatorname{co}(C_i)$$

where $|\mathcal{I}_x| \leq m$.

The Shapley-Folkman theorem has been used, for example, to deal with non-convexity in economic models of large but finite agents [] and to produce a priori bounds on the duality gap [7]. Quantitatively, the relation between the Minkowski sum of many nonconvex sets and its convex hull is possible in terms of a nonconvexity measure as the following, introduced by Cassels [15].

Definition 5.1 (Effective standard deviation) *For a set $C \subset \mathcal{H}$ we define the function $v_C^2 : \mathcal{H} \to \mathbb{R}_\infty$ as*

$$v_C^2(x) = \inf\left\{\sum_{i=1}^{k} \lambda_i \|x_i - x\|^2 : x_i \in \mathcal{H},\, k \geq 1,\, \lambda \in \Delta_k,\, \sum_{i=1}^{k} \lambda_i x_i = x\right\},$$

then the Effective standard deviation of C is $v^2(C) = \sup\{v^2(x) : x \in \operatorname{co} C\}$.

Equivalently, by the parallelogram's identity, the function v_C^2 can be expressed as

$$v_C^2 = \operatorname{co}\psi_C - \tfrac{1}{2}\|\cdot\|^2 \text{ with } \psi_C(x) := \delta_C + \tfrac{1}{2}\|\cdot\|^2. \tag{5.1}$$

The function ψ_C defined above is called Asplund's function (see [2]). This function reflects many topological and geometric properties of its associated set C as we show in the following result.

Proposition 5.1 *Let $C \subset \mathcal{H}$ be nonempty. Then the following assertions hold:*

(i) *ψ_C is convex if, and only if, C is convex.*

(ii) *ψ_C is lower semi-continuous if, and only if, C is closed.*

(iii) $\operatorname{cl}\psi_C$ *is convex if, and only if,* $\operatorname{cl} C$ *is convex.*

(iv) $\operatorname{cl}^w \psi_C$ *is convex only if* $\operatorname{cl}^w C$ *is convex.*

Proof. The "if" part of (i) and (ii) follows from noticing that $\operatorname{epi}\psi_C = (C \times \mathbb{R}_+) \cap \operatorname{epi}(\frac{1}{2}\|\cdot\|^2)$. To prove the "only if" part of (i) consider $x_1, x_2 \in C$, $\lambda \in [0,1]$ and $x_\lambda := \lambda x_1 + (1-\lambda)x_2$. Hence, by convexity of ψ_C,

$$\delta_C(x_\lambda) \leq \psi_C(x_\lambda) \leq \lambda\psi_C(x_1) + (1-\lambda)\psi_C(x_2) = \tfrac{\lambda}{2}\|x_1\|^2 + \tfrac{1-\lambda}{2}\|x_2\|^2 < \infty.$$

That is, $x_\lambda \in C$. To prove the "only if" part of (ii) consider $x_n \in C$ with $x_n \to x$. Hence, by lower semi-continuity of ψ_C,

$$\delta_C(x) \leq \psi_C(x) \leq \liminf_{n\to\infty} \psi_C(x_n) = \liminf_{n\to\infty} \tfrac{1}{2}\|x_n\|^2 = \tfrac{1}{2}\|x\|^2 < \infty.$$

That is, $x \in C$. Now (iii) follows from (i) upon noticing that $\operatorname{cl}\psi_C = \psi_{\operatorname{cl} C}$ thanks to the continuity of the norm. Finally, to prove (iv) we notice that $\delta_{\operatorname{cl}^w C} \leq \psi_{\operatorname{cl}^w C} \leq \operatorname{cl}^w \psi_C$. First we consider $x_1, x_2 \in C$, $\lambda \in [0,1]$ and $x_\lambda := \lambda x_1 + (1-\lambda)x_2$. Hence,

$$\delta_{\operatorname{cl}^w C}(x_\lambda) \leq \lambda \operatorname{cl}^w \psi_C(x_1) + (1-\lambda)\operatorname{cl}^w \psi_C(x_2) \leq \tfrac{\lambda}{2}\|x_1\|^2 + \tfrac{1-\lambda}{2}\|x_2\|^2 < \infty,$$

that is, $x_\lambda \in \operatorname{cl}^w C$. Now consider $x_1, x_2 \in \operatorname{cl}^w C$ and nets $(x_i^1)_i, (x_j^2)_j \subset C$ such that $x_i^1 \rightharpoonup x_1$ and $x_j^2 \rightharpoonup x_2$. By the first part of the argmunt, $\lambda x_i^1 + (1-\lambda)x_j^2 \in \operatorname{cl}^w C$ for all i, j. Hence by taking weak limits on i and j, we obtain $\lambda x_1 + (1-\lambda)x_2 \in \operatorname{cl}^w C$, concluding the proof.
□

Also, in relation to the Hausdorff distance, we always have that

$$d^2(C, \overline{\operatorname{co}}C) \leq v^2(C) \tag{5.2}$$

since

$$d_C^2(x) \leq v_C^2(x) \text{ for all } x \in \mathcal{H}. \tag{5.3}$$

This last inequality can be proven by observing that (see, e.g, [13, Exercise 2.3.20])

$$\tfrac{1}{2}d_C^2 = \tfrac{1}{2}\|\cdot\|^2 - \psi_C^* \text{ for all } x \in \mathcal{H} \tag{5.4}$$

and hence applying Fenchel's inequality.

The following Theorem is due to Wegmann [58]. A proof, in the case of C compact, can be found in [30].

Theorem 5.4 *Let $\mathcal{H} = \mathbb{R}^m$. If $v(C) = v(x_0)$ for some $x_0 \in \operatorname{ri} C$, then $d(C, \operatorname{co} C) = v(C)$.*
An important property of the effective standard deviation is its subadditivity (see [15, 30]), that is, for $C, D \subset \mathcal{H}$,

$$v^2(C + D) \leq v^2(C) + v^2(D), \tag{5.5}$$

which follows from the fact that for $x = x_1 + x_2 \in \operatorname{co}(C + D)$, we have that

$$v^2_{C+D}(x) \leq v^2_C(x_1) + v^2_D(x_2). \tag{5.6}$$

The following Theorem follows from the subadditivity property of the effective standard deviation stated above for arbitrary Hilbert spaces and from the Shapley-Folkman theorem for finite-dimensional Hilbert spaces.

Theorem 5.5 *[15, 30] Let $C_1, \ldots, C_k \subset \mathcal{H}$. Then*

$$d\left(\frac{1}{k}\sum_{i=1}^{k} C_i, \operatorname{co}\left(\frac{1}{k}\sum_{i=1}^{k} C_i\right)\right) \leq \tfrac{1}{\sqrt{k}} \max_{1 \leq 1 \leq k} v(C_i).$$

If $\mathcal{H} = \mathbb{R}^m$, then

$$d\left(\frac{1}{k}\sum_{i=1}^{k} C_i, \operatorname{co}\left(\frac{1}{k}\sum_{i=1}^{k} C_i\right)\right) \leq \tfrac{\sqrt{\min\{k,m\}}}{k} \max_{1 \leq 1 \leq k} v(C_i).$$

In particular, we can observe from the theorem above, that when $C_1 = \ldots = C_k = C$, $v(C) < \infty$, and $k \to \infty$

$$\tfrac{1}{k}(C + \ldots + C) \to \overline{\operatorname{co}}(C)$$

in the Hausdorff distance with rate $O(\frac{1}{k})$.

5.2. Weak projection and a partial positive answer to Klee's conjecture

In this section we consider a set $C \subset \mathcal{H}$ to be non-empty.

Definition 5.2 (weak projection) *Given a non-empty set $C \subset \mathcal{H}$, we define*

$$P^w_C(x) := \{y : \text{ there exists a net } y_i \in C \text{ such that } y_i \rightharpoonup y \text{ and } \|y_i - x\| \to d_C(x)\}$$

and call the set $P^w_C(x)$ the weak projection of x onto the set C.

Remark 5.1 *Clearly from the definition the set $P^w_C(x)$ is weakly closed and $P_C(x) \subset P^w_C(x)$ for all $x \in \mathcal{H}$. This last inclusion is (possibly) strict in the infinite-dimensional case. For example, consider $\mathcal{H} = \ell^2$, the space of square-summable sequences, and $C = \mathbb{S}_1(0)$ the unit* *sphere in* ℓ^2*. Then* $P_C(0) = C \subsetneq P_C^w(0) = \mathbb{B} = \mathrm{cl}^w C$*.*

The following proposition exposes the subdifferential of the conjugate and weakly lower semi-continuous hull of the Asplund function in terms of the weak projection.

Proposition 5.2 *Let* $C \subset \mathcal{H}$ *be non-empty. Then* ψ_C^* *is continuous on* $\mathcal{H}$ *and*

$$\partial\psi_C^*(x) = \overline{\mathrm{co}}(P_C^w(x)) \text{ and } (\partial \,\mathrm{cl}^w \psi_C)^{-1}(x) = P_C^w(x) \tag{5.7}$$

for all $x \in \mathcal{H}$*. Moreover, provided that* $\mathcal{H}$ *is finite-dimensional, we can omit the closure in the first equality.*

Proof. From the continuity of the distance function and (5.4) we obtain the continuity of ψ_C^* for all $x \in \mathcal{H}$. Now, $y \in (\partial \,\mathrm{cl}^w \psi_C)^{-1}(x)$ is equivalent to

$$\langle x, y\rangle = \mathrm{cl}^w \psi_C(y) + \psi_C^*(x) = \mathrm{cl}^w \psi_C(y) + \tfrac{1}{2}\|x\|^2 - \tfrac{1}{2}d_C^2(x)$$

Upon completing squares, this last equality is equivalent to the existence of a net $y_i \in C$ with $y_i \rightharpoonup y$ such that $\|y_i - x\| \to d_C(x)$. That is, $y \in P_C^w(x)$, yielding the second equality on (5.7). Finally, the first equality in (5.7) and the moreover part follows from the second together with Proposition 2.3. $\square$

The following two results resume the state of the art on Chebyshev sets. The latter one is our main result.

Theorem 5.6 *Let* $C \subset \mathcal{H}$ *be a nonempty set. Then the following are equivalent, and each implies the convexity of* $\mathrm{cl}^w C$*:*

(i) $P_C^w(x)$ *is convex for all* $x \in \mathcal{H}$*.*

(ii) $P_C^w(x)$ *is convex for all* x *in a convex dense subset of* $\mathcal{H}$*.*

(iii) *The function* $\mathrm{cl}^w \psi_C$ *is convex.*

Proof. That (i) implies (ii) is clear. If $P_C^w(x)$ is convex for all $x \in D$, where D is a convex dense subset of $\mathcal{H}$, then, by the second equality in (5.7) we have that $(\partial \,\mathrm{cl}^w \psi_C)^{-1}(x)$ is convex for all $x \in D$ and from Proposition 2.5 we obtain the convexity of $\mathrm{cl}^w \psi_C$. Now, if $\mathrm{cl}^w \psi_C$ is convex, then $\partial\psi_C^*(x) = (\partial \,\mathrm{cl}^w \psi_C)^{-1}(x)$ for all $x \in \mathcal{H}$. Hence, we obtain (i) by (5.7). Finally, from Proposition 2.5, each of the statements of the theorem implies the convexity of $\mathrm{cl}^w C$. $\square$

Theorem 5.7 *Let* $C \subset \mathcal{H}$ *be a proximinal set. Then the following are equivalent:*

(i) C *is convex.*

(ii) $P_C^w(x)$ *is a singleton for all* $x \in \mathcal{H}$*.*

(iii) d_C^2 *is Gâteaux differentiable on* $\mathcal{H}$*.*

(iv) d_C^2 *is Fréchet differentiable on* $\mathcal{H}$*.*

(v) *For all $x \in \mathcal{H}$, there exists a selection of P_C, norm-weak continuous at x.*

(vi) *For all $x \in \mathcal{H}$, there exists a selection of P_C, norm-norm continuous at x.*

Proof. Obviously, (vi) implies (v) and (iv) implies (iii). To prove (v) implies (iv), notice that any selection of $P_C(x)$ is also a selection of $\partial\psi_C^*(x)$, hence, by [60, Theorem 3.2.2], ψ_C^* is Fréchet differentiable. Thus, by formula (5.4) we get that d_C^2 is Fréchet differentiable. (iii) implies (ii) follows from (5.4) and the first equality in (5.7). Now, due to Proposition (5.7), (ii) implies that $\partial\psi_C^*(x) = P_C(x) = (\partial\psi_C)^{-1}(x)$ for all $x \in \mathcal{H}$. Thus, since $\psi_C^* \in \Gamma_0(\mathcal{H})$,

$$\partial\psi_C^{**}(x) = (\partial\psi_C^*)^{-1}(x) = \partial\psi_C(x) \text{ for all } x \in \mathcal{H}.$$

By proposition 2.4 we conclude that the (lower semi-continuous) function ψ_C is convex, hence, by Proposition 5.1, we obtain (i). The proof is complete since the convexity of the closed set C implies the continuity of the (single-valued) mapping $P_C(x)$ for all $x \in \mathcal{H}$ (see, e.g., [60]). $\square$

Remark 5.2 *It is still unknown whether we can replace the weak projection P_C^w in condition (ii) of the theorem 5.7 by P_C.*

5.3. Quantification of Klee's theorem

In this section we consider a set $C \subset \mathcal{H}$ to be non-empty and such that $v(C) < \infty$.

Theorem 5.8 *Let $C \subset \mathcal{H}$ be weakly closed. Then*

$$d^2(C, \overline{\text{co}}C) \le v^2(C) = \sup\left\{d_C^2(x) - d_{\overline{\text{co}}C}^2(x) : x \in \mathcal{H}\right\} = \sup\left\{\|P_C(x) - P_{\overline{\text{co}}C}(x)\|^2 : x \in \Pi\right\}, \tag{5.8}$$

where $\Pi := \{x \in \mathcal{H} : P_{\overline{\text{co}}C}(x) \in \overline{\text{co}}(P_C(x))\}$.

Proof. By (5.1), we can rewrite the effective standard deviation as the following

$$v^2(C) = -2\inf\left\{\delta_{\overline{\text{co}}C}(x) + \tfrac{1}{2}\|x\|^2 - \text{co}\,\psi_C(x) : x \in \mathcal{H}\right\}.$$

Now, Proposition 2.1 and Moreau-Rockafellar theorem leads us to

$$\begin{aligned} v^2(C) &= -2\inf\left\{\psi_C^*(x) - (\delta_{\overline{\text{co}}C} + \tfrac{1}{2}\|\cdot\|^2)^*(x) : x \in \mathcal{H}\right\} \\ &= -2\inf\left\{\psi_C^*(x) - \mathsf{e}_1\sigma_C(x) : x \in \mathcal{H}\right\}. \end{aligned}$$

By Moreau's decomposition (2.4) and since $\mathsf{e}_1\delta_{\overline{\text{co}}C} = \frac{1}{2}d_{\overline{\text{co}}C}^2$, it follows that

$$v^2(C) = -2\inf\left\{\psi_C^*(x) + \tfrac{1}{2}d_{\overline{\text{co}}C}^2(x) - \tfrac{1}{2}\|x\|^2 : x \in \mathcal{H}\right\}, \tag{5.9}$$

yielding the first equality in (5.8) by recalling (5.4). Now, suppose that the right-hand-side of (5.9) is attained at $\bar{x} \in \mathcal{H}$. Then $0 \in \partial\psi_C^*(\bar{x}) - P_{\overline{\text{co}}C}(\bar{x})$. That is, $\bar{x} \in \Pi$, by recalling (5.7).

Rewritting $P_{\overline{\text{co}}C}(\bar{x}) = \sum_{i=1}^k \lambda_i y_i$ with $y_i \in P_C(\bar{x})$ and $\lambda \in \Delta_k$, we notice that since

$$\langle \bar{x} - P_{\overline{\text{co}}C}(\bar{x}), y_i - P_{\overline{\text{co}}C}(\bar{x})\rangle \leq 0 \, \forall i = 1, \ldots, k \text{ and } \sum_{i=1}^{k} \langle \bar{x} - P_{\overline{\text{co}}C}(\bar{x}), \lambda_i y_i - P_{\overline{\text{co}}C}(\bar{x})\rangle = 0,$$

we have that $\langle \bar{x} - P_{\overline{\text{co}}C}(\bar{x}), y_i - P_{\overline{\text{co}}C}(\bar{x})\rangle = 0$ for all $i = 1, \ldots, k$. Consequently,

$$\begin{aligned} v^2(C) &= -2\psi_C^*(\bar{x}) - d^2_{\overline{\text{co}}C}(\bar{x}) + \|\bar{x}\|^2 = d_C^2(\bar{x}) - d^2_{\overline{\text{co}}C}(\bar{x}) = \|\bar{x} - P_C(\bar{x})\|^2 - \|\bar{x} - P_{\overline{\text{co}}C}(\bar{x})\|^2 \\ &= \|P_C(\bar{x}) - P_{\overline{\text{co}}C}(\bar{x})\|^2 \leq \sup\left\{\|P_C(x) - P_{\overline{\text{co}}C}(x)\|^2 : x \in \Pi\right\}. \end{aligned}$$

Moreover, by the first equality, we have that

$$\sup\left\{\|P_C(x) - P_{\overline{\text{co}}C}(x)\|^2 : x \in \Pi\right\} \leq \sup\left\{d_C^2(x) - d^2_{\overline{\text{co}}C}(x) : x \in \mathcal{H}\right\} = v^2(C),$$

concluding the proof. □

Theorem 5.8 entails the following quantification of the Blunt-Klee Theorem.

Corollary 5.1 *Let $C \subset \mathcal{H}$ be weakly closed. Then*

$$d(C, \overline{\text{co}}C) \leq \sup\left\{\text{diam}(P_C(x)) : x \in \mathcal{H}\right\}.$$

Proof. By the previous theorem, we have that

$$d(C, \overline{\text{co}}C) \leq \sup\left\{\|P_C(x) - P_{\overline{\text{co}}C}(x)\|^2 : x \in \Pi\right\}.$$

By the definition of Π, the elements in $x \in \Pi$ can be rewritten as $P_{\overline{\text{co}}C}(x) = \sum_{i=1}^k \lambda_i y_i$ where $y_i \in P_C(x)$ and $\lambda \in \Delta_k$, hence, the last equality yields

$$d(C, \overline{\text{co}}C) \leq \sup\left\{\sum_{i=1}^{k} \lambda_i \|P_C(x) - y_i\|^2 : x \in \Pi\right\} \leq \sup\left\{\text{diam}(P_C(x)) : x \in \Pi\right\},$$

from which we can conclude the proof. □

We present below an attainment result of the "dual" and "primal" problems of Theorem 5.8 and, as a corollary, we obtain Theorem 5.4.

Theorem 5.9 *Let $C \subset \mathcal{H}$ be weakly closed. Then the following two assertions are equivalent:*

(i) *$v^2(C)$ is attained at $\bar{x} \in \text{dom}\, \partial(\text{co}\, \psi_C)$.*

(ii) *$\sup\left\{\|P_C(x) - P_{\overline{\text{co}}C}(x)\|^2 : x \in \Pi\right\}$ is attained at $\bar{y} \in \Pi$ such that $P_{\overline{\text{co}}C}(\bar{y}) = \bar{x}$.*

Proof. Assume first (i). Since there exists some $\bar{y} \in \partial(\text{co}\, \psi_C)(\bar{x})$, by Proposition 5.4, we have that

$$\bar{x} \in \partial \psi_C^*(\bar{y}) = \overline{\text{co}}(P_C(\bar{y}))$$

as well as the equality $\langle \bar{x}, \bar{y}\rangle = \text{co}\, \psi_C(\bar{x}) + \psi_C^*(\bar{y})$. Therefore, using Theorem 5.8 and the fact that $\bar{x} \in \overline{\text{co}}C$, we obtain the following

$$\begin{aligned}
\sup\left\{\|P_C(x) - P_{\overline{\text{co}}C}(x)\|^2 : x \in \Pi\right\} &= v_C^2(\bar{x}) = \text{co}\,\psi_C(\bar{x}) - \tfrac{1}{2}\|\bar{x}\|^2 = \langle \bar{x}, \bar{y}\rangle - \psi_C^*(\bar{y}) - \tfrac{1}{2}\|\bar{x}\|^2 \\
&= \langle \bar{x}, \bar{y}\rangle - \tfrac{1}{2}\|\bar{y}\|^2 + \tfrac{1}{2}d_C^2(\bar{y}) - \tfrac{1}{2}\|\bar{x}\|^2 = \tfrac{1}{2}d_C^2(\bar{y}) - \tfrac{1}{2}\|\bar{x} - \bar{y}\|^2 \\
&\leq d_C^2(\bar{y}) - d_{\overline{\text{co}}C}^2(\bar{y}) \leq \sup\left\{d_C^2(x) - d_{\overline{\text{co}}C}^2(x) : x \in \mathcal{H}\right\} \\
&= v^2(C),
\end{aligned}$$

from which we deduce $d_{\overline{\text{co}}C}^2(\bar{y}) = \|\bar{y} - \bar{x}\|^2$, that is, $\bar{x} = P_{\overline{\text{co}}C}(\bar{y})$ and $\bar{y} \in \Pi$. Hence, assertion (ii) follows. Now, we assume (ii). We have that,

$$\sup\left\{\|P_C(x) - P_{\overline{\text{co}}C}(x)\|^2 : x \in \Pi\right\} = \|P_C(\bar{y}) - \bar{x}\|^2.$$

where $\bar{x} = P_{\overline{\text{co}}C}(\bar{y}) \in \overline{\text{co}}(P_C(\bar{y}))(= \partial\psi_C^*(\bar{y}))$. Then, again by Theorem 5.8, we have that

$$\begin{aligned}
v^2(C) &\leq \|P_C(\bar{y}) - P_{\overline{\text{co}}C}(\bar{y})\|^2 = \|P_C(\bar{y}) - \bar{y}\|^2 - \|\bar{y} - P_{\overline{\text{co}}C}(\bar{y})\|^2 \\
&= -2\psi_C^*(\bar{y}) + \|\bar{y}\|^2 - d_{\overline{\text{co}}C}^2(\bar{y}) = 2\overline{\text{co}}\psi_C(\bar{x}) - 2\langle \bar{x}, \bar{y}\rangle + \|\bar{y}\|^2 - d_{\overline{\text{co}}C}^2(\bar{y}) \\
&\leq 2\overline{\text{co}}\psi_C(\bar{x}) - \|\bar{x}\|^2 = v_C^2(\bar{x}).
\end{aligned}$$

That is, we have that v_C^2 is attained at $\bar{x}$, concluding the proof. □

Corollary 5.2 *Let $\mathcal{H} = \mathbb{R}^m$ and C be a closed set such that $v(C)$ is attained for some $\bar{x} \in \text{ri}(C)$. Then $v^2(C) = d^2(C, \overline{\text{co}}C)$.*

Proof. Assume now that $v_C(\bar{x}) = v(C)$ for some $\bar{x} \in \text{ri}(C)$. Then $v(C) = d(C, \overline{\text{co}}C)$. We may restrict ourselves to $\text{aff}\, C$. Since $\overline{\text{co}}C = \text{cl}(\text{dom}\,\text{co}\,\psi_C) = \overline{\text{co}}(\text{dom}\,\psi)$ and $\text{ri}(\overline{\text{co}}C) = \text{ri}(\text{dom}\,\text{co}\,\psi_C)$ we have that $\bar{x} \in \text{ri}(\text{dom}\,\text{co}\,\psi_C)$ and so $\bar{x} \in \text{dom}\,\partial(\text{co}\,\psi_C)$. Thus, according to Theorem 5.9, The supremum $\sup_{x\in\Pi}\|P_C(x) - P_{\overline{\text{co}}C}(x)\|$ is attained at $\bar{y} \in \Pi \cap \text{aff}\, C$ such that $P_{\overline{\text{co}}C}(\bar{y}) = \bar{x}$. More precisely, because $\bar{x} \in \text{ri}(\overline{\text{co}}C)$, we must have that $\bar{y} \in \overline{\text{co}}C$ and we obtain that

$$v^2(C) = \sup\left\{\|P_C(x) - P_{\overline{\text{co}}C}(x)\|^2 : x \in \Pi\right\} = \|P_C(\bar{y}) - \bar{y}\|^2 = d_C^2(\bar{y}) \leq d^2(C, \overline{\text{co}}C).$$

Since the opposite inequality always holds, we conclude the proof. □

Chapter 6

Conclusions and future work

In Chapter 3 we have suggested a regularization (3.6) of the probability function given in (3.1) employing the Moreau envelope. We have shown that this regularization inherits properties of the Moreau envelope itself, namely convergence to the original probability function. Under appropriate, yet mild conditions, convergence can be understood in the Painlevé-Kuratowski or Mosco sense. Furthermore, in a finite-dimensional setting, we established continuous differentiability of the regularized probability functions and asymptotic consistency of the resulting gradients which guarantees that accumulation points of sequences of critical points converge to critical points of the original probability function. Once again in infinite-dimensions, we managed to establish convergence of approximated optimization problems to original problems. Furthermore, the abstract initial conic formulation allows representing general inequality systems inside the probability function, for example, semidefinite constraints.

It is expected that our convergence results established in Section 3.4 provide the first steps in the development of general algorithms for solving probabilistic constraint programming problems. Besides, the available gradient formula given in Theorem 3.2 provides a suitable representation of the gradient to implement (nonlinear) first descent methods.

In Chapter 4 we considered an enlargement (4.3) of the probability function generated by a set-valued mapping (4.1). We proved that this enlargement is locally Lipschitz continuous and, by approximation, we established the local Lipschitz continuity of the probability function (4.1). In addition, we proved the local Lipschitz continuity of a joint probability function given by a inequality system with relaxed convexity assumption, and subsequently by assuming smooth data, its continuous differentiability with relaxed convexity assumption.

In Chapter 5 we gave a partial positive answer to Klee's conjecture in terms of a relaxed concept of the metric projection, by applying known variational characterizations of the convexity of a function to the Asplund function of a given set. Furthermore, via the nonconvexity measure called effective standard deviation, we achieved an estimate of the Hausdorff distance between a set and its closed convex envelope only in terms of the metric projection, and at the same time we obtained a quantification of Bunt-Klee's Theorem, all this under the condition that the effective standard deviation of the given set is finite.

Moreover, we expect to continue the research of the chapter 5 by defining the following truncated version of the effective standard deviation,

$$v_\rho(C) := \sup\{v_C(x) : x \in \overline{\mathrm{co}}C \cap \rho\mathbb{B}\} < +\infty$$

for all $\rho > 0$, similar to the truncated Hausdorff distance (see, e.g., [46]), and relax the condition of finite standard effective deviation.

Bibliography

[1] C. D. Aliprantis and K. C. Border, *Infinite dimensional analysis*, 3rd ed. Springer, Berlin, 2006, a hitchhiker's guide.

[2] E. Asplund, "Cebyshev sets in hilbert space," *Transactions of the American Mathematical Society*, vol. 144, pp. 235–240, 1969.

[3] H. Attouch, "Variational convergence for functions and operators," *Applicable Mathematics Series*, 1984.

[4] H. Attouch and D. Azé, "Approximation and regularization of arbitrary functions in Hilbert spaces by the Lasry-Lions method," *Ann. Inst. H. Poincaré Anal. Non Linéaire*, vol. 10, no. 3, pp. 289–312, 1993.

[5] H. Attouch, G. Buttazzo, and G. Michaille, *Variational analysis in Sobolev and BV spaces*, 2nd ed., ser. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2014, vol. 17, applications to PDEs and optimization.

[6] J. P. Aubin, "Lipschitz behavior of solutions to convex minimization problems," *Mathematics of Operations Research*, vol. 9, pp. 87–111, 02 1984.

[7] J. P. Aubin and I. Ekeland, "Estimates of the duality gap in nonconvex optimization," *Mathematics of Operations Research*, vol. 1, no. 3, pp. 225–245, 1976. [Online]. Available: http://www.jstor.org/stable/3689565

[8] J. P. Aubin and H. Frankowska, *Set-Valued Analysis*. Springer, 2009.

[9] E. J. Balder and A. R. Sambucini, "Fatou's lemma for multifunctions with unbounded values in a dual space," *J. Convex Anal.*, vol. 12, no. 2, pp. 383–395, 2005.

[10] H. Bauschke and P. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, ser. CMS Books in mathematics. Springer Berlin, 2011.

[11] G. Beer, *Topologies on Closed and Closed Convex Sets*, ser. Mathematics and Its Applications. Springer, Dordrecht, 1993, vol. 268.

[12] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, ser. Springer Series in Operations Research. Springer-Verlag, New York, 2000. [Online]. Available: https://doi.org/10.1007/978-1-4612-1394-9

[13] J. M. Borwein and J. D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, ser. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2010.

[14] L. N. H. Bunt, *Bijdrage tot de theorie der convexe puntverzamelingen*. Rijksuniversiteit

te Groningen, 1934.

[15] J. W. S. Cassels, "Measures of the non-convexity of sets and the shapley–folkman–starr theorem," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 78, no. 3, p. 433–436, 1975.

[16] A. Charnes, W. Cooper, and G. Symonds, "Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil," *Management Science*, vol. 4, pp. 235–263, 1958.

[17] R. Correa, A. Hantoute, and P. Pérez-Aros, "Characterizations of the subdifferential of convex integral functions under qualification conditions," *J. Funct. Anal.*, vol. 277, no. 1, pp. 227–254, 2019. [Online]. Available: https://doi.org/10.1016/j.jfa.2019.02.007

[18] R. Correa, Y. García, and A. Hantoute, "Integration formulas via the fenchel subdifferential of nonconvex functions," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 75, p. 1188–1201, 02 2012.

[19] R. Correa and A. Hantoute, "New formulas for the fenchel subdifferential of the conjugate function," *Set-Valued and Variational Analysis*, vol. 18, no. 3, pp. 405–422, 2010. [Online]. Available: https://doi.org/10.1007/s11228-010-0152-x

[20] R. Correa, A. Hantoute, and P. Pérez-Aros, "On the klee–saint raymond's characterization of convexity," *SIAM Journal on Optimization*, vol. 26, pp. 1312–1321, 01 2016.

[21] R. Correa, A. Hantoute, and P. Pérez-Aros, "Qualification conditions-free characterizations of the ε-subdifferential of convex integral functions," *Appl. Math. Optim.*, 2019. [Online]. Available: https://doi.org/10.1007/s00245-019-09604-y

[22] R. Correa, A. Hantoute, and P. Pérez-Aros, "Subdifferential calculus rules for possibly nonconvex integral functions," *SIAM Journal on Control and Optimization*, vol. 58, no. 1, pp. 462–484, 2020. [Online]. Available: https://doi.org/10.1137/18M1176476

[23] W. de Oliveira, "The abc of dc programming," *Set-Valued and Variational Analysis*, vol. 28, pp. 679–706, 2020.

[24] F. Deutsch, *Best Approximation in Inner-Product Spaces*, 01 2001.

[25] M. Fabian, P. Habala, P. Hájek, V. Montesinos, and V. Zizler, *Banach space theory*, ser. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011, the basis for linear and nonlinear analysis. [Online]. Available: https://doi.org/10.1007/978-1-4419-7515-7

[26] K. Fang, S. Kotz, and K. W. Ng, *Symmetric multivariate and related distributions*, 1st ed., ser. Monographs on Statistics and Applied Probability. Springer-Science, 1990, vol. 36.

[27] M. H. Farshbaf-Shaker, R. Henrion, and D. Hömberg, "Properties of chance constraints in infinite dimensions with an application to pde constrained optimization," *Set Valued and Variational Analysis*, vol. 26, no. 4, pp. 821–841, 2018.

[28] J. Fletcher and W. B. Moors, "Chebyshev sets," *Journal of the Australian Mathematical Society*, vol. 98, no. 2, p. 161–231, 2015.

[29] G. B. Folland, *Real analysis. Modern techniques and their applications.* Pure and Applied Mathematics, 1999.

[30] M. Fradelizi, M. Madiman, A. Marsiglietti, and A. Zvavitch, "The convexification effect of Minkowski summation," *arXiv e-prints*, p. arXiv:1704.05486, Apr. 2017.

[31] A. Hantoute, R. Henrion, and P. Pérez-Aros, "Subdifferential characterization of probability functions under Gaussian distribution," *Mathematical Programming*, vol. 174, no. 1-2, pp. 167–194, 2019.

[32] J.-B. Hiriart-Urruty, "Generalized differentiability, duality and optimization for problems dealing with differences of convex functions," in *Convexity and duality in optimization (Groningen, 1984)*, ser. Lecture Notes in Econom. and Math. Systems. Springer, Berlin, 1985, vol. 256, pp. 37–70. [Online]. Available: https://doi.org/10.1007/978-3-642-45610-7_3

[33] V. Klee, "Convexity of chebyshev sets," *Math. Ann.*, pp. 292–304, 1961.

[34] Z. M. Landsman and E. A. Valdez, "Tail conditional expectations for elliptical distributions," *North American Actuarial Journal*, vol. 7, no. 4, pp. 55–71, 2013.

[35] B. Miller and H. Wagner, "Chance constrained programming with joint constraints," *Operations Research*, vol. 13, pp. 930–945, 1965.

[36] B. S. Mordukhovich, *Variational analysis and applications*, ser. Springer Monographs in Mathematics. Springer, Cham, 2018. [Online]. Available: https://doi.org/10.1007/978-3-319-92775-6

[37] B. S. Mordukhovich and N. M. Nam, "Subgradient of distance functions with applications to Lipschitzian stability," *Math. Program.*, vol. 104, no. 2-3, Ser. B, pp. 635–668, 2005. [Online]. Available: https://doi.org/10.1007/s10107-005-0632-1

[38] B. S. Mordukhovich and P. Pérez-Aros, "Sensitivity analysis of stochastic constraint and variational systems via generalized differentiation," 2021. [Online]. Available: https://arxiv.org/abs/2112.05571

[39] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation I. Basic Theory.*, ser. Grundlehren der mathematischen Wissenschaften. Springer, Heidelberg, 2006, vol. 330.

[40] U. Mosco, "On the continuity of the young-fenchel transform," *Journal of Mathematical Analysis and Applications*, vol. 35, no. 3, pp. 518–535, 1971. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022247X71902009

[41] A. Pommelet, "Analyse convexe et théorie de morse," Ph.D. dissertation, Université de Paris IX, 1982.

[42] A. Prékopa, "On probabilistic constrained programming," in *Proceedings of the Princeton Symposium on Math. Prog.*, H. Kuhn, Ed., vol. 28, 1970, pp. 113–138.

[43] A. Prékopa, "Logarithmic concave measures with applications to stochastic programming," *Acta Scientiarium Mathematicarum (Szeged)*, vol. 32, pp. 301–316, 1971.

[44] A. Prékopa, "Probabilistic programming," in *Stochastic Programming*, ser. Handbooks in Operations Research and Management Science, A. Ruszczyński and A. Shapiro, Eds. Elsevier, Amsterdam, 2003, vol. 10, pp. 267–351.

[45] P. Pérez-Aros and E. Vilches, "Moreau envelope of supremum functions with applications to infinite and stochastic programming," *SIAM Journal on Optimization*, vol. 31, no. 3,

pp. 1635–1657, 2021. [Online]. Available: https://doi.org/10.1137/20M1373517

[46] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, ser. Grundlehren Math. Wiss. Springer-Verlag, Berlin, 1998, vol. 317.

[47] R. T. Rockafellar, *Convex Analysis*, 1st ed. Princeton University Press, 1970.

[48] J. Royset and E. Polak, "Implementable algorithm for stochastic optimization using sample average approximations," *Journal of Optimization Theory and Applications*, vol. 122, no. 1, pp. 157–184, 2004.

[49] R. Schneider, *Convex Bodies: The Brunn–Minkowski Theory*, 2nd ed., ser. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2013.

[50] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on stochastic programming*, 2nd ed., ser. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2014, vol. 9, modeling and theory.

[51] R. M. Starr, "Quasi-equilibria in markets with non-convex preferences," *Econometrica*, vol. 37, no. 1, pp. 25–38, 1969. [Online]. Available: http://www.jstor.org/stable/1909201

[52] W. van Ackooij, "A discussion of probability functions and constraints from a variational perspective," *Set-Valued and Variational Analysis*, vol. 28, no. 4, pp. 585–609, 2020.

[53] W. van Ackooij and R. Henrion, "Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions," *SIAM Journal on Optimization*, vol. 24, no. 4, pp. 1864–1889, 2014.

[54] W. van Ackooij and R. Henrion, "(Sub-) Gradient formulae for probability functions of random inequality systems under Gaussian distribution," *SIAM Journal on Uncertainty Quantification*, vol. 5, no. 1, pp. 63–87, 2017.

[55] W. van Ackooij and J. Malick, "Eventual convexity of probability constraints with elliptical distributions," *Mathematical Programming*, vol. 175, no. 1, pp. 1–27, 2019.

[56] W. van Ackooij and P. Pérez-Aros, "Generalized differentiation of probability functions acting on an infinite system of constraints," *SIAM J. Optim.*, vol. 29, no. 3, pp. 2179–2210, 2019. [Online]. Available: https://doi.org/10.1137/18M1181262

[57] W. van Ackooij and P. Pérez-Aros, "Generalized differentiation of probability functions: Parameter dependent sets given by intersections of convex sets and complements of convex sets," *Applied Mathematics & Optimization*, vol. 85, no. 1, p. 2, 2022. [Online]. Available: https://doi.org/10.1007/s00245-022-09844-5

[58] R. Wegmann, "Einige maßzahlen für nichtkonvexe mengen," *Archiv der Mathematik*, vol. 34, pp. 69–74, 1980.

[59] L. Zhou, "A simple proof of the shapley-folkman theorem," *Economic Theory*, vol. 3, no. 2, pp. 371–372, April 1993.

[60] C. Zălinescu, *Convex Analysis in General spaces*. World Scientific, 2002.