UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

Departamento de Matemática
Santiago-Chile

A variational approach to a cumulative distribution function estimation problem under stochastic ambiguity

Tesis presentada por:

Fernanda Paz Urrea Castillo

Como requisito parcial para optar al título profesional de Ingeniera Civil Matemática y al grado de Magíster en Ciencias Mención Matemática

Director de tesis:

Julio Deride Silva

30 de octubre de 2022

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
Departamento de Matemática
Santiago-Chile

A variational approach to a cumulative distribution function estimation problem under stochastic ambiguity

Memoria presentada por:

Fernanda Paz Urrea Castillo

Como requisito parcial para optar al título profesional de Ingeniera Civil Matemática y al grado de Magíster en Ciencias Mención Matemática

Director de tesis:
Julio Deride Silva

Examinadores:
Luis Briceño Arias
Johannes Royset

30 de octubre de 2022

TÍTULO DE LA TESIS:
A variational approach to a cumulative distribution function estimation problem under stochastic ambiguity.

AUTOR: Fernanda Paz Urrea Castillo.

TESIS, Como requisito parcial para optar al título de Ingeniería Civil Matemática y al grado de Magíster en Ciencias Mención Matemática.

COMISIÓN EVALUADORA:

Integrantes	Firma
Julio Deride Universidad Técnica Federico Santa María, Chile.	________________
Luis Briceño Arias Universidad Técnica Federico Santa María, Chile.	________________
Johannes Royset Naval Postgraduate School, Monterey, California, U.S .	________________

—

Santiago, 30 de octubre de 2022

Agradecimientos

Esta tesis fue parcialmente financiada por el proyecto FONDECYT de iniciación número 11190549.

Resumen

En esta tesis se propone un método para el problema de estimación de funciones de distribución de probabilidad bajo ambigüedad estocástica. La ambigüedad estocástica está representada por un conjunto de incertidumbre en el que se debe encontrar la función de distribución acumulada y el método propuesto considera a las funciones de distribución acumulada como un subconjunto de una clase más grande de funciones, el espacio de las funciones semicontinuas superior. Presentamos varios resultados para introducir las propiedades topológicas del espacio y también para justificar nuestra elección del espacio.

Las herramientas utilizadas para desarrollar este método se basan en la teoría del análisis variacional. En particular, trabajamos sobre el espacio de las funciones semicontinuas superior dotado de la distancia de Attouch Wets para las que se propone una aproximación y en conjunto con el uso de epi-splines (funciones polinomiales a trozos) impulsan un esquema de aproximación lineal a nuestro enfoque.

Implementamos un algoritmo para el caso bivariado que nos permite calcular soluciones al problema aproximado así como también incorporar información suave y condiciones de crecimiento al modelo. Enunciamos condiciones que garantizan la convergencia de los minimizadores cercanos de la sucesión de problemas aproximados en soluciones para el problema original y entregamos una clase de funciones que satisfacen aquella condición. Probamos nuestro algoritmo a modo de prueba de concepto con dos ejemplos distintos, donde analizamos varios parámetros y también realizamos experiencias numéricas para el problema de estimación de la posición de un vehículo submarino no tripulado dadas fuentes de información ruidosas.

Palabras claves: Análisis variacional · Topología de Attouch-Wets · Hipo-convergencia· Epi-convergencia · Funciones semicontiuas superior · Funciones de probabilidad acumulada

Abstract

In this thesis we propose a method to the problem of estimating cumulative distribution functions under stochastic ambiguity. The stochastic ambiguity is represented by a set of uncertainty in which the cumulative distribution function must be found and the proposed method considers cumulative distribution functions as a subset of a bigger class of functions, the space of upper semi-continuous functions. We present several results in order to introduce the topological properties of the space and also, to give justification of our choice.

The tools used to develop this method rely on the theory of variational analysis. In particular, we work on the space of upper semicontinuous functions endowed with the Attouch Wets distance for which an approximation is proposed and altogether with the use of epi-splines (piecewise polynomial functions) drive a linear approximating scheme to our setting.

We implemented an algorithm for the bivariate case that allowed us to compute solutions to the approximating problem as well as incorporating soft information and growing conditions to the model. We give conditions that guarantee the convergence of near-minimizers of the approximating sequence into solutions for the original problem and we hand out a class of functions that satisfy it. We test our algorithm as a proof of concept with two different examples, where we analyze various parameters and in addition, we performed numerical experiences for the problem of estimating the position of an unmanned underwater vehicle given noisy sources of information.

Key words: Variational analysis · Attouch-Wets topology · Set convergence · Hypo-convergence · Epi-convergence · Upper semi-continuous functions · Cumulative distribution functions.

Contents

Glossary

$\mathbb{N}$: Natural numbers without 0.

$\mathcal{N}_\infty^\#$: the set of all infinite collections of increasing numbers from $\mathbb{N}$.

$x^\nu \xrightarrow{N} x$: express that x is a cluster point to the subsequence specified by the index set N.

$\mathbb{R}_+$: $[0, +\infty)$

$\overline{\mathbb{R}}$: $\mathbb{R} \cup \{+\infty\}$

$\mathcal{B}$: Borelian set

$\mathscr{R}^k$: σ-algebra generated by the rectangles in $\mathbb{R}^k$

$\xrightarrow{\text{h}}$: Hypo-convergence

$\xrightarrow{\text{e}}$: Epi-convergence

$\xrightarrow{\text{p}}$: Point-wise convergence

$\rightarrow$: Convergence in $|\cdot|$

$\|\cdot\|_2$: Euclidean norm.

$\|\cdot\|_\infty$: Infinity norm.

$\xrightarrow{\text{S}}$: Set convergence.

$\mathbb{B}_\infty$: unit ball with the infinity norm.

$\mathbb{S}$: $\mathbb{S}(0, 1)$ (See Definition 2.3)

sgn : sign function

$\Delta_A F$: Difference of F over rectangle A (See Definition 2.31)

$\mathcal{M}$: Set of all probability measures

cd-fcns(S) : Set of cumulative distribution functions over S.

usc-fcns(S) : Set of upper semi continuous functions over S.

$\mathbb{1}_A$: Characteristic function over A (1 if $x \in A$, 0 o.w).

ι_C : Indicator function over C (0 if $x \in C$, $+\infty$ o.w)

$\mathcal{N}(\mu, \Sigma)$: Normal distribution of mean μ and covariance Σ

U$[a, b]$: Uniform distribution over $[a, b]$

Chapter 1

Introduction

1.1. Main problem and applications

Consider the euclidean space $(\mathbb{R}^m, \|\cdot\|_2)$ and let S be a closed subset. Let $(S, \mathcal{B}_S, \mathbb{P})$ be a probability space defined over S, where $\mathcal{B}_S$ is the associated Borel σ-algebra. Additionally, let $\mathscr{X} = \text{cl-sets}(S)$ be the hyper-space of closed subsets of S. We denote $X \sim F$ a random variable X on S that follows a cumulative distribution function F.

The main goal of this work is to develop a method to estimate cumulative distribution functions by studying a wider class of functions, this is, the set of upper semicontinuous functions endowed with the Attouch-Wets distance, represented by $d\!l$. The Attouch-Wets distance quantifies distances between sets, in this case hypographs and because of this, the name of hypo-distance is sometimes used. Definitions and properties of these last-mentioned elements are exhibited on Chapter 2. The work developed in this thesis is motivated by the following estimation problem.

Problem 1.1 Let $F_0 : R \to [0,1]$ and $G_0 : R \to [0,1]$ two cumulative distribution functions where $R \subset \mathbb{R}^m$ is a bounded rectangle. Consider $\mathscr{F}$ a set of functions where F_0 and G_0 belongs to, $C \subset \mathscr{F}$ a set of constraints, $\pi : \mathscr{F} \to (-\infty, +\infty]$ a penalty function and a parameter $\delta > 0$. The problem is to find $\hat{F} : R \to [0,1]$ in $C \subset \mathscr{F}$ such that

$$\hat{F} \in \text{argmin}_{F \in C \subset \mathscr{F}} \{ d\!l(F, F_0) + \pi(F) \;\mid\; d\!l(F, G_0) \leq \delta \} \tag{1.1}$$

The set of constraints C represents soft information of the problem that is being modelled. For example it can be used to include prior information about the problem, such as probability sets or to impose growing conditions to solutions. The penalty function π can be used to add sparseness or smoothness by setting π as some regularization function ℓ^1 or ℓ^2.

Several other metrics has been studied in the context of distribution functions (See [10], [6], [15]) different from the one used in this work, such as the Wasserstein distance ([16], [14]) the Prokhorov ([12]) and bounded Lipschitz metrics.

The problem we are interested in studying falls under a more general framework, the theory of set estimation, which deals with the statistical problem of estimating an unknown (usually compact) set S from sample points $\mathcal{X}$. To be precise the idea is to find a closed subset of S that minimizes a functional, and also satisfies a collection of constraints $\mathcal{C}$. Mathematically, let $\mathcal{C} \subset \mathscr{X}$ and $f : \mathscr{X} \to \mathbb{R}$ a lower semi-continuous cost function.

Problem 1.2 Find a subset C, such that

$$C \in \operatorname{argmin}\{f(C) \mid C \in \mathcal{C}, C \in \text{cl-sets}(S)\},$$

where the nature of the objective function, as the feasible set have stochastic nature. Note that we can also replace the subset C with a family of subsets $C_1, \ldots, C_m$. Under this setting we can study the following applications. Denote

$$\varphi : \mathscr{X} \to \overline{\mathbb{R}}, \quad \varphi(C) = \begin{cases} f(C) & C \in \mathcal{C} \\ +\infty & \text{o.w} \end{cases}$$

A more general formulation is stated as

$$\text{Find } C^* \text{ such that } C^* \in \operatorname{argmin}_{C \in \mathscr{X}} \varphi$$

Under the general problem presented above it relies our main problem, we are interested in Problem 1.2 when closed sets represent hypographs of functions that belong to a metric space where the set of cumulative distribution functions are part of.

Some applications of set estimation are related to different computer-intensive statistical methods as cluster analysis and level set estimation which can also be interpreted in terms of confidence sets (See [8], [1], [9], [4]). Set estimation has been studied with metrics instead of functionals in [2], where in particular they solve an statistical quality control problem, in which one has to decide if the distribution at stage $n + 1$ is different from the the one at the previous stage.

1.2. State of the art and objectives

To our knowledge there is no record of numerical results for an approximation of the Attouch-Wets distance. Nevertheless bounds for the hypo-distance are presented in [17] (for lower semicontinuous functions) and developed in [18] or in Sections 6J 7J of [23]. Approximations under the Attouch-Wets distance for the set of upper semi-continuous functions by a finite number of parameters are shown in [19]. This set of functions is known as epi-splines and it is extensively developed in [21]. Numerical results with the use of epi-splines in the context of statistical estimation are shown in [22] with the important difference that the work done there is centered around the estimation of probability density functions that solve non-parametric M-estimators.

From here, we aim to solve questions regarding well-posedness of Problem 1.1, minimum conditions for guaranteeing existence of a solution and an approximating sequence of problem for which the questions mentioned above we also aim to answer along with its convergence. This leads to Objective 1 and 2.

Objective 1: Present existence results and propose an approximation to Problem 1.1.

Objective 2: Provide the existence and convergence of the proposed approximation.

We will design and implement an algorithm that finds solutions for the proposed approximated problem and we will test the effectiveness and consistency of the results by studying two different numerical instances. Moreover, we will apply the proposed algorithm to a real data problem. These two challenges are summered up in the next and final objective.

Objective 3: Perform numerical results for the proposed algorithm and solve a problem with real data.

This thesis is organized as follows. In Chapter 2 we present the necessary theory and results to understand the tools that will be used to achieve the objectives. Chapter 3 furnishes the existence and approximation results for the proposed method and also and algorithm to compute solutions. In Chapter 4 we depict the numerical experiments obtained with the algorithm, among this, we present two sections, one in which as a proof of concept we deal with uniform distribution functions and in the other section we solve a position estimating problem. Finally, in Chapter 5 we recapitulate the results developed previously and propose future lines of work.

Chapter 2

Foundations

2.1. Set Convergence

The study of sets is central to most analysis of optimization and variational problems. There is great number of elements that are represented by sets, from sets of solutions and sets of subgradients to variational inequalities and generalized equations that are defined in terms of set-valued mappings. To develop a framework of approximations for such elements, a notion of convergence of sets becomes necessary. The theory of set convergence will provide ways of approximating for example extended-real valued functions through convergence of epigraphs or hypographs. Set-convergence can be understood and fully characterized using the notions of inner and outer limits. In the next part we recapitulate information available for deeper comprehension in [17] or in Section 4E of [23].

2.1.1. Inner and Outer Limits

Consider the euclidean space $(\mathbb{R}^m, \|\cdot\|_2)$, the symbol $\to$ will denote the convergence in this space and denote $\overline{\mathbb{R}}$ the set of extended real numbers, $\mathbb{R} \cup \{+\infty\}$. The issue of whether a sequence of subsets of $\mathbb{R}^m$ has a limit can best be approached through the study of two *semilimits* which always exist. The inner limit of $\{C^\nu \subset \mathbb{R}^m, \nu \in \mathbb{N}\}$, denoted by $\operatorname{LimInn} C^\nu$, is the collection of limit points to which sequences of points selected from the sets converge. Specifically,

Definition 2.1 (Inner limits) *For a sequence $\{C^\nu\}_{\nu\in\mathbb{N}}$ of subsets of $\mathbb{R}^m$, the inner limit is the set*

$$\operatorname{Lim\,Inn} C^\nu = \{x \in \mathbb{R}^m \mid \exists\ x^\nu \in C^\nu \to x\}$$

To understand the previous definition, consider the sequence of sets given by $C^\nu = \{0\}$ when ν is odd and $C^\nu = [0,1]$ when ν is even, then $\operatorname{Lim\,Inn} C^\nu = \{0\}$. The inner limit isn't $[0,1]$ because we can't construct $x^\nu \in C^\nu$ converging to $x \in (0,1]$. Before moving to the notion of outer limit, we need to introduce the following set

$$\mathcal{N}_\infty^\#, \text{ the set of all infinite collections of increasing numbers from } \mathbb{N}.$$

For example, $\{1,3,5,\ldots\} \in \mathcal{N}_\infty^\#$ and $\{2,4,6,\ldots\} \in \mathcal{N}_\infty^\#$. A subsequence of $\{x^\nu \in \mathbb{R}^m, \nu \in \mathbb{N}\}$ is then of the form $\{x^\nu \in \mathbb{R}^m, \nu \in N\}$ for some $N \in \mathcal{N}_\infty^\#$, with $x^\nu \xrightarrow{N} x$ expressing that x is a cluster point corresponding to the subsequence specified by the index set N.

The outer limit of $\{C^\nu \subset \mathbb{R}^m, \nu \in \mathbb{N}\}$, denoted by LimOut C^ν, is the collection of cluster points to which subsequences of points selected from the sets converge. To write this formally, using the notation above:

Definition 2.2 (Outer limits) *For a sequence $\{C^\nu\}_{\nu\in\mathbb{N}}$ of subsets of $\mathbb{R}^m$, the outer limit is the set*

$$\operatorname{LimOut} C^\nu = \left\{ x \in \mathbb{R}^m \mid \exists\ N \in \mathcal{N}_\infty^\# \text{ and } x^\nu \in C^\nu \underset{N}{\to} x \right\}$$

Hence, $x \in$ LimOut C^ν if we can select an index set $N \in \mathcal{N}_\infty^\#$ and points $\{x^\nu \in C^\nu, \nu \in N\}$ converging to x. In the example $C^\nu = \{0\}$ when ν is odd and $C^\nu = [0,1]$ when ν is even, LimOut $C^\nu = [0,1]$ because for any $x \in [0,1]$, we can rely on $N = \{2,4,\ldots\}$ and select $x^\nu = x$ for $\nu \in N$. The limit of the sequence exists if the outer and inner limit sets are equal:

$$\operatorname{Lim} C^\nu := \operatorname{LimOut} C^\nu = \operatorname{LimInn} C^\nu$$

Although Lim C^ν may not exist (as is clear from the "odd-even" example above), the inner limit and the outer limit always exist and are closed but they could be the empty set.

Definition 2.3 (Painlevé-Kuratowski convergence) *A sequence of sets in $\mathbb{R}^m$, $\{C^\nu\}_{\nu\in\mathbb{N}}$ is said to converge in the sense of Painlevé-Kuratowski to C if the* Lim C^ν *exists, meaning the outer and inner limits coincide and equals C, written*

$$C^\nu \xrightarrow{\text{S}} C.$$

In applications, a sequence of sets may not set-converge but useful insight can come from the inner and/or outer limits. For example, given a closed feasible set $C \subset \mathbb{R}^m$ that is approximated by $\{C^\nu \subset \mathbb{R}^m, \nu \in \mathbb{N}\}$, suppose that an algorithm generates $\{x^\nu \in C^\nu, \nu \in \mathbb{N}\}$, which we hope are near a feasible point in C. Indeed, if LimOut $C^\nu \subset C$, then every cluster point of $\{x^\nu, \nu \in \mathbb{N}\}$ is contained in C and is thus feasible for the actual problem. This might suffice as justification for the approximating sets C^ν. If $C^\nu \xrightarrow{\text{S}} C$, then we have in addition that every $x \in C$ can be approached by points in C^ν, i.e., there is $x^\nu \in C^\nu \to x$.

2.2. Upper and lower semi-continuous functions

Let $S \subset \mathbb{R}^m$ a closed set and endowed it with the euclidean norm $\|\ \|_2$. From now on we will work with functions that are defined over the metric space $(S, \|\cdot\|_2)$. For any $f : S \to \overline{\mathbb{R}}$, we adopt the notation

$$\liminf_{x'\to x} f\left(x'\right) := \lim_{\delta\downarrow 0}\left[\inf_{x'\in\mathbb{B}(x,\delta)} f\left(x'\right)\right], \qquad \limsup_{x'\to x} f\left(x'\right) := \lim_{\delta\downarrow 0}\left[\sup_{x'\in\mathbb{B}(x,\delta)} f\left(x'\right)\right],$$

where $\mathbb{B}(x,\delta)$ is a ball centered in $x \in S$ of radious $\delta > 0$ in the euclidean norm. Let $f : S \to \mathbb{R}$, the hypograph of f is the set

$$\text{hypo } f := \{(x, x_0) \in S \times \mathbb{R} : f(x) \geq x_0\}.$$

The function f is upper semi-continuous (usc) if it's hypograph is a closed set. Another characti-zation is that f is usc at $\bar{x}$ if

$$\limsup_{x \to \bar{x}} f(x) \leq f(\bar{x}),$$

The set of all upper semi-continuous functions that are define over S will be denoted by

$$\text{usc-fcns}(S;\mathbb{R}) := \{f : S \to \mathbb{R} : f \text{ is usc }\}. \tag{2.1}$$

The epigraph of a function $f : S \to \overline{\mathbb{R}}$ is the set

$$\text{epi } f := \{(x, x_0) \in S \times \mathbb{R} : f(x) \leq x_0\}.$$

The function f is lower semi-continuous (lsc) if it's epigraph is a closed set. Another charactization is that f is lsc at $\bar{x}$ if

$$\liminf_{x \to \bar{x}} f(x) \geq f(\bar{x}),$$

This notation in (2.1) assumes that the domain of the functions is S. To ease the notation it may also be refered just as usc-fcns(S). The hypograph of $f \in$ usc-fcns(S) is a subset of $\mathbb{R}^{m+1}$ and a possibility would be to adopt the euclidean norm $\| \|_2$ on $\mathbb{R}^{m+1}$ when measuring distances between points in $\mathbb{R}^{m+1}$. Here we choose a setup that distinguish between the last component of vectors in $\mathbb{R}^{m+1}$ and the other components. Thus, for $S \times \mathbb{R}$ we adopt the norm

$$\|(\xi, \xi_0)\|_{\mathbb{S}} := \max\{\|\xi\|_2, |\xi_0|\} \text{ for } (\xi, \xi_0) \in S \times \mathbb{R} \tag{2.2}$$

A ball under $\| \cdot \|_{\mathbb{S}}$, with radius r centered at $\bar{\xi} = (\xi, \xi_0) \in S \times \mathbb{R}$, is denoted by

$$\mathbb{S}(\bar{\xi}, r) := \left\{\bar{\zeta} \in S \times \mathbb{R} : \|\bar{\xi} - \bar{\zeta}\|_{\mathbb{S}} \leq r\right\} \tag{2.3}$$

and is actually a "hyper-cylinder," the symbol $\mathbb{S}$ will denote $\mathbb{S}(0, 1)$. The distance between $\bar{\xi} \in S \times \mathbb{R}$ and a set $C \subset S \times \mathbb{R}$ is given by

$$\text{dist}(\bar{\xi}, C) := \inf\left\{\|\bar{\xi} - \bar{\zeta}\|_{\mathbb{S}} : \bar{\zeta} \in C\right\} \tag{2.4}$$

Along this work, we will be pay special attention to functions that belong to usc-fcns $(S; [0, 1])$. Even though this space is not linear it is a pointed cone and as a consequence, convex. Hence it is still meaningful to consider the weighted averages $\lambda g + (1 - \lambda)g'$ for $g, g' \in$ usc-fcns $(S; [0, 1])$ and $\lambda \in [0, 1]$, defined in a pointwise manner.

Definition 2.4 *(convexity on* usc-fcns $(S; [0, 1])$*.)* *A set* $C \subset$ usc-fcns $(S; [0, 1])$ *is convex if*

$$\textit{for every } g, g' \in C \textit{ with } \lambda \in [0, 1], \lambda g + (1 - \lambda)g' \in C.$$

A function $\psi : C \to \mathbb{R}$ *is convex if* $C \subset$ usc-fcns $(S; [0, 1])$ *is a convex set and if*

$$\textit{for every } g, g' \in C \textit{ and } \lambda \in [0, 1],\ \psi\left(\lambda g + (1 - \lambda)g'\right) \leq \lambda\psi(g) + (1 - \lambda)\psi\left(g'\right).$$

2.2.1. Convergence in the space of upper (and lower) semi-continuous functions

Different types of convergence can be set up in the space of upper semi continuous functions. Naturally, choosing one set up over another will be beneficial in a certain manner but disadvantageous in others. For instance, it is an advantage the preservation of certain interesting features, like monotonicity or concavity but the richer the class of results that can be achieved, the reduced it is the family of functions that fulfills it. We would like to benefit from a topology that allows flexibility yet yields profitable results. Along these lines, the convergence of the hypographs will be a powerful tool for the aim of this work.

Definition 2.5 (Hypo-convergence) *The functions f^ν are said to hypo-converge to f, a condition symbolized by $f^\nu \overset{\mathrm{h}}{\to} f$. Thus,*

$$f^\nu \overset{\mathrm{h}}{\to} f \iff \operatorname{hypo} f^\nu \overset{\mathrm{S}}{\to} \operatorname{hypo} f.$$

where the convergence of the hypographs it's meant in the sense of Painlevé Kuratowski.

Definition 2.6 (Epi-convergence) *The functions f^ν are said to epi-converge to f, a condition symbolized by $f^\nu \overset{\mathrm{e}}{\to} f$. Thus,*

$$f^\nu \overset{\mathrm{e}}{\to} f \iff \operatorname{epi} f^\nu \overset{\mathrm{S}}{\to} \operatorname{epi} f.$$

where the convergence of the epigraphs it's meant in the sense of Painlevé Kuratowski.

Since Definition 2.5 and 2.6 are not easy to compute a much more straightforward characterization is needed. Since inner limits are contained in the corresponding outer limits, $f^\nu \overset{\mathrm{h}}{\to} f$ takes place if and only if $\operatorname{LimOut}(\operatorname{hypo} f^\nu) \subset \operatorname{hypo} f \subset \operatorname{Lim\,Inn}(\operatorname{hypo} f^\nu)$. The two inclusions translate into the next result.

Proposition 2.7 (characterization of hypo-limits) *Let $\{f^\nu\}_{\nu \in N}$ be any sequence of functions on S, and let x be any point of S. Then $f^\nu \overset{\mathrm{h}}{\to} f$ if and only if at each point x*

$$\begin{cases} \limsup_\nu f^\nu(x^\nu) \le f(x) & \textit{for every sequence } x^\nu \to x, \\ \liminf_\nu f^\nu(x^\nu) \ge f(x) & \textit{for some sequence } x^\nu \to x. \end{cases} \tag{2.5}$$

Proof. See [Proposition 7.2 [17]] □
Analagous for the epi-convergence:

Proposition 2.8 (characterization of epi-limits) *Let $\{f^\nu\}_{\nu \in N}$ be any sequence of functions on S, and let x be any point of S. Then $f^\nu \overset{\mathrm{e}}{\to} f$ if and only if at each point x*

$$\begin{cases} \limsup_\nu f^\nu(x^\nu) \le f(x) & \textit{for some sequence } x^\nu \to x, \\ \liminf_\nu f^\nu(x^\nu) \ge f(x) & \textit{for every sequence } x^\nu \to x. \end{cases} \tag{2.6}$$

What follows is mainly centered in hypo-convergence because upper semi-continuous functions are the core of this work, though the results in this chapter are also true for epi-convergence mutatis mutandis.

Hypoconvergence neither implies nor is implied by pointwise convergence. Nevertheless is possible to obtain a result of this fashion by introducing the following definition,

Definition 2.9 (equi-usc) *The sequence* $\{f^\nu\}_{\nu\in N}$ *is equi-usc at* $\bar{x}$ *relative to* S *if for every* $\varepsilon > 0$ *there exists* $\delta > 0$ *with*

$$f^\nu(x) \le f^\nu(\bar{x}) + \varepsilon \text{ for all } \nu \in \mathbb{N} \text{ when } x \in S, |x - \bar{x}| \le \delta.$$

The addition of the qualification asymptotically refers to the property holding for all tales of subsequences rather than for all subsequences.

Theorem 2.10 *(hypo-convergence versus pointwise convergence) Consider any sequence of usc functions* $f^\nu : S \to \overline{\mathbb{R}}$ *and a point* $\bar{x} \in S$. *If the sequence is asymptotically equi-usc at* $\bar{x}$, *then*

$$f^\nu \xrightarrow{\text{h}} f \text{ if and only if } f^\nu \xrightarrow{\text{p}} f,$$

where $\xrightarrow{\text{p}}$ *denotes pointwise convergence. More generally, relative to an arbitrary set* $C \subset S$ *containing* $\bar{x}$, *any two of the following conditions implies the third:*

(a) the sequence is asymptotically equi-lsc at $\bar{x}$ *relative to* C;

(b) $f^\nu \xrightarrow{\text{h}} f$ *at* $\bar{x}$ *relative to* C;

(c) $f^\nu \xrightarrow{\text{p}} f$ *at* $\bar{x}$ *relative to* C.

Proof. See [Theorem 7.10, [17]] □

Uniform convergence ensures hypo-convergence, but fails to handle extended real-valued functions satisfactory—a necessity in constrained optimization problems.

2.3. Attouch-Wets distance

We will use two basic measures of distance between upper semi-continuous functions. We define for every choice of the parameter $\rho \in \mathbb{R}_+$ and pair of functions $f, g \in$ usc-fcns(S), the ρ-hypo distance,

$$d\!l_\rho(f,g) := \max\left\{\left|\text{dist}(\bar{\xi},\ \text{hypo } f) - \text{dist}\left(\bar{\xi},\ \text{hypo } g\right)\right| \ :\ \|\bar{\xi}\|_{\mathbb{S}} \le \rho\right\} \quad \text{for } \rho \ge 0 \tag{2.7}$$

and the

$$\hat{d\!l}_\rho(f,g) := \inf_{\eta\ge 0}\{\ \eta \ \big|\ g(x+\eta\mathbf{1}) + \eta \ge \min\{f(x),\rho\} \text{ and} \tag{2.8}$$

$$f(x+\eta\mathbf{1}) + \eta \ge \min\{g(x),\rho\}\ \forall\ x \in \rho\mathbb{B}_\infty\} \tag{2.9}$$

where $\mathbb{B}_\infty$ is the unit ball on $\mathbb{R}^m$ relative to S with the infinity norm.

Remark 2.11 $d\!l_\rho$ and $\hat{d\!l}_\rho$ are non-decreasing functions of ρ.

Theorem 2.12 (Quantification of set convergence) *For each* $\rho \ge 0, d\!l_\rho$ *is a pseudo-metric on the space* usc-fcns$(\mathbb{R}^m)$, *but* $\hat{d\!l}_\rho$ *is not. Both families* $\{d\!l_\rho\}_{\rho\ge 0}$ *and* $\left\{\hat{d\!l}_\rho\right\}_{\rho\ge 0}$ *characterize hypo-convergence: for any* $\bar{\rho} \in \mathbb{R}_+$, *one has*

$$\begin{aligned} g^\nu \xrightarrow{\text{h}} g \quad &\iff \quad d\!l_\rho(g^\nu, g) \to 0 \text{ for all } \rho \ge \bar{\rho} \\ &\iff \quad \hat{d\!l}_\rho(g^\nu, g) \to 0 \text{ for all } \rho \ge \bar{\rho}. \end{aligned}$$

Proof. See Theorem 4.36, [17] □

The hypo-distance or Attouch-Wets distance between f and g, both in usc-fcns(S) is given by

$$d\!l\,(f,g) := \int_0^\infty d\!l_\rho\,(f,g)\,e^{-\rho}d\rho \tag{2.10}$$

Remark 2.13 From the Attouch-Wets distance expression we observe that weights $e^{-\rho}$ are assigned to each $d\!l_\rho$, thus for larger values of ρ, the measures of $d\!l_\rho(f,g)$ are not taken into consideration. Let G be the cumulative distribution of δ_0 (Dirac's distribution at 0, i.e 1 if $\xi = 0$ and 0 elsewhere)

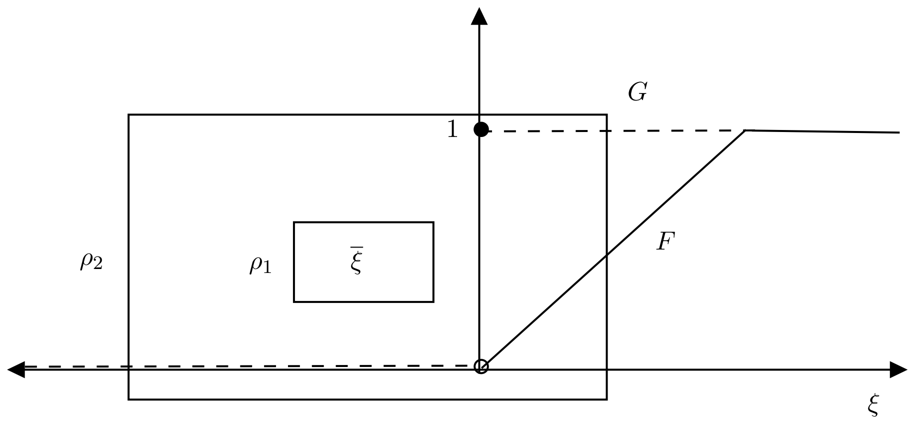


Figure 2.1: This example shows that for large values of ρ, $d\!l$,$d\!l_\rho$ and $\hat{d\!l}_\rho$ may have very different values.

and F be a uniform distribution over $[0,1]$ as is depicted in Figure 2.1. For rectangles centered in $\bar{\xi}$ of sizes $\rho_1 > 0$ and $\rho_2 > 0$ respectively, as in the figure, $d\!l_{\rho_1}(F,G) = 0$ (since F and G are both equal to 0 in this rectangle) and $d\!l_{\rho_2}(F,G) = 1$ (since the maximum difference between them is 1 in this rectangle) and the same is true for $\hat{d\!l}_{\rho_1}$, $\hat{d\!l}_{\rho_2}$, while $d\!l(F,G) = e^{-1}$ (this comes from computing the formula and the calculations above).

Theorem 2.14 (Theorem 4.42, [17]) *(usc-fcns$(S), d\!l)$ is a complete separable metric space.*

The next result states that (usc-fcns$(S), d\!l)$ is a locally compact metric space.

Corollary 2.15 (Corollary 4.43, [17]) *The metric space* (usc-fcns$\,(S)\,, d\!l)$ *has the property that for every one of its elements g and every $r > 0$ the ball*

$$\mathbb{B}_{d\!l}(g,r) := \{f \mid d\!l\,(f,g) \leq r\}$$

is compact.

Definition 2.16 (Escape to the horizon) *A sequence of sets, in this context, a sequence of* $\{\text{hypo}\, f^\nu\}_{\nu\in\mathbb{N}}$ *is said to escape to the horizon if*

$$\text{hypo}\, f^\nu \overset{\text{S}}{\to} \emptyset$$

or equivalently, $\text{OutLim}\,\text{hypo}\, f^\nu = \emptyset$.

Theorem 2.17 (Metric description of set convergence) *The expression* $d\!l$ *gives a metric on* $\text{usc-fcns}(\mathbb{R}^m)$ *which characterizes hypo-convergence:*

$$f^\nu \xrightarrow{\text{h}} f \quad \Longleftrightarrow \quad d\!l\,(f^\nu, f) \to 0.$$

Furthermore, $(\text{usc-fcns}(S), d\!l)$ *is a complete metric space in which a sequence* $\{\text{hypo}(f)^\nu\}_{\nu\in\mathbb{N}}$ *escapes to the horizon if and only if for some function* g *in this space (and then for every* g *) one has* $d\!l\,(f^\nu, g) \to \infty$*.*

2.3.1. Properties of hypo-limits

In this section, we present some properties that are preserved under hypo-convergence. The importance of close classes will be essential in view of Theorem 2.14 to retrieve that the following set of functions are again a complete separable metric space with the hypo-distance. The following definition will be needed for these purposes.

Definition 2.18 *A box in* $R \subset \mathbb{R}^m$ *is of the form* $R = [\alpha_1, \beta_2] \times \ldots \times [\alpha_m, \beta_m]$*, with* $-\infty \leq \alpha_i < \beta_i \leq \infty$*, where in the case of* $\alpha_i = -\infty$ *and* $\beta_i = \infty$ *the closed intervals are replaced by (half) open intervals. Its dimension is therefore* m*.*

We introduce the following class of functions

$$\text{usc-fcns}_+(S;[0,1]) = \{f : S \to [0,1] \mid f \text{ is usc and nondecreasing }\}.$$

$$\text{Lip-fcns}_\kappa(S;[0,1]) = \{f : S \to [0,1] \mid f \text{ is Lipschitz of modulus } \kappa.\}.$$

Proposition 2.19 (Monotonicity.) *For* $\{f^n\}_{n\in\mathbb{N}}, f \subset \text{usc-fcns}(S)$ *such that* $f^n \xrightarrow{\text{h}} f$*, we have:*

(i) If f^n *is nondecreasing in the sense that* $f^n(x) \leq f^n(y)$ *for* $x \in S, y \in \text{int}\, S$*, with* $x \leq y$*, then* f *is also nondecreasing in the same sense. If* S *is a box, then int* S *can be replaced by* S*.*

(ii) If f^n *is nonincreasing in the sense that* $f^n(x) \geq f^n(y)$ *for* $x \in \text{int}\, S, y \in S$*, with* $x \leq y$*, then* f *is also nonincreasing in the same sense. If* S *is a box, then int* S *can be replaced by* S*.*

Proof. See Proposition 4.3 [22]

Proposition 2.20 (Lipschitz continuity) *Suppose that* $\{f^n\}_{n\in\mathbb{N}} \subset \text{usc-fcns}(S)$*, with* $S \subset \mathbb{R}^m$ *closed and* $f^n \xrightarrow{\text{h}} f$*, and* $\{f^n\}_{n\in\mathbb{N}}$ *are Lipschitz continuous with common modulus* κ*. Then,* f *is also Lipschitz continuous with modulus* κ*.*

Proof. See Proposition 4.4 [22]

Remark 2.21 Proposition 2.19 and Proposition 2.20 imply that the classes of functions $\text{usc-fcns}_+(S;[0,1])$ and $\text{Lip-fcns}_\kappa(S;[0,1])$ are closed in the Attouch-Wets topology and in consequence are complete separable metric spaces. These two classes of functions will be beneficial to our purposes as we will see in the next Chapter.

The next result will be of particular importance when incorporating information about the uncertainty that the function we would like to estimate can have. This information will signify that our estimated function is unambiguously less (or more) risky than another. This concept is known as stochastic dominance.

Proposition 2.22 (Pointwise bounds) *Suppose that* $g : S \to \overline{\mathbb{R}}$*,* $h \in$ usc-fcns(S)*,* $\{f^n\}_{n\in\mathbb{N}} \subset$ usc-fcns(S)*, and* $f^n \xrightarrow{h} f$*. If* $g(x) \le f^n(x) \le h(x)$ *for all* $n \in \mathbb{N}$ *and* $x \in S$*, then*

$$g(x) \le f(x) \le h(x) \text{ for all } x \in S.$$

Proof. See Proposition 4.5 [22]

2.3.2. Distance bounds

This section is dedicated to give estimates for the hypo-distance in terms of functions $d\!\!l_\rho$ and $\hat{d\!\!l}_\rho$ given in equations 2.7 and 2.8. The bounds presented below will play a crucial role for the main goal of our work.

Proposition 2.23 *Let* $f, g \in$ usc-fcns(S; [0, 1])*, then* $d\!\!l_\rho(f, g) \le 1$ *for every* $\rho > 0$*.*

Proof. Let $(x, \alpha) \in \operatorname{hypo} f$, $(y, \beta) \in \operatorname{hypo} g$ and $\bar{\xi} = (\bar{\xi}_0, \bar{\xi}_1)$ such that $\|\bar{\xi}\|_{\mathbb{S}} \le \rho$

$$\begin{aligned}
\max\left\{\|x - \bar{\xi}_0\|_2, |\alpha - \bar{\xi}_1|\right\} &\le \max\left\{\|x - y\|_2 + \|\bar{\xi}_0 - y\|_2, |\alpha - \beta| + |\bar{\xi}_1 - \beta|\right\} \\
&\le \max\left\{\|x - y\|_2, |\alpha - \beta|\right\} + \max\left\{\|y - \bar{\xi}_0\|_2, |\beta - \bar{\xi}_1|\right\} \\
&\le \max\{\|x\|_2 + \|y\|_2, 1\} + \max\left\{\|y - \bar{\xi}_0\|_2, |\beta - \bar{\xi}_1|\right\},
\end{aligned}$$

since the values of f and g are between 0 and 1, we have that $|\alpha - \beta| \le 1$. Taking infimum over $(x, \alpha) \in \operatorname{hypo} f$ and using the fact that $0 \in \operatorname{dom} f$ we obtain that

$$\operatorname{dist}(\bar{\xi}, \operatorname{hypo} f) \le \max\{\|y\|_2, 1\} + \max\left\{\|y - \bar{\xi}_0\|_2, |\beta - \bar{\xi}_1|\right\}$$

The same argument follows for g and we obtain that

$$\operatorname{dist}(\bar{\xi}, \operatorname{hypo} f) - \operatorname{dist}(\bar{\xi}, \operatorname{hypo} g) \le 1$$

exchanging the roles of f ang g we obtain that

$$|\operatorname{dist}(\bar{\xi}, \operatorname{hypo} f) - \operatorname{dist}(\bar{\xi}, \operatorname{hypo} g)| \le 1$$

□

Remark 2.24 Under the setting above we also have that $d\!\!l(f, g) \le 1$ since $\int_0^{+\infty} e^{-\rho} d\rho = 1$.

Proposition 2.25 (Estimates of ρ-hypo distance) *Let* $f, g \in$ usc-fcns$_+$(S; [0, 1])*, then*

$$\hat{d\!\!l}_\rho(f, g) \le d\!\!l_\rho(f, g) \le \hat{d\!\!l}_{2\rho}(f, g) \quad \text{for } \rho \ge 0$$

Proof. Let $C, D \subset S \times \mathbb{R}$ two closed sets such that $0 \in C \cap D$. Let $\varepsilon > 0$, $\rho > 0$ and $\bar{\rho} \geq 2\rho$. We first show that

$$\operatorname{dist}(\cdot, D) \leq \operatorname{dist}(\cdot, C) + \varepsilon \text{ on } \rho\mathbb{S} \text{ implies that } C \cap \rho\mathbb{S} \subset D + \varepsilon\mathbb{S}. \tag{2.11}$$

For every $\bar{x} \in C \cap \rho\mathbb{S}$ that $\operatorname{dist}(\bar{x}, D) \leq \varepsilon$. As D is closed, we have that $C \cap \rho\mathbb{S} \subset D + \varepsilon\mathbb{S}$. Second, we establish that

$$C \cap \bar{\rho}\mathbb{S} \subset D + \varepsilon\mathbb{S} \text{ implies } \operatorname{dist}(\cdot, D) \leq \operatorname{dist}(\cdot, C) + \varepsilon \text{ on } \rho\mathbb{S}. \tag{2.12}$$

For any $\bar{x} \in \mathbb{R}^{n+1}$,

$$\begin{aligned}
\operatorname{dist}(\bar{x}, C \cap \bar{\rho}\mathbb{S}) &\geq \operatorname{dist}(\bar{x}, D + \varepsilon\mathbb{S}) \\
&= \inf\{\|(\bar{y} + \varepsilon\bar{z}) - \bar{x}\|_{\mathbb{S}} : \bar{y} \in D, \bar{z} \in \mathbb{S}\} \\
&\geq \inf\{\|\bar{y} - \bar{x}\|_{\mathbb{S}} - \varepsilon\|\bar{z}\|_{\mathbb{S}} : \bar{y} \in D, \bar{z} \in \mathbb{S}\} \\
&= \operatorname{dist}(\bar{x}, D) - \varepsilon
\end{aligned}$$

Thus, $\operatorname{dist}(\cdot, D) \leq \operatorname{dist}(\cdot, C \cap \bar{\rho}\mathbb{S}) + \varepsilon$ on $\mathbb{R}^{n+1}$. It remains to establish that

$$\operatorname{dist}(\bar{x}, C \cap \bar{\rho}\mathbb{S}) = \operatorname{dist}(\bar{x}, C) \text{ when } \bar{x} \in \rho\mathbb{S} \text{ and } \bar{\rho} \geq 2\rho. \tag{2.13}$$

Naturally we have $\operatorname{dist}(\bar{x}, C \cap \bar{\rho}\mathbb{S}) \geq \operatorname{dist}(\bar{x}, C)$, for the remainder let $\bar{x} \in \rho\mathbb{S}$ and $\bar{y} \in \operatorname{argmin}_{c \in C} \|\bar{x} - c\|_{\mathbb{S}}$, which exists since C is closed. Implication (2.13) is established if $\bar{y} \in \bar{\rho}\mathbb{S}$. This is indeed the case because

$$\|\bar{y}\|_{\mathbb{S}} \leq \|\bar{x}\|_{\mathbb{S}} + \|\bar{y} - \bar{x}\|_{\mathbb{S}}$$

with $\|\bar{y} - \bar{x}\|_{\mathbb{S}} = \operatorname{dist}(\bar{x}, C) \leq \operatorname{dist}(\bar{x}, 0) = \|\bar{x}\|_{\mathbb{S}}$ since $0 \in C$ and consequently

$$\|\bar{y}\|_{\mathbb{S}} \leq 2\|\bar{x}\|_{\mathbb{S}} \leq 2\rho \leq \bar{\rho}.$$

Considering $0 \in \operatorname{dom} f \cap \operatorname{dom} g$ implies that $(0,0) \in \operatorname{hypo}(f) \cap \operatorname{hypo} g$ since $f(x) \geq 0$ for all $x \in \operatorname{dom}(f)$ and in particular $f(0) \geq 0$, the same argument follows for g. Then applying the first implication (2.11) with $C = \operatorname{hypo} f$ and $D = \operatorname{hypo} g$ and then with $C = \operatorname{hypo} g$ and $D = \operatorname{hypo} f$, we obtain that for $\varepsilon = \eta > 0$,

$$|\operatorname{dist}(\cdot, \operatorname{hypo} f) - \operatorname{dist}(\cdot, \operatorname{hypo} g)| \leq \eta \text{ on } \rho\mathbb{S} \text{ implies } \begin{cases} \operatorname{hypo} f \cap \rho\mathbb{S} \subset \operatorname{hypo} g + \eta\mathbb{S} \text{ and} \\ \operatorname{hypo} g \cap \rho\mathbb{S} \subset \operatorname{hypo} f + \eta\mathbb{S} \end{cases}$$

which implies that $\hat{d\kern-0.4em l}_\rho(f,g) \leq d\kern-0.4em l_\rho(f,g)$. Repeating the same procedure with the second implication, we obtain that $\hat{d\kern-0.4em l}_\rho(f,g) \geq d\kern-0.4em l_{\bar{\rho}}(f,g)$ for $\bar{\rho} \geq 2\rho$. □

Theorem 2.26 (Estimates of hypo-distance) *For $f, g \in$ usc-fcns$_+(S; [0,1])$, we have that for any $\rho \in [0, \infty]$,*

$$\hat{d\kern-0.4em l}_\rho(f,g)e^{-\rho} \leq d\kern-0.4em l(f,g) \leq e^{-\rho} + \left(1 - e^{-\rho}\right)\hat{d\kern-0.4em l}_{2\rho}(f,g)$$

Proof.

$$d\kern-0.4em l(f,g) = \int_0^\rho d\kern-0.4em l_\tau(f,g)\, e^{-\tau} d\kern-0.4em l_\tau + \int_\rho^\infty d\kern-0.4em l_\tau(f,g)\, e^{-\tau} d\tau$$

Since $d\!\!l_\tau(f,g)$ is nondecreasing as τ increases and $0<\tau<\rho$, we have that

$$d\!\!l_0(f,g)\int_0^\rho e^{-\tau}d\tau \le \int_0^\rho d\!\!l_\tau(f,g)\,e^{-\tau}d\tau \le d\!\!l_\rho(f,g)\int_0^\rho e^{-\tau}d\tau \tag{2.14}$$

and

$$\int_\rho^\infty d\!\!l_\rho(f,g)\,e^{-\tau}d\tau \le \int_\rho^\infty d\!\!l_\tau(f,g)\,e^{-\tau}d\tau. \tag{2.15}$$

Because of $f,g\in \text{usc-fcns}_+(\mathrm{S};[0,1])$, $d\!\!l_\tau(f,g)\le 1$, then

$$\int_\rho^\infty d\!\!l_\tau(f,g)\,e^{-\tau}d\tau \le \int_\rho^\infty e^{-\tau}d\tau = e^{-\rho} \tag{2.16}$$

Carrying out the integrations on the right-hand sides of 2.14 and proceeding equally on the left sides of 2.14 and 2.15, we obtain that

$$\left(1-e^{-\rho}\right)\left|d^f-d^g\right| + e^{-\rho}d\!\!l_\rho(f,g) \le d\!\!l(f,g) \le \left(1-e^{-\rho}\right)d\!\!l_\rho(f,g)+e^{-\rho}, \tag{2.17}$$

where $d^f=\mathrm{dist}(0,\mathrm{hypo}\,f)$ (analogous for g). Observe that the first term to the left is non-negative and consequently Proposition 2.25 gives the result. □

The proofs of Proposition 2.25 and Theorem 2.26 are strongly based on similar results developed in [20].

Theorem 2.26 hints how could it be approximated the hypo-distance. Nevertheless since computing $\hat{d\!\!l}_\rho$ would take an infinite number of contraints, we will look for an approximation.

Definition 2.27 *For $\rho>0$, a box partition of $\rho\mathbb{B}_\infty\subset S$ is a collection $\mathcal{R}=\{R_1,\dots,R_N\}$ of subsets of the form $R_k=\left(l_1^k,u_1^k\right)\times\dots\times\left(l_m^k,u_m^k\right)$ with $R_k\cap R_{k'}=\emptyset$ for $k\neq k'$ and $\cup_{k=1}^N \mathrm{cl}R_k=\rho\mathbb{B}_\infty$. Its mesh-size* $\mathrm{msh}(\mathcal{R})=\max\left\{u_j^k-l_j^k \mid k=1,\dots,N, j=1,\dots,m\right\}$

For $f,g\in\text{usc-fcns}(S;[0,1])$, we introduce the following functions

$$\eta_\rho^-(f,g) = \inf_{\eta\ge 0}\left\{\eta \;\middle|\; \begin{array}{ll} g(l^k+\eta\mathbf{1})+\eta & \ge \min\left\{f(l^k),\rho\right\} \text{ and,} \\ f(l^k+\eta\mathbf{1})+\eta & \ge \min\left\{g(l^k),\rho\right\}\ \forall k=1,\dots,N \end{array}\right\} \tag{2.18}$$

$$\eta_\rho^+(f,g) = \inf_{\eta\ge 0}\left\{\eta \;\middle|\; \begin{array}{ll} g(l^k+\eta\mathbf{1})+\eta & \ge \min\left\{f(u^k),\rho\right\} \text{ and,} \\ f(l^k+\eta\mathbf{1})+\eta & \ge \min\left\{g(u^k),\rho\right\}\ \forall k=1,\dots,N \end{array}\right\} \tag{2.19}$$

Remark 2.28 The functions η_ρ^+ and η_ρ^- are well defined because the problem for which they are defined is a linear finite dimensional problem bounded below. Also since $\eta=1$ satisfies the constraints that define these functions then $\eta_\rho^+(f,g)\le 1$ and the same is valid for $\eta_\rho^-(f,g)$.

Theorem 2.29 (approximation of hat-distance) *For $f,g\in\text{usc-fcns}_+(S;[0,1])$, we have that for any $\rho\in(0,\infty)$ and box partition $\mathcal{R}=\{R_1,\dots,R_N\}$ of $\rho\mathbb{B}_\infty$*

$$\eta_\rho^-(f,g)\le \hat{d\!\!l}_\rho(f,g)\le \eta_\rho^+(f,g)$$

Proof. The lower bound comes from the fact that the consideration of a finite number of constraints is a relaxation in the minimization problem given in η_ρ^-. The upper bound follows by the fact that if the stated constraints hold, then for any $x \in \rho\mathbb{B}_\infty$ and $\eta \geq 0$ there exists a k such that

$$f(x+\eta\mathbf{1})+\eta \geq f\left(l_k+\eta\mathbf{1}\right)+\eta \geq \min\left\{g(u^k),\rho\right\} \geq \min\left\{g(x),\rho\right\}.$$

A similar argument holds with the roles of f and g reversed. Thus, if η satisfies the stated constraints, it will also satisfied the constraints in $\hat{d\!l}_\rho(f,g)$. □

Corollary 2.30 *Under the setting above, if f, g are Lipschitz continuous with modulus κ_1 and κ_2 respectively and define $\kappa = \max\{\kappa_1, \kappa_2\}$, then $\eta_\rho^+ - \eta_\rho^- \leq \kappa\,\mathrm{msh}(\mathcal{R})$.*

Proof. We consider the difference between the upper and lower bounds. Suppose that $\eta \geq 0$ satisfies

$$f(l^k+\eta\mathbf{1})+\eta \geq \min\left\{g(l^k),\rho\right\} \quad \forall\ k=1,\ldots,N$$

Since g is non-decreasing and Lipschitz of modulus κ

$$g(l^k) \geq g(u^k) - \kappa\,\mathrm{msh}(\mathcal{R}) \quad \forall\ k=1,\ldots,N,$$

we then also have that for $\eta' = \eta + \kappa\,\mathrm{msh}(\mathcal{R})$

$$\begin{aligned} f(l^k+\eta'\mathbf{1})+\eta' \geq f(l^k+\eta\mathbf{1})+\eta' &\geq \min\left\{g(l^k),\rho\right\} + \kappa\,\mathrm{msh}(\mathcal{R}) \\ &\geq \min\left\{g(l^k)+\kappa\,\mathrm{msh}(\mathcal{R}),\rho\right\} \quad (\rho \text{ large}) \\ &\geq \min\left\{g(u^k),\rho\right\} \end{aligned}$$

for all $k=1,\ldots,N$. A similar argument establishes the result with the roles of f and g reversed and the conclusion follows. □

2.4. Space of cumulative distribution functions

We view distribution functions as a subset of a space of upper semicontinuous functions as described next. The aim of this section is to present a series of results that facilitate the understanding and the analysis of distribution functions in this set up.

A well-known fact about distribution functions is that every probability measure P on $(S, \mathcal{B}_S)$ (where $\mathcal{B}_S$ denotes the Borelian set relative to S) defines a distribution function $F : S \to [0,1]$ through $F(\xi) = P\left(B_\xi\right)$ for $\xi \in \mathbb{R}^m$, where $B_\xi := \{\zeta \in S : \zeta \leq \xi\}$. Vector inequalities are understood componentwise.

Definition 2.31 (Cumulative distribution function) *A cumulative distribution function (cdf) F is*

(I) *it is upper semi-continuous;*

(II) *nondecreasing, i.e., $F(\zeta) \leq F(\xi)$ for $\zeta \leq \xi$*

(III) *it satisfies $F\left(\xi^\nu\right) \to 0$ whenever one of the components of ξ^ν tends to $-\infty$, with the others held fixed,and $F\left(\xi^\nu\right) \to 1$ if $\xi_i^\nu \to \infty$ for all i*

(IV) *and satisfies the distribution condition, i.e* $\Delta_A F \geq 0$ *for every rectangle, where*

$$\Delta_A F := \sum_{j=1}^{2^m} \left(\operatorname{sgn}_A v^j\right) F\left(v^j\right)$$

with $v^j, j = 1, \ldots, 2^m$*, being the vertices of* A*, and* $\operatorname{sgn}_A v^j = 1$ *if the number of components* v_i^j *at a lower bound of* A *is even and* $\operatorname{sgn}_A v^j = -1$ *if the number is odd.*

Remark 2.32 For the bivariate case ($m = 2$), condition (IV) of Definition 2.31 is simply given by $\Delta_A F = F(v_1) - F(v_2) + F(v_3) - F(v_4)$, following the notation on Figure 2.2.

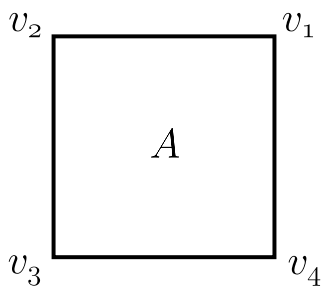


Figure 2.2: Condition (iv) in Definition 2.31 for $m = 2$.

For every $F : S \to \mathbb{R}$ that satisfy properties in Definition 2.31, there exists a unique probability measure P on $(S, \mathcal{B}_S)$ such that $P(A) = \Delta_A F$ for rectangles A, and $P(B_\xi) = F(\xi)$ for all $\xi \in B$.

The set of all probability measures on $(S, \mathcal{B}_S)$is denoted by $\mathcal{M}$. We denote by

$$\text{cd-fcns}\,(S) := \{F : S \to [0,1] : \exists P \in \mathcal{M} \text{ with } F(\xi) = P(S_\xi)\, \forall \xi \in S\}$$

the set of (cumulative) distribution functions. The set of all cumulative distribution functions is convex (in a pointwise manner, see Definition 2.4) as it is also usc-fcns(S). Furthermore, $\mathbb{d}(f,g) \leq 1$ for any functions $f, g \in$ cd-fcns(S) in view of Remark 2.24.

Definition 2.33 (Weak convergence) *A sequence* $\{\mu_n\}$ *of measures on* S *is said to converge weakly to a measure* μ *if*

$$\int_S g\ d\mu_n \to \int_S g\ d\mu,$$

for each bounded continuous function g *on* S*.*

Proposition 2.34 (Theorem 29.1, [3]) *Let* $\mathscr{R}^S$ *be the* σ*-algebra generated by all the rectangles in* S*. For probability measures* μ_n *and* μ *on* $\left(S, \mathscr{R}^S\right)$*, each of the following conditions is equivalent to the weak convergence of* μ_n *to* μ *:*

(I) $\lim_n \int f d\mu_n = \int f d\mu$ *for bounded continuous* f*;*

(II) $\limsup_n \mu_n(C) \leq \mu(C)$ *for every closed* C *in* $\mathscr{R}^S$*;*

(III) $\liminf_n \mu_n(G) \geq \mu(G)$ *for every open* G *in* $\mathscr{R}^S$*;*

(IV) $\lim_n \mu_n(A) = \mu(A)$ *for μ-continuity sets A (i.e sets whose boundary have 0 μ measure).*

Proposition 2.34, in particular, equivalence (II), helps to prove the equivalence of hypo-convergence and weak convergence.

Theorem 2.35 (Theorem 3.2, [20]) *For distribution functions $F^\nu, F \in \text{cd-fcns}(S)$ we have that $d\!l(F^\nu, F) \to 0$ if and only if F^ν converges weakly to F.*

The following fact, regarding an important consequence of the equivalence of hypo-convergence and weak convergence, was extracted from [20].

Remark 2.36 Since weak convergence establish the convergence of empirical measures (See [Theorem 11.4.1, [11]]), an immediate consequence of Theorem 2.35 is that if $\{\boldsymbol{\xi}^\nu\}_{\nu\in\mathbb{N}}$ is a sequence of independent and identically distributed random vectors with values in $\mathbb{R}^m$ that is defined on a probability space $(\Omega, \mathcal{A}, \mu)$ and $\xi^1 \sim F$, then the empirical distribution functions

$$F^\nu(\cdot, \omega) = \frac{1}{\nu}\sum_{j=1}^{\nu} \mathbb{1}\left(\xi^j(\omega) \le \cdot\right) \text{ hypo-converge to } F \text{ for } \mu\text{-almost every } \omega \in \Omega$$

The weak limit of a sequence of distribution functions might not be a distribution function. To see this it suffices to consider the sequence of cumulative distribution function given by $F_n(x) = 0$ if $x < n$ and 1 otherwise, which weak limit is $F(x) = 0$. An extra assumption has to be made in order to ensure this kind of results.

Definition 2.37 (Tight set) *A subset $C \subset$ cd-fcns (S) is tight if for all $\varepsilon > 0$ there exists a rectangle A such that $\Delta_A F \ge 1 - \varepsilon$ for all $F \in C$.*

Proposition 2.38 (Convergence to distribution function.) *If the sequence $\{F^\nu\}_{\nu\in\mathbb{N}} \subset \text{cd-fcns}(S)$ is tight, then the following hold:*

- *There exists a subsequence $\{F^{\nu_k}\}_{k\in\mathbb{N}}$ and a function $F \in \text{cd-fcns}(S)$ such that $d\!l(F^{\nu_k}, F) \to 0$ as $k \to \infty$*
- *If $d\!l(F^\nu, F) \to 0$ for some $F : S \to \mathbb{R}$, then $F \in \text{cd-fcns}(S)$.*

Proof. See Proposition 3.3, [20]. Nevertheless in view of Theorem 2.35 , the proof follows by Theorem 29.3 and its corollary [3]. □

Proposition 2.39 *(compactness, tightness.) For $C \subset \text{cd-fcns}(S)$, we have that*

(I) *if C is compact, then C is tight;*

(II) *if C is tight, then $\text{cl}\, C$ is compact and contained in $\text{cd-fcns}(S)$.*

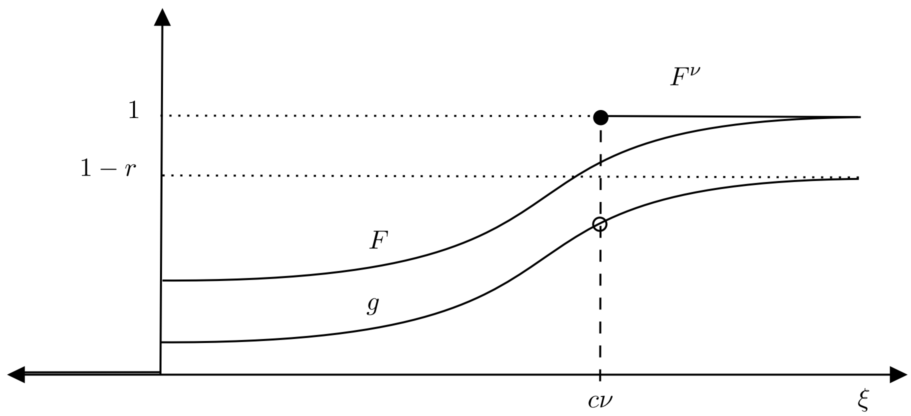
Remark 2.40 Because of Corollary 2.15, a ball in $\text{usc-fcns}(S; [0, 1])$, meaning the set

$$\mathbb{B}_{d\!l}(F, r) := \{G \in \text{usc-fcns}(S; [0, 1]) \mid d\!l(F, G) \le r, \}$$

with $F \in$ cd-fcns (S) is compact. However the subset $\mathbb{B}(F,r) \cap$ cd-fcns (S) is neither closed nor tight unless $r = 0$. To see this, let $g : S \to [0,1]$ be defined such that

$$g(\xi) = \max\{0, F(\xi) - r\} \text{ and } r > 0.$$

Then, $d\!l(g,F) \le r$. Let $A \subset S$ be such that $\Delta_A(F) \ge 1-r$. Set $c > 0$ such that $A \subset \{\xi \in S : \xi \le c\mathbf{1}\}$

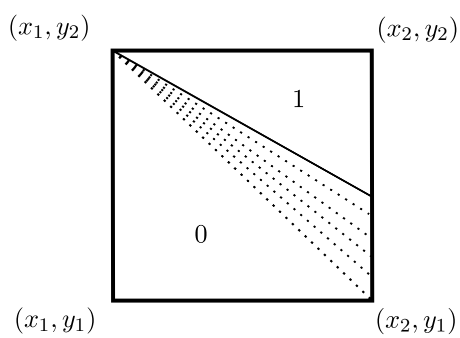


and construct the distribution functions $F^\nu : S \to [0,1]$ such that

$$F^\nu(\xi) = \begin{cases} 1 & \text{if } c\nu\mathbf{1} \le \xi \\ g(\xi) & \text{otherwise.} \end{cases}$$

Since $F(\xi) \ge 1 - r$ for ξ with $F^\nu(\xi) = 1$, we have that $|F^\nu(\xi) - F(\xi)| \le r$ for all ξ and thus $d\!l\,(F^\nu, F) \le r$. However, $\{F^\nu\}$ is not tight and does not tend to a distribution function. The only balls of (usc-fcns $(S;[0,1])$, $d\!l$) contained in cd-fcns (S) are those with zero radius, i.e., $\mathbb{B}(F,0)$. We observe that a setup centered on the metric space (cd-fcns(S), $d\!l$), instead of (usc-fcns $(S;[0,1])$, $d\!l$), is possible but has the disadvantage that the space is not complete.

Example 2.41 If a sequence $F^\nu : R :\to [0,1]$ of monotone upper semi continuous functions define over a bounded rectangular domain $R \subset \mathbb{R}^2$ that satisfy that $\Delta_A F^\nu \ge 0$ for every rectangle A and hypo-converge to some function F, it is not true that $\Delta_A F \ge 0$ for every rectangle A. To see this, let A be a rectangle on the upper right corner of the domain given by the following



and F^ν a sequence that $F^\nu(\xi) = 1$ if $\xi_1 \ge x_1$ and $\xi_2 \ge (1/\nu)y_2 + (1 - 1/\nu)y_1$, and 0 otherwise. This sequence is monotone, upper semi-continuous, $\Delta_A F^\nu = F^\nu(x_2,y_2) - F^\nu(x_1,y_2) + F^\nu(x_1,y_1) - F^\nu(x_2,y_1) = 1-1-0+0 = 0$ and hypo-converge to F that is 0 everywhere except in the upper triangle. We have that $F(x_2,y_1) = 1$, hence $\Delta_A F = -1$. This proofs that the subset of usc-fcns$_+(R;[0,1])$ that satisfy the distribution condition is not closed in the Attouch-Wets topology.

Proposition 2.42 *Let R of the form $[\alpha_1, \beta_1] \times, \ldots, \times [\alpha_m, \beta_m]$ on S and $\{F^\nu\}_{\nu \in \mathbb{N}}$ in* usc-fcns$(R; [0, 1])$ *equi-usc at the lower limit of R and*

$$\lim_{\xi \to +\infty} F^\nu(\xi) = 1 \text{ for all } \nu \tag{2.20}$$

$$\lim_{\xi \to -\infty} F^\nu(\xi) = 0 \text{ for all } \nu. \tag{2.21}$$

If $F^\nu \xrightarrow{h} F$ for some $F \in$ usc-fcns$(R; [0, 1])$ then F also satisfy (2.20)-(2.21).

Proof. Let $\xi \in R$ and $r > 0$ such that $|\xi_i| \leq r$ for every component ξ_i of ξ. For fixed ν, exists $0 < M < r$ for which

$$F^\nu(\xi) > 1 - \varepsilon$$

for all ξ that for some component i, $r > \xi_i > M$ and the hypo-convergence implies that

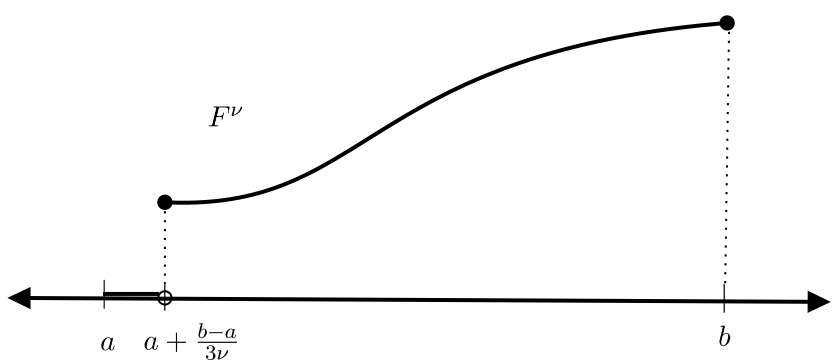
$$1 \geq F(\xi) \geq \limsup_{\nu \to \infty} F^\nu(\xi) > 1 - \varepsilon,$$

which proofs the first condition when $\varepsilon \to 0$. For what remains, let $a = (\alpha_1, \ldots, \alpha_m)$, the equi-usc implies by Theorem 2.10 that $F^\nu(a) \xrightarrow{p} F(a)$, thus $F(a) = 0$. □

Example 2.43 The equi-usc on every inferior point of the rectangle cannot be omitted. Suppose R is of the form $[\alpha_1, \beta_1] \times, \ldots, \times [\alpha_m, \beta_m]$. Let $\gamma_i = |b_i - a_i|$

$$F^\nu(\xi) = \begin{cases} 0 & \text{if } \xi_i \in [\alpha_i, \alpha_i + \gamma_i/\nu) \ \forall i = 1, \ldots, m \\ 1 & \text{otherwise} \end{cases}$$

$F^\nu \xrightarrow{h} F$, with F the function that is $F(\xi) = 1$ for every ξ with $\xi_i \neq \alpha_i$ for all i, and $F(\alpha) = 0$ if and only if $\lim_\nu F^\nu(\alpha) = F(\alpha)$. The following figure illustrates the previous phenomena by means of a similar one-dimensional setting.



Corollary 2.44 *Let $R \subset S$ be a bounded rectangular domain, A be any rectangle and a sequence $\{F^\nu\}_{\nu \in \mathbb{N}} \subset$ usc-fcns$(R; [0, 1])$ equi-usc at every vertex v_i of A and that $\Delta_A F^\nu \geq 0$. If F^ν hypo-converge to F, then $\Delta_A F \geq 0$ for every rectangle A.*

Proof. Let A be a rectangle in the mesh with vertices $\{v_i\}_{i=1}^{2m}$. Let's denote A^- the set of indexes such that $\operatorname{sign}_A v_j = -1$ and A^+ the ones with positive sign. The same arguments in Propositon 2.42 are true to prove now that

$$\sum_{j \in A^+} F(v_j) \geq \sum_{j \in A^-} F(v_j).$$

The left hand side is due to hypo-convergence,

$$\sum_{j\in A^+} F(v_j) \geq \limsup \sum_{j\in A^+} F^\nu(v_j) \geq \limsup \sum_{j\in A^-} F^\nu(v_j) \geq \liminf \sum_{j\in A^-} F^\nu(v_j) \geq \sum_{j\in A^-} \liminf F^\nu(v_j)$$

and the equi-usc in every $v_j \in A^-$ implies that $\liminf F^\nu(v_j) = F(v_j)$ which concludes the result. □

Remark 2.45 The equi-usc was used only in the vertices with with negative sign, for the remaining vertices suffice the hypoconvergence.

2.5. Epi-Splines

In this work we adapted the theory developed in [21] to the space of upper semi-continuous functions. Epi-splines are piecewise polynomial functions, that can approximate any lower semi-continuous function to an arbitrary level of accuracy. This section is an extract from [21] where it is developed a framework for epi-splines in the context of lower semicontinuous functions and although there must be some changes in the definition of epi-splines in order to obtain analogous the results for our case, we will still call them epi-splines. The definition for epi-splines given in [21] is the following:

Definition 2.46 (Lower semi-continuous epi-spline) *A (lsc) epi-spline* $s : S \to \overline{\mathbb{R}}$ *of order* $p \in \mathbb{N}$ *with partition* $\mathcal{R} = \{R_k\}_{k=1}^N$ *of a closed set* $B \subseteq \mathbb{R}^n$*, is a function that*

(I) *on each* $R_k, k = 1, \ldots, N$*, is polynomial of total degree* p*,*

(II) *has* $s(x) = \infty$ *for* $x \notin B$*,*

(III) *and for every* $x \in S$*, has* $s(x) = \liminf_{x' \to x} s(x')$*.*

Meanwhile our definition for epi-splines is,

Definition 2.47 (Upper semi-continuous epi-spline) *A (usc) epi-spline* $s : S \to \overline{\mathbb{R}}$ *of order* $p \in \mathbb{N}$ *with partition* $\mathcal{R} = \{R_k\}_{k=1}^N$ *of a closed set* $B \subseteq \mathbb{R}^n$*, is a function that*

(I) *on each* $R_k, k = 1, \ldots, N$*, is polynomial of total degree* p*,*

(II) *has* $s(x) = -\infty$ *for* $x \notin B$*,*

(III) *and for every* $x \in S$*, has* $s(x) = \limsup_{x' \to x} s(x')$*.*

The family of all (lsc) epi-splines is denoted by $\underline{\text{e}}\text{-spl}_n^p(\mathcal{R})$ and family of (usc) is denoted by $\overline{\text{e}}\text{-spl}_n^p(\mathcal{R})$.

Proposition 2.48 *For any partition* $\mathcal{R}$ *of a closed set* $S \subseteq \mathbb{R}^m, p \in N_0$*, and* $n \in \mathbb{N}$*,*

$$\underline{\text{e}}\text{-spl}_n^p(\mathcal{R}) \subset \text{lsc- fcns}(S) \subseteq \text{lsc- fcns}(\mathbb{R}^m)$$

and in the same way

$$\overline{\text{e}}\text{-spl}_n^p(\mathcal{R}) \subset \text{usc- fcns}(S) \subseteq \text{usc- fcns}(\mathbb{R}^m)$$

Remark 2.49 If $s: S \to \overline{\mathbb{R}}$ is an epi-spline (lsc or usc) of order p, with partition $\mathcal{R} = \{R_k\}_{k=1}^N$ of a closed set $B \subset \mathbb{R}^n$, then it involves N polynomials of total degree p and it is fully characterized by

$$n_e := Nn_p = \frac{N(n+p)!}{n!p!}$$

parameters.

Definition 2.50 (infinite refinement) *A sequence $\{\mathcal{R}^\nu\}_{\nu=1}^\infty$ of partitions of a closed set $B \subseteq \mathbb{R}^m$, with $\mathcal{R}^\nu = \{R_k^\nu\}_{k=1}^{N^\nu}$, is an infinite refinement if for every $x \in B$ and $\varepsilon > 0$, there exists $\bar{\nu} \in \mathbb{N}$ such that $R_k^\nu \subset \mathbb{B}(x, \varepsilon)$ for every $\nu \geq \bar{\nu}$ and k satisfying $x \in \mathrm{cl} R_k^\nu$.*

Example 2.51 In the case of a compact B, there are obvious choices of infinite refinements. A simple example of an infinite refinement on (unbounded) $\mathbb{R}$ is to take $N^\nu = 2\nu + 2$,

$$\begin{aligned} R_1^\nu &= (-\infty, -\sqrt{\nu}) \\ R_k^\nu &= ((k-\nu-2)/\sqrt{\nu}, (k-\nu-1)/\sqrt{\nu}) \text{ for } k = 2, 3, \ldots, 2\nu+1, \text{ and} \\ R_{2\nu+2}^\nu &= (\sqrt{\nu}, \infty) \end{aligned}$$

Then $\bar{\nu} > \max\left\{x^2, \varepsilon^{-2}\right\}$ satisfies the above condition. Obviously, much flexibility exists in constructing such infinite refinements.

Theorem 2.52 (dense approximation) *For any $p \in N_0$ and $\{\mathcal{R}^\nu\}_{\nu=1}^\infty$, an infinite refinement of a closed set $S \subseteq \mathbb{R}^m$,*

$$\bigcup_{\nu=1}^\infty \bar{\text{e}}\text{-spl}_n^p\left(\mathcal{R}^\nu\right) \textit{ is dense in } \text{usc-fcns}(S)$$

Proof. Let $f \in$ usc-fcns(S) and $\mathcal{R}^\nu = \{R_k^\nu\}_{k=1}^{N^\nu}$ an infinite refinement. It suffices to construct a sequence of (usc) epi-splines of order $p = 0$. For every $\nu \in N$ and $R_k^\nu, k = 1, 2, \ldots, N^\nu$. Let $\mathrm{cl} R_k^\nu$ be the closure of R_k^ν, we define

$$\sigma\left(R_k^\nu\right) := \begin{cases} \sup_{x \in \mathrm{cl} R_k^\nu} f(x) & \text{if } \sup_{x \in \mathrm{cl} R_k^\nu} f(x) \in [-\nu, \nu] \\ \nu & \text{if } \sup_{x \in \mathrm{cl} R_k^\nu} f(x) > \nu \\ -\nu & \text{otherwise} \end{cases}$$

and construct $s^\nu : S \to \overline{\mathbb{R}} \cup \{-\infty\}$ as follows:

$$s^\nu(x) := \begin{cases} \max\limits_{k=1,2,\ldots,N^\nu} \left\{\sigma\left(R_k^\nu\right) \mid x \in \mathrm{cl} R_k^\nu\right\}, & x \in S \\ -\infty & x \notin S. \end{cases}$$

Clearly, s^ν is constant on each $R_k^\nu, k = 1, 2, \ldots, N^\nu$ and satisfies

$$\limsup_{x' \to x} s^\nu\left(x'\right) = s^\nu(x) \text{ for all } x \in \mathbb{R}^m.$$

Hence, $s^\nu \in \bar{\text{e}}\text{-spl}_n^0\left(\mathcal{R}^\nu\right)$ and consequently also in $\bar{\text{e}}\text{-spl}_n^p\left(\mathcal{R}^\nu\right)$ for $p \in \mathbb{N}$. We next show that the two conditions of Proposition 2.7 holds. Let $x \in S$ be arbitrary. By upper semicontinuity of f, for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$f\left(x'\right) \leq f(x) + \varepsilon \text{ whenever } x' \in \mathbb{B}(x, \delta).$$

Since $\{\mathcal{R}^\nu\}_{\nu=1}^{\infty}$ is an infinite refinement, there also exists a $\bar{\nu}$ such that $R_k^\nu \subset \mathbb{B}(x,\delta)$ for every $\nu \geq \bar{\nu}$ and k satisfying $x \in \mathrm{cl} R_k^\nu$. Hence, there exists a neighborhood of x on which

$$s^\nu(x') \leq \min\left\{f(x') + \varepsilon, \nu\right\} \text{ for all } \nu \geq \bar{\nu} \text{ and } x' \in \mathbb{B}(x,\delta).$$

Thus, for every sequence $x^\nu \to x$,

$$\limsup_{\nu} s^\nu\left(x^\nu\right) \leq \limsup_{\nu} \min\left\{f(x) + \varepsilon, \nu\right\} = f(x) + \varepsilon$$

Since ε is arbitrary, $\limsup s^\nu\left(x^\nu\right) \leq f(x)$ and condition a) of Proposition 2.7 holds. For b), set $x^\nu = x$ for all ν. If $x \notin B$, then $s^\nu\left(x^\nu\right) = f(x) = -\infty$ and b) is satisfied. If $x \in B$ and $f(x) < +\infty$ then $s^\nu(x) \geq f(x)$ and $s^\nu(x) \geq -\nu$ for ν sufficiently large, so

$$s^\nu(x) \geq \max\left\{f(x), -\nu\right\} \text{ for all } \nu \text{ sufficiently large.}$$

If $x \in B$ and $f(x) = \infty$, then $s^\nu\left(x^\nu\right) = s^\nu(x) = \nu$. From the last two cases follows that,

$$\liminf_{\nu} s^\nu\left(x^\nu\right) = \liminf_{\nu} s^\nu(x) \geq f(x)$$

□

Remark 2.53 By considering only rational epi-splines of $\bar{\text{e}}\text{-spl}_n^0\left(\mathcal{R}^\nu\right)$ in the proof of Theorem 2.52, i.e., functions $s : S \to \overline{\mathbb{R}}$ with $s(x) = q_k$ for $x \in R_k^\nu$, q_k a rational constant, $k = 1, 2, \ldots, N^\nu$. Specifically, in that proof one can replace $\sigma\left(R_k\right) = \sup_{x \in \mathrm{cl}\, R_k^\nu} f(x)$ by $\sigma\left(R_k^\nu\right)$ equals any rational number in

$$\left[\max_{x \in \mathrm{cl}\, R_k^\nu} f(x), \min\left\{\nu, \max_{x \in \mathrm{cl}\, R_k^\nu} f(x) + 1/\nu\right\}\right]$$

and the next result follows.

Corollary 2.54 *(separability of the usc functions) For* $p \in N_0$ *and* $\{\mathcal{R}^\nu\}_{\nu=1}^{\infty}$*, an infinite refinement of a closed set* $S \subseteq \mathbb{R}^m$*, (usc-fcns*$(S), d\!l$*) is separable, with the rational epi-splines of*

$$\bigcup_{\nu=1}^{\infty} \bar{\text{e}}\text{-spl}_n^0\left(\mathcal{R}^\nu\right)$$

furnishing a countable dense subset.

Remark 2.55 The same results stated in Theorem 2.52 (and its Corollary 2.54) hold true for usc-fcns$_+\left(S; [0,1]\right)$ and Lip-fcns$_\kappa(S : [0,1])$ in virtue of Remark 2.21.

Chapter 3

Hypoestimation Problem

3.1. Introduction

The Attouch-Wets distance $d\!l$ given in (2.10) metrizes the hypo-convergence as it is stated in Theorem 2.17 but it is difficult to calculate. This metric is defined by the ρ-hypo distance $d\!l_\rho$ which even though is a pseudo-metric, it characterizes hypo convergence (Theorem 2.12). From these facts we can deduce that the topology induced by $d\!l$ is *almost equivalent* to the topology generated by $d\!l_\rho$. Even if the ρ-hypo distance $d\!l_\rho$ is easier to calculate than $d\!l$, we can still look for an easier way. Theorem 2.12 motivates the use of $\hat{d\!l}_\rho$ as an approximation of $d\!l_\rho$ and therefore to consider Problem 3.1, which is an approximation of Problem 1.1 when there is no penalty function π. For now, we will concentrate our efforts for when there is no penalty function since most laborious part of the work is the distance approximation.

Problem 3.1 Let F_0 and G_0 in usc-fcns$_+(\mathbb{R}^2; [0,1])$.

$$\text{Find} \quad \hat{F} \in \operatorname{argmin}_{F \in \mathscr{F}} \left\{ \hat{d\!l}_\rho\left(F, F_0\right) \;\middle|\; \hat{d\!l}_\rho(F, G_0) \leq \delta \right\}, \tag{3.1}$$

where $\mathscr{F}$ is a closed subset of usc-fcns$_+(\mathbb{R}^2; [0,1])$.

Observe that all the functions used before can have unbounded domain but because one of the objectives is to develop a computational purposes program we will consider a bounded rectangular domain $R \subset \mathbb{R}^2$ where all functions will be defined on.

Problem 3.2 Let F_0 and G_0 in usc-fcns$_+(R; [0,1])$.

$$\text{Find} \quad \hat{F} \in \operatorname{argmin}_{F \in \mathscr{F}} \left\{ \hat{d\!l}_\rho\left(F, F_0\right) \;\middle|\; \hat{d\!l}_\rho(F, G_0) \leq \delta \right\}, \tag{P}$$

where $\mathscr{F}$ is a closed subset of usc-fcns$_+(R; [0,1])$.

Our aim is to state the problem above as a problem that can be solved by a computational program. Let $\mathcal{R} = \{R_1, \ldots, R_N\}$ be a box partition of $\rho\mathbb{B}^\infty$ where each R_k is defined by $l^k = \left(l_1^k, l_2^k\right)$ and $u^k = \left(u_1^k, u_2^k\right)$. Consider functions $f, g \in$ usc-fcns$_+(R; [0,1])$ and the expressions

$$\eta_\rho^+(f, g) = \inf_{\eta \geq 0} \left\{ \eta \;\middle|\; \begin{array}{ll} g(l^k + \eta \mathbf{1}) + \eta & \geq \min\left\{f(u^k), \rho\right\} \text{ and,} \\ f(l^k + \eta \mathbf{1}) + \eta & \geq \min\left\{g(u^k), \rho\right\} \;\; \forall k = 1, \ldots, N \end{array} \right\} \tag{3.2}$$

$$\eta_\rho^-(f,g) = \inf_{\eta\geq 0}\left\{\eta \;\middle|\; \begin{array}{ll} g(l^k+\eta\mathbf{1})+\eta & \geq \min\left\{f(l^k),\rho\right\} \text{ and,} \\ f(l^k+\eta\mathbf{1})+\eta & \geq \min\left\{g(l^k),\rho\right\}\ \forall k=1,\dots,N \end{array}\right\} \tag{3.3}$$

Theorem 2.29 delivers upper and lower bounds for $\hat{d\!l}_\rho$ in terms of the expressions above and therefore we state the following problem

Problem 3.3

$$\text{Find } \hat{F} \in \operatorname{argmin}_{F\in\hat{\mathscr{F}}}\ \left\{\eta_\rho^+(F,F_0),\ \text{s.t } \eta_\rho^+(F,G_0)\leq\delta\right\},$$

where $\hat{\mathscr{F}}$ is a subset of $\bar{\text{e}}\text{-spl}_2^1(\mathcal{R})$.

Moreover, consider $\{\mathcal{R}^\nu\}_{\nu=1}^\infty$ an infinite refinement of $\rho\mathbb{B}^\infty$, this means that every $\mathcal{R}^\nu = \{R_1^\nu,\dots,R_{N^\nu}^\nu\}$ is a box partitions of $\rho\mathbb{B}^\infty$ and that the partition becomes finer as ν gets bigger. For $\rho>0$

$$\eta_\rho^+(f,g)^\nu = \inf_{\eta\geq 0}\left\{\eta \;\middle|\; \begin{array}{ll} g(l^k+\eta\mathbf{1})+\eta & \geq \min\left\{f(u^k),\rho\right\} \text{ and,} \\ f(l^k+\eta\mathbf{1})+\eta & \geq \min\left\{g(u^k),\rho\right\}\ \forall k=1,\dots,N^\nu \end{array}\right\} \tag{3.4}$$

$$\eta_\rho^-(f,g)^\nu = \inf_{\eta\geq 0}\left\{\eta \;\middle|\; \begin{array}{ll} g(l^k+\eta\mathbf{1})+\eta & \geq \min\left\{f(l^k),\rho\right\} \text{ and,} \\ f(l^k+\eta\mathbf{1})+\eta & \geq \min\left\{g(l^k),\rho\right\}\ \forall k=1,\dots,N^\nu \end{array}\right\} \tag{3.5}$$

and we will study the following sequences of problems

Problem 3.4

$$\text{Find } \hat{F} \in \operatorname{argmin}_{F\in\hat{\mathscr{F}}^\nu}\ \left\{\eta_\rho^+(F,F_0)^\nu,\ \text{s.t } \eta_\rho^+(F,G_0)^\nu\leq\delta\right\}, \tag{P$^\nu$}$$

where $\hat{\mathscr{F}}^\nu$ is a subset of $\bar{\text{e}}\text{-spl}_2^1(\mathcal{R}^\nu)$.

Proposition 3.5 (Existence of solutions) *Problems 3.1, 3.2, 3.3 and 3.4 are well defined and have solution.*

Proof. Problem 3.1 is proper because F_0 and G_0 belong to $(\text{usc-fcns}_+(\mathbb{R}^2;[0,1]),d\!l)$, besides $\hat{d\!l}_\rho$ is not a metric yet it still characterize hypo-convergence because of Theorem 2.12. The latter provides the lower semi-continuity of $\hat{d\!l}_\rho$ for the Attouch-Wets topology. Since $(\text{usc-fcns}_+(\mathbb{R}^2;[0,1]),d\!l)$ is a space where closed and bounded sets are compact, the set

$$\{F\in\text{usc-fcns}\left(\mathbb{R}^2;[0,1]\right)\ \mid \hat{d\!l}_\rho(F,G_0)\leq\delta\}\cap\text{usc-fcns}_+\left(\mathbb{R}^2;[0,1]\right)$$

is compact which concludes the existence. The same arguments are true for Problem 3.2 and still when this problem is defined for a closed subset of $(\text{usc-fcns}_+(\mathbb{R}^2;[0,1]),d\!l)$. Problem 3.3 and 3.4 have solution because of Remark 2.28.
□

3.2. Existence and convergence

Let $R\subset\mathbb{R}^2$ a bounded rectangular region and $\mathscr{F}\subset\text{usc-fcns}_+(R;[0,1])$ a closed set, then $(\mathscr{F},d\!l)$ is a complete metric separable space and there exists $\mathscr{F}^\nu\subset\mathscr{F}$ dense. Let $F_0,G_0\in\mathscr{F}$ and

$$\begin{aligned}\varphi : (\mathscr{F},d\!l) &\longrightarrow \overline{\mathbb{R}} \\ F &\longmapsto \varphi(F) := \hat{d\!l}_\rho(F,F_0)+\iota_C(F)\end{aligned}$$

where ι_C is an indicator function of C, i.e

$$\iota_C(F) = \begin{cases} 0 \text{ if } F \in C \\ +\infty \text{ otherwise} \end{cases}$$

with $C = \left\{ F \mid \hat{d\!l}_\rho(F, G_0) \leq \delta \right\}$ and the function

$$\begin{aligned} \varphi^\nu : (\mathscr{F}, d\!l) &\longrightarrow \overline{\mathbb{R}} \\ F &\longmapsto \varphi^\nu(F) := \eta_\rho^+(F, F_0)^\nu + \iota_{C^\nu}(F) + \iota_{\mathscr{F}^\nu}(F) \end{aligned}$$

where

$$C^\nu = \left\{ F \in \mathscr{F} \;\middle|\; \begin{array}{l} F\left(l^k + \delta \mathbf{1}\right) + \delta \geq \min\left\{G_0\left(u^k\right), \rho\right\}, \ \forall k = 1, \ldots, N^\nu \\ G_0\left(l^k + \delta \mathbf{1}\right) + \delta \geq \min\left\{F\left(u^k\right), \rho\right\}, \ \forall k = 1, \ldots, N^\nu \end{array} \right\}$$

Proposition 3.6 *For every $f \in \mathscr{F}$ and every sequence $f^\nu \in \mathscr{F}$ that hypo-convergence to f, we have that*

$$\liminf_\nu \varphi^\nu(f^\nu) \geq \varphi(f). \tag{3.6}$$

Proof. First observe that because of Theorem 2.29 we have that

$$\hat{d\!l}_\rho(f, g) \leq \eta_\rho^+(f, g)^\nu \quad \text{for any refinement } \mathcal{R}^\nu \text{ and } f, g \in \mathscr{F}.$$

and so every limit of subsequences $f^\nu \in C^\nu$ is in C, thus $\operatorname{LimOut} C^\nu \subset C$.

Let $f \in \mathscr{F}$ and $f^\nu \in \mathscr{F}$ that $f^\nu \xrightarrow{\text{h}} f$. If $f \notin C$ and $f^\nu \notin \mathscr{F}^\nu$ then (3.6) holds and is still true if $f \in C$ and $f^\nu \notin \mathscr{F}^\nu$. If $f \in C$, $f^\nu \in \mathscr{F}^\nu$ but $f^\nu \notin C^\nu$ then (3.6) holds. Let $f \in C$ and $f^\nu \in C^\nu \cap \mathscr{F}^\nu$ be any sequence that $f^\nu \xrightarrow{\text{h}} f$ (this is equivalent to $d\!l(f^\nu, f) \to 0$). We have that

$$\hat{d\!l}_\rho(f^\nu, F_0) \leq \eta_\rho^+(f^\nu, F_0)^\nu$$

because of Theorem (2.29). On the other hand for every $\overline{\rho} > 0$, $d\!l(f^\nu, f) \to 0 \iff \hat{d\!l}_\rho(f^\nu, f) \to 0$ for all $\rho \geq \overline{\rho}$, because of Theorem 2.12 and so

$$d\!l(f^\nu, F_0) \to d\!l(f, F_0) \iff \hat{d\!l}_\rho(f^\nu, F_0) \to \hat{d\!l}_\rho(f, F_0) \text{ for all } \rho \geq \overline{\rho}$$

and then

$$\liminf_\nu \varphi^\nu(f^\nu) = \liminf_\nu \left\{\eta_\rho^+(f^\nu, F_0)^\nu\right\} \geq \liminf_\nu \hat{d\!l}_\rho(f^\nu, F_0) = \lim_\nu \hat{d\!l}_\rho(f^\nu, F_0) = \hat{d\!l}_\rho(f, F_0)$$

$$\liminf_\nu \varphi^\nu(f^\nu) \geq \hat{d\!l}_\rho(f, F_0) = \varphi(f)$$

If $f \notin C$ and $f^\nu \in \mathscr{F}^\nu$ then it doesn't exist any $f^\nu \in C^\nu$ that hypo-converge to f. Suppose there is one $f^\nu \in C^\nu$ that does the trick. As $f \notin C$, $\hat{d\!l}_\rho(f, G_0) > \delta$ and

$$\liminf_{\nu \to \infty} \eta_\rho^+(f^\nu, G_0)^\nu \geq \liminf_{\nu \to \infty} \hat{d\!l}_\rho(f^\nu, G_0) = \hat{d\!l}_\rho(f, G_0) > \delta$$

and so there exists $\overline{\nu}$ such $\eta_\rho^+(f^\nu, G_0)^\nu > \delta$ for all $\nu \geq \overline{\nu}$ and this contradicts the fact that $f^\nu \in C^\nu$ for all ν. □

Assumption 3.7 Suppose that for every $f, F_0 \in \mathscr{F}$ and every f^ν that hypo-converge to f

$$\limsup_{\nu} \eta_\rho^+(f^\nu, F_0)^\nu - \eta_\rho^-(f^\nu, F_0)^\nu \leq 0 \tag{3.7}$$

The questions concerning when Assumption 3.7 is obtained will be discussed in the next section.

Assumption 3.8 Suppose that for $f \in \mathscr{F}$ such that $\varphi(f) < +\infty$ and $\hat{d\!l}_\rho(f, G_0) = \delta$ exists $f^n \in \mathscr{F}$ such that $f^n \xrightarrow{\text{h}} f$

$$\hat{d\!l}_\rho(f^n, G_0) < \delta, \quad \forall n \in \mathbb{N}. \tag{3.8}$$

Theorem 3.9 *Under Assumptions 3.7 and 3.8 for every $f \in \mathscr{F}$ there exists a sequence of functions $f^\nu \in \mathscr{F}$ that hypo-converge to f and*

$$\limsup_{\nu} \varphi^\nu(f^\nu) \leq \varphi(f). \tag{3.9}$$

Moreover $\varphi^\nu \xrightarrow{\text{e}} \varphi$ and C^ν set-converge to C.

Proof. If $f \notin C$ then $\varphi(f) = +\infty$ and (3.9) holds. Let $f \in C$ and for now suppose there exists $f^\nu \in \mathscr{F}^\nu$ that hypo-converge to f. From Assumption 3.7 we have that

$$0 \leq \limsup_{\nu \to \infty} \eta_\rho^+(f^\nu, F_0)^\nu - \eta_\rho^-(f^\nu, F_0)^\nu = 0 \tag{3.10}$$

$$\begin{aligned}
\limsup \eta_\rho^+(f^\nu, F_0)^\nu - \limsup \eta_\rho^-(f^\nu, F_0)^\nu &= \limsup \eta_\rho^+(f^\nu, F_0)^\nu + \liminf -\eta_\rho^-(f^\nu, F_0)^\nu & (3.11)\\
&\leq \limsup \eta_\rho^+(f^\nu, F_0)^\nu - \eta_\rho^-(f^\nu, F_0)^\nu & (3.12)\\
&\leq 0. & (3.13)
\end{aligned}$$

From Proposition 2.25 we have that

$$\eta_\rho^-(f^\nu, F_0)^\nu \leq \hat{d\!l}_\rho(f^\nu, F_0) \leq \eta_\rho^+(f^\nu, F_0)^\nu \tag{3.14}$$

then

$$\limsup \eta_\rho^-(f^\nu, F_0)^\nu \leq \limsup \eta_\rho^+(f^\nu, F_0)^\nu. \tag{3.15}$$

Combining the inequalities above we have that $\limsup \eta_\rho^+(f^\nu, F_0)^\nu = \limsup \eta_\rho^-(f^\nu, F_0)^\nu$ and

$$\limsup \eta_\rho^+(f^\nu, F_0)^\nu = \limsup \hat{d\!l}_\rho(f^\nu, F_0) = \lim \hat{d\!l}_\rho(f^\nu, F_0) = \hat{d\!l}_\rho(f, F_0) \tag{3.16}$$

where the last equality comes from the fact that $d\!l(f^\nu, f) \to 0$ and Theorem 2.12. In particular,

$$\limsup_{\nu} \eta_\rho^+(f^\nu, F_0)^\nu \leq \hat{d\!l}_\rho(f, F_0).$$

In order to have

$$\limsup_{\nu} \varphi^\nu(f^\nu) \leq \varphi(f),$$

we need that $f^\nu \in C^\nu$, so we are going to prove that for every $f \in C$ there exist $f^\nu \in C^\nu$ that $f^\nu \xrightarrow{\text{h}} f$. The latter is equivalent to have that $C \subset \operatorname{LimInn} C^\nu$ so by proving this we will have (3.9) and also the remaining condition for the set convergence of C^ν to C.

For $f \in C$, there are two cases. The first case is that $\hat{d\!l}_\rho(f, G_0) < \delta$. By density of $\mathscr{F}^\nu$ in $\mathscr{F}$, there exists $f^\nu \in \mathscr{F}^\nu$ that hypo-converge to f and by triangular inequality,

$$\hat{d\!l}_\rho(f^\nu, G_0) \leq \hat{d\!l}_\rho(f, G_0) + \hat{d\!l}_\rho(f^\nu, f) \quad \forall \nu,$$

and the hypo-convergence implies that for $\varepsilon = \delta - \hat{d\!l}_\rho(f, G_0) > 0$ there exists $\overline{\nu} \in \mathbb{N}$ that

$$\hat{d\!l}_\rho(f^\nu, G_0) \leq \delta \quad \forall \nu \geq \overline{\nu}.$$

The second case is that $\hat{d\!l}_\rho(f, G_0) = \delta$. By Assumption 3.8, it exists $f^n \in \mathscr{F}$ that $f^n \xrightarrow{h} f$ and satisfies (3.8). For fixed n there exists $g_n^\nu \in \mathscr{F}^\nu$ that $g_n^\nu \xrightarrow{h} f^n$ when $\nu \to \infty$. Repeating the arguments for the first case we have that for some $\overline{\nu}_n$,

$$\hat{d\!l}_\rho(g_n^\nu, G_0) < \delta \quad \text{for every } \nu \geq \overline{\nu}_n.$$

We choose f^ν as g_n^ν for $\nu \geq \overline{\nu}_n$. In both cases we obtain a sequence that hypo-converge to f when $\nu \to \infty$ and $\hat{d\!l}_\rho(f^\nu, G_0) \leq \delta$ for every ν but we need $\eta_\rho^+(f^\nu, G_0) \leq \delta$. However in view that $f^\nu \xrightarrow{h} f$, because of (3.16) we have that

$$\limsup_\nu \eta_\rho^+(f^\nu, G_0) = \hat{d\!l}_\rho(f^\nu, G_0) \leq \delta,$$

so for sufficiently large ν, $\eta_\rho^+(f^\nu, G_0) \leq \delta$ and in consequence $f^\nu \in C^\nu$ for sufficiently large ν.
□

Theorem 3.10 *Under Assumptions 3.7 and 3.8, every cluster point of sequences constructed from near minimizers of φ^ν in $\mathscr{F}$ is contained in* $\operatorname{argmin}_{F\in\mathscr{F}} \varphi(F)$ *provided that ε^ν vanishes, i.e*

$$\operatorname{LimOut}\left(\varepsilon^\nu\text{-}\operatorname{argmin}_{F\in\mathscr{F}} \varphi^\nu(F)\right) \subset \operatorname{argmin}_{F\in\mathscr{F}} \varphi(F)$$

Proof. Let

$$f^\star \in \operatorname{LimOut}\left(\varepsilon^\nu\text{-}\operatorname{argmin}_{F\in\mathscr{F}} \varphi^\nu(F)\right)$$

then it exists $\{\nu_k, k \in \mathbb{N}\}$ and $f^k \in \varepsilon^\nu\text{-}\operatorname{argmin}_{F\in\mathscr{F}^\nu} \varphi^\nu(F)$ that $f^k \xrightarrow{h} f^\star$. Let $g \in \operatorname{argmin}_{F\in\mathscr{F}} \varphi(F)$, from the epi-convergence of φ^ν to φ we have that there exists a sequence g^ν in $\mathscr{F}$ that $g^\nu \xrightarrow{h} g$ and

$$\limsup_\nu \varphi^\nu(g^\nu) \leq \varphi(g).$$

Since $g \in C$ and $C^\nu \xrightarrow{S} C$ we have that there exists $\overline{\nu} \in \mathbb{N}$ such that $\eta_\rho^+(g^\nu, G_0)^\nu \leq \delta$ for every $\nu \geq \overline{\nu}$ and so $\varphi^\nu(g^\nu) < +\infty$ for sufficiently large enough ν. Moreover,

$$\varphi(f^\star) \leq \liminf \varphi^{\nu_k}\left(f^k\right) \leq \liminf \left(\inf_{f\in\mathscr{F}^{\nu_k}} \varphi^{\nu_k}(f) + \varepsilon^{n_k}\right) \leq \limsup \varphi^{\nu_k}(g^{\nu_k}) \leq \varphi(g^\star) = \inf_{f\in\mathscr{F}} \varphi(f)$$

which proof that $f^\star \in \operatorname{argmin}_{F\in\mathscr{F}} \varphi(F)$. □

3.3. Lipschitz case

Suppose $F_0, G_0 \in \text{usc-fcns}_+(R, [0,1])$ are Lipschitz functions with Lipschitz constants κ_1 and κ_2 respectively. Define $\kappa = \max\{\kappa_1, \kappa_2\}$ and let $\text{Lip-fcns}_\kappa(R)$ the set of Lipschitz functions define over R with Lipschitz constant κ.

Remark 3.11 The set $\text{usc-fcns}_+(R;[0,1]) \cap \text{Lip-fcns}_\kappa(R)$ is closed under the Attouch-Wets topology in view of Remark 2.21 and Proposition 2.20 and in consequence is also a complete separable metric space. Furthermore because of Theorem 2.54 there exists a sequence $\mathscr{F}^\nu \subset \bar{\text{e}}\text{-spl}_2^1(\mathcal{R}^\nu)$ that approximates $\text{usc-fcns}_+(R;[0,1]) \cap \text{Lip-fcns}_\kappa(R)$.

Corollary 3.12 *Let* $\mathscr{F} = \text{usc-fcns}_+(R;[0,1]) \cap \text{Lip-fcns}_\kappa(R)$ *and suppose Assumption 3.8 is valid, then* $\varphi^\nu \xrightarrow{\text{e}} \varphi$ *and the conclusion in Theorem 3.10 holds.*

Proof. From Corollary 2.30 we have that

$$0 \leq \eta_\rho^+(f^\nu, F_0)^\nu - \eta_\rho^-(f^\nu, F_0)^\nu \leq \kappa \,\text{mesh}(\mathcal{R}^\nu) \tag{3.17}$$

so

$$\limsup_{\nu \to \infty} \eta_\rho^+(f^\nu, F_0)^\nu - \eta_\rho^-(f^\nu, F_0)^\nu = 0 \tag{3.18}$$

which delivers Assumption 3.7. □

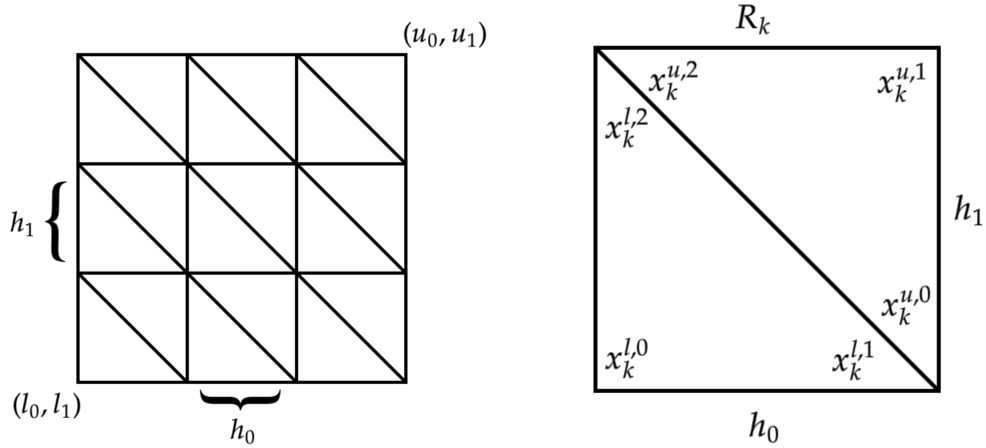
Remark 3.13 Future lines of research are focus on finding a more general class of functions than Lipschitz functions for which a similar bound such as equation 3.17 is satisfied. Lipschitz functions fit this condition because the Lipschitz constant bound uniformly the difference between images, no matter how small is the difference between two points.

3.4. Algorithm

The sequence of approximated problems (Problem 3.4) are defined by the function η_ρ^+ which is defined by a linear optimization problem. We take advantage of the linearity and we propose an algorithm that that for a fixed mesh, returns a function for which the objective value in Problem 3.3 is minimum.

Let $R = [l_0, u_0] \times [l_1, u_1]$ be a bounded rectangular domain and $\mathcal{R} = \{R_1, \ldots, R_\nu\}$ a partition of R. Each rectangular element R_k is defined by $l^k = (l_1^k, l_2^k)$ and $u^k = (u_1^k, u_2^k)$. We work on a triangular partition of the domain, this means that each rectangular element R_k is divided into two triangles, namely lower and upper, as described in Figures 3.1b.

Each function $F \in \bar{\text{e}}\text{-spl}_2^1(\mathcal{R})$ (piecewise affine functions) is represented by the values that it takes on vertices of each triangular region $x_k^l = (x_k^{l,0}, x_k^{l,1}, x_k^{l,2})$ and $x_k^u = (x_k^{u,0}, x_k^{u,1}, x_k^{u,2})$ (dropping the dependence on the rectangular element.



(a) Partition of the domain $[l_0, u_0] \times [l_1, u_1]$ (b) Description of the triangular elements

Figure 3.1: Triangular mesh of rectangular domains where epi-splines are constructed.

Fix $\eta > 0$. The algorithm consists on finding values $F(x_k^{u,2})$, $F(x_k + \eta\mathbf{1})$ and the minimum $s \geq 0$ that satisfy

$$F(x_k^{u,2}) \leq F_0\left(x_k^{l,1} + \eta\mathbf{1}\right) + \eta + s \quad \text{and} \tag{3.19}$$

$$F_0\left(x_k^{l,2}\right) \leq F\left(x_k + \eta\mathbf{1}\right) + \eta + s \quad \forall k = 1, \ldots, \nu \tag{3.20}$$

$$F\left(x_k^{u,2}\right) \leq G_0\left(x_k^{l,1} + \delta\mathbf{1}\right) + \delta \quad \text{and} \tag{3.21}$$

$$G_0\left(x_k^{l,2}\right) \leq F\left(x_k + \delta\mathbf{1}\right) + \delta \quad \forall k = 1, \ldots, \nu. \tag{3.22}$$

When there is no specification about the triangle (u or l) as in values $F(x_k + \eta\mathbf{1})$ in equations (3.20)-(3.22) is because it will depend on the new position $x_k + \eta$.

If the problem is feasible, we repeat the procedure with a smaller η, if not, we proceed with a larger η. We use a Bisection method on η to find the least slackness s. The function found by the

the previous procedure correpond to $\eta_\rho^+(f, F_0)^\nu$ s.t for $\eta_\rho^+(f, G_0)^\nu \leq \delta$ when ρ is large enough. The Algorithm is described as follows. First select the desired δ.

Algorithm 1 Hypo-estimate(η)

Solve

$$\begin{aligned}
\min_{s \geq 0, F} \quad & s \\
\text{s.t} \quad & F\left(x_k^{u,2}\right) \leq F_0\left(x_k^{l,1} + \eta \mathbf{1}\right) + \eta + s \\
& F_0\left(x_k^{l,2}\right) \leq F\left(x_k + \eta \mathbf{1}\right) + \eta + s \\
& F\left(x_k^{u,2}\right) \leq G_0\left(x_k^{l,1} + \delta \mathbf{1}\right) + \delta + s \\
& G_0\left(x_k^{l,2}\right) \leq F\left(x_k + \delta \mathbf{1}\right) + \delta + s \\
& F \in \text{usc-fcns}(R) \\
& F \text{ is non-decreasing}
\end{aligned}$$

return F, s

Algorithm 2 Binary Search

```
Set ε1, ε2 > 0, ηl = 0, ηu = 1.
procedure Binary Search(ηl, ηu)
    while |ηl − ηu| > ε2 do
        Set η = (ηl+ηu)/2
        F, s = Hypo-estimate(η)
        if s > ε1 then
            ηl = η
            F, s = Binary Search(ηl, ηu)
        else
            Set ηu = η
            F, s = Binary Search(ηl, ηu)
        end if
    end while
    return F, s, η
end procedure
```

The algorithm permits the use of a wide variety of state-of-the-art optimization to solve this linear problem and also there are some constraints that can be easily implemented in a linear manner, for instance pointwise bounds.

3.4.1. Constraints

For the rest of this section, we will consider vectors $z_k^l = (z_k^{l,0}, z_k^{l,1}, z_k^{l,2})$ and $z_k^u = (z_k^{u,0}, z_k^{u,1}, z_k^{u,2})$ to be the values of F on he upper a lower (respectively) triangle of the rectangular element k.

Upper/Lower semi continuity and monotonicity

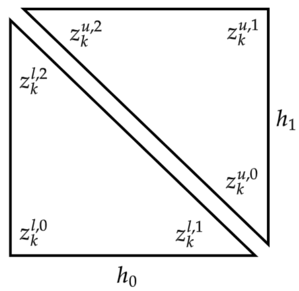


Figure 3.2: Values of F in rectangle R_k

The monotonicity constraint is implemented adding the following inequalities in addition to the corresponding inequalities with the neighbour rectangles.

$$\begin{aligned} z_k^{u,2} \leq z_k^{u,1} \qquad z_k^{u,0} \leq z_k^{u,1} \\ z_k^{l,0} \leq z_k^{l,1} \qquad z_k^{l,0} \leq z_k^{l,2} \end{aligned} \tag{3.23}$$

The upper semi-continuous constraint is represented by the following constraints

$$z_k^{l,1} \leq z_k^{u,0} \qquad z_k^{l,2} \leq z_k^{u,2}, \tag{3.24}$$

and the lower semi-continuous is represented by equation (3.24) with the reverse inequality.

Bounded growth constraint

Based on figure 3.2, the bounded growth constraint of parameter $L > 0$ is implemented on the lower triangle and the diagonal as follows,

$$\begin{aligned} z_k^{u,1} - z_k^{l,0} &\leq L\sqrt{h_0^2 + h_0^1} \\ z_k^{l,1} - z_k^{l,0} &\leq Lh_0 \\ z_k^{l,2} - z_k^{l,0} &\leq Lh_1. \end{aligned} \tag{3.25}$$

The same idea is replied to the upper triangle. This constraint sets that the height of epi-splines in each triangular element cannot be larger than L relative to the mesh or in other words, the rate of change in each triangular element cannot be larger than L.

Distribution Constraint

The distribution constraint for a function F defined over a subset of $\mathbb{R}^2$ is represented in Figure 2.2. For a function to satisfy the distribution constraint, it must satisfy it in all possible rectangles A. Implementing this constraint in all the possible rectangles would take $\mathcal{O}(N^4)$ while implementing it in the rectangles of the mesh is $\mathcal{O}(N^2)$, since for a mesh of size $N \times M$ the number of rectangles is

$$n_{\text{rectangles}} = \frac{N(N+1)M(M+1)}{4}. \tag{3.26}$$

The constraint for the upper right element of the mesh k is expressed as the following

$$z_k^{u,1} - z_k^{l,2} + z_k^{l,0} - z_k^{u,0} \geq 0.$$

The expression for the other rectangular elements changes slightly but the the idea is unchanged.

Chapter 4

Numerical experiments

In this section we show numerical results for Algorithm 2, we will refer as a *solution* to whatever results from Algorithm 2, is important to clarify this because a *real solution* to Problem 3.3 implies that the constraint related to the ambiguity set of size δ is satisfied, in other words, that $s = 0$ in the notation of Algorithm 2. We will study two different settings for the analysis of the parameters and it's effect on the solutions obtained. Also, we will solve the problem of estimating the position of an underwater vehicle given two different cumulative distributions functions assembled from noisy sources of information.

Algorithm 2 was implemented in Python and we used the Pyomo library ([5], [13]) to solve each iteration of it. In particular, we used Cplex ([7]) as a solver for Pyomo. All experiments were conducted on an Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz, 1.99 GHz with 16,0 GB RAM.

4.1. Estimation from two uniform distribution

This section is devoted to study how the solution to different settings changes as we consider different functions F_0, G_0 and values δ in Problem (3.3):

$$\text{Find } \hat{F} \in \text{argmin}_{F\in\hat{\mathscr{F}}} \left\{\eta_\rho^+(F, F_0),\ \text{s.t } \eta_\rho^+(F, G_0) \leq \delta\right\},$$

where $\hat{\mathscr{F}}$ is a subset of $\bar{\text{e}}\text{-spl}_2^1(\mathcal{R})$. We are going to study 2 settings that consists on uniform distribution functions F_0 and G_0, different on each setting. In each setting there is one function (F_0 or G_0) that dominates the other one and upon that we can examine the solutions' behaviour for each one of them. In addition, we present results for when the bounded growth constrained is included in non-smooth solutions and as well about the impact that implementing the distribution constraint in just a subset of rectangles has. All experiments were run with a mesh of 100 points per axis ($N = M = 100$) and tolerances $\varepsilon_1 = \varepsilon_2 =$ 1e-8. Next is detailed the experiments to be executed in each setting.

Theorem 2.29 states that η_ρ^+ is an upper bound for $\hat{d\!l}_\rho$ but we do not know if it is tightest upper bound. In practice, this could mean that solving Problem (3.3) may not be enough to obtain a solution of Problem 3.2. If $F_0 \in \mathscr{F}$ the solution of Problem 3.2 for large values of δ is F_0 and the optimal value is 0. The aim of this experiment is to check that for large values of δ, the algorithm retrieves the function in the objective function, F_0, in Problem (3.3) and also to check for small values of δ, how close the solution obtained with Algorithm (2) is to G_0.

The following notation will be used. The value δ represents the same as in Problem (3.3), the value η is value given by the Binary Search in Algorithm 2 and should be interpreted as an estimation for $\hat{d}_\rho(F_0, F)$, where F is a solution given by Algorithm 2. The value s, in the objective function of Algorithm 2, is meant to be less than a given tolerance, this value should be interpreted as a tolerance for the constraints involving G_0 to be violated. The expected value with respect to a function F is calculated as follows.

$$\mathbb{E}_F[f] = \int\int f(x,y)dF(x,y) \approx \sum_{R=[a,b]\times[c,d]\in\mathcal{R}} f(\bar{x})\mu([a,b]\times[c,d]) \tag{4.1}$$

where

$$\mu([a,b]\times[c,d]) = F(b,d) - F(a,d) - F(b,c) + F(a,c) \tag{4.2}$$

and $\bar{x}$ is the middle point of $[a,b]\times[c,d]$. Let $(x,y)\in\mathbb{R}^2$, the value $\mathbb{E}_x$ represents $\mathbb{E}_F[x]$ and $\mathbb{E}_y$ is $\mathbb{E}_F[y]$ in equation (4.1). the values are the expected value of a solution F with respect to the first and second variable respectively.

The distribution constraint is implemented just in rectangles on the mesh and this does not imply that it is true in all possible rectangles, so we will check how many of the total amount of rectangles fail the constraint. R_{true} is the number of rectangles that satisfy the distribution constraint and R_{false} is the number of rectangles that don't satisfy the constraint, both of them were computed with a tolerance of 1e-6. The fourth column presents the relative percentage error of the above. For a mesh of size $N = M = 100$ there are $n_{\text{rectangles}} = 25.502.500$ (in view of equation 3.26). The percentage error (% error) is obtained as

$$\%\text{error} = \frac{R_{false}}{R_{false} + R_{true}} \cdot 100.$$

Finally, we are going to incorporate the bounded growth constraint in Algorithm (2) and obtain solutions for different L values. This procedure will be applied to solutions found in Experiment 1 to be highly non-smooth.

4.1.1. Setting 1

In this setting we will work with two uniforms with the support of one inside the other one. Consider F_0 a uniform distribution over $[0,2] \times [0,2]$ and G_0 a uniform distribution over $[0.5, 1.5] \times [0.5, 1.5]$. The expected values of F_0 and G_0 by coordinates x and y calculated as in equation (4.1) are $\mathbb{E}_{F_0} = (1.0000, 1.0000)$ and $\mathbb{E}_{G_0} = (0.9998, 0.9998)$. These values will be important to do the comparison.

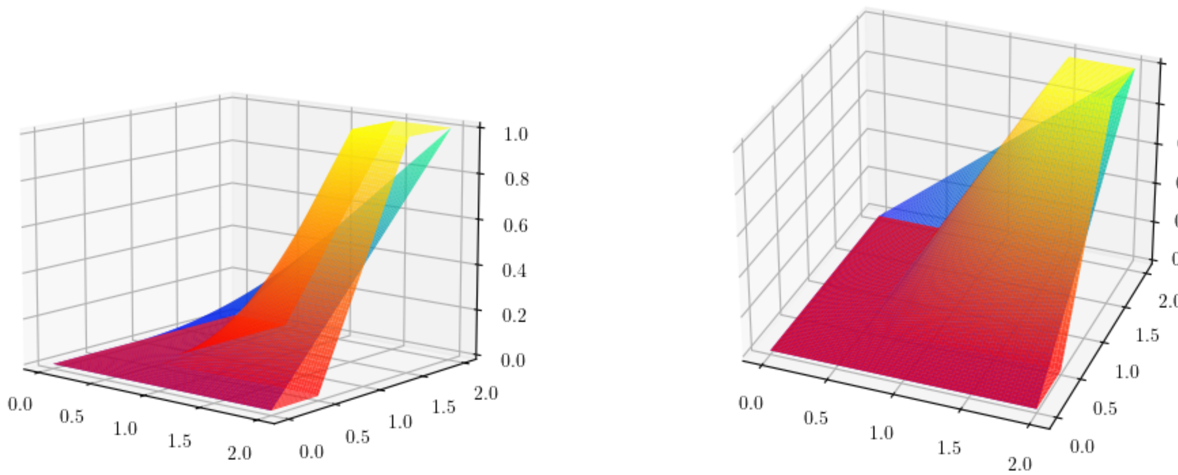


Figure 4.1: F_0 (cold colors) and G_0 (warm colors), note that $\hat{d}_\rho(F_0, G_0) = 1$

In view of Remark 2.28, $\delta = 1$ is large enough for the ambiguity set to be ignored. Table 4.1 shows realizations of the algorithm for Setting 1 with different values of the parameter δ. The corresponding expected value vectors are included and they should be compared with values $\mathbb{E}_{F_0}$ and $\mathbb{E}_{G_0}$. For $\delta = 1$ the expected value is underneath the one for F_0 and as δ decreases, this value increases, approaching an even overcoming G_0's expected value. For $\delta = 1.0, 0.5, 0.1$, the value s is (practically) 0, which means the solution obtained by Algorithm 2 is in fact at a distance at least δ to G_0, but as δ gets closer to 0, the distance constraint related to G_0 becomes hard to be satisfied at the solution obtained by Algorithm 2 is allows to violate this constraint in $s = 0.0298$ although the approximated distance is accurate ($\eta = 1$).

δ	s	η	$\mathbb{E}_x$	$\mathbb{E}_y$	R_{true}	R_{false}	% error	execution time $[s]$
1.0000	6.84e-09	0.0119	0.9785	0.9811	25239284	253096	0.9928	3122.02
0.5000	6.85e-09	0.0119	0.9697	1.0125	25088003	414497	1.6253	2174.22
0.1000	0.0	0.1602	0.8819	0.8580	24397261	1105239	4.3338	1271.60
0.0001	0.0298	1.0	1.0100	1.0052	25273719	228781	0.8970	471.25

Table 4.1: Different values of δ for Setting 1 with $N = M = 100$

Figure 4.2 (next page) shows solutions for $\delta = 1.0, 0.5, 0.0001$ which exhibit the general behaviour in the solutions as δ decreases. The first row shows the solution for $\delta = 1.0$ which functions are shown as almost the same. As the parameter δ decrease, the solution function F starts to reshape into G_0 from the inside. The last row shows how close the solution F gets to G_0.

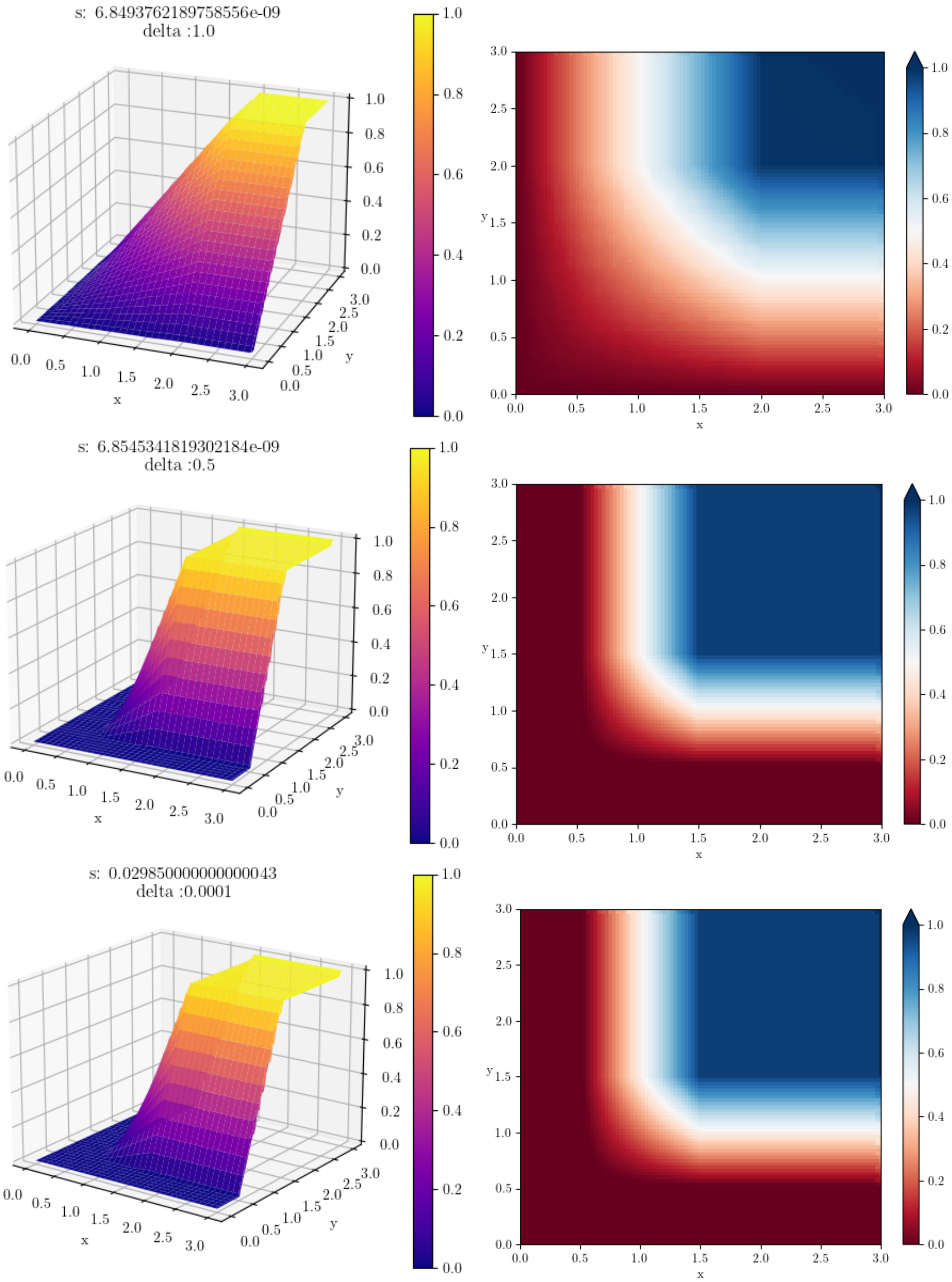


Figure 4.2: Evolution of solutions obtained for Setting 1 with $N = M = 100$. The solution F (left) and it's corresponding heatmap (right)

4.1.2. Setting 2

In this setting we will work with two uniforms whose supports are disjoint. Consider F_0 a uniform cumulative distribution over $[0,1]\times[0,1]$ and G_0 a uniform cumulative distribution over $[2,3]\times[2,3]$. The expected values of F_0 and G_0 by coordinates x and y calculated as in equation (4.1) are

$$\mathbb{E}_{F_0} = (0.5001, 0.5000) \quad \mathbb{E}_{G_0} = (2.4998, 2.4998)$$

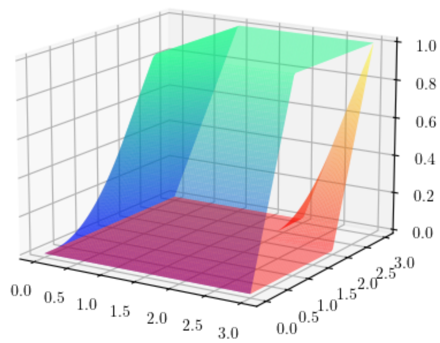


Figure 4.3: F_0 (cold colors) and G_0 (warm colors), note that $\hat{d\!l}_\rho(F_0, G_0) = 1$

The fact that $\hat{d\!l}_\rho(F_0, G_0) = 1$ implies that the approximated distance, represented by η, should be 1 for small enough δ and also that for any value of δ picked below 1, should return a solution with positive approximated distance η and should be 0 only when $\delta \geq 1$. Table 4.2 shows for $\delta = 1.0, 0.7, 0.1$ there was no violation of the constraints imposed so that the solution obtained with Algorithm 2 is at a distance of at least δ to G_0. For $\delta = 0.0001$, the value s means that the solution obtained can be assured to be just at a distance $\delta + s$, nonetheless, this value has been shown to decreased as the mesh gets finer. Moreover, for the same δ as before, the approximated distance, η, coincides with the theoretical value indicated in Figure 4.3. Note that the expected value at $\delta = 1$ is close to the expected value of F_0 and approaches to the expected value of G_0 as δ decreases. Note that all % errors are under 1%.

δ	s	η	$\mathbb{E}_x$	$\mathbb{E}_y$	R_{true}	R_{false}	% error	execution time $[s]$
1.0000	0.0	0.0199	0.4597	0.5021	25242254	250126	0.9811	3326.92
0.7000	0.0	0.3000	1.1400	1.1382	25403886	98614	0.3866	558.30
0.1000	0.0	0.8999	2.6385	2.6441	25483010	19490	0.0764	483.72
0.0001	0.05880	1.0	2.6385	2.6441	25499940	2560	0.0100	453.85

Table 4.2: Different values of δ for Setting 2 with $N = M = 100$

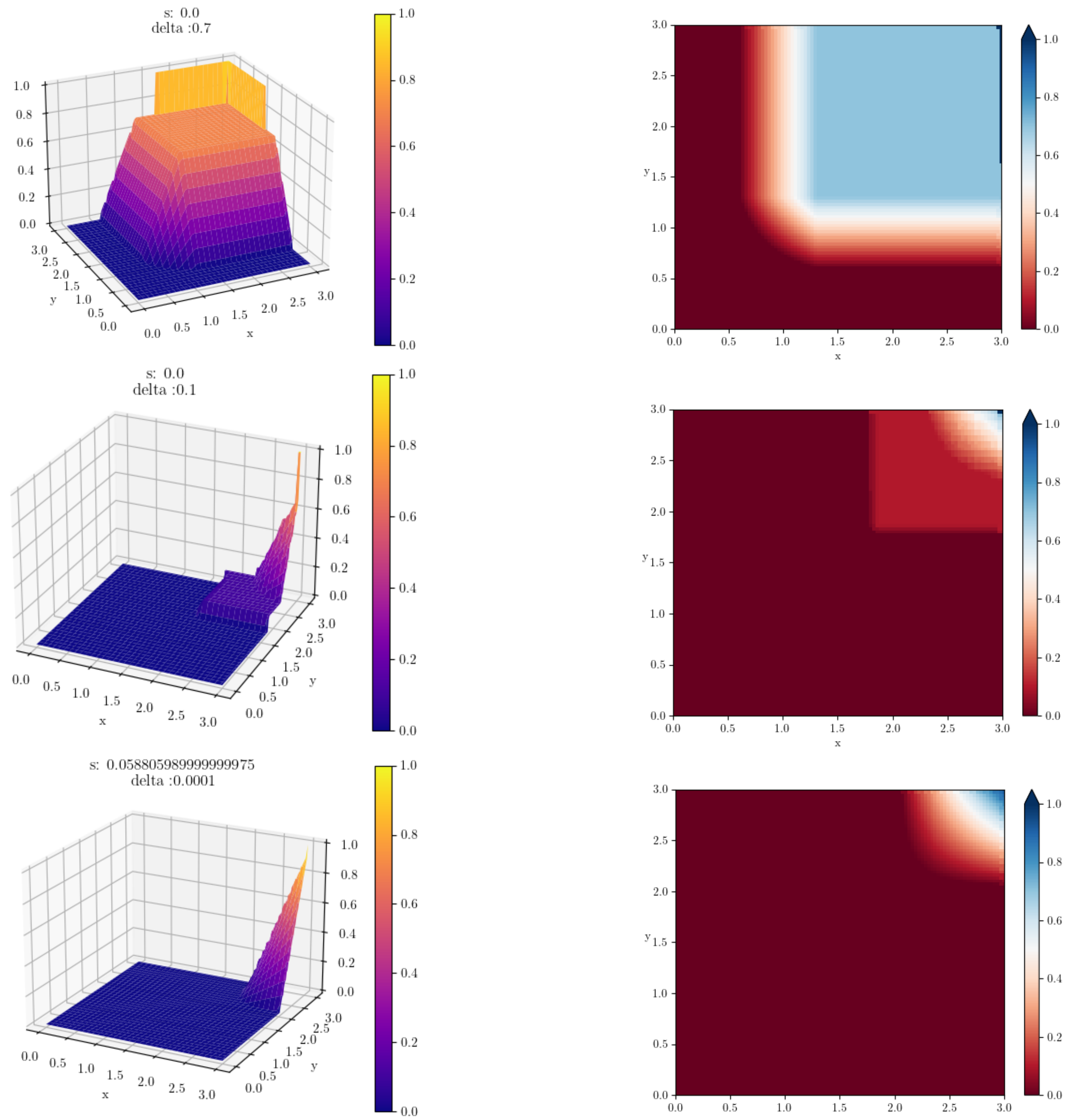


Figure 4.4: On the left is F and on the right is the corresponding heatmap

Figure 4.4 presents some of the solutions shown in Table 4.2. As δ decreases, Figure 4.4 shows that for $\delta = 0.7$, the solution follows the shape of F_0 and as δ decreases, it approaches steadily to G_0. The solution obtained by Algorithm 2 for $\delta = 0.0001$ should be compared with G_0 in Figure 4.3.

Bounded growth constraint

Solution for $\delta = 0.7$ and $\delta = 0.1$ show discontinuities. In this section we added the bounded growth constraint to obtain a softer solution as we can observe when comparing Figure 4.4 (first and middle row) with Figure 4.5 and Figure 4.6 respectively. Note that for $\delta = 0.7$, $s = 0.0$, thus this solution is in fact at a distance at least δ to G_0 as in opposed to $\delta = 0.1$ with $L = 0.85$ which can be assured to be just at a distance $s + \delta$ to G_0. The value s can be reduced by choosing a bigger constant L.

Figure 4.5: Solution for Setting 2 with $\delta = 0.7$ and $L = 1$

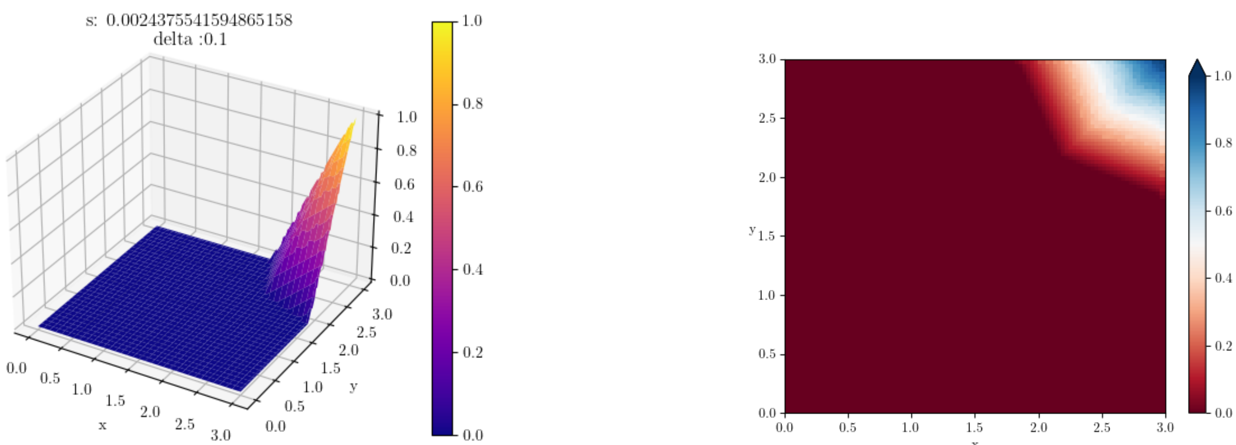


Figure 4.6: Solution for Setting 2 with $\delta = 0.1$ and $L = 0.85$

δ	L	s	η	$\mathbb{E}_x$	$\mathbb{E}_y$	execution time $[s]$
0.7000	1.0	0.0	0.2999	0.8017	0.7675	708.26
0.1000	0.85	0.0024	1.0	2.4030	2.4030	425.33

Table 4.3: Solutions for $\delta = 0.7, 0.1$ when adding a growth constraint constant L in Setting 2 with $N = M = 100$

Note that for $\delta = 0.1$ without the L constraint, the value $s = 0.0$ meanwhile by adding it, is now $s = 0.0024$. Thus, the new solution when the bounded growth condition is added is softer than the previous one but it violates the constraint of the size of the ambiguity set. The previous behaviour does not happen for $\delta = 0.7$ and $L = 1$ since $s = 0.0$ in both cases. Another thing that is interesting to notice is that the approximated distance in both cases (with and without L) in solution for $\delta = 0.7$ is almost exactly the same, meanwhile in the case of $\delta = 0.1$, the approximated distance increased when the bounded growth constraint is added.

4.2. Unmanned Underwater Vehicle Problem

Consider an unmanned underwater vehicle (uuv) that is returning to a docking station after a long mission. Although the uuv knows the location of the docking station on the map, it has only a vague idea about its own location because it has been underwater for a long time and has only used an inertial navigation system during that time. The docking station sends out pings that can be picked up by the uuv when close enough. The uuv can use these pings to improve the estimate of its own location. The uuv also has an accurate model (in the short term) of where it will be, given that it knows its initial condition.

Problem formulation

Let F_t be the cdf of the uuv location over the (x_1, x_2) coordinates based on the intertial navigation system and $y_t = (y_t^1, y_t^2)$ the ping data informed by the docking station at time t. If the ping data had no noise, y^t would be the true location of the uuv at time t. However, it has noise. Let f a function that models the location changes, this function is assumed to be a known model and it's called Dubin's Model.

Imagine we are at period t. The ping data available: $y^1, \ldots, y^t$ can be brought forward to the current time t using the Dubin's model by setting

$$z^t = y^t, \quad z^{t-1} = f(y^{t-1}), \quad z^{t-2} = f(f(y^{t-2})), \; \ldots, \; z^1 = f(\ldots f(y^1) \ldots)$$

We can view $z^1, \ldots, z^t$ as a sample of the location of the uuv at time t. Let G_t be the corresponding empirical distribution function. The Dubin's model f is given by:

$$x^{t+1} = x^t + \left[v \cos x_3^t, v \sin x_3^t, u^t \right] \tag{4.3}$$

where u^t is a scalar control input at time t and v is a constant velocity. The state x^t is three dimensional: x_1^t is the horizontal coordinate at time t, x_2^t is the vertical coordinate, and x_3^t is the heading at time t. Since we will only do this in two-dimensions, we will simply ignore the information related to heading, which anyhow the ping-data doesn't predict anyways. The scalar control u^t is the rate of change of the heading. If you set it to zero, then the uuv will go straight.

The problem is to find a cdf that models the position of the uuv at time t that is close to G_t because it represents the received data. However, the data is subject to uncertainty and not trustworthy for small t. So we give G_t an uncertainty set. We pick as estimate the cdf closest to F_t within that uncertainty set.

Instance construction

F_1 is the distribution of the initial positions and it's given by

$$F_1 \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5 & 0.25 \\ 0.25 & 0.25 \end{pmatrix} \right)$$

by sampling F_1 N times, we obtain N initial conditions. We randomly generate N initial headings as

$$x_3^{0,k} \sim \frac{1}{2} \left(-\frac{3}{20} + \frac{1}{3} \mathrm{U}[0,1] \right)$$

which have small variance and from these we generate u as

$$u^t = 0, \quad u^{10+t} = \frac{1}{100} + \mathrm{U}[0, 0.1], \quad u^{20+t} = \frac{1}{100} - \mathrm{U}[0, 0.1], \quad t = 1, \ldots, 10$$

This choise of u means driving straight at first, then turning slightly left and then turning slightly right. We set the speed as the constant value $v = \frac{1}{10}$. Each sample gives $\left(x_1^1, x_2^1, x_3^1\right)$, and we can use this to simulate forward using

$$x^{k+1} = x^k + \left[v \cos x_3^k, v \sin x_3^k, u^k\right], \quad k = 1, \ldots, t-1.$$

For each $k = 1, \ldots, t$, we have N data point about location of the uuv. The corresponding empirical cdfs define $F_1, \ldots, F_t$ after having dropped the third coordinate are bivariate. The true trajectory is considered as $\bar{x} = (\bar{x}^1, \ldots, \bar{x}^t)$, and the ping data is generated by $y^k \sim \mathcal{N}\left(\bar{x}^k, \Sigma\right)$, $k = 1, \ldots, t$. With this we compute

$$z^l = f^{(t-l)}\left(y^l\right), l = 1, \ldots, t$$

and construct G_k from it, for $k = 1, \ldots, t$, the empirical cdf based on $z^1, \ldots, z^k$ for $k = 1, \ldots, t$.

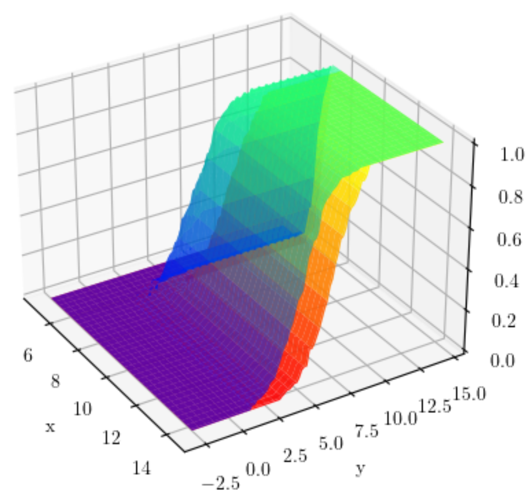


Figure 4.7: Functions F_t in cold colors and G_t in warm colors for $t = 90$ (total steps) and $N = 100$ (sample points).

The expected values of F_t and of G_t are

$$\mathbb{E}[F_t] = (8.8261, 1.0724), \quad \mathbb{E}[G_t] = (9.9965, 1.7468)$$

Finally, with Algorithm 1 we find a function $\hat{F}^t$ that minimize $\eta_\rho^+(F, F_t)^\nu$ with $F \in \mathscr{F}^\nu$ but $\eta_\rho^+(F, G_t)^\nu \leq \delta$. Table 4.4 illustrate the results obtained for different values of δ. This values represent how trustworthy is G_t, we have chosen three values that serve as three possible scenarios.

	s	η	$\mathbb{E}_x$	$\mathbb{E}_y$	execution time
$\delta_0 = 0.90$	6.0280e-09	0.0787	9.034	1.8908	969.19
$\delta_1 = 0.10$	0.0	0.3501	9.734	1.2908	1007.80
$\delta_2 = 0.01$	0.0824	1.0	10.3023	1.8908	1008.37

Table 4.4: Solution functions for the UUV problem.

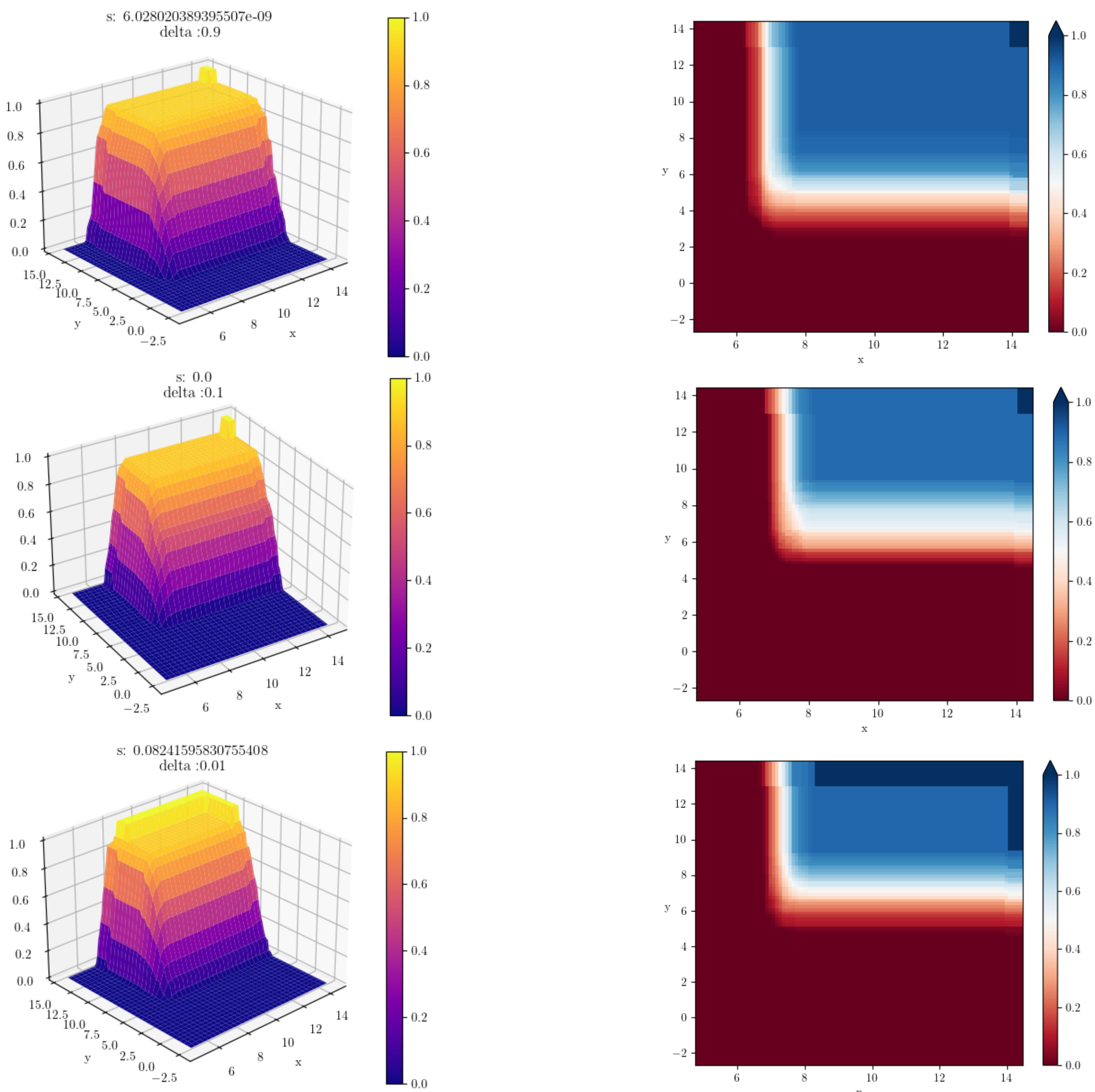


Figure 4.8: Solutions displayed in Table 4.4 (left) and it's corresponding heatmap (right).

The expected value of the functions displayed in Figure 4.8 can be used to give an answer to the actual position of the UUV. Figure 4.9 shows the expected values of the solutions found. This figure exhibit how close is the expected value of solution δ_i to functions F_t and G_t. The results show that the small it is the ambiguity set, the close it is to the expected value of G_t.

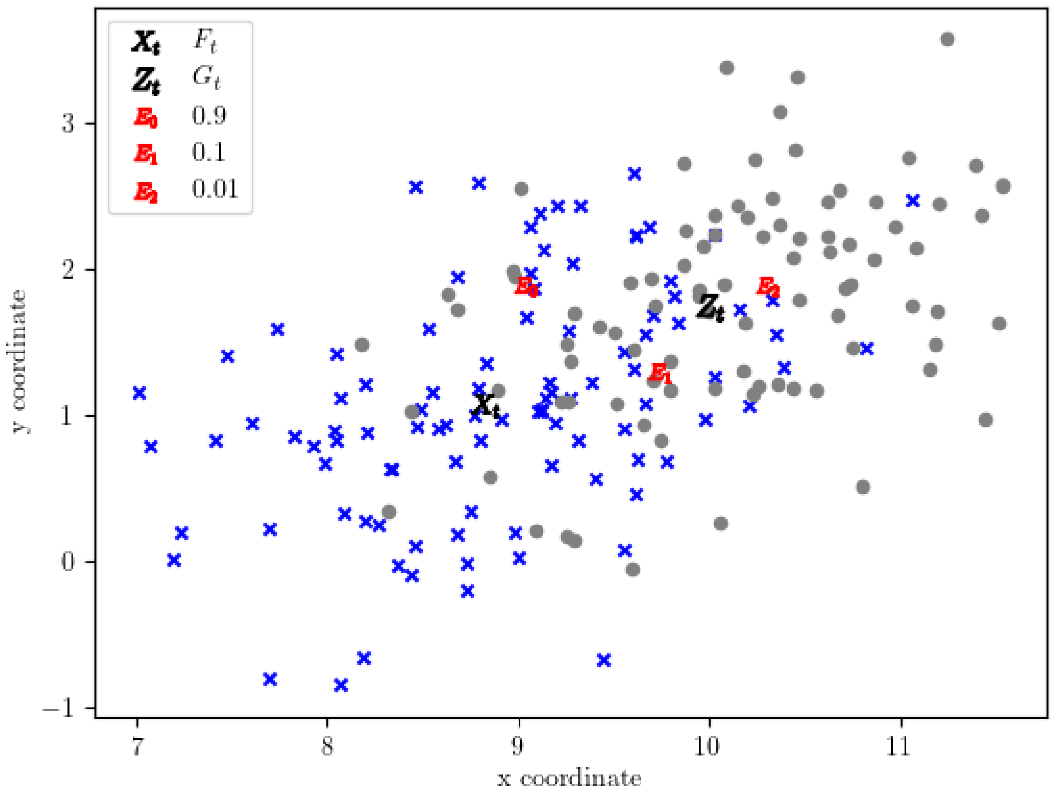


Figure 4.9: X_T and Z_T in black are the expected values of functions F_t and G_t respectively. The values E_i in red mark are the expected values for δ_i as in Table 4.4.

4.3. Returns funds A and E

We are interested in the estimation of the cdf between the real returns of pension funds A and E. Our ambiguity problem is stated with respect to a random variable that is uniformly distributed. This r.a models lack of information or skepticism towards data or correlation between these two funds.

The different types of pension funds are differentiated by the proportion of their resources invested in variable income financial securities, which are characterized by higher risk and higher expected return. Fund A has a higher proportion of its investments in equities, which decreases progressively up to Fund E, the latter being the one with the least amount.

The Pension Superintendence has data available of the real returns of each fund. The data was downloaded August 1st 2022.

From these data, we construct F_0, the empirical cdf of the monthly real returns of funds A and E deflected by UF (unidad de fomento).

	Fondo Tipo A	Fondo Tipo E
count	142	142
mean	-0.6426	0.1698
std	1.7202	1.3241
min	-4.4900	-4.3500
25%	-1.5250	-0.3175
50%	-0.3650	0.39500
75%	0.7000	0.7850
max	1.9300	4.4200

Table 4.5: data

Let G be a uniform distribution over $[-4.49, -4.35] \times [1.93, 4.42]$, then

$$\mathbb{E}_{G_0} = (-1.2800, 0.0349)$$

δ	s	$\mathbb{E}_x$	$\mathbb{E}_y$
0.0001	0.1406	-0.5408	0.5298
0.1	0.0847	-0.8781	-0.2476

Table 4.6: Expected values of F with $M = N = 300$

Chapter 5

Conclusions and further research

In this work, we provide a new method for the problem of estimating a cumulative distribution function. We set the problem in the space of upper semi continuous functions endowed with the Attouch Wets distance where the space of cumulative distribution functions is contained. We showed the topological advantages of setting the problem in the way it was done rather than pose it in the set of cumulative distribution functions endowed with the same metric. We presented an approximation for the hypo-distance that is easier to calculate than it and is defined by a linear finite dimensional problem which allowed the design of an algorithm that compute an approximate distance by means of state of art solvers. The importance of this approach is that opens the possibility to include soft information or to imposed regularity to solutions.

We provide an implementation and numerical results for the algorithm for the bivariate case from which we analysed through two different settings, the solutions for different sizes of the ambiguity set. The analysis contemplated comparison for the expected value, visual representation of the solutions, quantification of distribution constraint error and the effect of the bounded growth condition in highly non-soft solutions. Additionally, we proposed solution to the problem of estimating of the position of an unmanned underwater vehicle.

As far as we know, these are the first numerical results for the an approximated hypo-distance. Furthermore we provided theoretical guarantees for solutions to the approximated sequence of problems to belong to the set of minimizers of the original problem. We gave conditions for the latter to occur and we proved that the set of Lipschitz functions satisfy it.

Future work has different edges, the first potential line of research is on extending the theoretical guarantees to a more general class of functions as it is stated in Remark 3.13. The second line consists on adding a regularization function to obtain softer solutions and furthermore, the last potential line is the implementation of the multivariate case, since theoretical results are still valid in higher dimensions and we have focused on the bivariate case just to simplify the implementation.

The above will enlarge the variety of problems that can be solved through this approach and by this, we will be able to introduce new challenges, for instance, in the field of distributionally robust optimization. In particular, we are interested in proposing our approach to the robust Markowitz problem.

Bibliography

[1] Amparo Baíllo, Juan A Cuesta-Albertos, and Antonio Cuevas. Convergence rates in nonparametric estimation of level sets. *Statistics & probability letters*, 53(1):27–35, 2001.

[2] Amparo Baíllo, Antonio Cuevas, and Ana Justel. Set estimation and nonparametric detection. *Canadian Journal of Statistics*, 28(4):765–782, 2000.

[3] P. Billingsley. *Probability and Measure.* Wiley Ser. Probab. Math. Statist., 1995.

[4] Victor-Emmanuel Brunel. Concentration of the empirical level sets of tukey's halfspace depth. *Probability Theory and Related Fields*, 173(3):1165–1196, 2019.

[5] Michael L. Bynum, Gabriel A. Hackebeil, William E. Hart, Carl D. Laird, Bethany L. Nicholson, John D. Siirola, Jean-Paul Watson, and David L. Woodruff. *Pyomo–optimization modeling in python*, volume 67. Springer Science & Business Media, third edition, 2021.

[6] Yuhang Cai and Lek-Heng Lim. Distances between probability distributions of different dimensions. *IEEE Transactions on Information Theory*, 68(6):4020–4031, 2022.

[7] IBM ILOG Cplex. V12. 1: User's manual for cplex. *International Business Machines Corporation*, 46(53):157, 2009.

[8] Antonio Cuevas. Set estimation: Another bridge between statistics and geometry. *Bol. Estad. Investig. Oper*, 25(2):71–85, 2009.

[9] Antonio Cuevas, Wenceslao González-Manteiga, and Alberto Rodríguez-Casal. Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48(1):7–19, 2006.

[10] Richard M Dudley. Distances of probability measures and random variables. In *Selected Works of RM Dudley*, pages 28–37. Springer, 2010.

[11] Richard M Dudley. *Real analysis and probability.* CRC Press, 2018.

[12] Emre Erdoğan and Garud Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1):37–61, 2006.

[13] William E Hart, Jean-Paul Watson, and David L Woodruff. Pyomo: modeling and solving mathematical programs in python. *Mathematical Programming Computation*, 3(3):219–260, 2011.

[14] Daniel Kuhn, Peyman Mohajerin Esfahani, Viet Anh Nguyen, and Soroosh Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine

learning. In *Operations research & management science in the age of analytics*, pages 130–166. Informs, 2019.

[15] Viktor Kuzmenko and Stan Uryasev. Kantorovich Rubinstein distance minimization: Application to location problems. In *Large Scale Optimization in Supply Chains and Smart Manufacturing*, pages 59–68. Springer, 2019.

[16] Victor M. Panaretos and Yoav Zemel. Statistical aspects of Wasserstein distances. *Annual Review of Statistics and Its Application*, 6(1):405–431, 2019.

[17] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[18] Johannes O Royset. Approximations and solution estimates in optimization. *Mathematical Programming*, 170(2):479–506, 2018.

[19] Johannes O Royset. Approximations of semicontinuous functions with applications to stochastic optimization and statistical estimation. *Mathematical Programming*, 184(1):289–318, 2020.

[20] Johannes O Royset and Roger J-B Wets. Variational theory for optimization under stochastic ambiguity. *SIAM Journal on Optimization*, 27(2):1118–1149, 2017.

[21] Johannes O Royset and Roger JB Wets. Multivariate epi-splines and evolving function identification problems. *Set-Valued and Variational Analysis*, 24(4):517–545, 2016.

[22] Johannes O Royset and Roger JB Wets. Variational analysis of constrained m-estimators. *The Annals of Statistics*, 48(5):2759–2790, 2020.

[23] Johannes O Royset and Roger JB Wets. An optimization primer, 2021.