



**HAL**  
open science

# Descent dynamical systems and algorithms for tame optimization, and multi-objective problems

Guillaume Garrigos

► **To cite this version:**

Guillaume Garrigos. Descent dynamical systems and algorithms for tame optimization, and multi-objective problems. Modeling and Simulation. Université Montpellier; Universidad técnica Federico Santa María (Valparaiso, Chili), 2015. English. NNT: 2015MONTTS191 . tel-02023313

**HAL Id: tel-02023313**

**<https://theses.hal.science/tel-02023313>**

Submitted on 18 Feb 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

Pour obtenir le grade de **Docteur**

Délivrée par l'Université de Montpellier

En co-direction avec l'Universidad Técnica Federico Santa María

Préparée au sein de l'école doctorale **Information, Structure, Système**  
Et de l'unité de recherche **Institut Montpelliérain Alexander Grothendieck**

Spécialité: **Mathématiques**

Présentée par **Guillaume GARRIGOS**

**Descent dynamical systems and algorithms  
for tame optimization and multi-objective problems.**

Soutenue le 2 Novembre 2015 devant le jury composé de

M. HédY ATTOUCH	Prof.	Université de Montpellier	Directeur
M. Juan PEYPOUQUET	Prof.	Universidad Técnica Federico Santa María	Directeur
M. Aris DANIILIDIS	Prof.	Universidad de Chile	Rapporteur
M. Adrian D. LEWIS	Prof.	Cornell University	Rapporteur
M. Jalal FADILI	Prof.	École Nationale Supérieure d'Ingénieurs de Caen	Examineur
M. Pedro GAJARDO	Prof.	Universidad Técnica Federico Santa María	Examineur
M. Lionel THIBAUT	Prof.	Université de Montpellier	Examineur





## **Abstract.**

In a first part, we focus on gradient dynamical systems governed by nonsmooth but also nonconvex functions, satisfying the so-called Kurdyka-Łojasiewicz inequality. After obtaining preliminary results for a continuous steepest descent dynamic, we study a general descent algorithm. We prove, under a compactness assumption, that any sequence generated by this general scheme converges to a critical point of the function to be minimized. We also obtain new convergence rates both for the values and the iterates. The analysis covers alternating versions of the forward-backward method, with variable metric and relative errors. As an example, a nonsmooth and nonconvex version of the Levenberg-Marquardt algorithm is detailed. Applications to nonconvex feasibility problems, and to sparse inverse problems are discussed.

In a second part, the thesis explores descent dynamics associated to constrained vector optimization problems. For this, we adapt the classic steepest descent dynamic to functions with values in a vector space ordered by a closed convex cone with nonempty interior. It can be seen as the continuous analogue of various descent algorithms developed in the last years. We have a particular interest for multi-objective decision problems, for which the dynamic make decrease all the objective functions along time. We prove the existence of trajectories for this continuous dynamic, and show their convergence to weak efficient points. Then, we explore an inertial dynamic for multi-objective problems, with the aim to provide fast methods converging to Pareto points.



## Résumé.

Dans une première partie, nous nous intéressons aux systèmes dynamiques gradients gouvernés par des fonctions non lisses, mais aussi non convexes, satisfaisant l'inégalité de Kurdyka-Lojasiewicz. Après avoir obtenu quelques résultats préliminaires pour la dynamique de la plus grande pente continue, nous étudions un algorithme de descente général. Nous prouvons, sous une hypothèse de compacité, que tout suite générée par ce schéma général converge vers un point critique de la fonction à minimiser. Nous obtenons aussi de nouveaux résultats sur la vitesse de convergence, tant pour les valeurs que pour les itérés. Ce schéma général couvre en particulier des versions parallélisées de la méthode *forward-backward*, autorisant une métrique variable et des erreurs relatives. Cela nous permet par exemple de proposer une version non convexe non lisse de l'algorithme Levenberg-Marquardt. Enfin, nous proposons quelques applications de ces algorithmes aux problèmes de faisabilité, et aux problèmes inverses parcimonieux.

Dans une seconde partie, cette thèse développe une dynamique de descente associée à des problèmes d'optimisation vectoriels sous contrainte. Pour cela, nous adaptons la dynamique de la plus grande pente usuelle aux fonctions à valeurs dans un espace ordonné par un cône convexe fermé d'intérieur non vide. Cette dynamique peut être vue comme l'analogie continue de nombreux algorithmes développés ces dernières années. Nous avons un intérêt particulier pour les problèmes de décision multi-objectifs, pour lesquels cette dynamique de descente fait décroître toutes les fonctions objectif au cours du temps. Nous prouvons l'existence de trajectoires pour cette dynamique continue, ainsi que leur convergence vers des points faiblement efficaces. Finalement, nous explorons une nouvelle dynamique inertielle pour les problèmes multi-objectif, avec l'ambition de développer des méthodes rapides convergeant vers des équilibres de Pareto.

## Resumen.

En una primera parte, nos interesamos en sistemas dinámicos de tipo gradiente gobernados por funciones no diferenciables, también no convexas, que satisfacen la desigualdad de Kurdyka-Lojasiewicz. Después de obtener resultados preliminares para la dinámica continua del máximo descenso, estudiamos un algoritmo de descenso general. Demostramos, bajo una hipótesis de compacidad, que cualquier sucesión generada según este esquema general converge a un punto crítico de la función. Además, obtenemos nuevas estimaciones sobre la velocidad de convergencia para la secuencia y sus valores. Este análisis cubre versiones alternadas del método *forward-backward*, con métrica variable y errores aditivos. Nos permite por ejemplo de proponer una versión no diferenciable y no convexa del algoritmo de Levenberg-Marquardt. Por fin, proponemos algunas aplicaciones de esos algoritmos para resolver problemas de viabilidad no convexos, y problemas inversos.

En una segunda parte, la tesis explora una dinámica de descenso asociada con problemas de optimización vectoriales con restricción. Para esto, adaptamos la dinámica del máximo descenso clásica para las funciones con valores en un espacio vectorial ordenado por un cono sólido cerrado convexo. Puede ser visto como el análogo continuo de varios algoritmos de descenso desarrollados en los años pasados. Tenemos un interés particular por los problemas de decisión multiobjetivo, por lo cuales cada función objetivo decrece con el tiempo. Demostramos la existencia de trayectorias para esta dinámica continua, y mostramos su convergencia a puntos débilmente eficientes. Después, investigamos una nueva dinámica inercial para problemas multiobjetivo, con el propósito de proponer métodos rápidos que convergen a equilibrios de Pareto.



# Acknowledgements

Avant tout, je souhaiterais exprimer ma reconnaissance et toute mon amitié à mes directeurs de thèse, Hédý et Juan. Je n'aurais pu espérer trouver de personnes plus talentueuses et sympathiques pour m'encadrer. Comme un long discours ne saurait exprimer tout ce que je leur dois, je me contenterai donc ici de les remercier chaleureusement pour leur soutien, leur patience(!) et plus généralement pour tout ce qu'ils m'ont transmis et appris ces dernières années.

Mes remerciements vont aussi à Aris Daniilidis et Adrian Lewis, qui m'ont fait l'honneur de rapporter ce manuscrit, ainsi qu'à Lionel Thibault, Jalal Fadili et Pedro Gajardo qui ont accepté de faire partie de mon jury de thèse, ces deux derniers ayant fait un long (voire très long) chemin pour venir.

Durant ces trois années passées à l'ISM, j'ai pu profiter de la présence de moult personnes qui ont fait de cette thèse une succession de pauses cafés fort agréables. Merci Micka d'avoir fait quelques arrêts dans notre bureau pendant tes parties de cache-cache, et de m'avoir laissé te faire perdre ton temps quand j'en avais besoin. Un jour la valeur des services que tu as rendus à la communauté seront reconnus à leur juste valeur, et une statue à ton effigie trônera fièrement quelque part. Mes salutations vont aussi aux visiteurs réguliers du bureau 115: Christophe (qui nous a quittés bien trop tôt), Benjamin (merci pour ces discussions totalement vaines), Myriam (puisse ta bonne humeur bénéficier aux autres encore longtemps), mais aussi Gauthier, Joub, Julien, Etienne, Theo, Elsa, Antoine, et tous ceux dont les noms n'apparaissent pas ici mais qui font (ou ont fait) de ce labo un endroit très sympathique.

Ese tesis fue también la ocasión para mí de viajar a Chile y descubrir allá gente formidable. Primero entre todos, Cecio, hermano institucional, pero más que todo, de corazón. Grazie para tu buena onda comunicativa, perpetua e indefectible, y todas esas noches tontas que tuvimos. Un agradezco también a todos los Sansanitos, gracia a quien me sentí como en casa en Valpo: Pato (le plus génial de tous les canards), Andrei, Felipe, Laurent, Mathieu. Gracias a David y su hospitalidad, que me hizo descubrir un poco su capital. Doy también a un grande abrazo a Willy, Meike, Miriam y Martina, compañeros de Casa Viejas, así que a la gente de Casa Chanco y calle Perez.

Évidemment, je ne saurais oublier les copains, qui sont toujours là quand le besoin de passer un bon moment se fait sentir<sup>1</sup>. En premier lieu Laurent, qui m'a amicalement recueilli sur son canapé, et avec qui nous avons vécu des aventures toujours plus chocolatées. Puis bien entendu Brice, Marion, Julie, Kévin, Jérémy, Laurène, Renaud, et tous les autres Vinassanais de cœur. J'ai aussi une pensée chaleureuse pour ma famille, qui me supporte (dans tous les sens du terme) depuis toujours!

Enfin, mes dernières pensées (et non les moindres) vont à toi, Julie, qui est ma première supportrice (et relectrice!), d'un soutien sans réserve, et sans qui je n'aurais peut-être pas trouvé le courage de commencer/terminer cette thèse. Merci pour tout!

---

<sup>1</sup>Rectification: vous êtes TOUJOURS fourrés à la maison. De vrais parasites!





# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Variational analysis tools in Banach spaces</b>	<b>13</b>
2.1	Basic topology and differential calculus . . . . .	14
2.1.1	Convex sets . . . . .	14
2.1.2	Cones and duality . . . . .	15
2.1.3	Differential and directional derivative . . . . .	17
2.2	Nonsmooth analysis for extended-real-valued functions . . . . .	19
2.2.1	Fenchel analysis . . . . .	19
2.2.2	Bouligand analysis . . . . .	22
2.2.3	Fréchet analysis . . . . .	23
2.2.4	Clarke analysis . . . . .	25
2.2.5	A last few things . . . . .	26
2.3	Optimization for vector-valued functions . . . . .	27
2.3.1	Ordered space . . . . .	27
2.3.2	Base of a cone . . . . .	28
2.3.3	Vector optimization problem . . . . .	31
<b>I</b>	<b>Dynamics for Tame optimization</b>	<b>35</b>
<b>3</b>	<b>The Kurdyka-Łojasiewicz inequality and general descent methods</b>	<b>39</b>
3.1	The Kurdyka-Łojasiewicz inequality . . . . .	40
3.1.1	Introduction . . . . .	40
3.1.2	The Kurdyka-Łojasiewicz inequality . . . . .	42
3.1.3	Asymptotic behaviour of a subgradient differential inclusion . . . . .	43
3.2	Convergence of an abstract inexact descent method . . . . .	47
3.3	Rates of convergence for an abstract descent method . . . . .	52
3.3.1	Distance to the limit in terms of the gap . . . . .	53
3.3.2	Explicit rates when $\varphi(t) \propto t^\theta$ with $\theta \in ]0, 1]$ . . . . .	53
3.3.3	Sharper results for gradient-related methods . . . . .	55
3.4	Comments and perspectives . . . . .	57
<b>4</b>	<b>Splitting methods for KL functions</b>	<b>61</b>
4.1	Splitting methods with errors and variable metric . . . . .	63
4.1.1	The Alternating Forward-Backward (AFB) Method . . . . .	63
4.1.2	The AFB Method with errors . . . . .	64
4.1.3	Variable metric: Towards generalized newton methods . . . . .	68
4.2	Some problems arising in the KL context . . . . .	69
4.2.1	Nonconvex Feasibility . . . . .	69
4.2.1.1	Mixed feasibility problem . . . . .	72

4.2.1.2	The KL property for the penalization function . . . . .	76
4.2.2	Inverse problems and sparsity . . . . .	79
4.2.2.1	Compressed sensing . . . . .	79
4.2.2.2	Sparse and Low-rank Matrix Decomposition . . . . .	80
4.2.2.3	Numerical results on data compression . . . . .	81
4.3	Comments and perspectives . . . . .	84
<b>II</b>	<b>Dynamics for Vector optimization</b>	<b>89</b>
<b>5</b>	<b>An ordered subdifferential for vector-valued functions</b>	<b>93</b>
5.1	An ordered Clarke subdifferential for vector-valued functions . . . . .	98
5.1.1	Definition . . . . .	98
5.1.2	Properties of the ordered Clarke subdifferential . . . . .	100
5.2	Descent directions . . . . .	106
5.2.1	Descent direction and Fermat's rule . . . . .	106
5.2.2	Normals to level sets . . . . .	111
5.2.3	The steepest descent direction in Hilbert spaces . . . . .	112
5.3	Comments . . . . .	116
<b>6</b>	<b>The continuous steepest descent dynamic for vector optimization</b>	<b>121</b>
6.1	The steepest descent dynamic . . . . .	122
6.1.1	Definitions . . . . .	122
6.1.2	Qualitative properties of the trajectories . . . . .	123
6.1.3	Some basic examples . . . . .	125
6.1.4	Related dynamics . . . . .	130
6.1.5	About the uniqueness of the trajectories . . . . .	132
6.2	Asymptotic properties of the steepest descent dynamic . . . . .	134
6.3	The discrete steepest descent method, and numerical results . . . . .	138
6.3.1	Review of the existing algorithms, and links with (SD) . . . . .	138
6.3.2	The effective computation of the steepest descent direction . . . . .	142
6.3.3	Generation of Pareto fronts . . . . .	143
6.4	Comments and perspectives . . . . .	147
<b>7</b>	<b>The continuous steepest descent: existence of trajectories</b>	<b>153</b>
7.1	Existence in the smooth vector case . . . . .	154
7.2	Existence in the nonsmooth convex multiobjective case . . . . .	157
7.3	Existence in the nonsmooth convex vector case . . . . .	162
7.3.1	An equivalent description of $\partial F$ . . . . .	163
7.3.2	A Moreau-Yosida approximation for convex vector-valued functions . . . . .	165
7.3.3	Proof of the existence . . . . .	173
7.4	Comments . . . . .	179
<b>8</b>	<b>A continuous inertial approach to multi-objective optimization</b>	<b>181</b>
8.1	Existence of trajectories . . . . .	182
8.2	Qualitative study of the dynamic . . . . .	185
8.2.1	A dissipative system . . . . .	185
8.2.2	Energy estimations . . . . .	187
8.2.3	Convergence of the trajectories . . . . .	189
8.3	Examples and numerical results . . . . .	191
8.4	Comments and perspectives . . . . .	193

<b>Appendices</b>	<b>195</b>
<b>A Variational analysis tools in Banach spaces</b>	<b>197</b>
A.1 Basic topology and differential calculus . . . . .	197
A.2 Nonsmooth analysis for extended-real-valued functions . . . . .	201
A.3 Optimization for vector-valued functions . . . . .	204
<b>B Measurability and <math>L^p</math> spaces</b>	<b>209</b>
B.1 Measurability . . . . .	209
B.2 Strong measurability and $L^p$ spaces . . . . .	209
B.3 Weak* measurability and $L_{w^*}^\infty$ space . . . . .	213
<b>C Others</b>	<b>215</b>
<b>Bibliography</b>	<b>217</b>



# Chapter 1

## Introduction

The idea of using a descent method to solve an optimization problem goes back, at least, to Cauchy and its gradient method [96]. This algorithm has been the object of an intense research along the 20th century, considering for instance a line-search method for the choice of the stepsize [123, 178, 16, 324]. The development of convex analysis in the 60-70s opened the door to a nonsmooth version of the implicit gradient method, namely, the *proximal* algorithm, introduced by Martinet [236], and further developed in [287, 86]. It appeared clear at that time that the proximal algorithm is strongly connected to its continuous-time counterpart, the *steepest descent dynamic* [291, 170, 82, 83, 119]:

$$(1.1) \quad \dot{u}(t) + \partial f(u(t)) \ni 0,$$

where  $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower semi-continuous convex function defined on a Hilbert space  $H$ . Clearly, the explicit and implicit versions of the gradient method can both be derived from (1.1) after a time discretization. But the inter-connexions between the continuous and discrete dynamics go beyond this simple observation. For instance, the trajectories of (1.1) can be built from sequences generated by the proximal algorithm [119]. Furthermore, the equivalence between the asymptotic behaviour of the proximal algorithm and the trajectories of (1.1) has been recently established [13, 14, 269]. In practical, the connections between continuous and discrete dynamics are fruitful for the optimizer. For example, finding a Lyapunov function for a continuous dynamic is in general easier than for the discrete case. On the reverse, the discrete dynamics benefit from the fact that their analysis do not require any derivability assumption on the trajectory.

In this thesis, we develop descent dynamics, being discrete or continuous, for two optimization problems which gained in interest these last years. Firstly, we deal with a tame optimization problem, which consists in the minimization of a nonsmooth nonconvex function being *well-behaved*, in some way. The growing interest for tame problems comes from the fact that various problems in image and signal processing are tame, by nature. Secondly, we consider a multi-objective (also multi-criteria) optimization problem. Such problem involves a finite family of cost functions, which have to be minimized simultaneously, as far as possible. This kind of problem arises naturally in engineering (shape optimization), or in decision sciences. For these two problems, we adopt the same approach: building a descent dynamic, and proving its convergence to a solution. As much as we can, we shall underline the connexions between the continuous and discrete dynamics studied. We introduce now in detail the two parts of this manuscript, corresponding to the two aforementioned problems.

## Part I: Dynamics for tame optimization

In classic analysis, it is known for a long time that there exists generically wild functions, being for instance continuous but nowhere differentiable [45, 237]. In the more recent nonsmooth analysis setting, such pathologic functions are also known to be pre-eminent. Consider for instance, among the locally Lipschitz functions, the ones having a Clarke subdifferential constant to the unit ball [78]. On this subject, Grothendieck [181] condemns “the wildness carried [by the general topology] as an inevitable fate” and calls for a new “tame topology”, more adapted to “the context in which we live, breath, work”. Indeed, most of the functions/sets encountered in optimization problems (at least in finite dimensions) are not so much monstrous: least squares,  $\ell^p$  norms, functions being polynomial by parts, etc. In the literature, various notions can be found that ensure the good behaviour of a function, in a setting broader than the smooth or convex ones. Let us mention for instance the classes of semi-smooth [242], lower- $C^2$  [289, 193, 262], primal-lower-nice [273], prox-regular [274], or partly smooth functions [221], whose interest in optimization is not to be proved [278, 220, 306, 253, 226].

*Tame optimisation*, as mentioned in the title of this thesis, is a general term, referring to optimization of functions being pathologic-free. We intentionally do not give a precise meaning to the word ‘tame’, and prefer for now letting the reader keep in mind the picture of well-behaved objects<sup>1</sup>.

According to van den Dries [141], the o-minimal structures provide the first known framework in which Grothendieck’s tame topology could be developed. An o-minimal structure is a collection of sets, stable under usual operations, and a function is said to be definable in this structure, whenever its graph –or its epigraph– is a set lying in the structure. These structures have been modelled to generalize the properties of two known collections of sets: the semi-algebraic [318, 66, 140] and globally subanalytic sets [230, 166, 196]. While this theory was originally designed as a model theory in the 80s [271], it appeared soon that these o-minimal structures offer a good setting in which doing analysis. Indeed, the objects definable in these structures enjoy a lot of finitude properties. For instance, the sets definable in these structures admit a finite number of connected components. This implies for instance that the level sets of definable functions have a finite number of connected components, which excludes pathological functions like  $\frac{1}{x}\sin(x)$ . The definable sets also admit a Whitney stratification, which means that we can always decompose them into a finite number of smooth manifolds fitting together in a regular way. These examples are far to be exhaustive, and can be found for instance in [144].

Among all the properties satisfied by the functions definable in an o-minimal structure, there is one which will keep our attention: the so-called Kurdyka-Łojasiewicz property. It is a property describing the behaviour of function in the neighbourhood of its critical points. Its main feature is that it ensures the finite length for the trajectories of descent dynamics (being continuous or discrete). Thus, this Kurdyka-Łojasiewicz property is an excellent tool in optimization, in order to validate the convergence of numerical methods. Let us give some details on this property.

In 1984, Łojasiewicz [231] is interested by the asymptotic behaviour of the trajectories being solution of the steepest descent dynamic

$$(1.2) \quad \dot{u}(t) + \nabla f(u(t)) = 0,$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is an analytic function. Analyticity is not a too strong hypothesis, since it is known at this time that the gradient flow associated to a  $C^\infty$  function can not converge [257]. Łojasiewicz proves that any bounded trajectory of (1.2) has finite length, and converges

---

<sup>1</sup>We are aware of the fact that a notion of tame function is developed in [203, 204, 69], but we will not use it here.

to a critical point of  $f$ . The main tool in its proof is the following: for analytic functions, we have a particular estimate, previously proved by Łojasiewicz itself [230], linking the variations of the values  $f(u(t))$  and of the norm of  $\nabla f(u(t))$ . More exactly, for any critical point  $\bar{x}$  of  $f$ , there exists  $C > 0$  and  $\theta \in ]0, \frac{1}{2}]$ , such that

$$(1.3) \quad \forall x \sim \bar{x}, x \neq \bar{x}, \quad |f(x) - f(\bar{x})|^{1-\theta} \leq C \|\nabla f(x)\|.$$

The equation above is nowadays known as the *Łojasiewicz inequality*. If we assume that  $f(x) > f(\bar{x}) = 0$ , then the Łojasiewicz inequality can be rewritten as

$$(1.4) \quad \forall x \sim \bar{x}, x \neq \bar{x}, \quad 1 \leq \|\nabla(\varphi \circ f)(x)\|, \quad \text{with } \varphi(t) = \frac{C}{\theta} t^\theta.$$

In other words, up to a polynomial reparametrization,  $f$  is sharp around  $\bar{x}$ .

In parallel, Simon [296] studies the asymptotic behaviour of general nonlinear evolution equations. For example, they can take the form

$$\begin{cases} \dot{\mathbf{u}}(t, x) - \Delta \mathbf{u}(t, x) = \phi(\mathbf{u}(t, x)) & \text{on } \mathbb{R}_+ \times \Omega, \\ \mathbf{u}(t, x) = 0 & \text{on } \mathbb{R}_+ \times \partial\Omega, \end{cases}$$

where  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is a source term. Assuming that the source term  $\phi$  is analytic, and using the Łojasiewicz inequality (1.3), Simon proves under some conditions that the trajectories of its dynamic converge strongly to an equilibrium. This approach will be reused later for the asymptotic analysis of some PDE's involving an analytic feature. Let us mention, among others, some studies of the semilinear wave equation [187], degenerate parabolic equations [156], the Cahn-Hilliard equation [197], equations for phase transition [155] or quasi-linear parabolic equations [104]. See also [47] for an optimal control problem governed by a Schrödinger equation, and [10] for an inertial gradient-like system involving a Hessian-driven damping.

In 1998, Kurdyka [216] considers the steepest descent dynamic (1.2), and adapts Łojasiewicz's convergence result to any  $C^1$  function definable in an o-minimal structure. For this, he needs an appropriate extension of Łojasiewicz's inequality. Indeed, the function  $x \mapsto e^{-\frac{1}{x^2}}$ , which lies in some o-minimal structure, cannot satisfy the Łojasiewicz inequality, because it is exponentially flat around its minimum. That is why Kurdyka proves in a first time that, for any differentiable function definable in an o-minimal structure, the inequality (1.4) holds around critical points, for a general definable function  $\varphi$  (not necessarily a polynomial). This inequality, which can be rewritten as

$$\varphi'(f(x) - f(\bar{x})) \|\nabla f(x)\| \geq 1,$$

is now called the *Kurdyka-Łojasiewicz inequality*. More recently, Bolte, Daniilidis and Lewis [68] adapt the Łojasiewicz inequality to *nonsmooth* subanalytic functions, and derive a finite length property for the bounded trajectories of

$$(1.5) \quad \dot{u}(t) + \partial^F f(u(t)) \ni 0.$$

Finally, in 2007, a strong result of Bolte, Daniilidis, Lewis and Shiota [70] asserts that any lower semi-continuous function definable in an o-minimal structure satisfies a nonsmooth version of the Kurdyka-Łojasiewicz inequality.

The convergence analysis of descent algorithms for functions satisfying the Kurdyka-Łojasiewicz inequality is more recent. See [1] for gradient-related methods, and [253] for a nonsmooth subgradient-oriented descent method. The proximal method is investigated in [18, 71], as well as in [240, 174, 4] with a specific attention to applications for PDE's discretization. The celebrated Forward-Backward algorithm, a splitting method exploiting the nonsmooth/smooth structure of the objective function, has been studied in [20], and extended in [106] to take in



account a variable metric. Another splitting approach comes from Gauss-Seidel-like methods, which apply to functions with separated variables, and consist in doing a descent method relatively to each (block of) variable alternatively. See [19, 321] for a proximal alternating method, and [20] for a variable-metric version. Recent papers [321, 72, 107] propose to combine these two splitting approaches in order to exploit both the smooth/nonsmooth character and the separated structure of the function.

Let us now precise our contributions to this domain, which are presented in Part I. We consider a proper lower semi-continuous function  $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ , where  $H$  is a Hilbert space. Working in a Hilbert space setting allows to encompass the analytic functions, the functions definable in an o-minimal structure, or even some energies underlying some nonlinear PDE. Along all this part, we assume that  $f$  is a function satisfying the Kurdyka-Łojasiewicz inequality around each point  $\bar{x} \in H$ . Such function will be called a *KL function*.

Chapter 3 contains the main *theoretical* results of Part I. We start in Section 3.1 by giving some keys for the understanding of the Kurdyka-Łojasiewicz inequality, and its implications in descent dynamical systems. After giving a proper definition of the Kurdyka-Łojasiewicz inequality in the nonsmooth case, we revisit the result of [68] on the asymptotic behaviour of (1.5). Under general hypotheses on  $f$ , and a compactness assumption on  $u$ , we recover the finite length for the trajectories of this dynamic. The proof of this result for the continuous dynamic is instructive, since it provides a sketch for the proof in the discrete case. We also prove in Theorem 3.1.12 general convergence rates for both the trajectory and its image, in the line of [104].

In Section 3.2, we are interested in descent algorithms to solve

$$\min_{x \in H} f(x).$$

Most of the algorithms studied in the papers mentioned above share the same asymptotic behaviour: under a compactness assumption, the generated sequences converge strongly to critical points, and the affine interpolations have finite length. This is not surprising since the algorithms described in [18, 240, 71, 20, 19, 72], together with the ones of [106, 321] (without extrapolation step), fall into the general convergence result for abstract descent methods of Attouch, Bolte and Svaiter [20]. Besides, these methods essentially share the same hypotheses on the parameters with the abstract method of [20]: the step sizes (respectively the eigenvalues of the matrices underlying the metric) are required to remain in a compact subinterval of the positive numbers. Moreover, they have little flexibility regarding to the presence of computational errors. To our knowledge, vanishing step sizes (resp. unbounded eigenvalues) or sufficiently general errors have never been treated in the KL context. For these reasons, we introduce an abstract descent scheme, generating sequences  $(x_k)_{k \in \mathbb{N}} \subset H$  verifying for all  $k \in \mathbb{N}$ :

$$a_k \|x_{k+1} - x_k\|^2 \leq f(x_k) - f(x_{k+1}), \quad a_k > 0,$$

$$b_{k+1} \|\partial^F f(x_{k+1})\|_- \leq \|x_{k+1} - x_k\| + \varepsilon_{k+1}, \quad b_{k+1}, \varepsilon_{k+1} > 0.$$

This approach follows the ideas of [20], that we adapt to allow more flexibility for the parameters, and to introduce an error term  $\varepsilon_{k+1}$ . Theorem 3.2.2 proves, under a compactness assumption, that any sequence generated by this scheme has finite length and converges strongly to a critical point of  $f$ . Then, provided that the sequence  $(x_k)_{k \in \mathbb{N}}$  is initialized close enough to a minimum of  $f$ , Theorem 3.2.3 ensures that the sequence converges to a minimum of the function. Such a capture result is of particular importance, since we work with functions being non necessarily convex.

In section 3.3, we focus on the convergence rates of this abstract method. An interesting fact is that, in the literature, the convergence rates of several methods are essentially the same, and

depend on the KL inequality rather than the nature of the algorithm. We give a theoretical basis for this statement by proving general convergence rates in Theorem 3.3.2, under the assumption that  $f$  satisfies the Kurdyka-Łojasiewicz inequality with a polynomial function  $\varphi$ . Furthermore, in Theorem 3.3.4, we prove new convergence rates when a more general function  $\varphi$  is involved. This result can be seen as the discrete counterpart of the convergence rates obtained for the continuous steepest descent in Theorem 3.1.12.

Chapter 4 contains the most *practical* aspects of Part I. We introduce in Section 4.1 a particular instance of the model defined in Chapter 3, which provides further insight into a large class of known methods, and present some innovative variants. More exactly, we revisit the Alternating Forward-Backward algorithm, already considered in [72, 107, 224], but allowing inexact computation of the iterates, and a dynamic choice of metric. After a description of the method in Section 4.1.1, we prove its convergence to critical points in Section 4.1.2. As a by-product of the results of Chapter 4, we obtain in Theorems 4.1.6 and 4.1.8 a global and local convergence result, respectively. In Section 4.1.3, we exploit the flexibility given by the Alternating Forward-Backward method to propose a new projected-Newton-like algorithm, for solving constrained optimization problems.

Section 4.2 is devoted to two particular applications of the Alternating Forward-Backward method. In Section 4.2.1 a nonconvex feasibility problem, involving both ‘soft’ constraints (like convex, smooth, or more generally prox-regular) and ‘hard’ constraints (no particular assumption is made). In Section 4.2.2, we discuss some problems arising in signal processing, such that the compressed sensing and the low-rank and sparse matrix decomposition problem. In each case, we use the flexibility of the Alternating Forward-Backward method to design algorithm for solving these problems. Some of the proposed methods are new, like the hard shrinkage projection algorithm that we numerically illustrate in Section 4.2.2.3.

## Part II: Dynamics for vector optimization

The problems for which a decider have to manage simultaneously different costs are common in optimization: we want to maximize the correlation with respect to different data sets, together with the minimization of some energy, and being close to a desired state, etc. This kind of problem, called multi-objective (or multi-criteria) problem, arises a lot in shape optimization, where we have to design a shape submitted to different constraints [211, 227, 189]. But it also appears in various other domains, such that engineering [264], optimal control [229, 168], game theory [238], medical treatment [210], finance [320, 302], management, and more generally in decision sciences and operational research [41, 208]. Historically, it is in economy that multi-criteria equilibrium has been studied for the first time, in the context of the utility theory and welfare economics, introduced among others by Walras [314] and Edgeworth [152]. Pareto [258], and then Barone [43], gave the first definition of what we call nowadays a *Pareto equilibrium*, a state for which:

*“It must be impossible by any allocation of resources to enhance the welfare of one household without reducing that of another.”* E. Barone [43]

The reader looking for more details on the historical development of these notions in mathematics and economy can consult the complete survey of Stadler [300].

The mathematical formulation of this notion under its actual form is due to Debreu [131, 132]. Basically, we consider a finite family  $\{f_1, \dots, f_m\}$  of functions from  $\mathbb{R}^n$  to  $\mathbb{R}$ , called objective functions, and we look for points  $\bar{x} \in \mathbb{R}^n$  for which no objective function can be improved without penalizing an other. More precisely, points  $\bar{x}$  for which there is no  $x \in \mathbb{R}^n$  such that

$$f_i(x) \leq f_i(\bar{x}) \text{ for all } i \in \{1, \dots, m\} \text{ and } (f_1(x), \dots, f_m(x)) \neq (f_1(\bar{x}), \dots, f_m(\bar{x})).$$

Multi-objective optimization problems ((MOP) for short) are generally written under the form

$$(MOP) \quad \underset{x \in \mathbb{R}^n}{\text{MIN}} (f_1(x), \dots, f_m(x)).$$

These multi-objective optimization problems are part of a more general setting, the one of vector optimization problems:

$$(VOP) \quad \underset{x \in X}{\text{MIN}} F(x),$$

where  $F$  is a function from some vector space  $X$  to an ordered vector space  $(Y, \leq_Y)$ . In these problems, we look for the *efficient points* of (VOP), that is, the points  $\bar{x} \in X$  for which there is no  $x \in X$  such that

$$F(x) \leq_Y F(\bar{x}) \text{ and } F(x) \neq F(\bar{x}).$$

The multi-objective problem can then be recovered by taking  $F := (f_1, \dots, f_m) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\mathbb{R}^m$  being equipped with the usual partial order  $\leq_{\mathbb{R}^m}$  induced by the positive orthant  $\mathbb{R}_+^m$ .

These vector equilibrium and optimization problems were introduced by Debreu [131] to model the *preferences* of a decider, and have been the object of various studies (see the survey [128] and its references, or more recently [133, 245, 46]). From the applications point of view, there is few vector optimization problems which are not simply multi-objective, except in economy. Let us mention for instance some recent developments on equilibrium in financial markets [7, 8].

When we face a multi-objective problem, one of the purposes is to detect the set of Pareto points. A well-known approach is the weighting method [323, 171], which minimizes some convex combination of the objective functions. Its popularity comes from its easy implementation, since it reduces to a classical optimization problem. It works well in the convex setting, since it is known in that case that the minimizers of all the convex combinations cover the Pareto points [139]. Nevertheless, in the general nonconvex case, this method does not recover well the nonconvex parts of the Pareto front [126, 241]. Moreover, we cannot use the weighting method if we look for a cooperative approach: for instance, in shape optimization, if we are given an initial shape, we might want to reach a Pareto equilibrium by improving all the involved criteria along time. But the objective functions corresponding to the weakest weightings will suffer from consequent variations. Furthermore, this method requires the choice of an *a priori* convex combination, which is not always easy when the objectives are not correlated.

Of course, other approaches exist, like the hierarchical minimization [158] which first minimizes one objective function, then an other under the constraint of remaining in the minimizers of the first, and so on. This hierarchy asks to make a choice among the objectives, and it is easy to see on simple examples that the Pareto set cannot be entirely recovered like this. The  $\varepsilon$ -constraint method [183] relies more or less on the same idea: we minimize one objective, under a sublevel constraint concerning the other objective functions. Let us finally mention other empiric methods, called evolutive or genetic methods [161, 129, 130], for which there is no theoretical guarantee of convergence .

We detail now some works published in the 70s, which influenced our own work and were precursor for multi-objective descent methods. These works have in common that they have been written for economical purposes, in particular for modeling allocation of resources in planned procedures.

Smale<sup>2</sup> is the first to be interested in a continuous dynamic promoting the simultaneous decrease of a finite family of objective functions  $\{f_1, \dots, f_m\}$ , without introducing a subjective choice or combination [297]. Its approach consists in defining a trajectory whose derivative is, at each point, a common descent direction for all objective functions  $f_i$ . Given a convex constraint

---

<sup>2</sup>Who was a co-worker of Debreu at the University of California, Berkeley.

$C \subset X = \mathbb{R}^n$ , he defines at each  $x \in C$  the cone<sup>3</sup> of common descent directions  $D(x)$ , as the intersection of the cones of admissible descent directions for each objective function  $f_i$ , that is:

$$D(x) := \{d \in T_C(x) \mid \langle \nabla f_i(x), d \rangle < 0, \forall i \in \{1, \dots, n\}\}.$$

He defines then a *Pareto critical point* as a point  $x \in C \subset \mathbb{R}^n$  such that  $D(x)$  is empty. Roughly speaking, it is a point such that, at the first order, we cannot improve strictly all the objective functions simultaneously, and remain in the constraint. It is a weaker notion than the one of Pareto optimality (strictly weaker in general), and is well-known at this time [258, 293, 315]. Note that this first-order optimality condition is exactly the usual notion of critical point that we are used to work with, when  $m = 1$ . Thus, Smale is interested in trajectories  $u : C \rightarrow \mathbb{R}^n$  satisfying the following differential inclusion:

$$\dot{u}(t) \in \begin{cases} D(u(t)) & \text{if } u(t) \text{ is not Pareto critical,} \\ \{0\} & \text{else.} \end{cases}$$

As it can easily be verified, such trajectories make each objective function decrease along time. Moreover, this dynamic is defined in such a way that each stationary point is a Pareto critical point (and so, candidate to be a Pareto point). These results are effectively proved by Cornet [115] some years later. This last author proves also, under some (restrictive) conditions, the existence of a smooth selection, that is, a function  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  verifying  $\psi(x) \in D(x)$  for all  $x \in \mathbb{R}^n$ , and such that  $\dot{u}(t) = \psi(u(t))$  admits a unique solution of class  $C^1$ .

In its thesis, Cornet [116] defines *explicitly* a cooperative dynamic, which enters in the setting of Smale's, and generalizes to the multi-objective setting the classic steepest descent dynamic. This dynamic writes as follows:

$$(SD) \quad \dot{u}(t) + (N_C(u(t)) + \text{co}\nabla f_i(u(t)))^0 = 0,$$

where  $\text{co}\nabla f_i(u(t))$  denotes the convex hull of the family of gradients  $\{\nabla f_i(u(t))\}_{i \in \{1, \dots, m\}}$ , and we use the notation  $A^0$  to indicate the element of minimal norm of a closed convex set  $A$ . Cornet observes that the vector field governing its dynamic can be interpreted as an analogue of the usual *steepest descent direction*. Indeed, if we note  $s(u) = -(N_C(u(t)) + \text{co}\nabla f_i(u(t)))^0$ , then

$$\frac{s(u)}{\|s(u)\|} = \underset{\substack{d \in T_C(u) \\ \|d\|=1}}{\text{argmin}} \max_{i \in \{1, \dots, m\}} \langle \nabla f_i(u), d \rangle,$$

which generalizes a well-known property of the gradient in classic optimization. The author gives also a proof of the existence of global trajectories (based on the Kakutani-Fan fixed point theorem), together with their convergence to Pareto critical points when  $t \rightarrow +\infty$ , in the convex case. A very interesting aspect of this dynamic is that, at each instant  $t$ , there exists a convex combination  $(\theta_i(t))_{i \in \{1, \dots, m\}}$  such that

$$\dot{u}(t) + N_C(u(t)) + \nabla \left( \sum_{i=1}^m \theta_i(t) f_i \right) (u(t)) = 0, \quad \sum_{i=1}^m \theta_i(t) = 1, \quad \theta_i(t) \geq 0.$$

In other words, the dynamic behaves like the steepest descent dynamic associated to a convex combination of the objective functions, except that this combination evolves *dynamically*. The weighting is chosen endogenously by the dynamic, to promote the common decrease of the objective functions.

This idea of extending the steepest descent dynamic to vector optimization problems appears also in the works of Malivert [234] and Pascoletti-Serafini [259]. Nevertheless, there has been almost no works in this direction until the end of the century.

<sup>3</sup>In fact a strict cone, as defined later in Chapter 2, since it does not contain the origin.

In the 2000s, in a paper of Fliege and Svaiter [159], we observe a new interest for the steepest descent method for multi-criteria optimization problems. These authors propose an algorithm to solve an unconstrained smooth multi-objective optimization problem. For this, they define on  $X = \mathbb{R}^n$  the following vector field:

$$(1.6) \quad s(x) = \operatorname{argmin}_{d \in \mathbb{R}^n} \left\{ \frac{1}{2} \|d\|^2 + \max_{i \in \{1, \dots, m\}} \langle \nabla f_i(x), d \rangle \right\}.$$

This direction  $s(x)$  being a common descent<sup>4</sup> direction at  $x$  for each objective function  $f_1, \dots, f_m$ , we can define a descent algorithm

$$x_{n+1} = x_n + t_n s(x_n),$$

where  $t_n$  is a stepsize chosen by an Armijo-like rule. As we will see later in Chapter 5,  $s(x)$  is nothing but the steepest descent direction at  $x$ , in the sense of Cornet, that is  $s(x) = -(\operatorname{co} \nabla f_i(u(t)))^0$ . Thus, we can see the algorithm of Svaiter and Fliege as a discretization in time of the steepest descent dynamic, and we recover (up to some differences) the same asymptotic behaviour.

In the next years, a series of papers will extend this steepest descent method to multi-objective optimization problems with constraint, taking into account relative errors, and/or adapting these methods to more general vector optimization problems [173, 176, 248, 63, 62, 165, 60]. Since this steepest descent method extends the usual gradient descent method to the multi-objective case, some authors adapt other well-known optimization methods. Let us mention for instance the Newton method [160, 175], or quasi-Newton [279, 277], the interior point method [312, 313], the proximal algorithm [73, 99, 98, 61], and subgradient methods [109, 59].

In parallel, other authors focus again on the continuous steepest descent dynamic. Schäffler-Schultz-Weinzierl [295] consider<sup>5</sup> a stochastic version of the steepest descent dynamic (SD). Through the notion of vector *pseudo-gradient*, Miglierina [247] recovers the (SD) dynamic in the unconstrained case. At the same time, Recchionni [282] introduces a dynamic close to (SD), considering an interior penalization for a box constraint. This author is the first to verify numerically that this steepest descent approach is efficient for recovering nonconvex Pareto fronts. Miglierina-Molho-Recchionni [248] study a similar dynamic, and show that their approach is more efficient than the weighting method. More recently, Attouch-Goudou [28] revisit also (SD) in the convex case under a general convex constraint. They propose a new constructive proof of the existence of the trajectories, based on Peano's Theorem and a Moreau-Yosida regularization of the nonsmooth operator  $N_C$ .

Let us now describe our main contributions and the contents of Part II.

In a first place, we studied the (SD) dynamic by considering the multi-objective optimization problem associated to Lipschitz continuous functions  $\{f_1, \dots, f_m\}$ . The idea was to replace in (SD) the gradients of the objective functions by their Clarke subdifferentials:

$$(SD) \quad \dot{u}(t) + (\operatorname{co} \partial^c f_i(u(t)))^0 = 0.$$

We studied this dynamic, but also the properties of the operator governing it, namely  $x \mapsto \operatorname{co} \partial^c f_i(x)$ .

<sup>4</sup>Just take  $d = 0$  in (1.6).

<sup>5</sup>Pay attention to the fact that one of their main lemmas is false. Indeed, the authors prove that the steepest descent vector field  $s : \mathbb{R}^n \rightarrow \mathbb{R}^n$  defined in (1.6) is locally Lipschitz continuous. We will see in Chapter 6, Section 6.1.3, that this is false, through a counter-example.

This guided us to think about the links between this operator  $x \mapsto \text{co } \partial^C f_i(x)$ , and the family of objective functions  $\{f_1, \dots, f_m\}$ . Indeed, why are involved the gradients (or subgradients) of all convex combinations

$$(1.7) \quad \left\{ f_\theta := \sum_{i=1}^m \theta_i f_i \mid \sum_{i=1}^m \theta_i = 1, \theta_i(t) \geq 0 \right\},$$

when we only look for the minimization of a finite family of functions? Why do we have to call this larger family of functions?

By taking the more general point of view of vector optimization, we understood that the minimization of  $F = (f_1, \dots, f_m) : X \rightarrow (\mathbb{R}^m, \leq_{\mathbb{R}^m})$  expresses naturally as the simultaneous minimization of the convex combinations in (1.7). Even better, we understood that the central role given to the objective functions  $\{f_1, \dots, f_m\}$  is essentially due to the extremely favorable geometrical properties of the ordering cone  $\mathbb{R}_+^m$ . This brought us to the study of descent dynamics associated to general vector optimization problems, which seems to be the good setting to work in.

Our aim, in the second part of this thesis, is to give a new look on vector optimization problems. For this, we place at the center of our analysis the notion of *steepest descent direction*, and the dynamic(s) that it induces. The setting in which we will work is the following: we consider a function  $F : X \rightarrow Y$  between two Banach spaces, the space  $Y$  being equipped with a partial order noted  $\leq$ . We will always assume that the order on  $Y$  is induced by a closed convex cone with nonempty interior. Given a nonempty closed convex set  $C \subset X$  which models the constraints, we study the optimization problem

$$(VOP) \quad \text{MIN}_{x \in C} F(x).$$

Taking  $X$  as a Banach space allows to cover situations appearing in decision science, engineering or economy, where the decision space can be taken as a space of integrable functions [281, 261]. Taking  $Y$  as an ordered Banach space allows of course to cover the multi-objective case, but also to consider functions with values in a space of continuous functions, or  $L^\infty([0, 1], \mathbb{R})$ . This might lead to new modelisations in economy, or applications in optimal control.

In Chapter 5, we study the notions of descent direction for  $F$ , and of solution(s) for (VOP). For this, we introduce in Section 5.1 the *ordered (Clarke) subdifferential*,

$$\partial^C F : X \rightrightarrows X^*,$$

which will give us some first-order local information on the behaviour of  $F$  with respect to the order on  $Y$ . When we take place in the multi-objective setting with  $F = (f_1, \dots, f_m)$ ,  $\partial^C F(x)$  reduces to  $\text{co } \partial^C f_i(x)$ , and we recover the operator introduced by Cornet. After giving some calculus rules and basic properties for this ordered subdifferential, we show in Section 5.2 that  $\partial^C F(x)$  has interesting properties for the analysis of (VOP):

- It verifies a Fermat's rule, giving a first-order necessary condition for the solutions of (VOP), see Theorem 5.2.10.
- Like in the scalar case ( $Y = \mathbb{R}$ ), under some regularity conditions,  $\partial^C F(x)$  spans the normal cone of the sublevel set of  $F$  at  $x$ , see Theorem 5.2.12.
- When  $X$  is a Hilbert space, the ordered subdifferential gives us *descent directions*. It suffices to consider at each point  $x \in X$  the element of minimal norm of  $-\partial^C F(x)$ , see Propositions 5.2.18 and 5.2.8. In fact, this direction is the *steepest*, in the sense given by Theorem 5.2.20.

For this two last properties, the key point is that, for  $x \in X$ , the ordered subdifferential  $\partial^c F(x)$  is a subset of  $X^*$ , related to the decision space  $X$ . This is in clear opposition with the usual differentiability notions, or their generalizations to the nonsmooth setting, which give a subset of  $L(X, Y)$  (see Remark 5.3.2 for more discussion on this subject).

In Chapter 6, we assume that  $X$  is a Hilbert space, and we use  $\partial^c F$  to define the *steepest descent* dynamic

$$(SD) \quad \dot{u}(t) + (N_C(u(t)) + \partial^c F(x))^0 = 0, \quad t \geq 0.$$

In the smooth multi-objective case, we recover the dynamic introduced by Cornet. Furthermore, a simple discretization in time of (SD) gives the descent algorithms developed in [159, 173, 176, 248, 63, 62, 165, 60], or even the proximal methods of [73, 99, 98, 61]. Being able to compare the continuous and discrete dynamics is always fruitful, since they often share the same asymptotic behaviour (see for instance the works of Peypouquet-Sorin [269] and Alvarez-Peypouquet [14]). We discuss into more details this comparison in Section 6.3.1. Given a global solution of (SD) (in a sense to precise), we prove in Proposition 6.1.7 that it has a descent property:

$$\forall s, t \in [0, +\infty[, \quad s \leq t \Rightarrow F(u(t)) \leq F(u(s)).$$

Then, we study the asymptotic behaviour of a trajectory  $u(\cdot)$  in Theorem 6.2.6. We prove in particular in the convex case that, when  $t$  goes to  $+\infty$ , the trajectory  $u(t)$  weakly converges to a weak efficient point of (VOP). Weaker results are obtained in the more general quasi-convex setting. Section 6.3 is devoted to some numerical simulations of this dynamic in the multi-objective setting. In particular, we agree with [248] which observe that the steepest descent dynamic is efficient to recover nonconvex Pareto fronts. The fact that a convex combination of the objective functions is chosen dynamically in (SD) seems to be an advantage, contrary to the weighting method.

Chapter 7 is entirely devoted to the question of the existence of trajectories for (SD), in the nonsmooth convex case. Our proof relies essentially on two ingredients, like in [28]: a Moreau-Yosida regularization for  $\partial^c F$ , and Peano's theorem. In particular, because of Peano's existence theorem, we have to assume that  $X$  has finite dimension. We cannot hope to use the Cauchy-Lipschitz theorem, because the vector field governing (SD) is not locally Lipschitz continuous (see Remark 6.1.16). As a consequence, for now, the uniqueness of the trajectories remain an open problem. In Section 7.1, we give a first result of existence in the smooth case, using an abstract result from [28]. In Section 7.2, we prove the existence of trajectories of (SD) in the nonsmooth convex multi-objective case. In this favorable setting, we only deal with a finite number of objective functions, on which we can apply a Moreau-Yosida regularization. In Section 7.3, we focus on the general vector case, and the main difficulty lies in the definition of an appropriate regularization for  $\partial^c F$ . This technical question (which has its own interest) is addressed in Sections 7.3.1 and 7.3.2. This being done, we are able to prove Theorem 7.0.1 which establishes the existence of global trajectories for (SD).

Finally, in Chapter 8, we open a completely new road: an inertial approach to multi-objective optimization problems. Indeed, generating a set of Pareto points asks the simulation of a consequent number of trajectories (using the steepest descent dynamic (SD), or a weighting method). The problem is that these methods are quite slow to converge, and we look for fast methods, which arise in general by considering *second-order* methods. It already exists methods involving second-order informations in *space*, using the curvature (second derivatives) of the objective functions. These are the Newton (or quasi-Newton) methods for multi-criteria optimization [160, 175, 277]. Nevertheless, they can be expensive, due to the necessity to solve a quadratic problem at each iteration.

In scalar-valued optimization, inertial methods, which involve second-order information in *time*, are very popular because they are quick to converge, and easy to implement. In the same

spirit than [276, 29], we look for an *inertial* version of the steepest descent dynamic, called the Inertial Steepest Descent with Friction:

$$\text{(ISDF)} \quad \ddot{u}(t) + \gamma\dot{u}(t) + \partial^c F(u(t))^0 = 0,$$

where  $\gamma > 0$  can be interpreted as a friction parameter. We restrict ourselves in Chapter 8 to the multi-objective case, and assume that  $F = (f_1, \dots, f_m)$  is convex and of class  $C^2$ . We prove the existence of global trajectories for (ISDF) in Theorem 8.1.5, and study the asymptotic behaviour of its solutions. In particular, we show in Theorem 8.2.8 that the trajectories of (ISDF) weakly converge to weak Pareto points. We think that this first approach opens the road to future studies of Nesterov-like algorithms for multi-objective optimization problems.

We end this introduction with some practical comments on the structure of the thesis. Since it is separated in two independent parts, we chose not to write a general conclusion. Instead, we placed at the end of each chapter a small concluding section, in which we discuss in detail particular points of the chapter, or present some directions of research for the future.

This thesis resulted in two published papers [163, 27], and a preprint which should be submitted soon [26].

The article [163] comes from a joined work with P. Frankel and J. Peypouquet. It has been split in two parts, covering the Sections 3.2 and 3.3 of Chapter 3, and Section 4.1 of Chapter 4. The rest of Chapters 3 and 4 are unpublished, and have been mostly written during the redaction of [163]. Most of the contents of Chapters 5 to 7 for the multi-objective optimization problems, results from a collaboration with H. Attouch and X. Goudou, published in [27]. Chapter 8 comes from a joint work with H. Attouch, whose preprint should be submitted soon [26]. The notions and results of Chapter 5 to 7, concerning the general vector optimization problem are new, in particular Section 7.3. Some developments for the multi-objective setting are also new, such as Sections 5.2.2 or 6.3.





## Chapter 2

# Variational analysis tools in Banach spaces

In this introducing chapter, we present the basic notions and tools which will be needed along this thesis. It is also the occasion to fix some notations. As much as we could, we avoided the proofs in this chapter, referring properly to some textbooks. If a result is given without reference, this means that its proof is putted into Appendix A. The idea is that the reader can quickly browse through this introductory chapter, and focus only on what he could not know.

The content of this chapter is the following: In Section 2.1 we recall classic properties of the convex sets and cones, which can be found in the textbooks of Aliprantis and Border [5] and Fabian and al. [153]. In Section 2.2, we present some of the most standard tools used in nonsmooth optimization, such as the Fréchet, limiting or Clarke subdifferentials, and the corresponding tangent and normal cones. See the books of Clarke [111], Mordukhovich [244] and Penot [263] for a wide account on this subject. Section 2.3 is devoted to the notions involved in vector optimization problems, in particular the important notion of base for a cone. The majority of the results there comes from the books of Dinh [139] and Jahn [206].

In this thesis, we will consider functions  $F : X \rightarrow Y$ , where  $X$  will often be a Hilbert space. While in Part I we will consider extended-real-valued functions, in Part II we will deal with functions taking values in a general Banach spaces. This is why most of this introductory chapter considers objects in a Banach space. Moreover, we will need to state some results which applies for both the norm and weak topologies in  $X$ , or the weak\* topology in  $X^*$ . This is why, in this chapter, we will sometimes work in the general setting of a topological vector space  $(X, \tau)$ . In the interest of the sanity of the reader, we will not recall here all the basic necessary notions about topological vector spaces, and refer for this material to the excellent book [5].

Let us fix some notations which will be used throughout this thesis. Given a real Banach space  $(X, \|\cdot\|_X)$ , we note  $X^*$  its topological dual space, which is also a Banach space once equipped with its operator norm  $\|\cdot\|_{X^*}$ . If there is no ambiguity on the space we work with, we will just note  $\|x\|$  (resp.  $\|x^*\|$ ) instead of  $\|x\|_X$  (resp.  $\|x^*\|_{X^*}$ ). We note  $\mathbb{B}_X$  and  $\mathbb{B}_{X^*}$  the unit balls of  $X$  and  $X^*$ , respectively.

The primal space  $X$  can be equipped with the *weak topology*  $w(X, X^*)$ , which is the weakest (or coarser) topology that makes all the linear forms  $x^* \in X^*$  continuous. We will also equip the dual  $X^*$  with the *weak-star topology*  $w^*(X^*, X)$ , which is the restriction to  $X^*$  of the pointwise convergence topology of  $\mathbb{R}^X$ . In all this document, we will write  $w^*$  instead of  $w^*(X^*, X)$ , and  $w$  instead of  $w(X, X^*)$ .

If we have to mention some topological property (such as closedness, compactness, convergence, continuity, ...), we shall precise for what topology it refers to, except for the case of the norm topology. For instance, a set will be said weakly closed if it is closed for the weak topology. If we just mention that it is closed, it means that we implicitly refer to the norm topology of

the ambient space.

Given a topological vector space  $(X, \tau)$ , we will write  $\text{cl}_\tau A$  for the closure of  $A \subset X$  with respect to the topology  $\tau$ . Whenever a sequence  $(x_n)_{n \in \mathbb{N}}$  is  $\tau$ -convergent to some  $x$ , we note  $x_n \xrightarrow[n \rightarrow +\infty]{\tau} x$ . For such topological space, we also note  $X^*$  its topological dual, and note  $w^*$  the corresponding weak\* topology on  $X^*$ . We shall use the fact that, in that case, the topological dual of  $(X^*, w^*)$  is  $X$ .

Since we deal with weak topologies in Banach spaces, which are not sequential, we are going to use nets to characterize continuity or closedness. Generically, we will use nets  $(x_\alpha)_{\alpha \in \mathcal{A}}$  indexed by a directed set  $\mathcal{A}$ . Sometimes, using for instance an argument involving weak\* compactness in  $X^*$ , we will make use of subsets. A subset of  $(x_\alpha)_{\alpha \in \mathcal{A}}$  is  $(x_{f(\beta)})_{\beta \in \mathcal{B}}$ , for some directed set  $\mathcal{B}$  and a final monotone function  $f : \mathcal{B} \rightarrow \mathcal{A}$ . To lighten the notations, we will abusively write  $(x_\beta)_{\beta \in \mathcal{B}}$  instead of  $(x_{f(\beta)})_{\beta \in \mathcal{B}}$  when referring to such subset.

## 2.1 Basic topology and differential calculus

### 2.1.1 Convex sets

Let  $(X, \tau)$  be a Banach space. We recall that  $C \subset X$  is said to be *convex* if

$$\text{for all } x \in C, y \in C, \lambda \in [0, 1], \quad \lambda x + (1 - \lambda)y \in C.$$

According to Hahn-Banach separation theorem [325, Theorem 1.1.5], closed convex sets are exactly the weakly closed convex sets. In particular, for any convex set  $C \subset X$ , the norm closure and weak closure coincide:  $\text{cl} C = \text{cl}_w C$ . An other consequence of the Hahn-Banach theorem is that the convex sets in  $X$  can be written as the intersection of all closed half-spaces which contain it [5, Corollary 5.83]. When a set  $C$  is defined as a finite intersection of closed half-spaces, we say that  $C$  is a *polyhedron*.

Given  $A \subset X$ , its *convex hull* is defined by

$$(2.1) \quad \text{co}A = \left\{ \sum_{i=1}^m \theta_i a_i \mid m \in \mathbb{N}, (\theta_i) \in \Delta_m, (x_i) \subset A \right\},$$

where  $\Delta_m$  denotes the unit simplex in  $\mathbb{R}^m$

$$\Delta_m := \left\{ \theta = (\theta_i) \in \mathbb{R}^m \mid \sum_{i=1}^m \theta_i = 1 \text{ and } \theta_i \geq 0 \forall i \in \{1, \dots, m\} \right\}.$$

If  $A = \bigcup_{i=1}^m A_i$ , we just note its convex hull  $\text{co}A_i$  instead of  $\text{co} \left\{ \bigcup_{i=1}^m A_i \right\}$ , if there is no ambiguity on the index  $i$ . If the sets  $(A_i)_{i \in \{1, \dots, m\}}$  are convex, the convex hull of  $\bigcup_{i=1}^m A_i$  admits the more simple representation (see [5, Lemma 5.29]):

$$(2.2) \quad \text{co}A_i = \left\{ \sum_{i=1}^m \lambda_i a_i \mid (\lambda_i) \in \Delta_m, a_i \in A_i \right\}.$$

The convex hull of a finite set is called a *polytope*. Polytopes are always compact [5, Corollary 5.30], and coincide with polyhedrons whether  $X$  has finite dimension<sup>1</sup> [5, Theorem 7.79].

Are the topological properties of a set conserved after taking its convex hull? The answer is in general negative:

---

<sup>1</sup>In fact, the fact that all polytopes are polyhedrons characterizes the finite dimension, since polyhedrons cannot be compact in an infinite dimensional space.

**Example 2.1.1.** The convex hull of a closed set is not necessary closed. Take for instance  $A = \{(0, 0)\} \cup \mathbb{R} \times \{1\}$  in  $\mathbb{R}^2$ .

**Example 2.1.2.** The convex hull of a compact set is not necessary compact. Take for instance  $A = \{0\} \cup \{u_1, u_2, \dots\}$  in  $\ell^2(\mathbb{N})$ , where  $u_n = \frac{1}{n}e_n$ . Here,  $(e_n)_{n \in \mathbb{N}}$  is the usual canonical basis. Then it can be shown that  $A$  is compact but its convex hull is not even closed, for the norm topology (see [5, Example 5.34]).

Nevertheless, the answer is yes in some cases. For instance the convex hull of a finite set (hence compact) is compact. More generally, if  $A$  is structured as a finite union of convex compact sets (see also [5, Theorems 5.35 and 6.35]):

**Proposition 2.1.3** ([5, Lemma 5.29]). Let  $(X, \tau)$  be a topological vector space. Suppose that  $A = \bigcup_{i=1}^m A_i$  with the  $A_i$  being convex and compact. Then  $\text{co}A$  is compact.

We note  $\overline{\text{co}}A$  the *closed convex hull* of  $A \subset X$ , defined as the closure of  $\text{co}A$  for the norm topology (or equivalently, the weak topology). It can also be defined as the intersection of all closed convex sets containing  $A$ . Equivalently, we note  $\overline{\text{co}}^*A := \text{cl}_{w^*} \text{co}A$  the weakly\* closed convex hull of a set  $A \subset X^*$ .

Given a nonempty  $A \subset X^*$ , define the *support function* of  $A$  as

$$\begin{aligned} \sigma_A : X &\longrightarrow \mathbb{R} \cup \{+\infty\} \\ x &\longmapsto \sup_{x^* \in A} \langle x^*, x \rangle, \end{aligned}$$

where  $\langle \cdot, \cdot \rangle$  denotes the dual pairing between  $X^*$  and  $X$ . It is lower semi-continuous and sublinear, and even continuous if  $A$  is weakly\* compact (see Proposition A.3.4). Clearly,  $\sigma_A = \sigma_{\overline{\text{co}}^*A}$ , but we have a stronger and very useful characterization:

**Theorem 2.1.4.** For all  $A, B \subset X^*$  nonempty,  $\sigma_A = \sigma_B \Leftrightarrow \overline{\text{co}}^*A = \overline{\text{co}}^*B$ .

This theorem is a direct consequence of the bijection between weakly\* closed convex sets and lower semi-continuous sublinear functions (see [5, Theorem 7.51], or [111, Proposition 2.1.4]).

A last interesting notion about convex sets are the *extreme points*. Given a convex set  $C$ , we say that  $x \in C$  is an extreme point of  $C$  if  $C \setminus \{x\}$  is still a convex set. The Krein-Milman theorem [5, Theorem 7.68] ensures that in a locally convex separated topological vector space, any convex compact set can be written as the closed convex hull of its extreme points. The polytopes are exactly the sets having a finite number of extreme points [5, Lemma 7.76].

## 2.1.2 Cones and duality

A set  $K \subset X$  is said to be a *cone* if  $\mathbb{R}_+K \subset K$ , where  $\mathbb{R}_+ := [0, +\infty[$ . A set  $S \subset X$  is said to be a *strict cone* if  $\mathbb{R}_{++}S \subset S$  and  $0 \notin S$ , where  $\mathbb{R}_{++} := ]0, +\infty[$ . Once given an arbitrary set  $A$ , one can consider its *conical hull* as  $\mathbb{R}_+A$ . As before, a natural question is to know whether if a set remains closed by taking its cone hull (it is obviously false for the compact property since a nontrivial cone is always unbounded). It is a question of interest, see for instance a direct application in Theorem 5.2.12.

**Example 2.1.5.** Take  $X = \mathbb{R}^2$  and consider  $A = \{(x, y) \in \mathbb{R}_+^2 \mid xy = 1\}$ . It is a closed set, but  $\mathbb{R}_+A = \mathbb{R}_{++}^2 \cup \{(0, 0)\}$  is not.

**Example 2.1.6.** Take  $X = \mathbb{R}^2$  and take  $A$  as the closed disc of center  $(1, 0)$  and radius 1 in  $\mathbb{R}^2$ . It is a closed (and even compact) set, but  $\mathbb{R}_+A = (\mathbb{R}_{++} \times \mathbb{R}) \cup \{(0, 0)\}$  is not closed.

We see here that the conical hull of a closed set can fail to be closed. In Example 2.1.5, it comes from a lack of boundedness of  $A$ . In Example 2.1.6, it comes from the infinite number of rays arising from  $0 \in \text{bd } A$ . In fact a positive result for compact sets can be obtained under a finiteness assumption (see [5, Corollary 5.25]), or if  $0 \notin A$  (the proof is left in Appendix):

**Proposition 2.1.7.** In a Hausdorff locally convex topological vector space  $(X, \tau)$ , let  $A \subset X$  be a compact set not containing the origin. Then  $\mathbb{R}_+ A$  is closed.

The end of this section is devoted to dual objects. Given a Banach space  $(X, \|\cdot\|)$  and a nonempty  $A \subset X$ , we define the *polar set* of  $A$  in  $X^*$  by

$$(2.3) \quad A^* := \{x^* \in X^* \mid \langle x^*, x \rangle \leq 1 \ \forall x \in A\}.$$

One sees immediately that  $A^*$  is a weakly\* closed convex set containing the origin.

**Example 2.1.8.** Let  $A$  be the unit ball  $\mathbb{B}_X$  of a normed space  $X$ , then  $A^*$  is the unit ball  $\mathbb{B}_{X^*}$  of  $X^*$ . Indeed,  $x^* \in (\mathbb{B}_X)^* \Leftrightarrow \sup_{\|x\| \leq 1} \langle x^*, x \rangle \leq 1 \Leftrightarrow \|x^*\| \leq 1 \Leftrightarrow x^* \in \mathbb{B}_{X^*}$ .

If  $K$  is a cone, its polar is also a cone, called the *polar cone*, which can be equivalently defined by

$$(2.4) \quad K^* := \{x^* \in X^* \mid \langle x^*, x \rangle \leq 0 \ \forall x \in K\}.$$

It can be also useful to consider its *dual cone*  $K^+$ , which is just the opposite of the polar cone

$$K^+ := -K^* = \{x^* \in X^* \mid \langle x^*, x \rangle \geq 0 \ \forall x \in K\}.$$

When  $X$  is a Hilbert space, we say that a cone is *self-dual* whenever  $K = K^+$ .

**Example 2.1.9.**

- If  $A$  is a closed linear subspace of a Hilbert space, then  $A^*$  is its orthogonal.
- If  $A = \mathbb{R}_+ \times \{0\}$  in  $\mathbb{R}^2$  then  $A^* = \mathbb{R}_- \times \mathbb{R}$ .
- If  $A$  is the orthant cone  $\mathbb{R}_+^m := \{x \in \mathbb{R}^m \mid x_i \geq 0\}$  then  $A^* = -A$ . It is an example of self-dual cone.
- The cone of symmetric positive real matrices  $\mathcal{S}_+^m(\mathbb{R})$  in  $\mathcal{M}^m(\mathbb{R})$ , or the cone of almost-everywhere positive functions  $L_+^2([0, 1], \mathbb{R})$  in  $L^2([0, 1], \mathbb{R})$  are also examples of self-dual cones.

Given a set  $B \subset X^*$ , we can naturally<sup>2</sup> define its polar in  $X$  by

$$(2.5) \quad B^* := \{x \in X \mid \langle x^*, x \rangle \leq 1 \ \forall x^* \in B\}.$$

It is a natural question to compare a set  $A$  and its *bipolar*  $(A^*)^*$ . For instance, a cone  $K \subset X$  is closed and convex if and only if  $K = (K^*)^*$ . Equivalently, a cone  $K \subset X^*$  is weakly\* closed and convex if and only if  $K = (K^*)^*$  (see [153, Theorems 3.38, 3.45]).

Let us give a word on the particular case of  $(X, \langle \cdot, \cdot \rangle)$  being a Hilbert space. In that setting, a cone and its dual can be considered to live in the same space, and we can derive a direct sum decomposition result, due to Moreau [246] (see also [51, Theorem 6.29]):

<sup>2</sup>In fact, given a general topological vector space  $(X, \tau)$ , we can define the polar of  $A \subset X$  as a set in  $(X, \tau)^*$ . Hence, when considering a set  $A$  in the dual of a Banach space  $X^*$ , it is enough to endow it with the weak\* topology in order to obtain  $A^*$  in the primal  $X$ , since  $(X^*, w^*)^* = X$ .

**Theorem 2.1.10.** (Moreau) Let  $X$  be a Hilbert space and  $K \subset X$  a closed convex cone. Then, for all  $x \in X$ , there exists a unique couple  $(x_K, x_{K^*}) \in K \times K^*$  such that

$$(2.6) \quad x = x_K + x_{K^*} \text{ and } \langle x_K, x_{K^*} \rangle = 0.$$

In that case,  $x_K = \text{proj}_K(x)$  and  $x_{K^*} = \text{proj}_{K^*}(x)$ . In particular,  $\|x\|^2 = \|\text{proj}_K(x)\|^2 + \|\text{proj}_{K^*}(x)\|^2$ .

In this Theorem, we used the *projection operator*  $\text{proj} : X \rightrightarrows X$ , which is defined as follows: for a nonempty set  $A$  in a Banach space  $X$  and  $x \in X$ , the projection of  $x$  onto  $A$  is the set noted  $\text{proj}_A(x)$  defined as

$$\text{proj}_A(x) := \underset{a \in A}{\text{argmin}} \|x - a\|.$$

If  $X$  is a Hilbert space and  $A \subset X$  is a nonempty closed convex set, then  $\text{proj}_A(x)$  reduces to a single point, which is the situation occurring in Moreau's decomposition theorem.

We end with *polyhedral* cones in a topological space  $(X, \tau)$ . A cone  $K$  is said to be polyhedral if there exists a finite family  $\{x_1, \dots, x_m\} \subset X$  such that

$$K = \mathbb{R}_+ \text{co}\{a_1, \dots, a_m\}.$$

In other words, a polyhedral cone is the cone hull of a polytope. According to a well-known result (see [154, 317, 309, 102] or more recently [77, Theorem 5.1.7]), when  $X$  has finite dimension, a cone  $K$  is polyhedral if and only if it is a polyhedron. Equivalently, if there is a finite number of linear forms  $\{x_1^*, \dots, x_m^*\} \subset X^*$  such that

$$K = \{x \in X \mid \langle x_i^*, x \rangle \leq 0, \forall i \in \{1, \dots, m\}\}.$$

Exploiting both definitions, it is easy to see that in the finite dimensional setting, a cone  $K$  is polyhedral if and only if its polar cone  $K^*$  is polyhedral. Note that the finite dimension of  $X$  is guaranteed as soon as  $K$  is a polyhedral cone with a nonempty interior.

### 2.1.3 Differential and directional derivative

We set here some differentiability notions for a function  $F : X \longrightarrow Y$ , where  $X$  and  $Y$  are two Banach spaces.

We say that  $F$  is *directionally derivable* at  $\bar{x} \in X$  in the direction  $d \in X$  if the following limit exists in  $(Y, \|\cdot\|)$

$$\lim_{t \downarrow 0} \frac{F(\bar{x} + td) - F(\bar{x})}{t}.$$

In such a case, this limit is noted  $DF(\bar{x}; d)$ , and called the directional derivative of  $F$  at  $\bar{x}$  in the direction  $d$ . We shall say that  $F$  is directionally derivable at  $\bar{x}$  if  $DF(\bar{x}; d)$  exists for all  $d \in X$ . We directly see from the definition that, if  $F : X \longrightarrow Y$  is directionally derivable at  $\bar{x}$ , and  $A \in L(Y, Z)$  for some Banach space  $Z$ , then  $A \circ F$  is also directionally derivable at  $\bar{x}$ , with

$$\forall d \in X, D(A \circ F)(\bar{x}; d) = A(DF(\bar{x}; d)).$$

The function  $F$  is said to be Gateaux differentiable at  $\bar{x} \in X$  if it is directionally derivable at  $\bar{x}$ , and if the application

$$\begin{aligned} DF(\bar{x}) : X &\longrightarrow Y \\ d &\longmapsto DF(\bar{x}; d) \end{aligned}$$

is linear and continuous (for the norm topologies of  $X$  and  $Y$ ). In that case, we say that  $DF(\bar{x})$  is the differential (or Gateaux derivative) of  $F$  at  $\bar{x}$ . We will also note  $D^*F(\bar{x}) \in L(Y^*, X^*)$  the

adjoint of  $DF(\bar{x})$ . If  $F$  is Gateaux differentiable at each point of an open set  $U \subset X$ , we can consider the *differential* of  $F$

$$DF : U \longrightarrow L(X, Y).$$

We say that  $F$  is *strictly Gateaux* differentiable at  $\bar{x}$  if there exists  $A \in L(X, Y)$  which satisfies the stronger property

$$\lim_{\substack{x \rightarrow \bar{x} \\ t \downarrow 0}} \frac{F(x + td) - F(x)}{t} = A(d).$$

In that case,  $A$  is unique and equals  $DF(\bar{x})$ . Since we ask the point  $x$  to ‘move’ around  $\bar{x}$ , we can interpret this stronger notion as a form of continuity of  $DF(\cdot, d)$  around  $\bar{x}$ . More precisely, we have:

**Proposition 2.1.11.** Let  $F : X \longrightarrow Y$  be Gateaux differentiable on an open set  $U \subset X$ . Then, for all  $\bar{x} \in U$ , the following is equivalent:

- i)  $F$  is strictly Gateaux differentiable at  $\bar{x}$ ,
- ii)  $DF : U \longrightarrow L(X, Y)$  is *pointwise continuous* at  $\bar{x}$ , i.e. continuous with respect to the norm topology of  $X$  and the pointwise<sup>3</sup> topology of  $L(X, Y)$ . In other words, for all converging net  $(x_\alpha)_{\alpha \in \mathcal{A}} \subset U$  converging to  $\bar{x}$ , we have  $DF(x_\alpha; d) \xrightarrow[\alpha \in \mathcal{A}]{\|\cdot\|_Y} DF(\bar{x}; d)$  for all  $d \in X$ .

We will also use the differentiability in the sense of Fréchet.  $F$  is said to be *Fréchet* differentiable at  $\bar{x}$  if there exist a continuous operator  $A \in L(X, Y)$  such that

$$\lim_{x \rightarrow \bar{x}} \frac{F(x) - F(\bar{x}) - A(x - \bar{x})}{\|x - \bar{x}\|} = 0.$$

Similarly,  $F$  is *strictly Fréchet* differentiable at  $\bar{x}$  if there exists  $A \in L(X, Y)$  such that

$$(2.7) \quad \lim_{\substack{x \rightarrow \bar{x} \\ x' \rightarrow \bar{x}}} \frac{F(x) - F(x') - A(x - x')}{\|x - x'\|} = 0.$$

In both cases,  $F$  is Gateaux differentiable and the operator  $A$  is equal to the differential  $DF(\bar{x})$ . Here also the strict Fréchet differentiability is somehow equivalent to the continuity of  $DF$ :

**Proposition 2.1.12.** Let  $F : X \longrightarrow Y$  be Fréchet differentiable on an open set  $U \subset X$ . Then, for all  $\bar{x} \in U$ , the following is equivalent:

- i)  $F$  is strictly Fréchet differentiable at  $\bar{x}$ ,
- ii)  $DF : U \longrightarrow L(X, Y)$  is *strongly continuous* at  $\bar{x}$ , i.e. continuous with respect to the norm topology of  $X$  and the usual operator norm topology of  $L(X, Y)$ .

We say that a function  $F : U \subset X \longrightarrow Y$  is of class  $C^1$  if it is Gateaux differentiable on  $U$  and  $DF : U \longrightarrow L(X, Y)$  is strongly continuous. According to the Proposition above, functions of class  $C^1$  on  $U$  are exactly the strictly Fréchet differentiable functions on  $U$ . We say moreover that a function is of class  $C^{1,1}$  on  $\mathcal{U}$  if its derivative  $DF : U \longrightarrow L(X, Y)$  is Lipschitz continuous. Functions of class  $C^{1,1}$  enjoy the following property, very useful when studying descent methods (see Chapter 4). Its proof can be found for instance in [268, Lemma 1.30].

**Lemma 2.1.13.** [Descent lemma] Let  $U$  be an open subset of  $X$ . Let  $f : X \longrightarrow \mathbb{R}$  be a function of class  $C^1$ , its derivative being  $L$ -Lipschitz continuous on  $U$ . Then, for all  $x_1, x_2 \in X$ ,

$$f(x_2) - f(x_1) - \langle Df(x_1), x_2 - x_1 \rangle \leq \frac{L}{2} \|x_2 - x_1\|^2.$$

<sup>3</sup>The pointwise topology is the locally convex topology  $\tau_{pw}$  on  $L(X, Y)$  defined by the family of seminorms  $\{p_x\}_{x \in X}$ , where  $p_x : A \in L(X, Y) \mapsto \|Ax\|_Y$ . This topology is also called the weak operator topology.

## 2.2 Nonsmooth analysis for extended-real-valued functions

All along this section,  $X$  is a Banach space and we consider a function  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ . The *domain* of  $f$  is

$$\text{dom } f := \{x \in X \mid f(x) \in \mathbb{R}\}.$$

We say that  $f$  is *proper* whenever  $\text{dom } f \neq \emptyset$ . The *epigraph* of  $f$  is

$$\text{epi } f := \{(x, \lambda) \in X \times \mathbb{R} \mid f(x) \leq \lambda\}.$$

A *convex* function is a function for which the epigraph is convex. Equivalently, a function  $f$  satisfying

$$\forall x, x' \in X, \forall t \in [0, 1], f(tx + (1-t)x') \leq tf(x) + (1-t)f(x').$$

For  $A \subset X$ , we define the Lipschitz modulus of  $f$  on  $A$  by

$$\text{Lip}(f, A) := \sup_{x \neq x' \in A} \frac{|f(x') - f(x)|}{\|x' - x\|} \in [0, +\infty].$$

We say that  $f$  is Lipschitz continuous on  $A$  if  $\text{Lip}(f, A) < +\infty$ . If  $A = X$ ,  $f$  is said to be globally Lipschitz continuous. We say that  $f$  is locally Lipschitz continuous at  $\bar{x} \in X$  if there exists some neighborhood  $U$  of  $\bar{x}$  on which  $f$  is Lipschitz continuous. In that case, we will note  $\text{Lip}(f, \bar{x})$  instead of  $\text{Lip}(f, U)$ .

### 2.2.1 Fenchel analysis

Given a nonempty set  $\Omega \subset X$  and  $x \in \Omega$ , we define respectively the *radial tangent cone* and the *admissible tangent cone* by

$$\begin{aligned} T_{\Omega}^r(x) &:= \mathbb{R}_+(\Omega - x) = \{d \in X \mid \exists t > 0, x + td \in \Omega\}, \\ T_{\Omega}^{ad}(x) &:= \{d \in X \mid \exists \epsilon > 0, \forall t \in ]0, \epsilon[, x + td \in \Omega\}. \end{aligned}$$

An equivalent way to define the radial cone  $T_{\Omega}^r(x)$  is to see it as the cone hull of  $\Omega - x$ :

$$T_{\Omega}^r(x) = \mathbb{R}_+(\Omega - x).$$

According to Proposition 2.1.7 there is no particular reason for  $T_{\Omega}^r(x)$  to be closed, since  $0 \in \Omega - x$ .

#### Example 2.2.1.

- Let  $\Omega = \mathbb{R}_+^2 \cup \mathbb{R}_-^2 \subset \mathbb{R}^2$ , then  $T_{\Omega}^r((0,0)) = T_{\Omega}^{ad}((0,0)) = \Omega$ . In particular these cones can fail to be convex.
- Let  $\Omega = \{(x, y) \in \mathbb{R}^2 \mid (x-1)^2 + (y)^2 \leq 1\}$  be a disc in the plane. Then  $T_{\Omega}^r((0,0)) = T_{\Omega}^{ad}((0,0)) = \{(x, y) \in \mathbb{R}^2 \mid x > 0\}$ . We see here that these cones are not closed in general, even if  $\Omega$  is closed.
- Let  $\Omega = \{(x, y) \in \mathbb{R}^2 \mid (x-1)^2 + (y)^2 = 1\}$  be a circle in the plane. Then  $T_{\Omega}^r((0,0)) = \{(x, y) \in \mathbb{R}^2 \mid x > 0\}$ , while  $T_{\Omega}^{ad}((0,0)) = \emptyset$ . Hence, these cones doesn't coincide in general.

It is clear that  $T_{\Omega}^{ad}(x) \subset T_{\Omega}^r(x)$  always holds, and the equality between these cones will characterize a geometrical property of  $\Omega$ , as shown below (see Proposition A.2.1 in Appendix A.2). We say that  $\Omega$  is *radial (or absorbing) at  $x$*  if for all  $y \in \Omega$ , there exists a neighborhood  $U$  of  $x$  such that  $[x, y] \cap U \subset \Omega$  holds.



**Proposition 2.2.2.** Let  $x \in \Omega \subset X$ . Then  $T_\Omega^{ad}(x) = T_\Omega^r(x)$  if and only if  $\Omega$  is radial at  $x$ . In particular  $T_C^{ad}(x) = T_C^r(x)$  for all  $x \in C$  when  $C$  is convex.

Given the radial cone, we define the *Fenchel tangent cone*  $T_\Omega(x)$  as the closure of the radial cone  $T_\Omega^r(x)$ , with respect to the norm topology of  $X$ . In other words,

$$T_\Omega(x) := \text{cl } \mathbb{R}_+(\Omega - x).$$

We also define the *Fenchel normal cone*  $N_\Omega(x)$  as the polar cone of  $T_\Omega(x)$ , which is convex and weakly\* closed. It is quite immediate from the definitions to see that this normal cone admits the equivalent formulation, involving an obtuse angle property:

$$(2.8) \quad N_\Omega(x) = \{x^* \in X^* \mid \forall y \in \Omega, \langle x^*, y - x \rangle \leq 0\}.$$

Consider now a function  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ . Given  $x \in \text{dom } f$  and  $d \in X$ , we define the *lower Dini directional derivative* of  $f$  at  $x$  in the direction  $d$  by

$$d^D f(x; d) := \liminf_{t \downarrow 0} \frac{f(x + td) - f(x)}{t},$$

which can possibly take infinite values.

We are now in position to define the *Fenchel subdifferential* of  $f$  at  $x \in \text{dom } f$ . For this, consider the following sets:

- i)  $\{x^* \in X^* \mid \forall y \in X, f(y) - f(x) - \langle x^*, y - x \rangle \geq 0\}$ ,
- ii)  $\{x^* \in X^* \mid x \text{ is a global minimum of } y \mapsto f(y) - \langle x^*, y \rangle\}$ ,
- iii)  $\{x^* \in X^* \mid (x^*, -1) \in N_{\text{epi } f}(x, f(x))\}$ ,
- iv)  $\{x^* \in X^* \mid \forall d \in X, \langle x^*, d \rangle \leq d^D f(x; d)\}$ .

Hence the first three sets always coincide, and we call it the *convex subdifferential* of  $f$  at  $x$ , noted  $\partial f(x)$ . Moreover, the fourth set always contains  $\partial f(x)$ , with equality when  $f$  is convex (see Appendix A.2). We pose  $\partial f(x) = \emptyset$  when  $x \notin \text{dom } f$ , and define the domain of  $\partial f$  as

$$\text{dom } \partial f := \{x \in X \mid \partial f(x) \neq \emptyset\}.$$

**Remark 2.2.3.** These different presentations of  $\partial f(x)$  reunify almost all of the different classical approaches to build a subdifferential in a nonconvex setting. In general, we will replace a global property (characteristic of convex functions) by a local one.

The first set describes the elements of  $\partial f(x)$  as the slope of some global exact affine minorant. In the nonconvex setting, we will drop the global assumption and ask for some local/asymptotically affine minorant (Fréchet [244], Lipschitz-Smooth [36] subdifferentials). Instead of asking for *exact* affine minorant, we can also consider affine minorant, which are exact up to a small constant  $\epsilon$ , leading to the so-called approximate subdifferentials (approximate Fréchet [244], Hadamard subdifferentials[201]).

In the second set, we can see the viscosity approach: we take  $x^*$  as a subgradient at  $x$  if and only if there exists some ‘smooth’ function  $g$  defined in the neighbourhood of  $x$  satisfying  $\nabla g(x) = x^*$  and  $x$  is a local minimum of  $f - g$ . For the convex subdifferential, we restrict ourselves to linear functions, but with this general approach we could consider a broader class of functions. The most important point is the following: what do we mean by ‘smooth’ function? We can take for definition the Fréchet differentiability, or the Gateaux differentiability, but more generally we can consider differentiability relatively to a given *bornology*. This construction leads to the

family of the so-called viscosity subdifferentials, sometimes called bornology subdifferentials or  $\beta$ -subdifferentials (in particular the Fréchet or Gateaux subdifferentials, see [202]).

The third set corresponds to what we call a *geometric* subdifferential. We built a geometric tool which gives local informations of a set, namely a normal cone, and we derive local property of  $f$  by looking at the local property of its epigraph in  $X \times \mathbb{R}$ . Here the convex normal cone is built as the polar cone of a tangent cone, but one can consider other ways to define it, in particular if one wants a normal cone not necessarily convex. It can be defined through a local obtuse angle property like in (2.8) (see the Fréchet or Lipschitz-Smooth normals), or using the subdifferential of the distance function (see the Clarke-Rockafellar normals [111], and also the proximal normals in Hilbert spaces [202]).

The fourth set corresponds to what could be called an *analytic* subdifferential. We compute some directional derivative and take  $\partial f(x)$  as the set of its linear minorants. See the family of Dini derivative [36], Hadamard/Bouligand derivative [202], or the Clarke-Rockafellar derivative [111, 38].

From its definition, the Fenchel subdifferential is clearly convex and weakly\* closed, and satisfies the following Fermat's rule:

$$x \text{ is a global minimum of } f \Leftrightarrow 0 \in \partial f(x).$$

Moreover, the Fenchel subdifferential is consistent with the Fenchel normal cone. Indeed, for any set  $\Omega \subset X$  and  $x \in X$ , the Fenchel normal cone  $N_\Omega(x)$  is exactly  $\partial \delta_\Omega(x)$ , the Fenchel subdifferential of  $\delta_\Omega$  the indicator function of  $\Omega$ . It is also an easy exercise to show the following closure properties of the graph of  $\partial f : X \rightrightarrows X^*$ .

**Proposition 2.2.4.** Let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be lower semi-continuous at  $x \in X$ . Let  $(x_\alpha, x_\alpha^*)_{\alpha \in \mathcal{A}}$  be a bounded net in  $X \times X^*$  such that

$$x_\alpha \xrightarrow{\|\cdot\|} x \text{ and } x_\alpha^* \xrightarrow{w^*} x^* \text{ with } x_\alpha^* \in \partial f(x_\alpha).$$

Then  $x^* \in \partial f(x)$ . Suppose now that  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  is weakly lower semi-continuous at  $x \in X$ . If  $(x_\alpha, x_\alpha^*)_{\alpha \in \mathcal{A}}$  is a bounded net in  $X \times X^*$  such that

$$x_\alpha \xrightarrow{w} x \text{ and } x_\alpha^* \xrightarrow{\|\cdot\|} x^* \text{ with } x_\alpha^* \in \partial f(x_\alpha),$$

then  $x^* \in \partial f(x)$ .

**Remark 2.2.5.** In the nonconvex setting, a common practical consists in enlarging a given subdifferential by closing its graph, sequentially or topologically (see for instance the limiting Fréchet/Mordukhovitch subdifferential [244], or the limiting Dini/Approximate subdifferential [201]). These limiting subdifferentials are in general more robust and have better calculus rules, see [244, 263].

The Fenchel subdifferential is essentially useful if  $f$  is convex<sup>4</sup>, in which case it satisfies the most interesting properties. For instance if  $f$  is convex, then the Gateaux differentiability at  $\bar{x} \in \text{int dom } f$  implies

$$\partial f(x) = \{DF(x)\}.$$

The reverse is also true as soon as we assume  $f$  being continuous at  $x$  [325, Theorem 2.4.4, Corollary 2.4.10]. Still in the convex case, the Fenchel subdifferential enjoys the following sum rule.

<sup>4</sup>In fact, we can even show that the Fenchel subdifferential of  $f$  coincides on  $\text{dom } \partial f$  with the Fenchel subdifferential of its closed convex envelope, see [325, Theorem 2.4.1].

**Proposition 2.2.6** ([21, Theorem 9.5.4]). Let  $f, g : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be two proper lower semi-continuous convex functions. Then,

$$\text{for all } x \in X, \partial f(x) + \partial g(x) \subset \partial(f + g)(x).$$

If moreover  $f$  is continuous at some  $\bar{x} \in \text{dom } g$ , then

$$\partial f(\bar{x}) + \partial g(\bar{x}) = \partial(f + g)(\bar{x}).$$

In the convex case, we also have a nice relationship between the Fenchel subdifferential and the sublevel sets (see [263, Proposition 5.48] together with Proposition 2.1.7):

**Proposition 2.2.7.** Let  $f : X \rightarrow \mathbb{R}$  be a continuous convex function. Let  $\bar{x} \in X$  be such that  $0 \notin \partial f(\bar{x})$ , and note  $[f \leq f(\bar{x})] := \{x \in X \mid f(x) \leq f(\bar{x})\}$  be the sublevel set of  $f$  at  $\bar{x}$ . Then,

$$N_{[f \leq f(\bar{x})]}(\bar{x}) = \mathbb{R}_+ \partial f(\bar{x}).$$

**Remark 2.2.8.** The fact that the Fenchel subdifferential is normal to the sublevel sets is important in Hilbert spaces, since it helps to find descent directions. For instance, we can show that the element of minimal norm of  $-\partial f(\bar{x})$  is a descent direction for  $f$  at  $\bar{x}$  (see Chapter 5).

In the nonconvex setting, there has been some attempts to define a subdifferential satisfying this relationship with the sublevel sets. This is the approach in [75, 37, 38], and it is well adapted to the study of quasiconvex functions (whose sublevel sets are convex, see Section 2.2.5).

We will introduce in Chapter 5 a subdifferential for functions with values in ordered vector spaces, which satisfies this same nice property.

## 2.2.2 Bouligand analysis

Given a nonempty set  $\Omega \subset X$ , we define the *Bouligand* tangent cone (also called *contingent* tangent cone) by:

$$(2.9) \quad T_{\Omega}^B(x) := \{d \in X \mid \exists d_n \rightarrow d, \exists t_n \downarrow 0, x + t_n d_n \in \Omega\} \text{ if } x \in \Omega, T_{\Omega}^B(x) = \emptyset \text{ else.}$$

It is a closed cone (see Proposition A.2.4), but not necessarily convex<sup>5</sup>. We compare the Bouligand tangent cone to the previously seen radial, admissible and Fenchel tangent cones:

**Proposition 2.2.9.** Let  $x \in \Omega \subset X$ . Then:

- i)  $\text{cl } T_{\Omega}^{ad}(x) \subset T_{\Omega}^B(x) \subset \text{cl } T_{\Omega}^r(x) = T_{\Omega}(x)$ ,
- ii) If  $\Omega$  is radial at  $x$ , then the above inclusions become equalities. In particular  $T_{\Omega}^B(x) = T_{\Omega}(x)$ .

Given the Bouligand tangent cone, we can introduce the corresponding *Bouligand normal cone* by taking its polar cone:

$$(2.10) \quad N_{\Omega}^B(x) := \{x^* \in X^* \mid \langle x^*, d \rangle \leq 0 \ \forall d \in T_{\Omega}^B(x)\}.$$

It is clearly a weakly\* closed convex cone in  $X^*$ . Observe that, since  $T_{\Omega}^B(x)$  is not convex, the polar cone of  $N_{\Omega}^B(x)$  is generally strictly bigger than  $T_{\Omega}^B(x)$ .

Now, through a classic ‘geometric’ procedure using the normal cone to the epigraph, we define the *Bouligand subdifferential* of any function  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  at  $x \in X$  as:

$$(2.11) \quad \partial^B f(x) := \{x^* \in X^* \mid (x^*, -1) \in N_{\text{epi } f}^B(x, f(x))\} \text{ if } x \in \text{dom } f, \partial^B f(x) = \emptyset \text{ else.}$$

<sup>5</sup>Take  $\Omega = \mathbb{R}_+^2 \cup \mathbb{R}_-^2 \subset \mathbb{R}^2$ , such that  $T_{\Omega}^B((0,0)) = \Omega$ .

It is a weakly\* closed convex set, and its definition is consistent with the previously defined Bouligand normal cone, i.e.  $N_{\Omega}^B(x) = \partial^B \delta_{\Omega}(x)$  (see [263, Proposition 4.13]). Moreover, according to Proposition 2.2.9, the Bouligand subdifferential coincides with the Fenchel one whenever  $f$  is a convex function.

Define now the *Bouligand directional derivative* (also called *contingent*, *lower Hadamard* or *lower Dini-Hadamard* directional derivative) of  $f$  at  $x \in \text{dom } f$  in the direction  $d \in X$ :

$$(2.12) \quad d^B f(x; d) := \liminf_{d' \rightarrow d, t \downarrow 0} \frac{f(x + td') - f(x)}{t}.$$

From its definition,  $d^B f(x; \cdot)$  is lower semi-continuous, but not convex in general<sup>6</sup>. As we can expect from the notations, this partial derivative is directly related to the previous Bouligand subdifferential (see [263, Corollary 4.15]):

$$(2.13) \quad \partial^B f(x) = \{x^* \in X \mid \forall d \in X, \langle x^*, d \rangle \leq d^B f(x; d)\}.$$

**Remark 2.2.10.** The property  $\partial^B f(x) = \{x^* \in X \mid \forall d \in X, \langle x^*, d \rangle \leq d^B f(x; d)\}$  is generally called *geometric consistency* of the subdifferential. As a direct consequence, one has

$$(2.14) \quad \sup_{x^* \in \partial^B f(x)} \langle x^*, d \rangle \leq d^B f(x; d).$$

Equality instead of inequality in (2.14) would be called *analytic consistency*, but it does not hold in general. Take for instance  $f : x \in \mathbb{R} \mapsto -|x| \in \mathbb{R}$ , then  $d^B(0, d) = -|d|$  while  $\partial^B f(0) = \emptyset$ , so  $\sup_{x^* \in \partial^B f(0)} \langle x^*, d \rangle = -\infty$ . In the light of [5, Theorem 7.51], and since  $\partial^B f(x)$  is defined as the set of linear minorants of  $d^B f(x; \cdot)$ , one can see that analytic consistency holds if and only if  $d^B f(x; \cdot)$  is sublinear. See the example above where  $d^B f(x; \cdot)$  is not even convex.

The Bouligand subdifferential enjoys the following chain rule, which will be useful in our future study of dynamical systems (see Proposition A.2.5):

**Proposition 2.2.11.** Let  $u : I \subset \mathbb{R} \rightarrow X$  and  $f \circ u : I \rightarrow \mathbb{R} \cup \{+\infty\}$ , where  $I$  is an open interval of  $\mathbb{R}$ . Suppose that both  $u$  and  $f \circ u$  are derivable at  $t \in I$ , and that  $u(t) \in \text{dom } \partial^B f$ . Then,

$$(f \circ u)'(t) = \langle x^*, \dot{u}(t) \rangle, \quad \forall x^* \in \partial^B f(u(t)).$$

### 2.2.3 Fréchet analysis

Given a function  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ , its *Fréchet subdifferential* at  $\bar{x} \in \text{dom } f$  is defined as the set  $\partial^F f(\bar{x})$  of elements  $x^* \in X^*$  such that :

$$(2.15) \quad \liminf_{\substack{x \rightarrow \bar{x} \\ x \neq \bar{x}}} \frac{f(x) - f(\bar{x}) - \langle x^*, x - \bar{x} \rangle}{\|x - \bar{x}\|} \geq 0.$$

As usual its domain  $\text{dom } \partial^F f$  is the (possibly strict or empty) subset of  $\text{dom } f$  of points  $x$  such that  $\partial^F f(x)$  is nonempty. In the particular case of the indicator function of a closed set  $\Omega \subset H$ , we denote by  $N_{\Omega}^F := \partial^F \delta_{\Omega}$  the Fréchet normal cone of  $F$ , which reduces to

$$N_{\Omega}^F(\bar{x}) = \left\{ x^* \in X^* \mid \limsup_{\substack{x \xrightarrow{\Omega} \bar{x} \\ x \neq \bar{x}}} \frac{\langle x^*, x - \bar{x} \rangle}{\|x - \bar{x}\|} \leq 0 \right\}.$$

<sup>6</sup>Take for instance  $f : x \in \mathbb{R} \mapsto -|x| \in \mathbb{R}$ , such that  $d^B(0, d) = -|d|$ .

For all  $x \in \text{dom } f$ ,  $\partial^F f(x)$  is convex and closed for the norm topology, maybe empty<sup>7</sup>.

It is easy to see from the definitions that the Fenchel subdifferential  $\partial f$  is included in the Fréchet one  $\partial^F f$ , and equality holds between them if  $f$  is convex [244, Theorem 1.93]. Now, by taking  $x = \bar{x} + td'$  in (2.15), we easily obtain the following dual relation between the Fréchet subdifferential and Bouligand derivative:

$$(2.16) \quad \text{for all } d \in X, \quad \sup_{x^* \in \partial^F f(\bar{x})} \langle x^*, d \rangle \leq d^B f(x; d).$$

We deduce directly from this inequality and Proposition 2.13 that, for all  $x \in X$ ,  $\partial^F f(x) \subset \partial^B f(x)$ . In fact, Fréchet and Bouligand subdifferential coincide if  $X$  has finite dimension<sup>8</sup> [290, Proposition 6.5].

Let us give some calculus rules for this Fréchet subdifferential. We first observe that this subdifferential is related to the Fréchet differentiability. Indeed, for  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  Fréchet differentiable at  $\bar{x}$  we have ([212, Proposition 1.1])

$$\partial^F f(\bar{x}) = \{Df(\bar{x})\}.$$

We now look at a sum rule. Given two functions  $f, g : X \rightarrow \mathbb{R} \cup \{+\infty\}$ , we have for all  $\bar{x} \in X$

$$\partial^F(f + g)(\bar{x}) \supset \partial^F f(\bar{x}) + \partial^F g(\bar{x}).$$

This comes immediately from the definition of  $\partial^F$  (see also [263, Proposition 4.34]). If we assume moreover that  $f$  is Fréchet differentiable at  $\bar{x} \in \text{dom } g$ , then we have equality ([263, Corollary 4.35])

$$\partial^F(f + g)(\bar{x}) = \{Df(\bar{x})\} + \partial^F g(\bar{x}).$$

We introduce now the *limiting Fréchet subdifferential* (or just *limiting subdifferential*) by taking the sequential closure of the graph of  $\partial^F f$ . More precisely, given  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ , it consists in the set  $\partial^L f(x)$  of elements  $x^* \in X^*$  for which there exists:

- a sequence  $(x_n)_{n \in \mathbb{N}}$  in  $X^*$  such that  $x_n \xrightarrow[n \rightarrow +\infty]{f} x$  with  $x_n \in \text{dom } \partial^F f$ ,
- a sequence  $(x_n^*)_{n \in \mathbb{N}}$  in  $X^*$  such that  $x_n^* \xrightarrow[n \rightarrow +\infty]{w^*} x^*$  with  $x_n^* \in \partial^F f(x_n)$ .

We used in this definition the notation  $x_n \xrightarrow[n \rightarrow +\infty]{f} x$ , which denotes the  $f$ -attentive convergence of  $x_n$  to  $x$ , and means that  $x_n \xrightarrow[n \rightarrow +\infty]{\|\cdot\|} x$ , together with  $f(x_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{R}} f(x)$ . As previously, we define the domain  $\text{dom } \partial^L$ , and the related *limiting normal cone*  $N_\Omega^L := \partial^L \delta_\Omega$ . The limiting subdifferential  $\partial^L f(x)$  is not necessarily weakly\* closed<sup>9</sup> neither convex<sup>10</sup>. Clearly, the Fréchet subdifferential is included in the limiting one. Using Proposition 2.2.4, we also see that, if  $f$  is a lower semi-continuous convex function, then its limiting subdifferential equals to the Fenchel subdifferential.

**Remark 2.2.12.** It is important to notice that the Limiting subdifferential introduced here is not the same than the one used for instance in the book of Mordukhovich [244]. In that case, it involves *approximate* Fréchet subgradients in its definition, but both constructions coincide in reflexive spaces [244, Theorem 2.34], which is far enough for our needs in Part I.

<sup>7</sup>Consider for instance  $f(x) = -|x|$ , for which  $\partial^F f(0) = \emptyset$ .

<sup>8</sup>In fact, in reflexive spaces, the Fréchet subdifferential can be seen as a *weakly Bouligand subdifferential*, since it is exactly the set of linear minorants of a weak Bouligand derivative  $d^B w f(x; d)$ , see [212, Proposition 1.17] or [244, Theorem 1.10].

<sup>9</sup>See [244, Example 1.7].

<sup>10</sup>Take  $f(x) = -|x|$ , whose limiting subdifferential at zero is  $\partial^L f(x) = \{-1; +1\}$ .

If  $f$  is Fréchet differentiable at  $\bar{x}$ , then  $DF(\bar{x}) \in \partial^L f(\bar{x})$ , but the limiting subdifferential can be strictly larger (see [290, p. 304]). This is essentially because  $\partial^L f(\bar{x})$  “collects” some information on  $f$  *around*  $\bar{x}$  and not *at*  $\bar{x}$ . The strict Fréchet differentiability of  $f$  at  $\bar{x}$  guarantees that  $\partial^L f(\bar{x}) = \{DF(\bar{x})\}$  (see [244, Corollary 1.82]). In the same spirit, we have the following sum rule:

**Proposition 2.2.13** ([244, Proposition 1.107]). Let  $f, g : X \rightarrow \mathbb{R} \cup \{+\infty\}$ , with  $X$  being reflexive. Assume that  $f$  is strictly Fréchet differentiable at  $\bar{x} \in \text{dom } g$ . Then,

$$\partial^L(f + g)(\bar{x}) = \{Df(\bar{x})\} + \partial^L g(\bar{x}).$$

## 2.2.4 Clarke analysis

Given a locally Lipschitz continuous function  $f : X \rightarrow \mathbb{R}$ , its *Clarke directional derivative* at  $\bar{x} \in X$  in the direction  $d \in X$ , is given by

$$(2.17) \quad d^C f(\bar{x}; d) = \limsup_{x \rightarrow \bar{x}, t \downarrow 0} \frac{f(x + td) - f(x)}{t}.$$

Since  $f$  is locally Lipschitz continuous, for each  $\bar{x} \in X$  the function  $d^C f(\bar{x}; \cdot) : X \rightarrow \mathbb{R}$  is globally Lipschitz continuous. In particular the Clarke directional derivative has always finite values in this context. Moreover,  $d^C f(\bar{x}; \cdot)$  is sublinear and, in particular, convex, and  $d^C f(\cdot; \cdot) : X \times X \rightarrow \mathbb{R}$  is upper semi-continuous [111, Proposition 2.1.1].

As a dual notion, let us introduce the *Clarke subdifferential*  $\partial^C f(x)$  which is the subset of  $X^*$  defined for any  $\bar{x} \in X$  by

$$(2.18) \quad \partial^C f(\bar{x}) = \{x^* \in X^* \mid \langle x^*, d \rangle \leq d^C f(\bar{x}; d) \quad \forall d \in X\}.$$

From (2.18) it is easily verified that this is a weakly\* compact convex set, and Hahn-Banach’s separation theorem guarantees that it is nonempty [111, Proposition 2.1.2]. Once again, if  $f$  is convex it reduces to the Fenchel subdifferential [111, Proposition 2.2.7]. It must be noticed that the Clarke directional derivative can be recovered from the Clarke subdifferential, since

$$(2.19) \quad d^C f(\bar{x}; d) = \max \{\langle x^*, d \rangle \mid x^* \in \partial^C f(\bar{x})\}.$$

This *max formula* can be easily deduced from Theorem 2.1.4, see also Remark 2.2.10.

Let us give some calculus rules for the Clarke subdifferential. First of all, using basic properties of the lim sup, we see that for two locally Lipschitz functions  $f, g : X \rightarrow \mathbb{R}$ , it always holds

$$(2.20) \quad \text{for all } \bar{x} \in X \text{ and all } d \in X, \quad d^C(f + g)(\bar{x}; d) \leq d^C f(\bar{x}; d) + d^C g(\bar{x}; d).$$

By duality in (2.18), we also have

$$(2.21) \quad \text{for all } \bar{x} \in X, \quad \partial^C(f + g)(\bar{x}) \subset \partial^C f(\bar{x}) + \partial^C g(\bar{x}).$$

Similarly to the limiting subdifferential, the Clarke subdifferential is sensitive to the strict Gateaux differentiability [111, Proposition 2.2.4]:

$$\partial^C f(\bar{x}) = \{x^*\} \Leftrightarrow f \text{ is strictly Gateaux differentiable at } \bar{x} \text{ and } Df(\bar{x}) = x^*.$$

The Clarke subdifferential enjoys also an exact sum rule under a strict differentiability assumption. If  $f, g : X \rightarrow \mathbb{R}$  are two locally Lipschitz functions such that  $f$  is strictly Gateaux differentiable at  $\bar{x} \in X$ , then [111, Proposition 2.3.3]:

$$(2.22) \quad \partial^C(f + g)(\bar{x}) = \{Df(\bar{x})\} + \partial^C g(\bar{x}).$$

### 2.2.5 A last few things

Let us make some links between the subdifferentials we presented below. First note that, if  $f : X \rightarrow \mathbb{R}$  is locally Lipschitz continuous, we have for all  $\bar{x} \in X$ , and all  $d \in X$ ,

$$\liminf_{\substack{d' \rightarrow d \\ t \downarrow 0}} \frac{f(\bar{x} + td') - f(\bar{x})}{t} = \liminf_{t \downarrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t},$$

in other words,

$$(2.23) \quad d^B f(\bar{x}; d) = d^D f(\bar{x}; d).$$

Since we can easily compare lower Dini derivatives and Clarke derivatives, we deduce using (2.18) and (2.13) that in the locally Lipschitz case,

$$(2.24) \quad d^B f(\bar{x}; \cdot) \leq d^C f(\bar{x}; \cdot) \text{ and } \partial^F f(\bar{x}) \subset \partial^B f(\bar{x}) \subset \partial^C f(\bar{x}).$$

In the Lipschitz case, we can also bound these subdifferentials [111, Proposition 2.1.1]:

**Proposition 2.2.14.** Let  $f : X \rightarrow \mathbb{R}$  be locally Lipschitz continuous at  $\bar{x}$ . Then

$$\partial^F f(\bar{x}) \subset \partial^B f(\bar{x}) \subset \partial^C f(\bar{x}) \subset \text{Lip}(f, \bar{x})\mathbb{B}.$$

We say that a locally Lipschitz continuous function  $f$  is Clarke regular at  $\bar{x}$  if  $f$  is directionally derivable at  $\bar{x}$ , and for all  $d \in X$ ,

$$Df(\bar{x}; d) = d^C f(\bar{x}; d).$$

Using (2.23) and (2.24), we see that  $f$  is Clarke regular at  $\bar{x}$  if and only if

$$d^B f(\bar{x}; \cdot) = d^C f(\bar{x}; \cdot),$$

which is also equivalent to

$$\partial^B f(\bar{x}) = \partial^C f(\bar{x}).$$

We easily see in this Lipschitz case that strictly Gateaux differentiable functions are Clarke regular. But it is also true for convex functions, and any positive linear combination of Clarke regular functions [111, Proposition 2.3.6].

Now, let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be an arbitrary function. In [307, Theorem 6.1], Treiman showed<sup>11</sup> that for all  $\bar{x} \in X$ ,

$$(2.25) \quad \overline{\text{co}}^* \partial^L f(\bar{x}) \subset \partial^C f(\bar{x}),$$

the inclusion being an equality if  $X$  is reflexive. This implies in particular that

$$\partial^F f(\bar{x}) \subset \partial^L f(\bar{x}) \subset \partial^C f(\bar{x}).$$

An other concept of regularity is the *lower regularity*. We say that a function  $f : X \rightarrow \mathbb{R}$  is lower regular at  $\bar{x}$  if  $\partial^F f(\bar{x}) = \partial^L f(\bar{x})$ . If  $X$  is reflexive and  $f$  locally Lipschitz, with the representation formula (2.25) we can deduce that the lower regularity at  $\bar{x}$  is equivalent to ask  $\partial^F f(\bar{x}) = \partial^C f(\bar{x})$ . In that case, lower regularity implies Clarke regularity. Moreover, if  $X$  has finite dimension, lower regularity and Clarke regularity coincide, because Bouligand and Fréchet subdifferentials both coincide.

<sup>11</sup>Note that Treiman defines the limiting subdifferential through  $\varepsilon$ -Fréchet subgradients, which gives a bigger set than our, so the inclusion remains valid. For the equality in reflexive spaces, recall Remark 2.2.12.

Equivalently, a closed set  $\Omega \subset H$  is said *lower regular* at  $\bar{x}$  if  $N_{\Omega}^F(\bar{x}) = N_{\Omega}^L(\bar{x})$ . For instance (still with  $X$  being reflexive),  $C^2$  manifolds, prox-regular sets, strictly Fréchet differentiable functions, primal-lower-nice functions [118, Corollary 3.1] are lower regular.

It follows directly from its definition that the Fréchet subdifferential enjoys this Fermat's rule:

$$\text{if } \bar{x} \text{ is a local minimum of } f, \text{ then } 0 \in \partial^F f(\bar{x}).$$

Of course, since  $\partial^F$  is included in all other subdifferentials, the limiting and Clarke subdifferentials enjoy also the same Fermat's rule.

We end this section with a few words about quasi-convex functions. A function  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  is said to be *quasi-convex* if for all  $x \in \text{dom } f$ , the sublevel set  $[f \leq f(x)]$  is convex. Equivalently,

$$\forall x, x' \in \text{dom } f, \forall t \in [0, 1], f(tx + (1-t)x') \leq \max\{f(x), f(x')\}$$

These functions enjoy the following nice property, which generalizes a common fact for convex functions (see Appendix A.2):

**Proposition 2.2.15.** Let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be a quasi-convex function such that  $f(x_2) \leq f(x_1)$ . Then  $\langle x^*, x_2 - x_1 \rangle \leq 0$  for all  $x^* \in \partial^B f(x_1) \supset \partial^F f(x_1)$ .

A similar property exists for the limiting and Clarke subdifferentials, under a slightly stricter assumption:

**Proposition 2.2.16.** Let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be a quasi-convex function such that  $f(x_2) < f(x_1)$ . Then  $\langle x^*, x_2 - x_1 \rangle \leq 0$  for all  $x^* \in \partial^L f(x_1)$ . If  $f$  is locally Lipschitz then  $\partial^L$  can be replaced by  $\partial^C$ .

## 2.3 Optimization for vector-valued functions

### 2.3.1 Ordered space

We say that a relation  $\mathcal{R}$  on  $X$  is

- *reflexive* if  $\{x \in X \mid x\mathcal{R}x\} = X$ ,
- *irreflexive* if  $\{x \in X \mid x\mathcal{R}x\} = \emptyset$ ,
- *transitive* if  $\forall x, y, z \in X, x\mathcal{R}y$  and  $y\mathcal{R}z$  implies  $x\mathcal{R}z$ ,
- *antisymmetric* if  $\forall x, y \in X, x\mathcal{R}y$  and  $y\mathcal{R}x$  implies  $x = y$ .

We call *order* a reflexive and transitive relation<sup>12</sup>, *strict order* an irreflexive and transitive relation. Observe that each order  $\leq$  induces a strict order  $\preceq$  defined by  $x \preceq y$  iff  $x \leq y$  and  $x \neq y$ . In a similar way, each strict order  $<$  induces an order  $\preceq$  defined by  $x \preceq y$  iff  $x < y$  and  $x = y$ .

If  $X$  is a vector space, we say that an order  $\leq$  *respects the vectorial structure* of  $X$  if :

- $\forall x_1, x_2, y_1, y_2 \in X, x_1 \leq y_1$  and  $x_2 \leq y_2$  implies  $x_1 + x_2 \leq y_1 + y_2$ ,
- $\forall x, y \in X, \lambda \in \mathbb{R}_+, x \leq y$  implies  $\lambda x \leq \lambda y$ .

<sup>12</sup>In some books, reflexive transitive relations are called preorders, and antisymmetry is asked to be an order.



In the same way, we say that a strict order  $<$  respects the vectorial structure of  $X$  whether it holds for the associated order  $\preceq$ . In the following, we will *always* assume that the considered orders (resp. strict orders) respect the vectorial structure of  $X$ , without mentioning it.

In vector spaces, there is a natural correspondence between orders and convex cones. Consider the following :

$$\text{Given a relation } \mathcal{R}, \text{ define } K_{\mathcal{R}} := \{x \in X \mid 0\mathcal{R}x\},$$

$$\text{Given a set } K, \text{ define } x\mathcal{R}_K y \Leftrightarrow y - x \in K.$$

Then, it can be easily verified that the above constructions define a one-to-one correspondence between the orders (resp. strict orders) on  $X$  and the convex cones (resp. convex strict cones) of  $X$ . This is why we will say in the following that  $X$  is ordered by a given convex cone  $K$ , referring implicitly to the corresponding order  $\leq_K$ , and will note this ordered vector space by the couple  $(X, K)$ .

### Example 2.3.1.

- Let  $X = \mathbb{R}$ , equipped with the usual order  $\leq$ . It is characterized by the convex cone of positive reals  $\mathbb{R}_+$ . The corresponding strict order  $<$  is characterized by the convex strict cone of strictly positive reals  $\mathbb{R}_{++}$ . Note that there is no other orders in  $\mathbb{R}$  than the ones corresponding to the cones  $\mathbb{R}_+$ ,  $\mathbb{R}_-$  and  $\{0\}$ .
- Let  $X = \mathbb{R}^m$ ,  $m \in \mathbb{N}^*$ . It is usually ordered by  $a \leq b \Leftrightarrow a_i \leq b_i$  in  $\mathbb{R}$ , for all  $i \in \{1, \dots, m\}$ . The corresponding convex cone is the *positive orthant*  $\mathbb{R}_+^m := (\mathbb{R}_+)^m$ . A common strict order on  $\mathbb{R}^m$  is  $a < b$  iff  $a_i < b_i$  in  $\mathbb{R}$ , for all  $i \in \{1, \dots, m\}$ . The corresponding convex strict cone is  $\mathbb{R}_{++}^m := (\mathbb{R}_{++})^m = \text{int } \mathbb{R}_+^m$ .  $\mathbb{R}^m$  can be equipped with other orders, like the *lexicographical order*, which is characterized when  $m = 2$  by the convex cone  $\mathbb{R}_{++}^2 \cup \{0\} \times \mathbb{R}_+$ . Let us also mention the order induced by the second order cone in  $\mathbb{R}^{m+1}$ , which is just the epigraph of the euclidean norm of  $\mathbb{R}^m$ .
- Let  $X = \mathcal{S}^m(\mathbb{R})$ , the vector space of real symmetric matrices of size  $m \in \mathbb{N}^*$ . In general, it is ordered by the convex cone  $\mathcal{S}_+^m(\mathbb{R})$  of *positive symmetric matrices*, which consists in symmetric positive matrices.
- Consider some set  $\Omega$  and  $X = \mathcal{F}(\Omega, \mathbb{R})$  the vector space of functions on  $\Omega$  with real values. It is ordered by the convex cone of *positive functions*  $\mathcal{F}_+(\Omega, \mathbb{R}) := \{f \in \mathcal{F}(\Omega, \mathbb{R}) \mid f(x) \geq 0 \forall x \in \Omega\}$ . It induces a similar order on any subspace  $Y$  of  $\mathcal{F}(\Omega, \mathbb{R})$  by taking the intersection  $Y \cap \mathcal{F}_+(\Omega, \mathbb{R})$ .
- Suppose now that  $(\Omega, \Sigma, \mu)$  is a  $\sigma$ -finite measured space and consider  $\mathcal{L}^0(\Omega, \Sigma)$  the space of measurable functions on  $\Omega$  with real values. We can quotient this space by the equivalence relation  $f \sim g \Leftrightarrow f(x) = g(x)$  for  $\mu$ -a.e.  $x \in \Omega$ , to obtain the vector space  $L^0(\Omega, \Sigma, \mu)$ . It is ordered by the convex cone of  $\mu$ -almost positive functions  $L_+^0(\Omega, \Sigma, \mu) := \{f \in \mathcal{L}^0(\Omega, \Sigma, \mu) \mid f(x) \geq 0 \text{ for } \mu\text{-a.e. } x \in \Omega\}$ . Here again, this cone induces a similar order on any subspace, including all  $L^p(\Omega, \Sigma, \mu)$  spaces, for  $p \in [1, +\infty]$ .

Of course, this list is non exhaustive, and we could also have define positive convex cones in spaces of measures for instance. But it contains the ordered spaces which will be the most interesting for us in the sequel.  $\mathbb{R}_+^m$  will be our first object of study, because it is particularly easy to handle and benefits from good “finite” properties : it is a finite-dimensional object, and it is polyhedral.  $(\mathcal{S}^m(\mathbb{R}), \mathcal{S}_+^m(\mathbb{R}))$  is also finite-dimensional, but has a more complicated geometrical nature: it is for instance not polyhedral. Finally,  $(L^\infty(\Omega, \Sigma, \mu), L_+^\infty(\Omega, \Sigma, \mu))$  is an example of a “good” infinite-dimensional ordered space since, as we will see later,  $L_+^\infty(\Omega, \Sigma, \mu)$  has a nonempty interior.

### 2.3.2 Base of a cone

We focus now on the concept of base for a cone. Its interest is that the knowledge of the order structure on  $(X, K)$  is equivalent to know a base of the dual cone  $K^+$  (see Theorem 2.3.7 below).

**Definition 2.3.2.** Let  $S \subset (X, \tau)$  be a nonempty strict cone. We say that  $\Theta \subset S$  is a *base* of  $S$  if for all  $d \in S$ , there exists a unique  $(\lambda, \theta) \in \mathbb{R}_{++} \times \Theta$  such that  $d = \lambda\theta$ . We say that  $\Theta$  is a base of a cone  $K$ , if  $\Theta$  is a base of  $K \setminus \{0\}$ .

Observe that, by definition, a base do not contain the origin. Note also that any cone admits a base, as a direct consequence of the axiom of choice. We give now some examples.

#### Example 2.3.3.

- Let  $(X, \|\cdot\|)$  be a normed space, and note  $\mathbb{S}_X$  its unit sphere. Then  $K \cap \mathbb{S}_X$  is a base for any cone  $K \subset X$ .
- $\mathbb{R}_+^m$  admits as a base the unit simplex  $\Delta_m := \{\theta \in \mathbb{R}_+^m \mid \sum_{i=1}^m \theta_i = 1\}$ .
- $\mathcal{S}_+^m(\mathbb{R})$  admits as a base the set  $\{M \in \mathcal{S}_+^m(\mathbb{R}) \mid \text{tr}(M) = 1\}$ .
- Let  $(\Omega, \Sigma, \mu)$  be a finite measured space and consider  $L_+^p(\Omega, \Sigma, \mu)$  for  $p \in [1, +\infty]$ . It admits as a base  $\{f \in L_+^p(\Omega, \Sigma, \mu) \mid \int_{\Omega} f \, d\mu = 1\}$ .

Note that the bases presented here share a common structure : it is the intersection of a cone  $K$  with the level set of some function. The first example deals with the norm while in the three others we intersect with an hyperplane. It can be also noticed that the last base can equivalently be written as the intersection of  $L_+^p(\Omega, \Sigma, \mu)$  with the unit sphere of  $L^1(\Omega, \Sigma, \mu)$ .

In Part II, we will need for our analysis to work with a weakly\* compact and convex base of  $K^+$ . We follow then with some necessary and sufficient conditions on  $K$  for the existence of such a base (see Appendices A.3 for a proof).

**Theorem 2.3.4.** *Let  $K$  be a closed convex cone in a Banach space  $X$ . Then  $K$  has a nonempty interior if and only if  $K^+$  admits a convex weakly\* compact base  $\Theta$ . In that case, there exists  $e \in \text{int } K$  such that*

$$\Theta = \{x^* \in K^+ \mid \langle x^*, e \rangle = 1\}.$$

**Example 2.3.5.** Let us briefly convince ourselves that this result is optimal. Take in  $X = \mathbb{R}^2$  the closed convex cone  $K = \mathbb{R}_+ \times \{0\}$ , which has an empty interior. Then  $K^+ = \mathbb{R}_+ \times \mathbb{R}$ , and it does not admit any convex base since such a base would contain the origin.

**Example 2.3.6.** We have already seen in Example 2.3.3 bases of convex cones built on the model  $\Theta = \{x^* \in K^+ \mid \langle x^*, e \rangle = 1\}$  (that we will note  $\Theta_e$  for short).

- The positive orthant  $\mathbb{R}_+^m \subset \mathbb{R}^m$  is self-dual for the euclidean scalar product, and  $\text{int } \mathbb{R}_+^m = \mathbb{R}_{++}^m$ . By taking  $e = (1, \dots, 1) \in \mathbb{R}_{++}^m$ , we recognize  $\Theta_e$  as the unit simplex  $\Delta_m$ . It is convex and compact.
- The cone of positive symmetric matrices is self-dual for the scalar product  $\langle X, Y \rangle = \text{tr } {}^tXY$ . By taking  $e = Id \in \text{int } \mathcal{S}_+^m(\mathbb{R}) = \mathcal{S}_{++}^m(\mathbb{R})$ , we recognize  $\Theta_e$  as the set of positive symmetric matrices of trace one. It is convex and compact.
- Let  $(\Omega, \Sigma, \mu)$  be a finite measured space and consider the Banach space  $L^p(\Omega, \Sigma, \mu)$  for  $p \in [1, +\infty[$ . Let  $q \in ]1, +\infty]$  be the conjugate<sup>13</sup> of  $p$ . The dual cone of  $L_+^p(\Omega, \Sigma, \mu)$  is

<sup>13</sup>That is, satisfying  $\frac{1}{p} + \frac{1}{q} = 1$  with the convention  $\frac{1}{+\infty} = 0$ .

$L_+^q(\Omega, \Sigma, \mu)$ . We consider  $e = \mathbb{1}_\Omega$ , i.e., the function being constant to 1 on all  $\Omega$ . It is clear that  $e \in L_+^p(\Omega, \Sigma, \mu)$ , since  $\Omega$  is finite and  $\langle e, f \rangle_{L^q \times L^p} = \int_\Omega f d\mu$  for all  $f \in L^p(\Omega, \Sigma, \mu)$ . Then we recover the base  $\Theta_e$  presented in Example 2.3.3. Nevertheless, note that it is not bounded, because the  $L^1$  ball is not bounded in such  $L^p$ . This is not surprising in the light of Theorem 2.3.4, since it is known that the interior of  $L_+^p(\Omega, \Sigma, \mu)$  is empty for  $p \in [1, +\infty[$ .

- Let  $(\Omega, \Sigma, \mu)$  be a finite measured space and consider the Banach space  $L^\infty(\Omega, \Sigma, \mu)$ . Its dual is  $ba(\Omega, \Sigma, \mu)$ , the Banach space of bounded additive measures which are absolutely continuous with  $\mu$  (see [5, Chapter 10.10]). The dual cone of  $L_+^\infty(\Omega, \Sigma, \mu)$  is  $ba_+(\Omega, \Sigma, \mu)$ , the convex cone of such positive measures. By taking again  $e = \mathbb{1}_\Omega$ , it is easy to see, for the same reasons as above, that  $e \in \text{int } L_+^\infty(\Omega, \Sigma, \mu)$ . It gives us  $\Theta_e = \{\pi \in ba_+(\Omega, \Sigma, \mu) \mid \pi(\Omega) = 1\}$ . This is nothing but the set of additive probability measures on  $\Omega$  absolutely continuous with respect to  $\mu$ , which is weakly\* compact and convex.

As announced, a base  $\Theta$  of the dual  $K^+$  controls the monotonic properties in  $(X, K)$ . In fact, the extreme points of  $\Theta$  are enough to characterize the monotonicity.

**Theorem 2.3.7.** *Let  $K$  be a closed convex cone with nonempty interior in a Banach space  $X$ . Let  $\Theta$  be a weakly\* compact convex base of a  $K^+$ , and let  $\Xi$  be the set of its extreme points. Note  $\leq$  the order on  $X$  induced by  $K$ . Then for all  $x, x' \in X$ :*

- i)  $x \leq x'$  if and only if  $\langle \theta, x \rangle \leq \langle \theta, x' \rangle$  for all  $\theta \in \Theta$ ,
- ii)  $x \leq x'$  if and only if  $\langle \xi, x \rangle \leq \langle \xi, x' \rangle$  for all  $\xi \in \Xi$ ,
- iii) If  $x \leq x'$  then  $\sigma_\Theta(x) \leq \sigma_\Theta(x')$ . (monotonicity)

Note  $<$  the strict order on  $X$  induced by  $\text{int } K$ . Then for all  $x, x' \in X$ :

- iv)  $x < x'$  if and only if  $\langle \theta, x \rangle < \langle \theta, x' \rangle$  for all  $\theta \in \Theta$ ,
- v)  $x < x'$  if and only if  $\langle \xi, x \rangle < \langle \xi, x' \rangle$  for all  $\xi \in \Xi$ ,
- vi) If  $x < x'$  then  $\sigma_\Theta(x) < \sigma_\Theta(x')$ . (strict monotonicity)

The proof of this result is left in the Appendix A.3. Let us give an exact formulation of  $\sigma_\Theta$  for some important examples :

**Example 2.3.8.**

- If  $X = \mathbb{R}^m$  and  $\Theta = \Delta_m$  the unit simplex, then it is easy to see that  $\sigma_\Theta$  is the max function, that is  $\sigma_\Theta(x) = \max_{i \in \{1, \dots, m\}} x_i$ .
- If  $X = \mathcal{S}^m(\mathbb{R})$  and  $\Theta = \{M \in \mathcal{S}_+^m(\mathbb{R}) \mid \text{tr}(M) = 1\}$ , then we can demonstrate that  $\sigma_\Theta(A)$  is the greatest eigenvalue of  $A \in \mathcal{S}_+^m$ , that we note  $\lambda_{\max}(A)$ . Indeed, by compactness of  $\Theta$ , there exists some  $M \in \Theta$  such that  $\sigma_\Theta(A) = \text{tr}(MA)$ . Using Ky Fan's inequality, we easily see that this is bounded from above by  $\lambda_{\max}(A)$ . On the other hand,  $\lambda_{\max}(A) = \langle Av, v \rangle$  for some eigenvector  $v \in \mathbb{R}^m$  such that  $\|v\| = 1$ . Since  $\langle Av, v \rangle = \text{tr}(v {}^t v A)$ , and  $v {}^t v \in \Theta$ , the conclusion follows.
- Take  $X = L^\infty(\Omega, \Sigma, \mu)$  as introduced in Example 2.3.6. Let  $\Theta = \{\lambda \in ba_+(\Omega, \Sigma, \mu) \mid \lambda(\Omega) = 1\}$ . Then we can show, analogously to the previous examples, that  $\sigma_\Theta$  corresponds to the essential supremum on  $L^\infty(\Omega, \Sigma, \mu)$ . For a given  $\phi \in L^\infty(\Omega, \Sigma, \mu)$ , use the weak\* compactness of  $\Theta$  to obtain a probability  $\pi \in \Theta$  such that  $\sigma_\Theta(\phi) = \langle \pi, \phi \rangle_{ba, L^\infty} := \int_\Omega \phi d\pi$ .

Since  $\phi$  is bounded from above by  $\text{supess } \phi$  for  $\mu$ -a.e.  $x \in \Omega$ , and  $\pi$  is absolutely continuous with respect to  $\mu$ , then  $\phi(x) \leq \text{supess } \phi$  for  $\pi$ -a.e.  $x \in \Omega$ . Moreover  $\pi(\Omega) = 1$ , then it follows that  $\int_{\Omega} \phi d\pi \leq \text{supess } \phi$ . On the other hand, first suppose that  $\phi$  reach its essential supremum on non  $\mu$ -negligeable  $A \subset \Omega$ . Then we can see that  $\pi := \frac{\mathbb{1}_A}{\mu(A)}$  lies in  $\Theta$  and verify  $\langle \pi, \phi \rangle_{ba, L^\infty} = \text{supess } \phi$ . In the general case, write  $\phi$  as the uniform limit of a sequence of step functions  $\phi_n$ . Each step function attains its essential supremum on a non negligible set and we can define  $\pi_n$  as previously such that  $\langle \pi_n, \phi_n \rangle_{ba, L^\infty} = \text{supess } \phi_n$ . Passing to the limit, using the uniform convergence of  $\phi_n$  in  $L^\infty(\Omega, \Sigma, \mu)$ , and the strong continuity of  $\text{supess}$ , gives  $\text{supess } \phi \leq \lim_n \sigma_{\Theta}(\phi_n)$ . Using the strong continuity of  $\sigma_{\Theta}$  gives the conclusion.

We end with a final result concerning the bases of polyhedral cones (see Appendix A.3)

**Theorem 2.3.9.** *Let  $K$  be a closed convex cone with nonempty interior in a finite-dimensional Banach space  $X$ . Let  $\Theta$  be a weakly\* compact convex base of a  $K^+$ . Then the following statements are equivalent:*

- i)  $K$  is polyhedral,
- ii)  $\Theta$  is a polytope,
- iii)  $\Theta$  has a finite number of extremal points.

### 2.3.3 Vector optimization problem

Let  $X$  and  $Y$  be two Banach spaces, and  $F : X \rightarrow Y$ . Suppose that  $Y$  is equipped with an order induced by a closed convex cone  $K$ , with nonempty interior. In the following,  $\leq$  will denote the order induced by  $K$ , and  $\preceq$  the associated strict order induced by  $K \setminus \{0_Y\}$ . Moreover, we will note  $<$  the strict order induced by the strict cone  $\text{int } K$ . We recall from Section 2.3.1 that this is equivalent to say that

- $y_1 \leq y_2$  if and only if  $y_2 - y_1 \in K$ ,
- $y_1 \preceq y_2$  if and only if  $y_2 - y_1 \in K \setminus \{0\}$ ,
- $y_1 < y_2$  if and only if  $y_2 - y_1 \in \text{int } K$ .

We consider a nonempty closed convex set  $C \subset X$ , which will model the constraints. The associated *vector optimization problem* ((VOP) for short) consists in solving

$$(VOP) \quad \text{MIN}_{x \in C} F(x),$$

where a precise meaning shall be given to  $\text{MIN } F$ .

Clearly, it would not be very useful to define a solution of (VOP) as a point  $\bar{x} \in C$  satisfying  $F(\bar{x}) \leq F(x)$  for all  $x \in C$ . Indeed, the order  $\leq$  is not total, and there is few chances that all the elements of  $C$  can be compared. Instead, we say that  $\bar{x}$  is an *efficient solution* of (VOP) (or that  $\bar{x}$  is an efficient point of  $F$  on  $C$ ) if there is no  $x \in C$  such that  $F(x) \preceq F(\bar{x})$ . We note  $\text{ARGMIN}_{x \in C} F$  the set of such efficient points. An other way to define the efficient points is to say that it is the points  $\bar{x} \in C$  such that

$$[F \preceq F(\bar{x})] \cap C = \emptyset,$$

where  $[F \preceq F(\bar{x})]$  denotes a *sublevel set* of  $F$  at  $\bar{x}$ , defined by

$$[F \preceq F(\bar{x})] := \{x \in X \mid F(x) \preceq F(\bar{x})\}.$$

If, instead of considering  $\preceq$ , we consider the strict order  $<$  induced by  $\text{int } K$ , we obtain a weaker notion of solution for (VOP). Indeed, we say that  $\bar{x}$  is a *weak efficient solution* of (VOP) (or that  $\bar{x}$  is a weak efficient point of  $F$  on  $C$ ) if there is no  $x \in C$  such that  $F(x) < F(\bar{x})$ . We note  $\text{ARGMIN}_{w, x \in C} F$  the set of such efficient points. Equivalently,  $\bar{x} \in C$  is weakly efficient if and only if

$$[F < F(\bar{x})] \cap C = \emptyset.$$

Since  $\text{int } K \subset K \setminus \{0\}$ , all efficient points are in particular weakly efficient, the converse being false in general. When facing a vector optimization problem, we usually look for its efficient points, but usual methods generally provide only weak efficient points (see Section 6.3 of Chapter 6 for a discussion on this question).

Observe that, in the scalar-valued case (i.e.  $(Y, K) = (\mathbb{R}, \mathbb{R}_+)$ ), both notions of efficiency and weak efficiency coincide with the usual notion of minimum. In other words, (VOP) reduces to a classic optimization problem. An other essential case is the *multi-objective* optimization problem ((MOP) for short), or *Pareto* optimization problem, where  $(Y, K) = (\mathbb{R}^m, \mathbb{R}_+^m)$ . In that case, there exists a finite family  $(f_1, \dots, f_m)$  of real-valued functions on  $X$  such that

$$\text{for all } x \in X, F(x) = (f_1(x), \dots, f_m(x)).$$

With this notation, the multi-objective optimization problem writes as

$$\text{(MOP)} \quad \text{MIN}_{x \in C} (f_1(x), \dots, f_m(x)),$$

and we can revisit the notions of efficiency (reps. weak efficiency). We say that a point  $\bar{x} \in C$  is:

- *Pareto efficient* if there is no  $x \in C$  such that  $f_i(x) \leq f_i(\bar{x})$  for all  $i \in \{1, \dots, m\}$ , and  $f_I(x) < f_I(\bar{x})$  for some  $I \in \{1, \dots, m\}$ ,
- *weakly Pareto efficient* if there is no  $x \in C$  such that  $f_i(x) < f_i(\bar{x})$  for all  $i \in \{1, \dots, m\}$ .

The guideline of this thesis is to use descent methods to solve optimization problems. For solving (VOP), the idea essentially reduces to the following: given a current state  $x \in C$ , can we improve the value  $F(x)$ ? That is, can we find some  $x' \in C$  such that  $F(x') \preceq F(x)$  (resp.  $F(x') < F(x)$ )? If so,  $x'$  becomes our new current state, otherwise  $x$  is by definition an efficient (resp. weakly efficient) solution of (VOP).

Then it is clear that being able to compare two points  $x, x'$  is essential, and that is why Proposition 2.3.7 is really important. Suppose that we choose some  $e \in \text{int } K$ , and take

$$\Theta = \{y^* \in K^+ \mid \langle y^*, e \rangle = 1\}.$$

According to Theorem 2.3.4 and Proposition 2.3.7, we have for all  $x, x' \in X$

$$(2.26) \quad F(x') \preceq F(x) \Leftrightarrow \forall \theta \in \Theta, (\theta \circ F)(x') \leq (\theta \circ F)(x),$$

where  $\theta \circ F : X \rightarrow \mathbb{R}$  is the composition between the linear form  $\theta \in \Theta \subset Y^*$  and  $F$ . We will generally use the notation  $f_\theta := \theta \circ F$ , and call these functions the *cost* functions associated<sup>14</sup>

<sup>14</sup>Of course these functions do not depend only on  $F$ , but also on  $K$  and the choice of a base for  $K^+$  (that is, according to Theorem 2.3.4, the choice of an element  $e \in \text{int } K$ ). But note that 1) we will always work with a fixed cone  $K$  and a fixed base  $\Theta$ , 2) the choice of a base for  $K^+$  has no influence on the main concepts which will be used later (see Chapter 5).

to  $F$ . What (2.26) means, is that controlling the descent of  $F$  is equivalent to control the simultaneous descent of its cost functions  $\{f_\theta\}_{\theta \in \Theta}$ . What is interesting in this approach is that we went back from vector values to scalar values, so we will be able to use the tools we are familiar with for scalar optimization.

Proposition 2.3.7 says also that it is not necessary to control all the cost functions  $\{f_\theta\}_{\theta \in \Theta}$ , but only the ones corresponding to  $\Xi$ , the set extreme points of  $\Theta$ :

$$(2.27) \quad F(x') \leq F(x) \Leftrightarrow \forall \xi \in \Xi, (\xi \circ F)(x') \leq (\xi \circ F)(x).$$

We will call these functions  $\{f_\xi := \xi \circ F\}_{\xi \in \Xi}$  the *extreme* cost functions associated to  $F$ . They are of particular interest when  $K$  is polyhedral, since it is exactly the case for which  $\Theta$  has a finite number of extreme points (see Theorem 2.3.9). This means a finite number of extreme cost functions to manage. Take for instance the multi-objective case  $F = (f_1, \dots, f_m) : X \rightarrow \mathbb{R}^m$ ,  $\mathbb{R}^m$  being ordered by the usual positive orthant  $\mathbb{R}_+^m$ . We choose  $e = (1, \dots, 1) \in \text{int } \mathbb{R}_+^m = \mathbb{R}_{++}^m$ , whose corresponding basis is  $\Delta_m$  the unit simplex (see Example 2.3.6), and the extreme points of  $\Delta_m$  corresponds to the canonical basis  $\Xi = \{e_1, \dots, e_m\}$ . The corresponding extreme cost functions are exactly the family  $\{f_1, \dots, f_m\}$ . In that case, (2.26) says that the decrease of  $F$  is governed by the decrease of all convex combinations (the cost functions)

$$f_\theta = \sum_{i=1}^m \theta_i f_i, \quad \theta = (\theta_i) \in \Delta_m,$$

while (2.27) tells us that it is enough to have the decrease of the finite family  $\{f_1, \dots, f_m\}$  (the extreme cost functions).

We now discuss the convexity of a vector-valued function  $F : X \rightarrow (Y, K)$  with respect to  $K$ . Keep the notations that we used above. We say that  $F$  is *convex* (with respect to  $K$ ) if

$$\forall x, x' \in X, \forall t \in [0, 1], F(tx + (1-t)x') \leq tF(x) + (1-t)F(x').$$

Applying the monotonicity Theorem 2.3.7, we see that  $F$  is convex if and only if each cost function  $\{f_\theta\}_{\theta \in \Theta}$  is convex. It is in fact sufficient for the extreme cost functions  $\{f_\xi\}_{\xi \in \Xi}$  to be convex, which is a quite useful characterization in the polyhedral case. We also say that  $F$  is *strictly convex* (with respect to  $K$ ) if

$$\forall x \neq x' \in X, \forall t \in ]0, 1[, F(tx + (1-t)x') < tF(x) + (1-t)F(x').$$

Convex functions enjoy a useful characterization of their weakly efficient points, which is the base of all scalarization techniques mentioned in the introduction.

**Theorem 2.3.10.** *Let  $F : X \rightarrow (Y, K)$  be a convex function between two Banach spaces. Suppose that  $C \subset X$  is convex and  $K$  is a closed convex cone with nonempty interior. Then,*

$$\text{ARGMIN}_w F(x) = \bigcup_{\theta \in \Theta} \underset{x \in C}{\text{argmin}} f_\theta(x).$$

This result can be found in the book of Dihn [139, Proposition IV.2.10], or in [73, Theorem 2.1]. We will see in Chapter 5 how this result can be simply interpreted as a Fermat's rule for functions with values in an ordered vector space.

We say that  $F : X \rightarrow (Y, K)$  is *quasiconvex* (with respect to  $K$ ) if

$$\forall x, x', \bar{x} \in X, \forall t \in [0, 1], F(x) \leq F(\bar{x}) \text{ and } F(x') \leq F(\bar{x}) \Rightarrow F(tx + (1-t)x') \leq F(\bar{x}).$$

Equivalently,  $F$  is quasiconvex if and only if

$$\forall \bar{x} \in X, [F \leq F(\bar{x})] \text{ is convex,}$$

where  $[F \leq F(\bar{x})] = \{x \in X \mid F(x) \leq F(\bar{x})\}$  denotes as before a sublevel set of  $F$  at  $F(\bar{x})$ . We say moreover that  $F$  is *scalarly quasiconvex* whenever all cost functions  $\{f_\theta\}_{\theta \in \Theta}$  are quasiconvex. Scalar quasiconvexity entails quasiconvexity since we can write (use Theorem 2.3.7)

$$\forall \bar{x} \in X, [F \leq F(\bar{x})] = \bigcap_{\theta \in \Theta} [f_\theta \leq f_\theta(\bar{x})].$$

But contrary to the what happens in the convex setting, the reverse is not true in general.

We end with a basic notion for functions with values in an ordered space. We say that  $F : E \rightarrow Y$  is *bounded from below* if there exists  $y \in Y$  such that  $y \leq F(x)$  for all  $x \in E$ . In our setting, where  $K$  has a nonempty interior, the boundedness of  $F$  is equivalent to the uniform boundedness of the cost functions  $\{f_\theta\}_{\theta \in \Theta}$ :

**Proposition 2.3.11.** Let  $F : X \rightarrow Y$ , let  $e \in \text{int } K$  and  $\theta$  the corresponding base of  $K^+$ , noting  $\Xi$  its set of extremal points. The following are equivalent :

- i)  $F$  is bounded from below,
- ii)  $\exists m \in \mathbb{R}$  such that  $me \leq F(x)$  for all  $x \in E$ ,
- iii)  $\exists m \in \mathbb{R}$  such that  $m \leq f_\xi(x)$  for all  $x \in E$  and  $\xi \in \Xi$ ,
- iv)  $\exists m \in \mathbb{R}$  such that  $m \leq f_\theta(x)$  for all  $x \in E$  and  $\theta \in \Theta$ .

*Proof.* Suppose that i) holds. Then there exists some  $y \in Y$  such that, for all  $x \in E$  and  $\theta \in \Theta$ ,  $\langle \theta, y \rangle \leq f_\theta(x)$ . Let  $M > 0$  be a bound for  $\Theta$  in  $Y^*$ , then by using Cauchy-Schwarz's inequality, we obtain  $-M\|y\| \leq f_\theta(x)$ . So item iv) holds, with  $m = -M\|y\|$ . The implication iv)  $\Rightarrow$  iii)  $\Rightarrow$  ii) follows Theorem 2.3.7 and the fact that  $\langle \theta, e \rangle = 1$  for all  $\theta \in \Theta$ . Implication ii)  $\Rightarrow$  i) is obvious. ■

## Part I

# Dynamics for Tame optimization





Most analysts spend half their time  
hunting through the literature for  
inequalities they want to use, but cannot  
prove.

---

Harald Bohr



## Chapter 3

# The Kurdyka-Łojasiewicz inequality and general descent methods

Along this chapter, we consider a Hilbert space  $(\mathbb{H}, \langle \cdot, \cdot \rangle)$ , and a proper lower semi-continuous function  $f : \mathbb{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ . We want to solve the associated optimization problem

$$\min_{x \in \mathbb{H}} f(x).$$

For this, we look for descent dynamical systems corresponding to  $f$ , being discrete or continuous. Our main tool in this chapter is the assumption that  $f$  satisfies the so-called *Kurdyka-Łojasiewicz inequality*.

This notion is introduced and illustrated in Section 3.1: after an informal introduction in Section 3.1.1, we properly define it in Section 3.1.2, and give some examples of functions satisfying this inequality. In Section 3.1.3, we illustrate one of the main consequences of the Kurdyka-Łojasiewicz inequality in Theorem 3.1.8. Considering the subgradient differential inclusion

$$(3.1) \quad \dot{u}(t) + \partial^F f(u(t)) \ni 0,$$

any of its trajectories has finite length and converges towards a critical point of  $f$ , provided that  $u$  satisfies some compactness assumption, and that  $f$  has the Kurdyka-Łojasiewicz property. Then, we detail in Theorem 3.1.12 some rates of convergence for the trajectories.

In Section 3.2, we adapt this convergence result to a general abstract descent algorithm. It is inspired by the work of [20], but we extend its setting in order to account for additive computational errors, and more versatility in the choice of the parameters. The strong convergence of the iterates towards a critical point of  $f$ , with a *finite-length* condition is obtained in Theorem 3.2.2. A local convergence result to global minimum is also provided in Theorem 4.1.8, under certain hypotheses. Then, in Section 3.3, we prove new and interesting general convergence rates. They are similar to the ones obtained in the literature for various numerical methods, see e.g. [18, 19, 321, 72, 107]. Surprisingly, an explicit form of the algorithm terminates in a finite number of iterations in several cases, and shares the same convergence rates with the continuous-time dynamical system.

## 3.1 The Kurdyka-Łojasiewicz inequality

### 3.1.1 Introduction

In a first time, we propose an informal and naive approach to the Kurdyka-Łojasiewicz inequality<sup>1</sup>.

Suppose that  $f : H \rightarrow \mathbb{R}$  is a continuously differentiable function, and consider the corresponding steepest descent dynamic:

$$(SD) \quad \dot{x}(t) + \nabla f(x(t)) = 0.$$

Given a classic global solution  $x : [0, +\infty[ \rightarrow H$  of (SD), we focus on its asymptotic behaviour, when  $t$  goes to  $+\infty$ . What we aim to prove here is that the trajectory has finite length, i.e.

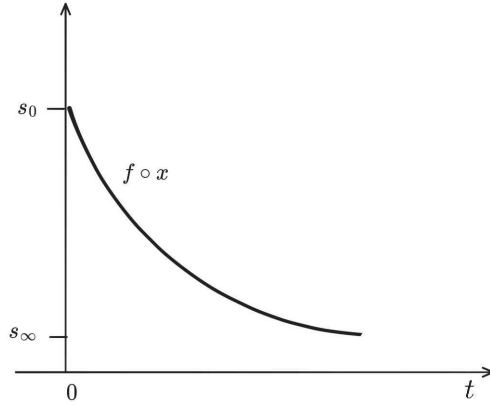
$$\int_0^{+\infty} \|\dot{x}(t)\| dt < +\infty.$$

This is a strong property, which implies in particular that the trajectory strongly converges when  $t \rightarrow +\infty$ .

First observe that the function  $f$  is Lyapunov for the (SD) dynamic. Indeed, it suffices to take the scalar product of (SD) with  $\dot{x}(t)$ , and to use a chain rule, to obtain

$$(3.2) \quad \frac{d}{dt}(f \circ x)(t) = -\|\dot{x}(t)\|^2 \leq 0.$$

If we assume that  $\dot{x}(t)$  is nonzero on  $[0, +\infty[$ , (3.2) means that  $f \circ x$  is strictly decreasing on  $[0, +\infty[$ . If we assume moreover that  $f$  is bounded from below, we obtain that  $f \circ x$  realizes a diffeomorphism between  $[0, +\infty[$  and  $]s_\infty, s_0]$ , where  $s_0 := f(x(0))$  and  $s_\infty := \inf_{t \geq 0} f(x(t))$ .



Given the diffeomorphism  $f \circ x : [0, +\infty[ \rightarrow ]s_\infty, s_0]$ , consider the change of variables  $s = (f \circ x)(t)$ , and the associated trajectory

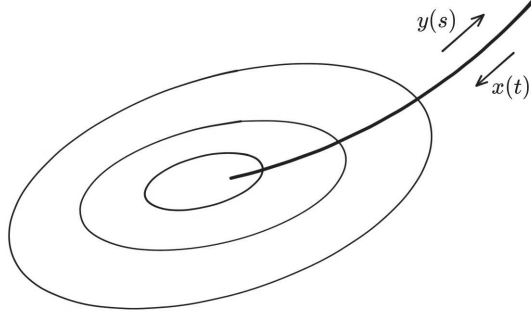
$$\forall s \in ]s_\infty, s_0], y(s) := x((f \circ x)^{-1}(s)).$$

This trajectory verifies

$$\dot{y}(s) - \frac{\nabla f(y(s))}{\|\nabla f(y(s))\|^2} = 0,$$

<sup>1</sup>Most of the contents presented here are strongly inspired by two oral presentations given by A. Daniilidis, given at the SOMACHI 2014 and ISMP 2015.

in other words,  $y(\cdot)$  describes the same curve than  $x(\cdot)$ , but at reverse and with a different speed.



We can use this reparametrized trajectory to rewrite the length of  $x(\cdot)$ :

$$(3.3) \quad \int_0^{+\infty} \|\dot{x}(t)\| dt = \int_{s_\infty}^{s_0} \frac{1}{\|\nabla f(y(s))\|} ds.$$

Observe that the quantity in (3.3) is finite, as soon as  $\|\nabla f(y(s))\| \geq C > 0$  for all  $s \in [s_\infty, s_0]$ . In other words, the question of the finite length of  $x(\cdot)$  is essentially dependent on how  $\nabla f$  behaves around its zeroes, namely, the critical points of  $f$ .

Suppose now that there exists some smooth increasing function  $\varphi : [s_\infty, s_0] \rightarrow \mathbb{R}_+$ , such that  $\varphi(s_\infty) = 0$  and

$$(3.4) \quad \frac{1}{\|\nabla f(y(s))\|} \leq \varphi'(s) \quad \text{for all } s \in [s_\infty, s_0].$$

Then, it follows from (3.3) that  $x(\cdot)$  has finite length:

$$\int_0^{+\infty} \|\dot{x}(t)\| dt \leq \int_{s_\infty}^{s_0} \varphi'(s) ds = \varphi(s_0) < +\infty.$$

Clearly, the key point here is the inequality (3.4), which can be rewritten, after using the change of variables  $s = f(x(t))$ , as

$$\varphi'(f(x(t)))\|\nabla f(x(t))\| \geq 1, \quad \forall t \geq 0.$$

As we said before, the question of the finite length of the trajectory only matters around the critical points (points  $x$  such that  $\nabla f(x) = 0$ ).

Roughly speaking, the Kurdyka-Łojasiewicz inequality asks that, for any critical point  $\bar{x} \in \mathbb{H}$ , there exists a smooth and increasing function  $\varphi : [0, +\infty[ \rightarrow [0, +\infty[$  such that  $\varphi(0) = 0$  and

$$(3.5) \quad \varphi'(f(x) - f(\bar{x}))\|\nabla f(x)\| \geq 1, \quad \text{for all } x \sim \bar{x} \text{ such that } f(\bar{x}) < f(x).$$

Assuming that  $f(\bar{x}) = 0$  and using a chain rule, we see that the Kurdyka-Łojasiewicz inequality becomes

$$(3.6) \quad \|\nabla(\varphi \circ f)(x)\| \geq 1, \quad \forall x \sim \bar{x}.$$

In other words, we ask for the possibility to locally reparametrize  $f$ , in order to make it become *sharp* around its critical points. That is why, in general,  $\varphi$  is called the *desingularizing function* for  $f$  at  $\bar{x}$ . Indeed, the more  $f$  is flat around its critical point, the more  $\varphi$  has to be steep around 0.

### 3.1.2 The Kurdyka-Łojasiewicz inequality

In the previous discussion, we took for convenience a smooth function  $f : H \rightarrow \mathbb{R}$ . Consider now a real-extended valued function  $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ . In that case, the Kurdyka-Łojasiewicz inequality will set a relation between the variations of  $f$  and the norm of its subgradients. More exactly, we will consider the limiting subgradients of  $f$ , noted  $\partial^L f$ , introduced in Section 2.2.3. We say that  $x \in H$  is a limiting critical point of  $f$  if  $0 \in \partial^L f(x)$ . We also consider the *lazy slope* of  $f$  at  $x$ :

$$\|\partial^L f(x)\|_- := \inf_{x^* \in \partial^L f(x)} \|x^*\|.$$

Let  $\eta > 0$ . We say that  $\varphi : [0, \eta[ \rightarrow [0, +\infty[$  is a *desingularizing function* if

- i)  $\varphi(0) = 0$ ,
- ii)  $\varphi$  is continuous on  $[0, \eta[$  and of class  $C^1$  on  $]0, \eta[$ ,
- iii)  $\varphi'(t) > 0$  for all  $t \in ]0, \eta[$ .

Typical examples of desingularizing functions are the functions  $\varphi(t) = \frac{C}{\theta} t^\theta$ , for  $C > 0$  and  $\theta \in ]0, 1]$ .

**Definition 3.1.1.** Let  $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ . We say that  $f$  satisfies the Kurdyka-Łojasiewicz property at  $\bar{x} \in H$  if there exists a neighbourhood  $B(\bar{x}, \delta)$  of  $\bar{x}$ , and a concave desingularizing function  $\varphi : [0, \eta[ \rightarrow [0, +\infty[$ , such that the *Kurdyka-Łojasiewicz inequality*

$$(KL) \quad \varphi'(f(x) - f(\bar{x})) \|\partial^L f(x)\|_- \geq 1$$

holds, for all  $x$  in the strict local upper level set

$$\Gamma_\eta(\bar{x}, \delta) := \{x \in B(\bar{x}, \delta) \mid f(\bar{x}) < f(x) < f(\bar{x}) + \eta\}.$$

A proper lower-semicontinuous function having the Kurdyka-Łojasiewicz property (KL property for short) at any  $x \in H$  is said to be a *KL function*.

**Remark 3.1.2.** In the following, we will give a special attention to the functions satisfying the KL inequality with a desingularizing function of the form  $\varphi(t) = \frac{C}{\theta} t^\theta$ ,  $\theta \in ]0, 1]$ . In that case,  $f$  will be said to be *KL with the Łojasiewicz exponent  $\theta$* .

**Remark 3.1.3.** As mentioned before, for functions of class  $C^1$ , the KL inequality holds automatically at any regular point (i.e. being not limiting critical points), with a desingularising function  $\varphi(t) = Ct$ . This property is also verified by all proper lower semi-continuous functions, when  $H$  has finite dimension. Indeed, assume by contradiction that for some regular  $\bar{x} \in H$  there exists  $(x_k)_{k \in \mathbb{N}}$  in  $H$  such that

$$\forall k \in \mathbb{N}, \|x_k - \bar{x}\| \leq \frac{1}{k}, \quad f(\bar{x}) < f(x_k) < f(\bar{x}) + \frac{1}{k} \quad \text{and} \quad \|\partial^L f(x_k)\|_- < \frac{1}{k}.$$

This sequence would satisfy  $x_k \xrightarrow[k \rightarrow +\infty]{f} \bar{x}$  and  $\|\partial^L f(x_k)\|_- \xrightarrow[k \rightarrow +\infty]{} 0$ . But in finite dimension, the limiting subdifferential  $\partial^L f$  has a  $f$ -closed graph (see [290, Proposition 8.7]), so it would follow that  $0 \in \partial^L f(\bar{x})$ , which contradicts the noncritical assumption on  $\bar{x}$ . In more general Hilbert spaces, the limiting subdifferential is no more guaranteed to have a  $f$ -closed graph (see [244, Example 1.7]). In that case, we can assume that  $f$  is lower-regular, in which case the argument above still works.

We present now some known classes of KL functions.

**Example 3.1.4** (Analytic functions). Łojasiewicz [230] proved that, for any critical point  $\bar{x}$  of an analytic function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , there exists an open ball  $B(\bar{x}, \delta)$  such that

$$|f(x) - f(\bar{x})|^{1-\theta} \leq C \|\nabla f(x)\|,$$

for some  $C > 0$  and  $\theta \in ]0, \frac{1}{2}]$ . This proves that analytic functions are KL functions, with a Łojasiewicz exponent  $\theta \in ]0, \frac{1}{2}]$ . Let us mention that Łojasiewicz's result applies only in finite dimension. Indeed, in infinite-dimensional spaces, analytic functions need not have the KL property, see [188] for a counter-example. In a similar fashion to Łojasiewicz, Bolte-Daniilidis-Lewis show that continuous subanalytic functions on  $\mathbb{R}^n$ , with closed domain, are KL with a Łojasiewicz exponent  $\theta \in ]0, 1]$ , see [68, Theorem 3.1].

**Example 3.1.5** (Functions definable in o-minimal structures). In the finite-dimensional setting, the *o-minimal structures* provide a huge class of functions satisfying the KL property. O-minimal structures were developed around the 90's to generalize the good stability and regularity properties of the class of semi-algebraic sets [271, 144]. These o-minimal structures includes the semi-algebraic structure, in which are definable the polynomials, the linear maps, the  $\ell^p$  norms with  $p \in \mathbb{Q}^*$ , and the counting function, among others. It is known that the lower semi-continuous functions lying in this semi-algebraic structure are KL, with a Łojasiewicz exponent  $\theta \in ]0, 1]$  [70, Corollary 16]. There also exists bigger o-minimal structures, containing for instance the exponential function [143, 319], the  $\ell^p$  norms with  $p \in \mathbb{R}^*$  [243] and the primitives of all the aforementioned functions [145, 146, 298]. Note that besides being huge, these class of functions are stable by usual operations which are very useful in optimization: addition, composition, supremum, restriction to an o-minimal set. Furthermore, the distance function to an o-minimal set and the indicator function of an o-minimal set are also o-minimal functions (this list is far from being exhaustive). The fact that any lower semi-continuous function definable in an o-minimal structure is a KL function is a strong result, due to Bolte, Daniilidis, Lewis and Shiota [70, Corollary 15]. It extends a previous result of Kurdyka [216] on functions of class  $C^1$  definable on an o-minimal structures. Observe that, in the general case, a definable function may not be KL for any Łojasiewicz exponent  $\theta \in ]0, 1]$ . Indeed, a function such that  $x \mapsto e^{-\frac{1}{x^2}}$  needs a desingularizing function  $\varphi$  with an exponential growth around the origin. For more information about o-minimal structures, its definition and the properties they enjoy, the reader shall consult [142, 141, 204].

**Example 3.1.6** (KL inequality versus convexity). Let us mention the relationships between convexity and the KL property. It is easy to check that strongly convex functions are KL with a Łojasiewicz exponent  $\theta = \frac{1}{2}$  (see [19, Section 4.2]). But in general, convex functions are not KL: see [71, Section 4.3], where the authors coined a counter-example of a convex function not being KL. Nevertheless, it requires such a twisted construction that we can reasonably say that 'common' convex functions are KL. For instance, a convex function is KL as soon as it verifies some growth condition around its critical points [71, Theorem 30].

Let us finally mention that the KL property can be derived from appropriate assumptions on the Hessian  $D^2f(\bar{x})$  ([103, 188]). It is also worth noticing that the KL inequality has strong connexions with metric regularity, see [19, 71].

### 3.1.3 Asymptotic behaviour of a subgradient differential inclusion

Given a proper lower-semicontinuous function  $f : \mathbb{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ , consider the following *subgradient differential inclusion*:

$$(3.7) \quad \dot{u}(t) + \partial^F f(u(t)) \ni 0.$$

When  $f$  is smooth, this dynamical system specialises in the usual steepest descent dynamic:

$$(3.8) \quad \dot{u}(t) + \nabla f(u(t)) = 0.$$



We saw in the previous section –at least formally– that in this smooth case, the trajectories converge to a critical point of  $f$ , provided that  $f$  is a KL function. We shall verify below that the subgradient dynamic (3.7) shares the same properties that the steepest descent one (3.8): it is a *descent* dynamic, and its trajectories converge to critical points whenever  $f$  is a KL function.

First of all, let us precise the notion of solution we will use for (3.7). Indeed, this dynamic is governed by the set-valued mapping  $x \mapsto \partial^F f(x)$  which is not continuous, so we cannot use a classical notion of solution. We recall here the definition of absolutely continuous functions (see the monograph of Brezis [83, Appendix] for more details).

**Definition 3.1.7.** Given  $T \in \mathbb{R}_+$ , a function  $u : [0, T] \rightarrow \mathbb{H}$  is said to be absolutely continuous if one of the following equivalent properties holds:

i) there exists an integrable function  $g : [0, T] \rightarrow \mathbb{H}$  such that

$$u(t) = u(0) + \int_0^t g(s) ds \quad \forall t \in [0, T];$$

ii)  $u$  is continuous and its distributional derivative belongs to the Lebesgue space  $L^1([0, T]; \mathbb{H})$ ;

iii) for every  $\epsilon > 0$ , there exists  $\eta > 0$  such that for any finite family of intervals  $I_k = ]a_k, b_k[$ ,  $I_k \cap I_j = \emptyset$  for  $k \neq j$  and  $\sum_k |b_k - a_k| \leq \eta \implies \sum_k \|u(b_k) - u(a_k)\| \leq \epsilon$ .

We can now make precise the notion of solution for the subgradient differential inclusion. We say that  $u : [0, +\infty[ \rightarrow \mathbb{H}$  is a *strong global solution* of (3.7) if it is absolutely continuous on each bounded interval  $[0, T]$ ,  $T < +\infty$ , and if it satisfies the inclusion (3.7) for a.e.  $t \in [0, +\infty[$ . We also introduce the notion of *f-precompactness* (or *f-attentive precompactness*) for a trajectory  $(u(t))_{t \geq 0}$ , meaning that there exists a sequence  $t_n \rightarrow +\infty$  and some  $u_\infty \in \mathbb{H}$  such that

$$u(t_n) \xrightarrow[n \rightarrow +\infty]{f} u_\infty, \text{ meaning that } u(t_n) \xrightarrow[n \rightarrow +\infty]{} u_\infty \text{ and } f(u(t_n)) \xrightarrow[n \rightarrow +\infty]{} f(u_\infty).$$

We are now ready to prove the announced result, which is a simple extension of the one obtained by Bolte-Daniilidis-Lewis [68] for continuous subanalytic functions. A very similar result has been obtained by Ioffe [204] in  $\alpha$ -minimal structures, with a different proof.

**Theorem 3.1.8.** *Let  $f : \mathbb{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semi-continuous function, bounded from below. Let  $u : [0, +\infty[ \rightarrow \mathbb{H}$  be a strong global solution of (3.7), and assume that  $f \circ u$  is absolutely continuous on each bounded interval  $[0, T]$ ,  $T < +\infty$ . Then, the following holds:*

i) (Descent property) *For a.e.  $t \in [0, +\infty[$ ,  $\frac{d}{dt}(f \circ u)(t) = -\|\dot{u}(t)\|^2$ .*

ii) (Laziness) *For a.e.  $t \in [0, +\infty[$ ,  $\dot{u}(t) + \partial^F f(u(t))^0 = 0$ , where  $\partial^F f(u(t))^0$  denotes the element of minimal norm of  $\partial^F f(u(t))$ .*

*Assume moreover that  $f$  is a KL function, and that  $(u(t))_{t \geq 0}$  is  $f$ -precompact, then:*

iii) (Strong convergence) *The trajectory has finite length and converges, when  $t \rightarrow +\infty$ , to some  $u_\infty$ . This limit point is  $\partial^L$ -critical, in the sense that  $0 \in \partial^L f(u_\infty)$ .*

**Remark 3.1.9.** Let us briefly discuss the hypotheses involved in Theorem 3.1.8. The absolute continuity of  $f \circ u$  can be directly derived from the absolute continuity of  $u$ , whenever  $f$  is locally Lipschitz continuous on its domain. The  $f$ -precompactness assumption can be reduced to simple precompactness, provided that  $f$  is continuous on its domain. It is worth noticing that precompactness comes from boundedness if  $\mathbb{H}$  is finite-dimensional. More generally, according to the decrease property i), precompactness is guaranteed as soon as  $f$  has compact level sets. In finite dimensions, the latter is equivalent for  $f$  to be coercive.

**Remark 3.1.10.** We do not address here the question of the existence of such solutions. It is well-known for  $C^{1,1}$  functions (it is the Cauchy-Lipschitz theorem), or for lower-semi continuous convex functions [83]. For an existence result involving a wider class of functions, the reader can consult [134, 235].

**Remark 3.1.11.** To prove the convergence of the trajectories, we will follow Lojasiewicz's argument [231]. Roughly speaking, the proof is divided in three steps: in a first time, use a compactness argument to prove that the trajectory will enter into  $\Gamma_\eta(u_\infty, \delta)$  of  $u_\infty$ , at some time  $t \geq 0$ . In a second time, we prove that the length of the trajectory, when it lies in  $\Gamma_\eta(u_\infty, \delta)$ , is independent of the time. Finally, prove that the set  $\Gamma_\eta(u_\infty, \delta)$  captures the trajectory: once the trajectory enters  $\Gamma_\eta(u_\infty, \delta)$ , it cannot escape from it. Once combined, these three facts entails the finite length of the trajectory.

*Proof.* Items i) and ii) are quite immediate from the definition of the dynamic. For item ii), use a chain rule (Proposition 2.2.11) together with the absolute continuity of  $f \circ u$  and  $u$  to write for a.e.  $t \in [0, +\infty[$ :

$$(3.9) \quad \frac{d}{dt}(f \circ u)(t) = \langle x^*, \dot{u}(t) \rangle, \quad \forall x^* \in \partial^F f(u(t)).$$

Since by definition we have  $-\dot{u}(t) \in \partial^F f(u(t))$ , the decrease property  $\frac{d}{dt}(f \circ u)(t) = -\|\dot{u}(t)\|^2$  follows immediately. For the lazy property, let us start by recalling that in a Hilbert space, the Fréchet subdifferential is closed and convex, so its element of minimal norm is well-defined, provided it is nonempty. This being said, it suffices to show, for a.e.  $t \in [0, +\infty[$ , that  $\|\dot{u}(t)\| = \|\partial^F f(u(t))\|_-$ . The inequality  $\|\dot{u}(t)\| \geq \|\partial^F f(u(t))\|_-$  comes directly from  $-\dot{u}(t) \in \partial^F f(u(t))$ . For the reverse inequality, use (3.9) together with the Cauchy-Schwarz inequality to obtain

$$\forall x^* \in \partial^F f(u(t)), \quad \|\dot{u}(t)\|^2 = \langle x^*, \dot{u}(t) \rangle \leq \|x^*\| \|\dot{u}(t)\|.$$

Thus,  $\|\dot{u}(t)\| \leq \|x^*\|$  for all  $x^* \in \partial^F f(u(t))$ , and item i) follows.

We turn now on item iii). First of all, recall that the  $f$ -precompactness gives a sequence  $t_n \rightarrow +\infty$  such that

$$u(t_n) \xrightarrow[n \rightarrow +\infty]{} u_\infty \quad \text{and} \quad f(u(t_n)) \xrightarrow[n \rightarrow +\infty]{} f(u_\infty).$$

Moreover, we supposed that  $f$  is bounded from below, so we deduce from item i) that  $f(u(t)) \downarrow f(u_\infty)$ . We can assume, without loss of generality, that  $f(u(t)) > f(u_\infty)$  for all  $t \geq 0$ . Otherwise, we would have from item i) that  $u(t) \equiv u_\infty$  for all  $t \geq T$ , for some  $T \geq 0$ . We invoke now the Kurdyka-Lojasiewicz inequality at  $u_\infty$ , which holds on the local strict upper level set  $\Gamma_\eta(u_\infty, \delta)$ :

$$\forall x \in \Gamma_\eta(u_\infty, \delta), \quad \varphi'(f(x) - f(u_\infty)) \|\partial^L f(x)\|_- \geq 1.$$

The key element to obtain the convergence of the trajectory is the introduction of the Lyapunov function  $h(t) := \varphi(f(u(t)) - f(u_\infty))$ . By definition of the desingularizing function, it is an absolutely continuous function on bounded intervals, positive, nonincreasing, such that  $h(t) \downarrow 0$  when  $t \rightarrow +\infty$ .

For an arbitrary pair  $0 \leq t_1 < t_2 \leq +\infty$ , we claim that

$$(3.10) \quad u(]t_1, t_2[) \subset \Gamma_\eta(u_\infty, \delta) \Rightarrow \int_{t_1}^{t_2} \|\dot{u}(t)\| \, dt \leq h(t_1).$$

To see this, use item ii) together with the KL inequality to write for a.e.  $t \in ]t_1, t_2[$ :

$$\|\dot{u}(t)\| = \|\partial^F f(u(t))\|_- \leq \varphi'(f(u(t)) - f(u_\infty)) \|\partial^F f(u(t))\|_-^2.$$

Since  $\|\partial^F f(u(t))\|_-^2 = \|\dot{u}(t)\|^2 = -\frac{d}{dt}(f \circ u)(t)$ , we deduce that

$$\|\dot{u}(t)\| \leq -\varphi'(f(u(t)) - f(u_\infty)) \frac{d}{dt}(f \circ u)(t) = -\frac{d}{dt} h(t).$$

Recalling that  $h$  is positive, (3.10) is proved after integrating the above equation on  $]t_1, t_2[$ .

Now, we deduce from  $u(t_n) \xrightarrow{n \rightarrow +\infty} u_\infty$ ,  $f(u(t)) \downarrow f(u_\infty)$  and  $h(t) \downarrow 0$ , that there exists some  $t_1 \geq 0$  such that

$$(3.11) \quad \|u(t_1) - u_\infty\| \leq \frac{\delta}{3} \text{ and } \forall t \geq t_1, h(t) \leq \frac{\delta}{3}, f(u(t)) < f(u_\infty) + \eta.$$

We also define

$$t_2 := \inf\{t \geq t_1 \mid u(t) \notin \Gamma_\eta(u_\infty, \delta)\} \in ]t_1, +\infty[.$$

By definition of these two numbers, and according to the continuity of  $u$  together with the monotonicity of  $(f(u(t)))_{t \geq 0}$ , we have  $u(]t_1, t_2[) \subset \Gamma_\eta(u_\infty, \delta)$  and  $u(t_2) \notin \Gamma_\eta(u_\infty, \delta)$ . The claim (3.10), together with (3.11), imply

$$(3.12) \quad \int_{t_1}^{t_2} \|\dot{u}(t)\| dt \leq \frac{\delta}{3}.$$

If we assume that  $t_2 < +\infty$ , we would obtain from (3.11) and (3.12):

$$\|u(t_2) - u_\infty\| \leq \|u(t_2) - u(t_1)\| + \|u(t_1) - u_\infty\| \leq \int_{t_1}^{t_2} \|\dot{u}(t)\| dt + \|u(t_1) - u_\infty\| \leq \frac{2\delta}{3},$$

which contradicts  $t_2 \notin \Gamma_\eta(u_\infty, \delta)$  and  $f(u(t_2)) \leq f(u(t_1)) < \eta$ .

Thus, we deduce that  $t_2 = +\infty$ , and obtain from (3.12) that  $u$  has finite length. In particular, this means that the trajectory converges strongly to  $u_\infty$ . Furthermore, item ii) combined with (3.12) gives

$$\int_0^\infty \|\partial^F f(u(t))\|_- dt < +\infty.$$

The latter implies the existence of some sequence  $t_m \xrightarrow{m \rightarrow +\infty} +\infty$  and a corresponding sequence of Fréchet subgradients  $x_m^* \in \partial^F f(u(t_m))$  such that  $\|x_m^*\| \xrightarrow{m \rightarrow +\infty} 0$ . This means, by definition of the limiting subdifferential, that  $0 \in \partial^L f(u_\infty)$ .  $\blacksquare$

We can also derive for the solutions of (3.7) an estimate of the rate of convergence, depending only on the behaviour of the desingularizing function around the origin. Next result generalises [104, Theorem 2.7] to the nonsmooth case, and we recover the rates of [68] in the subanalytic case.

**Theorem 3.1.12.** *Let  $f : \mathbb{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semi-continuous function. Let  $u(\cdot)$  be a strong global solution of (3.7), which  $f$ -converges when  $t \rightarrow +\infty$  to some  $u_\infty \in \mathbb{H}$ . We assume that  $f$  has the KL property at  $u_\infty$ , with a desingularizing function  $\varphi : [0, \eta[ \rightarrow \mathbb{R}_+$ . Let  $\Phi : ]0, \eta[ \rightarrow \mathbb{R}$  be any primitive of  $-\varphi'^2$ .*

*i) If  $\lim_{t \rightarrow 0} \Phi(t) \in \mathbb{R}$ , then the algorithm converges in a finite number of steps.*

*ii) If  $\lim_{t \rightarrow 0} \Phi(t) = +\infty$ , then there exists some  $t_1 \in \mathbb{R}$  such that:*

*ii.a)  $f(u(t)) - f(u_\infty) = O(\Phi^{-1}(t - t_1))$ , and*

*ii.b)  $\|u(t) - u_\infty\| = O(\varphi \circ \Phi^{-1}(t - t_1))$ .*

*Proof.* Since we suppose that  $u(t)$   $f$ -converges to  $u_\infty$ , there exists some  $t_0 \in \mathbb{R}_+$  such that, for all  $t \geq t_0$ ,  $u(t)$  lies in the local strict upper level set  $\Gamma_\eta(u_\infty, \delta)$  where the KL inequality holds. Assume that, for all  $t \geq t_0$ ,  $f(u(t)) > f(u_\infty)$ . Otherwise, as we seen in the previous proof, the trajectory would stop in finite time. Define  $r(t) := f(u(t)) - f(u_\infty)$  and as before  $h(t) := \varphi(r(t))$ . Observe that we can take  $t_0$  big enough, so that  $r(t) < \eta$  for all  $t \geq 0$ , making  $h$  well-defined on  $[t_0, +\infty[$ .

We saw in the previous proof that, for all  $t \geq t_0$ ,  $\|\dot{u}(t)\| \leq -h'(t)$ . By integrating this inequality, we obtain for all  $T \geq t \geq t_0$ :

$$\|u(t) - u(T)\| \leq \int_t^T \|\dot{u}(\tau)\| \, d\tau \leq - \int_t^T h(\tau) \, d\tau = h(t) - h(T) \leq h(t).$$

After taking the limit when  $T \rightarrow +\infty$ , we deduce that

$$(3.13) \quad \|u(t) - u_\infty\| \leq h(t) = \varphi(f(u(t)) - f(u_\infty)).$$

Let now  $\Phi$  be a primitive of  $-\varphi'^2$ . Then, for all  $t \geq t_0$ , we can use Theorem 3.1.8 to write

$$(\Phi \circ r)'(t) = -\varphi'^2(r(t))(f \circ u)'(t) = \varphi'^2(f(u(t)) - f(u_\infty)) \|\partial^F f(u(t))\|_-^2 \geq 1.$$

Integrate the inequality above to obtain for all  $t \geq t_0$ :

$$(3.14) \quad \Phi(r(t)) - \Phi(r(t_0)) = \int_{t_0}^t (\Phi \circ r)'(\tau) \, d\tau \geq t - t_0.$$

Observe that, by definition, the function  $\Phi$  is strictly decreasing on  $]0, \eta[$ , hence invertible on this interval. We consider now two cases. If  $\lim_{r \rightarrow 0} \Phi(r) = \Phi_0 \in \mathbb{R}$ , then we would obtain from (3.14) that  $\Phi_0 \geq +\infty$  by taking the limit when  $t \rightarrow +\infty$ . This is a clear contradiction, meaning that our assumption made in the beginning is false. As a consequence, the trajectory converge in finite time. On the other hand, if  $\lim_{r \rightarrow 0} \Phi(r) = +\infty$ , we can assume that  $t_0$  is big enough so that  $t - t_0 + \Phi(r(t_0))$  lies in the domain of  $\Phi^{-1}$ . Thus, we can deduce from (3.14) that

$$(3.15) \quad f(u(t)) - f(u_\infty) \leq \Phi^{-1}(t - t_1), \text{ where } t_1 := t_0 - \Phi(r(t_0)).$$

The conclusion follows from the combination of (3.15) and (3.13). ■

## 3.2 Convergence of an abstract inexact descent method

Throughout this section,  $f : \mathbb{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper function, lower semi-continuous for the strong topology. We consider a sequence  $(x_k)_{k \in \mathbb{N}}$ , computed by means of an abstract algorithm satisfying the following hypotheses:

$\mathbf{H}_1$  (Sufficient decrease): For each  $k \in \mathbb{N}$ , for some  $a_k > 0$ ,

$$f(x_{k+1}) + a_k \|x_{k+1} - x_k\|^2 \leq f(x_k).$$

$\mathbf{H}_2$  (Relative error): For each  $k \in \mathbb{N}$ , for some  $b_{k+1} > 0$  and  $\varepsilon_{k+1} \geq 0$ ,

$$b_{k+1} \|\partial^F f(x_{k+1})\|_- \leq \|x_{k+1} - x_k\| + \varepsilon_{k+1}.$$

$\mathbf{H}_3$  (Parameters): The sequences  $(a_k)_{k \in \mathbb{N}}$ ,  $(b_k)_{k \in \mathbb{N}}$  and  $(\varepsilon_k)_{k \in \mathbb{N}}$  satisfy:

- (i)  $a_k \geq \underline{a} > 0$  for all  $k \geq 0$ .
- (ii)  $(b_k)_{k \in \mathbb{N}} \notin \ell^1$ ;
- (iii)  $\sup_{k \in \mathbb{N}^*} \frac{1}{a_k b_k} < +\infty$ ;
- (iv)  $(\varepsilon_k)_{k \in \mathbb{N}} \in \ell^1$ .

In Section 4.1, we complement this axiomatic description of descent methods by providing a large class of implementable algorithms that produce sequences verifying hypotheses  $\mathbf{H}_1$ ,  $\mathbf{H}_2$  and  $\mathbf{H}_3$ . A simple example is:

**Example 3.2.1.** If  $f$  is differentiable, a *gradient-related* method (see [65, p.36]) is an algorithm whose each iteration has the form

$$x_{k+1} = x_k + \lambda_k d_k,$$

where  $\lambda_k > 0$  is a stepsize, and  $d_k \in \mathbf{H}$  is a direction *agreeing* with the steepest descent direction  $-\nabla f(x_k)$ , in the sense that

$$\langle d_k, \nabla f(x_k) \rangle + C_1 \|\nabla f(x_k)\|^2 \leq 0 \text{ and } \|d_k\| \leq C_2 \|\nabla f(x_k)\|,$$

with  $C_1, C_2 > 0$ . If  $\nabla f$  is  $L$ -Lipschitz continuous on  $\mathbf{H}$ , we can use a classic descent lemma 2.1.13 to obtain, after basic calculus, that  $\mathbf{H}_1$  and  $\mathbf{H}_2$  are satisfied with  $a_k = \frac{C_1}{C_2 \lambda_k} - \frac{C_2 L}{2}$ , and  $b_{k+1} = (L + \frac{1}{C_1 \lambda_k})^{-1}$ . We easily deduce that  $\mathbf{H}_3$  is satisfied if we assume for instance that  $\lambda_k \in [\underline{\lambda}, \bar{\lambda}]$ , with  $\underline{\lambda} > 0$  and  $\bar{\lambda} < \frac{C_1}{C_2^2 L}$ .

Sequences generated by the procedure described above converge strongly to critical points of  $f$ , and the piecewise linear curve obtained by interpolation has *finite length*. By finite length for a sequence  $(x_k)_{k \in \mathbb{N}}$ , we mean that

$$\sum_{k=0}^{+\infty} \|x_{k+1} - x_k\| < +\infty.$$

This is stated in the main convergence result of this chapter:

**Theorem 3.2.2.** *Let  $f : \mathbf{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a KL function and let  $\mathbf{H}_1$ ,  $\mathbf{H}_2$  and  $\mathbf{H}_3$  hold. If the sequence  $(x_k)_{k \in \mathbb{N}}$  is  $f$ -precompact, then it has finite length, and  $f$ -converges to a  $\partial^L$ -critical point of  $f$ .*

It is possible in Theorem 3.2.2 to drop the  $f$ -precompactness assumption and obtain a *capture result*, near a global minimum of  $f$ . To simplify the notation, for  $x^* \in \mathbf{H}$ ,  $\eta \in ]0, +\infty]$  and  $\delta > 0$ , define the *local upper level set* by

$$(3.16) \quad \underline{\Gamma}_\eta(\bar{x}, \delta) := \{x \in \mathbf{H} : \|x - \bar{x}\| < \delta \text{ and } f(\bar{x}) \leq f(x) < f(\bar{x}) + \eta\}.$$

We have then:

**Theorem 3.2.3.** *Let  $f : \mathbf{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  having the KL property at a global minimum  $\bar{x}$  of  $f$ . Let  $(x_k)_{k \in \mathbb{N}}$  be a sequence satisfying  $\mathbf{H}_1$ ,  $\mathbf{H}_2$  and  $\mathbf{H}_3$ , with  $\epsilon_k \equiv 0$ . Then, there exist  $\delta > 0$  and  $\eta > 0$  such that, if  $x_0 \in \underline{\Gamma}_\eta(\bar{x}, \delta)$ , then the sequence  $(x_k)_{k \in \mathbb{N}}$  has finite length, and  $f$ -converges to a global minimum of  $f$ .*

As mentioned in [20], Theorem 3.2.3 admits a more general formulation. For instance, if  $x^*$  is a local minimum of  $f$ , where a growth property is locally satisfied (see [20, Remark 2.11]).

The proofs of Theorems 3.2.2 and 3.2.3 rely on the Kurdyka-Lojasiewicz inequality, and we adapt for this the Lojasiewicz's argument used in Section 3.1.3 for the continuous case. In a first time, assuming that the sequence remains in a local strict upper level set  $\underline{\Gamma}_\eta(\bar{x}, \delta)$ , where the KL inequality holds, we show that it has finite length. In a second time, we prove the existence

of such a set  $\Gamma_\eta(\bar{x}, \delta)$  capturing the sequence. The adaptation of this argument to the discrete case goes back to Absil, Mahony and Andrews [1], who studied algorithms satisfying

$$a\|x_{k+1} - x_k\| \|\nabla f(x_k)\| \leq f(x_k) - f(x_{k+1}).$$

Since then, this proof has been refined and adapted to different methods and settings [218, 18, 19, 71, 240, 4, 20, 72, 321, 106, 107]. From a technical point of view, our proof follows closely the arguments in [20], adapted to the presence of errors and the variability of the parameters.

First, we prove below Proposition 3.2.4, which summarizes the first step mentioned above. In view to express precisely the capture/stability property we need in  $\Gamma_\eta(\bar{x}, \delta)$ , let us introduce an auxiliary property. Given a point  $\bar{x} \in \mathbf{H}$ , and two radius  $0 < r < R < +\infty$ , we say that  $\mathbf{S}(\bar{x}, r, R)$  holds if:

- i) for each  $k \in \mathbb{N}$ ,  $\{x_0, \dots, x_k\} \subset \underline{\Gamma}_\eta(\bar{x}, r)$  implies  $x_{k+1} \in \underline{\Gamma}_\eta(\bar{x}, R)$ ,
- ii) the initial point  $x_0$  belongs to  $\Gamma_\eta(\bar{x}, r)$  and

$$(3.17) \quad \|\bar{x} - x_0\| + 2\sqrt{\frac{f(x_0) - f(\bar{x})}{a_0}} + M\varphi(f(x_0) - f(\bar{x})) + \sum_{i=1}^{+\infty} \epsilon_i < r.$$

Basically, assuming  $\mathbf{S}(\bar{x}, r, R)$  means that the initialization  $x_0$  is close enough to  $\bar{x}$ , and that we can control how can  $x_{k+1}$  escape from  $\underline{\Gamma}_\eta(\bar{x}, r)$ .

**Proposition 3.2.4.** Let  $f : \mathbf{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a lower semi-continuous function, satisfying the KL inequality in  $\Gamma_\eta(\bar{x}, R)$ . Let  $(x_k)_{k \in \mathbb{N}}$  be a sequence such that  $\mathbf{H}_0$ ,  $\mathbf{H}_1$ ,  $\mathbf{H}_2$  and  $\mathbf{S}(\bar{x}, r, R)$  hold. Then:

- i)  $\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| < \infty$ .
- ii) The sequence  $(x_k)_{k \in \mathbb{N}}$  remains in the local upper level set  $\underline{\Gamma}_\eta(\bar{x}, r)$ , and converges strongly to some  $x_\infty \in \text{cl} B(\bar{x}, r)$ .
- iii)  $\liminf_{k \rightarrow +\infty} \|\partial^F f(x_k)\|_- = 0$ .
- iv)  $f(x_\infty) \leq f(\bar{x}) = \lim_{k \rightarrow +\infty} f(x_k)$ .

*Proof of Proposition 3.2.4.* The key point is obtaining the finite length property of item i). For this, we will prove estimations on the sequence by using the involved hypotheses.

We start with the following claim: if  $x_k$  and  $x_{k+1}$  belong to  $\underline{\Gamma}_\eta(\bar{x}, r)$ , then

$$(3.18) \quad 2\|x_{k+1} - x_k\| \leq \|x_k - x_{k-1}\| + \frac{1}{a_k b_k} [\varphi(f(x_k) - f(\bar{x})) - \varphi(f(x_{k+1}) - f(\bar{x}))] + \epsilon_k.$$

If  $x_{k+1} = x_k$  this inequality holds trivially, so we can assume that  $x_{k+1} \neq x_k$ . Using  $\mathbf{H}_1$  and  $x_{k+1} \in \underline{\Gamma}_\eta(\bar{x}, r)$  gives us  $f(x_k) > f(x_{k+1}) \geq f(\bar{x})$ . With  $x_k \in \Gamma_\eta(\bar{x}, r) \subset \Gamma_\eta(\bar{x}, R)$ , we are allowed to use the KL inequality, together with  $\mathbf{H}_2$ , to obtain:

$$(3.19) \quad \varphi'(f(x_k) - f(\bar{x})) \geq \frac{1}{\|\partial^F f(x_k)\|_-} \geq \frac{b_k}{\|x_k - x_{k-1}\| + \epsilon_k} > 0.$$

As a desingularizing function,  $\varphi$  is concave and  $\varphi'$  is positive, which implies with  $\mathbf{H}_1$  and (3.19):

$$\begin{aligned} \varphi(f(x_k) - f(\bar{x})) - \varphi(f(x_{k+1}) - f(\bar{x})) &\geq \varphi'(f(x_k) - f(\bar{x}))(f(x_k) - f(x_{k+1})) \\ &\geq \varphi'(f(x_k) - f(\bar{x})) a_k \|x_{k+1} - x_k\|^2 \\ &\geq a_k b_k \frac{\|x_{k+1} - x_k\|^2}{\|x_k - x_{k-1}\| + \epsilon_k}. \end{aligned}$$

Whence

$$\|x_{k+1} - x_k\|^2 \leq [\|x_k - x_{k-1}\| + \epsilon_k] \left[ \frac{1}{a_k b_k} [\varphi(f(x_k) - f(\bar{x})) - \varphi(f(x_{k+1}) - f(\bar{x}))] \right].$$

Taking the square root on both sides, and using the fact that  $2\sqrt{\alpha\beta} \leq \alpha + \beta$ , we obtain (3.18) as desired.

We are now ready to estimate the length of the sequence  $(x_k)_{k \in \mathbb{N}}$ . For this, we claim that for all  $K \in \mathbb{N}^*$ ,  $x^K$  belongs to  $\underline{\Gamma}_\eta(\bar{x}, r)$ , and

$$(3.20) \quad \sum_{k=1}^K \|x_{k+1} - x_k\| + \|x_{K+1} - x_K\| \leq \|x_1 - x_0\| + M [\varphi(f(x_1) - f(\bar{x})) - \varphi(f(x_{K+1}) - f(\bar{x}))] + \sum_{k=1}^K \epsilon_k,$$

where  $M := \sup_{k \in \mathbb{N}^*} \frac{1}{a_k b_k}$ . We proceed for this by induction on  $K$ .

*Initial step:* Since it is assumed in  $\mathbf{S}(\bar{x}, r, R)$ (ii) that  $x_0 \in \Gamma_\eta(\bar{x}, r)$ , we prove with  $\mathbf{S}(\bar{x}, r, R)$ (i) that  $x_1 \in \underline{\Gamma}_\eta(\bar{x}, R)$ . Using in particular  $f(x_1) \geq f(\bar{x})$ , coupled with  $\mathbf{H}_1$ , we can write

$$(3.21) \quad \|x_1 - x_0\| \leq \sqrt{\frac{f(x_0) - f(x_1)}{a_0}} \leq \sqrt{\frac{f(x_0) - f(\bar{x})}{a_0}}.$$

Using  $\mathbf{S}(\bar{x}, r, R)$ (ii) together with the triangle inequality and (3.21), we deduce that

$$\|\bar{x} - x_1\| \leq \|\bar{x} - x_0\| + \|x_0 - x_1\| \leq \|\bar{x} - x_0\| + \sqrt{\frac{f(x_0) - f(\bar{x})}{a_0}} < r,$$

and so  $x_1 \in B(\bar{x}, r)$ . Now we have  $x_0, x_1 \in \underline{\Gamma}_\eta(\bar{x}, r)$ , so we can use (3.18) with  $k = 1$ , and  $\mathbf{H}_3$ (iii), to obtain

$$2\|x_2 - x_1\| \leq \|x_1 - x_0\| + M [\varphi(f(x_1) - f(\bar{x})) - \varphi(f(x_2) - f(\bar{x}))] + \epsilon_1.$$

*Induction step:* Suppose now that  $x_1, \dots, x_K \in \underline{\Gamma}_\eta(\bar{x}, R)$ , and assume that (3.20) holds. It is immediate from  $\mathbf{S}(\bar{x}, r, R)$ (i) that  $x_{K+1} \in \underline{\Gamma}_\eta(\bar{x}, R)$ . Using successively the triangular inequality, the equations (3.17) and (3.21), and  $\mathbf{S}(\bar{x}, r, R)$ (ii), we obtain:

$$(3.22) \quad \begin{aligned} & \|\bar{x} - x_{K+1}\| \\ & \leq \|\bar{x} - x_0\| + \|x_0 - x_1\| + \sum_{k=1}^K \|x_{k+1} - x_k\| \\ & \leq \|\bar{x} - x_0\| + 2\|x_0 - x_1\| + M [\varphi(f(x_1) - f(\bar{x})) - \varphi(f(x_{K+1}) - f(\bar{x}))] + \sum_{k=1}^K \epsilon_k \\ & \leq \|\bar{x} - x_0\| + 2\sqrt{\frac{f(x_0) - f(\bar{x})}{a_0}} + M [\varphi(f(x_1) - f(\bar{x})) - \varphi(f(x_{K+1}) - f(\bar{x}))] + \sum_{k=1}^K \epsilon_k \\ & < r. \end{aligned}$$

This last inequality (3.22) gives in particular  $x_{K+1} \in B(\bar{x}, r)$ , and so  $x_{K+1} \in \underline{\Gamma}_\eta(\bar{x}, r)$ . Furthermore, if we write (3.18) with  $k = K + 1$

$$2\|x_{K+2} - x_{K+1}\| \leq \|x_{K+1} - x_K\| + M [\varphi(f(x_{K+1}) - f(\bar{x})) - \varphi(f(x_{K+2}) - f(\bar{x}))] + \epsilon_{K+1},$$

and add it to (3.22), we obtain the desired result.

We just proved that the sequence  $(x_k)_{k \in \mathbb{N}}$  remains in the local level set  $\underline{\Gamma}_\eta(\bar{x}, r)$ . Furthermore, we deduce from (3.20) and the positivity of  $\varphi$  that, for all  $K \in \mathbb{N}^*$ :

$$\sum_{k=1}^K \|x_{k+1} - x_k\| \leq \|x_1 - x_0\| + M\varphi(f(x_1) - f(\bar{x})) + \sum_{k=1}^K \epsilon_k.$$

Because of the assumption  $\mathbf{H}_3$ (iv), we can deduce by taking the limit when  $K \rightarrow +\infty$  that

$$\sum_{k=1}^{+\infty} \|x_{k+1} - x_k\| \leq \|x_1 - x_0\| + M\varphi(f(x_1) - f(\bar{x})) + \sum_{k=1}^{+\infty} \epsilon_k,$$

which proves item i). In particular, this finite length property entails the strong convergence of the sequence to some  $x_\infty \in \mathbf{H}$ . Since we proved that the sequence remains in  $\underline{\Gamma}_\eta(\bar{x}, r)$ , item ii) follows. To prove item iii), use  $\mathbf{H}_2$  to write

$$\sum_{k=1}^{\infty} b_{k+1} \|\partial^F f(x_{k+1})\|_- \leq \sum_{k=1}^{\infty} \|x_{k+1} - x_k\| + \sum_{k=1}^{\infty} \epsilon_{k+1} < +\infty,$$

and conclude with the assumption  $(b_k) \notin \ell^1$  in  $\mathbf{H}_3$ . To prove item iv), observe in  $\mathbf{H}_1$  that  $f(x_k)$  is decreasing, and use the lower semi-continuity of  $f$ , to deduce that

$$f(x_\infty) \leq \lim_{k \rightarrow +\infty} f(x_k).$$

Defining  $\ell := \lim_{k \rightarrow +\infty} f(x_k)$ , we need to show that  $\ell = f(\bar{x})$ . First observe that  $(x_k)_{k \in \mathbb{N}} \subset \underline{\Gamma}_\eta(\bar{x}, r)$  entails directly  $f(\bar{x}) \leq \ell$ . Moreover, if we assume by contradiction that  $f(\bar{x}) < \ell$ , the KL inequality together with the fact that  $\varphi'$  is decreasing would give

$$\varphi'(\ell - f(\bar{x})) \|\partial^F f(x_k)\|_- \geq \varphi'(f(x_k) - f(\bar{x})) \|\partial^F f(x_k)\|_- \geq 1$$

for all  $k \in \mathbb{N}$ . This would contradict item iii), so  $\ell = f(\bar{x})$ .  $\blacksquare$

We are now in position to prove Theorems 3.2.2 and 3.2.3. For this, all we need is verifying the stability assumption  $\mathbf{S}(\bar{x}, r, R)$ , and apply Proposition 3.2.4. In both cases, the fact that the abstract method described in  $\mathbf{H}_1$  and  $\mathbf{H}_2$  is a *descent* method for  $f$  is essential.

*Proof of Theorem 3.2.2.* Recall that a sequence is  $f$ -precompact if it admits a  $f$ -convergent subsequence. Let  $x_{n_k} \xrightarrow[k \rightarrow +\infty]{} \bar{x}$  with  $f(x_{n_k}) \xrightarrow[k \rightarrow +\infty]{} f(\bar{x})$ . Since  $f(x_k)$  is nonincreasing and admits a limit point, we deduce that  $f(x_k) \downarrow f(\bar{x})$ . In particular, we have  $f(\bar{x}) \leq f(x_k)$  for all  $k \in \mathbb{N}$ . The function  $f$  satisfies the KL inequality on  $\Gamma_\eta(\bar{x}, R)$  for some  $\eta \in ]0, +\infty]$  and  $\underline{a} > 0$ , with desingularizing function  $\varphi$ . Let  $K_0 \in \mathbb{N}$  be sufficiently large so that  $f(x_{K_0}) - f(\bar{x}) < \min\{\eta, \underline{a}R^2\}$ , and pick  $r \in ]0, R[$  such that  $f(x_{K_0}) - f(\bar{x}) < \underline{a}(R - r)^2$ . Hence, for all  $k \geq K_0$ ,

$$f(\bar{x}) \leq f(x_{k+1}) < f(\bar{x}) + \eta$$

and

$$\|x_{k+1} - x_k\| \leq \sqrt{\frac{f(x_k) - f(x_{k+1})}{a_k}} \leq \sqrt{\frac{f(x_{K_0}) - f(\bar{x})}{\underline{a}}} < R - r.$$

Now take  $K \geq K_0$  such that

$$\|\bar{x} - x_K\| + 2\sqrt{\frac{f(x_K) - f(\bar{x})}{a_{n_K}}} + M\varphi(f(x_K) - f(\bar{x})) + \sum_{k=K+1}^{+\infty} \epsilon_k < r.$$

Then, the sequence  $(y_k)_{k \in \mathbb{N}}$  defined by  $y_k = x_{K+k}$  satisfies  $\mathbf{S}(\bar{x}, r, R)$ , and so, the hypotheses of Proposition 3.2.4. In particular, the sequence  $(x_k)_{k \in \mathbb{N}}$  converges to some  $x_\infty$ , and Proposition 3.2.4 gives us  $\liminf_{k \rightarrow +\infty} \|\partial^F f(x_k)\|_- = 0$ . We conclude from the definition of the limiting subdifferential that  $0 \in \partial^L f(x_\infty)$ .  $\blacksquare$



**Remark 3.2.5.** In the previous proof, it suffices for  $f$  to have the KL property at any  $\bar{x}$  such that  $x_{n_k} \xrightarrow[k \rightarrow +\infty]{} \bar{x}$  and  $f(x_{n_k}) \xrightarrow[k \rightarrow +\infty]{} f(\bar{x})$ .

*Proof of Theorem 3.2.3.* Since  $f$  has the KL property in  $\bar{x}$ , there is a strict local upper level set  $\Gamma_\eta(\bar{x}, R)$ , where the KL inequality holds with  $\varphi$  as a desingularising function. Take  $r = \frac{2}{3}R$  and, if necessary, shrink  $\eta$  so that

$$(3.23) \quad 2\sqrt{\frac{\eta}{a}} + M\varphi(\eta) < \frac{r}{2}$$

This is possible since  $\varphi$  is continuous in 0, with  $\varphi(0) = 0$ .

Choose now to take  $x_0 \in \underline{\Gamma}_\eta(\bar{x}, \frac{r}{2})$ . It suffices to verify that  $\mathbf{S}(\bar{x}, r, R)$  is fulfilled and to use Proposition 3.2.4. It is easy to see, under our assumptions on  $x_0$  and  $\eta$ , that  $\mathbf{S}(\bar{x}, r, R)$ (ii) is verified. Indeed, use the monotonicity of  $\varphi$  and  $x_0 \in \underline{\Gamma}_\eta(\bar{x}, \frac{r}{2})$ , together with (3.23), to write

$$\|x_0 - \bar{x}\| + 2\sqrt{\frac{f(x_0) - f(\bar{x})}{a_0}} + M\varphi(f(x_0) - f(\bar{x})) < \frac{r}{2} + 2\sqrt{\frac{\eta}{a}} + M\varphi(\eta) < r.$$

For item i), let us assume that  $x_0, \dots, x_k$  lie in  $\underline{\Gamma}_\eta(\bar{x}, r)$  and prove that  $x_{k+1} \in \underline{\Gamma}_\eta(\bar{x}, R)$ . Since  $\bar{x}$  is a global minimum, from  $\mathbf{H}_1$  and the fact that  $(f(x_k))_{k \in \mathbb{N}}$  is decreasing, we obtain

$$f(\bar{x}) + a\|x_{k+1} - x_k\|^2 \leq f(x_{k+1}) + a\|x_{k+1} - x_k\|^2 \leq f(x_k) \leq f(x_0) < f(\bar{x}) + \eta.$$

In other words,

$$\|x_{k+1} - x_k\| \leq \sqrt{\frac{\eta}{a}} < \frac{r}{2}.$$

It follows, using the inequality above and the triangular inequality, that

$$\|x_{k+1} - \bar{x}\| \leq \|x_{k+1} - x_k\| + \|x_k - \bar{x}\| < \frac{r}{2} + r = R,$$

and so  $x_{k+1} \in \underline{\Gamma}_\eta(\bar{x}, R)$ . ■

### 3.3 Rates of convergence for an abstract descent method

Along this section, it is assumed that  $(x_k)_{k \in \mathbb{N}}$  is a sequence  $f$ -converging to a point  $x_\infty$ , where  $f$  verifies the KL property with a desingularizing function  $\varphi$ . We assume that  $\mathbf{H}_1$ ,  $\mathbf{H}_2$  and  $\mathbf{H}_3$  hold, and for simplicity and precision, we restrict ourselves to the case where  $\varepsilon_k \equiv 0$ . We study three types of convergence rate results, depending on the nature of the desingularising function  $\varphi$ :

- i) Theorem 3.3.1 establishes the relationship between the distance to the limit  $\|x_k - x_\infty\|$  and the gap  $f(x_k) - f(x_\infty)$ , for a generic desingularising function. It is similar to the result in [71, Theorem 24] for the proximal method in the convex case.
- ii) Theorem 3.3.2 gives explicit convergence rates in terms of the parameters – both for the distance and the gap – when the desingularising function is of the form  $\varphi(t) = \frac{C}{\theta}t^\theta$  with  $C > 0$  and  $\theta \in ]0, 1]$ . Several results obtained in the literature for various methods are recovered. In particular, according to [240], these convergence rates are optimal for  $\theta \in ]0, \frac{1}{2}]$ .
- iii) Finally, Theorem 3.3.4 provides convergence rates when  $\mathbf{H}_2$  is replaced by a slightly different hypothesis that holds for certain explicit schemes, namely gradient-related methods. This result is valid for a generic desingularising function  $\varphi$ . However, when  $\varphi$  is of the form  $\varphi(t) = \frac{C}{\theta}t^\theta$  ( $C > 0$ ,  $\theta \in ]0, 1]$ ) the prediction is considerably better than the one provided by Theorem 3.3.2.

### 3.3.1 Distance to the limit in terms of the gap

**Theorem 3.3.1.** *Set  $\tilde{\varphi}(t) := \max\{\varphi(t), \sqrt{t}\}$ . Then,  $\|x_\infty - x_k\| = O(\tilde{\varphi}(f(x_{k-1}) - f(x_\infty)))$ .*

*Proof.* By assumption,  $x_k \xrightarrow[k \rightarrow +\infty]{f} x_\infty$  and  $f$  satisfies the KL inequality on some  $\Gamma_\eta(x_\infty, \delta)$ . Let  $r_k := f(x_k) - f(x_\infty) \geq 0$ . We may suppose that  $r_k > 0$  for all  $k \in \mathbb{N}$ , because otherwise the algorithm terminates in a finite number of steps. For  $K$  large enough, we have  $x_k \in \Gamma_\eta(x_\infty, \delta)$  for all  $k \geq K$ . Then, we can use the estimation (3.18) to write (recall that  $\epsilon_k \equiv 0$ ):

$$2\|x_{k+1} - x_k\| \leq \|x_k - x_{k-1}\| + M[\varphi(r_k) - \varphi(r_{k+1})].$$

Summing this inequality for  $k = K, \dots, N$ , we obtain

$$\sum_{k=K}^N \|x_{k+1} - x_k\| \leq \|x_K - x_{K-1}\| + M\varphi(r_K).$$

Using the triangle inequality and passing to the limit, we get

$$\|x_\infty - x_K\| \leq \sum_{k=K}^{\infty} \|x_{k+1} - x_k\| \leq \|x_K - x_{K-1}\| + M\varphi(r_K) \leq \frac{\sqrt{f(x_{K-1}) - f(x_K)}}{\sqrt{a_K}} + M\varphi(r_K)$$

by  $\mathbf{H}_1$ . Then, using  $\mathbf{H}_0$ , along with the fact that  $f(x_K) \geq f(x_\infty)$  and that  $(r_k)$  is decreasing, we deduce that  $\|x_\infty - x_K\| \leq \frac{1}{\sqrt{a}} \sqrt{r_{K-1}} + M\varphi(r_{K-1})$ , which gives the desired result.  $\blacksquare$

### 3.3.2 Explicit rates when $\varphi(t) \propto t^\theta$ with $\theta \in ]0, 1]$

We assume now that the desingularizing function of  $f$  at  $x_\infty$  has the form  $\varphi(t) = \frac{C}{\theta} t^\theta$  for some  $C > 0$ ,  $\theta \in ]0, 1]$ . This holds for instance when  $f$  is semi-algebraic, or strongly convex (see the end of Section 3.1.2). Theorem 3.3.2 below is qualitatively analogous to the results in [18, 240, 19, 321, 72, 107]: we prove convergence in a finite number of steps if  $\theta = 1$ , exponential convergence if  $\theta \in [\frac{1}{2}, 1[$  and polynomial convergence if  $\theta \in ]0, \frac{1}{2}[$ . In the general convex case, finite-time termination of the proximal point algorithm was already proved in [287] and [157] (see also [266]).

**Theorem 3.3.2.** *Assume  $\varphi(t) = \frac{C}{\theta} t^\theta$  for some  $C > 0$ ,  $\theta \in ]0, 1]$ .*

*i) If  $\theta = 1$  and  $\inf_{k \in \mathbb{N}} a_k b_{k+1}^2 > 0$ , then  $x_k$  converges in finite time.*

*ii) If  $\theta \in [\frac{1}{2}, 1[$ ,  $\sup_{k \in \mathbb{N}} b_k < +\infty$  and  $\inf_{k \in \mathbb{N}} a_k b_{k+1} > 0$ , there exist  $c > 0$  and  $k_0 \in \mathbb{N}$  such that:*

*ii.a)  $f(x_k) - f(x_\infty) = O\left(\exp\left(-c \sum_{n=k_0}^{k-1} b_{n+1}\right)\right)$ , and*

*ii.b)  $\|x_\infty - x_k\| = O\left(\exp\left(-\frac{c}{2} \sum_{n=k_0}^{k-2} b_{n+1}\right)\right)$ .*

*iii) If  $\theta \in ]0, \frac{1}{2}[$ ,  $\sup_{k \in \mathbb{N}} b_k < +\infty$  and  $\inf_{k \in \mathbb{N}} a_k b_{k+1} > 0$ , there is  $k_0 \in \mathbb{N}$  such that:*

*iii.a)  $f(x_k) - f(x_\infty) = O\left(\left(\sum_{n=k_0}^{k-1} b_{n+1}\right)^{\frac{-1}{1-2\theta}}\right)$ , and*

*iii.b)  $\|x_\infty - x_k\| = O\left(\left(\sum_{n=k_0}^{k-2} b_{n+1}\right)^{\frac{-\theta}{1-2\theta}}\right)$ .*

**Remark 3.3.3.** Note that a simple sufficient – yet not necessary – condition for  $\inf_{k \in \mathbb{N}} a_k b_{k+1}^2 > 0$  and  $\inf_{k \in \mathbb{N}} a_k b_{k+1} > 0$  is that  $\inf_{k \in \mathbb{N}} b_k > 0$ .

*Proof.* As before, we use the notation  $r_k := f(x_k) - f(x_\infty) \geq 0$ . We can assume that  $r_k > 0$  for all  $k \in \mathbb{N}$ , because otherwise the algorithm would terminate in a finite number of steps. Since  $x_k$  converges to  $x_\infty$ , there exists  $k_0 \in \mathbb{N}$  such that, for all  $k \geq k_0$ ,  $x_k$  remains in  $\Gamma_\eta(x_\infty, \delta)$ , where the KL inequality holds. Using successively  $\mathbf{H}_1$ ,  $\mathbf{H}_2$  and the KL inequality, we obtain

$$\forall k \geq k_0, \quad \varphi'^2(r_{k+1})(r_k - r_{k+1}) \geq \varphi'^2(r_{k+1})a_k b_{k+1}^2 \|\partial^F f(x_{k+1})\|_-^2 \geq a_k b_{k+1}^2.$$

Equivalently, using the fact that  $\varphi(t) = \frac{C}{\theta} t^\theta$ ,

$$(3.24) \quad \forall k \geq k_0, \quad C r_{k+1}^{2\theta-2} (r_k - r_{k+1}) \geq a_k b_{k+1}^2.$$

Let us now consider different cases for  $\theta$ :

Case  $\theta = 1$ : Recall that we assume that  $r_k > 0$  for all  $k \in \mathbb{N}$ , and deduced from that (3.24), rewritten here with  $\theta = 1$ :

$$\forall k \geq k_0, \quad C^2 (r_k - r_{k+1}) \geq a_k b_{k+1}^2 \geq \inf_{k \in \mathbb{N}} a_k b_{k+1}^2 > 0.$$

Since  $r_k$  converges, we must have  $\inf_{k \in \mathbb{N}} a_k b_{k+1}^2 = 0$ , which is a contradiction. Therefore, it must exist some  $k \in \mathbb{N}$  such that  $r_k = 0$ , which means that the algorithm terminates in a finite number of steps.

Case  $\theta \in ]0, 1[$ : This case covers both items ii) and iii). We only have to prove the convergence rates on the values, and the rates for the iterates will follow Theorem 3.3.1. Write  $\bar{b} := \sup_{k \in \mathbb{N}} b_k$ ,

$m := \inf_{k \in \mathbb{N}} a_k b_{k+1}$  and  $c = \frac{m}{C^2(1+\bar{b})}$  and, for each  $k \in \mathbb{N}$ ,  $\beta_k := \frac{b_k m}{C^2}$ . For each  $k \geq k_0$ , (3.24) can be rewritten, after dividing by  $C^2 r_{k+1}^{2\theta-2}$ :

$$(3.25) \quad (r_k - r_{k+1}) \geq \frac{a_k b_{k+1}^2 r_{k+1}^{2-2\theta}}{C^2} \geq \beta_{k+1} r_{k+1}^{2-2\theta}.$$

Subcase  $\theta \in [\frac{1}{2}, 1[$ : Since  $r_k$  tends to zero and  $0 < 2 - 2\theta \leq 1$ , we may assume, by enlarging  $k_0$  if necessary, that  $r_{k+1}^{2-2\theta} \geq r_{k+1}$  for all  $k \geq k_0$ . Inequality (3.25) implies  $(r_k - r_{k+1}) \geq \beta_{k+1} r_{k+1}$  or, equivalently,  $r_{k+1} \leq r_k \left( \frac{1}{1 + \beta_{k+1}} \right)$  for all  $k \geq k_0$ . By induction, we obtain

$$r_{k+1} \leq r_{k_0} \left( \prod_{n=k_0}^k \frac{1}{1 + \beta_{n+1}} \right) = r_{k_0} \exp \left( \sum_{n=k_0}^k \ln \left( \frac{1}{1 + \beta_{n+1}} \right) \right)$$

for all  $k \geq k_0$ . Moreover, using a classic estimation gives

$$\ln \left( \frac{1}{1 + \beta_{n+1}} \right) \leq \frac{-\beta_{n+1}}{1 + \beta_{n+1}} \leq \frac{-1}{1 + \bar{b}} \beta_{n+1}.$$

We can deduce the desired convergence rate by using the definitions of  $\beta_k$  and  $c$ :

$$r_{k+1} \leq r_{k_0} \exp \left\{ \sum_{n=k_0}^k \left( \frac{-1}{1 + \bar{b}} \beta_{n+1} \right) \right\} = r_{k_0} \exp \left( -c \sum_{n=k_0}^k b_{n+1} \right).$$

Subcase  $\theta \in ]0, \frac{1}{2}[$ : Recall from inequality (3.25) that  $r_{k+1}^{2\theta-2} (r_k - r_{k+1}) \geq \beta_{k+1}$ .

Set  $\phi(t) := \frac{C}{1-2\theta} t^{2\theta-1}$ . Then  $\phi'(t) = -C t^{2\theta-2}$ , and

$$\phi(r_{k+1}) - \phi(r_k) = \int_{r_k}^{r_{k+1}} \phi'(t) dt = C \int_{r_{k+1}}^{r_k} t^{2\theta-2} dt \geq C (r_k - r_{k+1}) r_k^{2\theta-2}.$$

We distinguish now two cases. On one hand, if we suppose that  $r_{k+1}^{2\theta-2} \leq 2r_k^{2\theta-2}$ , then

$$(3.26) \quad \phi(r_{k+1}) - \phi(r_k) \geq \frac{C}{2}(r_k - r_{k+1})r_{k+1}^{2\theta-2} \geq \frac{C}{2}\beta_{k+1}.$$

On the other hand, assume that  $r_{k+1}^{2\theta-2} > 2r_k^{2\theta-2}$ . Since  $2\theta - 2 < 2\theta - 1 < 0$ , we have  $\frac{2\theta-1}{2\theta-2} > 0$ . Thus  $r_{k+1}^{2\theta-1} > qr_k^{2\theta-1}$ , where  $q := 2^{\frac{2\theta-1}{2\theta-2}} > 1$ . Therefore,

$$\phi(r_{k+1}) - \phi(r_k) = \frac{C}{1-2\theta}(r_{k+1}^{2\theta-1} - r_k^{2\theta-1}) > \frac{C}{1-2\theta}(q-1)r_k^{2\theta-1} \geq C',$$

with  $C' := \frac{C}{1-2\theta}(q-1)r_{k_0}^{2\theta-1} > 0$ . Since  $\beta_{k+1} \leq \frac{bm}{C^2}$ , we can write

$$(3.27) \quad \phi(r_{k+1}) - \phi(r_k) \geq \frac{C'C^2}{bm}\beta_{k+1}.$$

Setting  $c := \min\{\frac{C}{2}, \frac{C'C^2}{bm}\} > 0$  we can write by using (3.26) and (3.27)

$$\forall k \geq k_0, \quad \phi(r_{k+1}) - \phi(r_k) \geq c\beta_{k+1}.$$

This implies, after using  $\phi(r_{k_0}) \geq 0$ ,

$$\phi(r_{k+1}) \geq \phi(r_{k+1}) - \phi(r_{k_0}) = \sum_{n=k_0}^k \phi(r_{n+1}) - \phi(r_n) \geq c \sum_{n=k_0}^k \beta_{n+1},$$

which is precisely  $r_{k+1} \leq D \left( \sum_{n=k_0}^k b_{n+1} \right)^{\frac{-1}{1-2\theta}}$  with  $D = \left( \frac{cm(1-2\theta)}{C^3} \right)^{\frac{-1}{1-2\theta}}$ . ■

### 3.3.3 Sharper results for gradient-related methods

Convergence rates for the continuous subgradient differential inclusion

$$(3.28) \quad \dot{u}(t) + \partial^F f(u(t)) \ni 0,$$

are given in Theorem 3.1.12. For any desingularizing function  $\varphi$  on  $]0, \eta[$ , this result states that for some  $t_1 \in \mathbb{R}$ ,

$$\text{i) } f(u(t)) - f(u_\infty) = O(\Phi^{-1}(t - t_1)), \text{ and}$$

$$\text{ii) } \|u_\infty - u(t)\| = O(\varphi \circ \Phi^{-1}(t - t_1)),$$

where  $\Phi$  is any primitive of  $-(\varphi')^2$  on  $]0, \eta[$ . If the desingularising function  $\varphi$  is of the form  $\varphi(t) = \frac{Ct^\theta}{\theta}$ , we recover (see Remark 3.3.5 later) convergence in finite time if  $\theta \in ]\frac{1}{2}, 1]$ , exponential convergence if  $\theta = \frac{1}{2}$ , and polynomial convergence if  $\theta \in ]0, \frac{1}{2}[$ . The same conclusion was established in [68, Theorem 4.7] for a nonsmooth version of (3.28) when  $f$  is any subanalytic function on  $H = \mathbb{R}^n$ . This prediction is better than the one given by Theorem 3.3.2 above, as well as the results in [18, 240, 19, 321, 72, 107], since it guarantees convergence in finite time for  $\theta > \frac{1}{2}$ . We shall prove that, for certain algorithms including gradient-related methods, this better estimation remains true. To this end, consider the following variant of hypothesis  $\mathbf{H}_2$ :

$\mathbf{H}'_2$  (Relative error): For each  $k \in \mathbb{N}$ ,  $b_{k+1}\|\partial^F f(x_k)\|_- \leq \|x_{k+1} - x_k\|$ .

**Theorem 3.3.4.** *Let condition  $\mathbf{H}'_2$  be satisfied instead of  $\mathbf{H}_2$ , and assume  $m := \inf_{k \in \mathbb{N}} a_k b_{k+1} > 0$ .*

*Let  $\Phi : ]0, \eta[ \rightarrow \mathbb{R}$  be any primitive of  $-(\varphi')^2$ .*

i) If  $\lim_{t \rightarrow 0} \Phi(t) \in \mathbb{R}$ , then the algorithm converges in a finite number of steps.

ii) If  $\lim_{t \rightarrow 0} \Phi(t) = +\infty$ , then there exists  $k_0 \in \mathbb{N}$  such that:

ii.a)  $f(x_k) - f(x_\infty) = O\left(\Phi^{-1}\left(m \sum_{n=k_0}^{k-1} b_{n+1}\right)\right)$ , and

ii.b)  $\|x_\infty - x_k\| = O\left(\tilde{\varphi} \circ \Phi^{-1}\left(m \sum_{n=k_0}^{k-1} b_{n+1}\right)\right)$ , with  $\tilde{\varphi}(t) := \max\{\varphi(t), \sqrt{t}\}$ .

*Proof.* The following proof is inspired by the one of [104] in the continuous case. First, we claim that, if  $r_k := f(x_k) - f(x_\infty) > 0$  for all  $k \in \mathbb{N}$ , then there is  $k_0 \in \mathbb{N}$  such that

$$(3.29) \quad \forall k \geq k_0, \quad \Phi(r_{k+1}) \geq \Phi(r_{k_0}) + m \sum_{n=k_0}^k b_{n+1}.$$

To see this, let  $k_0$  be large enough to have  $x_k \in \Gamma_\eta(x_\infty, \delta)$ , where the KL inequality holds, for all  $k \geq k_0$ . We apply successively  $\mathbf{H}_1$ ,  $\mathbf{H}'_2$ , the KL inequality and  $\mathbf{H}_3$  to obtain

$$\varphi'(r_k)^2 (r_k - r_{k+1}) \geq \varphi'(r_k)^2 a_k b_{k+1}^2 \|\partial^F f(x_k)\|_-^2 \geq a_k b_{k+1}^2 \geq b_{k+1} m.$$

Let  $\Phi$  be a primitive of  $-(\varphi')^2$  on  $]0, \eta[$ . Then, because  $\varphi$  is increasing and concave, and using the assumption  $r_k > 0$ , we can write

$$\Phi(r_{k+1}) - \Phi(r_k) = \int_{r_{k+1}}^{r_k} \varphi'(t)^2 dt \geq (r_k - r_{k+1}) \varphi'(r_k)^2 \geq b_{k+1} m.$$

Therefore,

$$\Phi(r_{k+1}) - \Phi(r_{k_0}) = \sum_{n=k_0}^k \Phi(r_{n+1}) - \Phi(r_n) \geq m \sum_{n=k_0}^k b_{n+1}$$

as claimed. Let us now distinguish the two cases:

For item i), we argue by contradiction. If  $r_k > 0$  for all  $k \in \mathbb{N}$ , then (3.29), together with the hypothesis  $(b_k)_{k \in \mathbb{N}} \notin \ell^1$  in  $\mathbf{H}_3$ , imply  $\lim_{k \rightarrow +\infty} \Phi(r_{k+1}) = +\infty$ . This contradicts the fact that  $\lim_{t \rightarrow 0} \Phi(t) \in \mathbb{R}$ . Hence,  $r_k = 0$  for some  $k \in \mathbb{N}$  and  $\mathbf{H}_1$  ensures that the sequence remains stationary there.

For ii), we may suppose that  $r_k > 0$  for all  $k \in \mathbb{N}$  (otherwise the algorithm stops in a finite number of steps), and so (3.29) holds for all  $k \in \mathbb{N}$ . Since  $\lim_{k \rightarrow +\infty} \Phi(r_k) = +\infty$ , we can take

$k_0$  large enough to have  $\Phi(r_{k_0}) > 0$ , whence  $\Phi(r_{k+1}) \geq m \sum_{n=k_0}^k b_{n+1}$ . Since  $(b_n) \notin \ell^1$ , for all

sufficiently large  $k$ ,  $m \sum_{n=k_0}^k b_{n+1}$  is in the domain of  $\Phi^{-1}$  and we obtain the first estimation, namely:

$$(3.30) \quad r_{k+1} \leq \Phi^{-1}\left(m \sum_{n=k_0}^k b_{n+1}\right).$$

For the second one, since  $\varphi$  is concave and differentiable, we have by  $\mathbf{H}_1$

$$\varphi(r_k) - \varphi(r_{k+1}) \geq \varphi'(r_k)(r_k - r_{k+1}) \geq \varphi'(r_k) a_n \|x_{k+1} - x_k\|^2.$$

The KL property and  $\mathbf{H}'_2$  then give  $\varphi(r_k) - \varphi(r_{k+1}) \geq m \|x_{k+1} - x_k\|$ , which in turn yields

$$\|x_\infty - x_k\| \leq \frac{1}{m} \sum_{n=k}^{\infty} [\varphi(r_n) - \varphi(r_{n+1})] \leq \frac{1}{m} \varphi(r_k).$$

We conclude by using (3.30). ■

**Remark 3.3.5.** Let us analyse the results of Theorem 3.3.4 when  $\varphi(t)$  has the form  $\frac{C}{\theta}t^\theta$ . We assume that  $C = 1$ , for simplicity. In that case, we have  $\varphi'^2(t) = \frac{-1}{t^{2-2\theta}}$ .

- If  $\theta \in ]\frac{1}{2}, 1[$ , then we can write  $\varphi'^2(t) = -t^{\alpha-1}$  with  $\alpha := 2\theta - 1 \in ]0, 1[$ . Hence, we can take  $\Phi(t) = 1 - \frac{1}{\alpha}t^\alpha$ , such that  $\Phi(t) \xrightarrow[t \downarrow 0]{} 1$ . This entails the convergence of the sequence in a finite number of steps, which clearly improves Theorem 3.3.2. This convergence in finite time is natural, if one considers that the convergence rate is governed by  $\Phi^{-1}$ , which, in that case, goes to zero in finite time.
- If  $\theta = \frac{1}{2}$ , then we have  $\varphi'^2(t) = \frac{-1}{t}$ . In that case, take  $\Phi(t) = -\ln(t) \xrightarrow[t \downarrow 0]{} +\infty$ . Thus, the convergence rates of the values and the iterates are respectively governed by

$$\Phi^{-1}(t) = \exp(-t) \text{ and } (\varphi \circ \Phi)(t) = \exp\left(-\frac{1}{2}t\right).$$

- If  $\theta \in ]0, \frac{1}{2}[$ , we write  $\varphi'^2(t) = \frac{-1}{t^{1+\alpha}}$  with  $\alpha := 1 - 2\theta \in ]0, 1[$ . So, consider  $\Phi(t) = \frac{1}{\alpha}t^\alpha$ , which tends to  $+\infty$  when  $t \downarrow 0$ . We can deduce that, in that case, the convergence rates of the values and the iterates are respectively governed by

$$\Phi^{-1}(t) = \alpha^{\frac{-1}{\alpha}} t^{\frac{-1}{\alpha}} = O\left(t^{\frac{-1}{1-2\theta}}\right) \text{ and } (\varphi \circ \Phi)(t) = O\left(t^{\frac{-\theta}{1-2\theta}}\right).$$

We clearly recover in the last two points the estimates of Theorem 3.3.2.

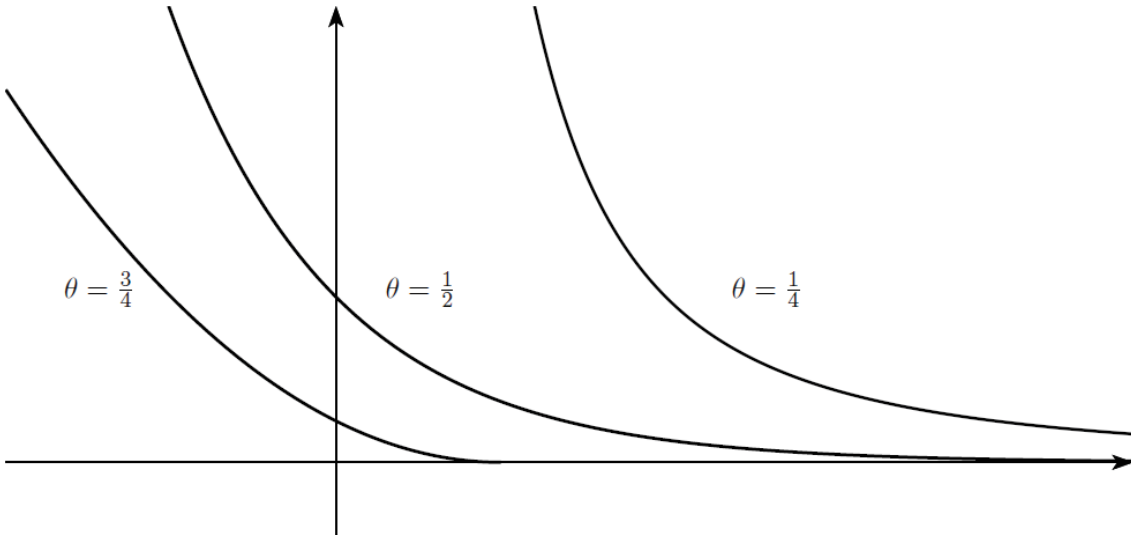


Figure 3.1: Asymptotic behaviour of  $\Phi^{-1}$ , for different values of  $\theta$

### 3.4 Comments and perspectives

**Remark 3.4.1** (On the convergence in a finite number of iterations). In Theorem 3.3.4, we obtained the convergence in finite time of  $x_k$  to  $x_\infty$ , under the assumption that  $\mathbf{H}_2$  is replaced by  $\mathbf{H}'_2$ . It is clear that such hypothesis is verified if the sequence  $(x_k)_{k \in \mathbb{N}}$  is generated by the gradient method

$$(3.31) \quad x_{k+1} = x_k - \lambda_k \nabla f(x_k),$$

or more generally by gradient-related methods (see Example 3.2.1). At first look,  $\mathbf{H}'_2$  seems to refer to *explicit* methods, in opposition with implicit schemes (like the proximal method).

Obtaining convergence in finite time for an explicit scheme is quite surprising, since it is a property that we usually know for implicit methods, under some conditions. Take for instance  $f = \delta_C$ , the indicator function of a closed convex set  $C$ , for which the proximal algorithm converges in one single step. More generally, it is known that the proximal method converges in a finite number of steps whenever  $f$  is a lower semi-continuous function, which is *sharp* around its minimum points (see [287, 157, 266]):

$$\exists \sigma > 0, \forall \bar{x} \in \operatorname{argmin} f, \forall x \in \mathbf{H}, \quad f(x) - f(\bar{x}) \geq \sigma \|x - \bar{x}\|.$$

In fact, it must be remembered that we also need the  $\mathbf{H}_1$  assumption. For explicit methods like in (3.31), we usually derive  $\mathbf{H}_1$  from the descent lemma, which asks the function  $f$  to be of class  $C^{1,1}$ . So the question is the following: is it compatible for a function to be at the same time of class  $C^{1,1}$ , and satisfying the KL inequality with a Łojasiewicz coefficient  $\theta > \frac{1}{2}$ ? We exhibit a simple example illustrating that these two properties are hardly compatible.

Let  $\alpha \geq 1$  and  $f_\alpha : \mathbb{R} \rightarrow \mathbb{R}$  be defined by  $f_\alpha(x) := \alpha|x|^\alpha$ . Its derivative is given by  $f'_\alpha(x) = \alpha|x|^{\alpha-1}$ . Clearly,

- if  $\alpha \geq 2$ ,  $f_\alpha$  is of class  $C^\infty$ ,
- if  $1 \leq \alpha < 2$ ,  $f_\alpha$  is of class  $C^1$  but  $f'_\alpha$  is not Lipschitz continuous around zero.

Thus, satisfying  $\mathbf{H}_1$  would require  $\alpha \geq 2$ . On an other hand, let us discuss the KL property around zero. We look for a constant  $C > 0$  and a Łojasiewicz exponent  $\theta \in ]0, 1]$  such that

$$\forall x \sim 0, \quad f_\alpha(x)^{1-\theta} \leq C |f'_\alpha(x)|,$$

or equivalently,

$$\forall x \sim 0, \quad \alpha|x|^{\alpha(1-\theta)} \leq C|x|^{\alpha-1}.$$

This can be rewritten as

$$\forall x \sim 0, \quad |x|^{1-\alpha\theta} \leq \frac{C}{\alpha}$$

This boundedness assumption on  $|x|^{1-\alpha\theta}$  around zero is satisfied if and only if  $\alpha\theta \leq 1$ . Hence, if we want the KL inequality to hold with  $\theta > \frac{1}{2}$ , we would have in particular that  $\alpha < 2$ .

From this example, and the fact that explicit gradient methods are not used to converge in a finite number of steps, one can conjecture that the functions satisfying the KL inequality with a Łojasiewicz exponent  $\theta > \frac{1}{2}$  cannot have a Lipschitz continuous gradient<sup>2</sup>.

**Remark 3.4.2** (On the asymptotic equivalence between continuous and discrete dynamics). Doing the comparison between a continuous dynamic and its discrete counterpart is always fruitful, since each of them enjoys its own advantages/difficulties. In the convex setting, the correspondence of the asymptotic behaviour between the proximal algorithm and the steepest descent dynamic has been recently established [265, 13, 14, 269]. In our KL setting, we observed in Section 3.3 that the convergence rates between the continuous and discrete dynamics are the same. Furthermore, it is clear that the proofs of Theorems 3.1.8 and 3.2.2 work on the same arguments. It is thus natural to wonder if some asymptotic equivalence can be established in the KL context, like in the convex case.

In fact, such a correspondence does exist. In [71, Theorem 39], the authors consider a KL convex function of class  $C^{1,1}$   $f : \mathbf{H} \rightarrow \mathbb{R}$ , which has compact level sets. Thus, they coined a link between the finite length of the trajectories of

$$\dot{u}(t) + \nabla f(u(t)) = 0$$

---

<sup>2</sup>I became aware only very recently of a work of Bégout, Bolte and Jendoubi, in which the authors verify partially this conjecture [55, Proposition 2.8].

and the sequences satisfying the abstract descent scheme

$$a\|x_{k+1} - x_k\| \|\nabla f(x_k)\| \leq f(x_k) - f(x_{k+1}).$$

In [71, Remark 40], the authors observe that the convexity assumption can be dropped, provided that the desingularizing function  $\varphi$  of  $f$  is concave. Recall that this concavity is required in our Definition 3.1.1, so this convex hypothesis on  $f$  is not necessary in our setting. It is an interesting problem for the future to investigate whether this asymptotic equivalence can be extended to nonsmooth functions.





## Chapter 4

# Splitting methods for KL functions

As stressed in [20], the abstract scheme developed in Chapter 3 covers, among others, the gradient-related methods (a wide variety of schemes based on the gradient method sketched in [96], see also Example 3.2.1), the proximal algorithm [236, 86, 287], and the forward-backward algorithm (a combination of the preceding, see [228, 260]). This last one is a splitting method, used to solve structured optimization problems with the form

$$(4.1) \quad \underset{x \in \mathbf{H}}{\text{minimize}} \quad f(x) = g(x) + h(x),$$

where  $g$  is a nonsmooth, proper and lower semi-continuous function, and  $h$  is differentiable with a  $L$ -Lipschitz gradient. Basically, the forward-backward method consists in performing a gradient descent step with respect to  $h$ , followed by a proximal step with respect to  $g$ . It has been studied in the nonsmooth and nonconvex setting in [20], and the algorithm was stated as follows: starting with  $x_0 \in \mathbf{H}$ , consider  $(\lambda_k)_{k \in \mathbb{N}} \subset [\underline{\lambda}, \bar{\lambda}]$  with  $0 < \underline{\lambda} \leq \bar{\lambda} < \frac{1}{L}$ , and  $\forall k \in \mathbb{N}$

$$(4.2) \quad x_{k+1} \in \text{prox}_{\lambda_k g}(x_k - \lambda_k \nabla h(x_k)).$$

It satisfies  $\mathbf{H}_1$ ,  $\mathbf{H}_2$  and  $\mathbf{H}_3$  (see [20, Theorem 5.1]) and falls into the setting of Theorem 3.2.2. In section 4.1, we shall extend this class of algorithms in different directions.

We allow the consideration of an alternative choice of metric for the ambient space, which may vary at each step (see [11, 12] and the references therein). Let  $S_{++}(\mathbf{H})$  denote the space of bounded, uniformly elliptic and self-adjoint operators on  $\mathbf{H}$ . Each  $A \in S_{++}(\mathbf{H})$  induces a metric on  $\mathbf{H}$  by the inner product  $\langle x, y \rangle_A := \langle Ax, y \rangle$ , and the norm  $\|x\|_A := \sqrt{\langle x, x \rangle_A}$ . Thus, the proximal operator of  $f$  in the metric induced by  $A$  is the set-valued mapping  $\text{prox}_f^A : \mathbf{H} \rightrightarrows \mathbf{H}$ , defined as

$$(4.3) \quad \text{prox}_f^A(x) := \underset{y \in \mathbf{H}}{\text{argmin}} \quad \left\{ f(y) + \frac{1}{2} \|y - x\|_A^2 \right\}.$$

Observe that  $\text{prox}_f^A(x) \neq \emptyset$  if  $f$  is weakly lower semi-continuous and bounded from below [21, Theorem 3.2.5], which holds in many relevant applications. If  $f$  is the indicator function of a set, then  $\text{prox}_f^A(x)$  is the *projection mapping* relatively to the metric induced by  $A$ . If  $A = \frac{1}{\lambda} \text{id}_{\mathbf{H}}$ , we just note  $\text{prox}_{\lambda f}$  instead of  $\text{prox}_f^{\frac{1}{\lambda} \text{id}_{\mathbf{H}}}$ . Considering metrics induced by a sequence  $(A_k)_{k \in \mathbb{N}}$  in  $S_{++}(\mathbf{H})$ , the forward-backward method becomes

$$(4.4) \quad x_{k+1} \in \text{prox}_g^{A_k}(x_k - A_k^{-1} \nabla h(x_k)).$$

Observe that (4.4) can be rewritten as

$$(4.5) \quad x_{k+1} \in \underset{y \in \mathbf{H}}{\text{argmin}} \quad g(y) + h(x_k) + \langle y - x_k, \nabla h(x_k) \rangle + \frac{1}{2} \langle y - x_k, A_k(y - x_k) \rangle.$$

At each step, an approximation of  $f$ , replacing its smooth part  $h$  by a quadratic model, is minimized. See [106] for a similar algorithm called Variable Metric Forward-Backward, and [253] for an approach considering more general models. Note that, when  $A_k = \frac{1}{\lambda_k} \text{id}_{\mathbb{H}}$ , one recovers (4.2). Allowing variable metric can improve convergence rates, help to implicitly deal with certain constraints, or compensate the effect of ill-conditioning. Rather than simply giving a convergence result for a general choice of  $A_k$ , we handle, in Subsection 4.1.3, a detailed method to select these operators, using a second-order information on  $h$ . It can be seen as a *generalized Levenberg-Marquardt algorithm*, a Newton-like method adapted for nonconvex and nonsmooth functions.

We also want to effectively solve structured problems as

$$(4.6) \quad \underset{x_1 \in \mathbb{H}_1, x_2 \in \mathbb{H}_2}{\text{minimize}} \quad f(x_1, x_2) = g_1(x_1) + g_2(x_2) + h(x_1, x_2),$$

where  $g_1, g_2$  are nonsmooth, proper and lower semi-continuous functions, and  $h$  is differentiable with Lipschitz gradient. One approach is the regularized Gauss-Seidel method, which exploits the fact that the variables are separated in the nonsmooth part of  $f$  [20, 19, 321]. It consists in minimizing alternatively a regularized version of  $f$  with respect to each variable. In other words, it is an alternating proximal algorithm, of the form:

$$\begin{cases} x_{1,k+1} \in \text{prox}_{f(\cdot, x_{2,k})}(x_{1,k}) \\ x_{2,k+1} \in \text{prox}_{f(x_{1,k+1}, \cdot)}(x_{2,k}). \end{cases}$$

However this algorithm does not exploit the smooth nature of  $h$ . An alternative is to use an alternating minimization method which can deal with the nonsmooth character, while it benefits from the smooth features.

We present such a method, considering variable metrics, in Section 4.1. More exactly, we revisit the *Alternating Forward-Backward* methods, already considered in [72, 107, 224], but we point out that our setting differs from these two works on the following points:

- We allow more flexibility in the choice of parameters, accounting, in particular, for vanishing step sizes or unbounded eigenvalues for the metrics.
- We allow additive errors. Indeed, the computation of  $\tilde{x}_k := x_k - A_k^{-1} \nabla h(x_k)$  and  $x_{k+1} \in \text{prox}_g^{A_k}(\tilde{x}_k)$  often require solving some subroutines, which may produce  $\tilde{x}_k$  and  $x_{k+1}$  inexactly. To take these errors into account we introduce two sequences  $(r_k)_{k \in \mathbb{N}}$  and  $(s_k)_{k \in \mathbb{N}}$ , and consider

$$(4.7) \quad x_{k+1} - s_{k+1} \in \text{prox}_g^{A_k}(x_k - A_k^{-1} \nabla h(x_k) + r_k).$$

Convergence of the Alternating Forward-Backward method with errors is given in Theorem 4.1.6. This result is the consequence of the convergence for the abstract scheme studied in Section 3.2, as it is stated in Proposition 4.1.3.

The Alternating Forward-Backward method provides a framework suitable for the numerical resolution of a wide variety of structured problems. Section 4.2 is devoted to a discussion on some applications of this method. In Section 4.2.1, we adapt the AFB method to Feasibility problems involving eventually nonconvex constraints, considering semi-algebraic constraints, or doing a regularity hypothesis on the intersection (see also [222, 20]). Consider for instance the problems arising in image processing and data compression, which are generally semi-algebraic by nature [149, 150, 127]. Indeed, they generally involve the semi-algebraic *counting function*  $\|x\|_0 := \#\{i \mid x_i \neq 0\}$ , whose proximal operator is easily implementable (see Section 4.2.2.1). Such problems can also involve the rank function, like the sparse and low-rank matrix decomposition problem (see Section 4.2.2.2). They are also well suited for our analysis since the proximal operator of the rank function admits an explicit formulation [194]. We end with some numerical results in Section 4.2.2.3, solving the data compression problem with a seemingly new Hard Shrinkage Projection algorithm.

## 4.1 Splitting methods with errors and variable metric

### 4.1.1 The Alternating Forward-Backward (AFB) Method

Let  $H_1, \dots, H_p$  be Hilbert spaces, each  $H_i$  provided with its own inner product  $\langle \cdot, \cdot \rangle_{H_i}$  and norm  $\| \cdot \|_{H_i}$ . If there is no ambiguity, we will just note  $\|x_i\|$  instead of  $\|x_i\|_{H_i}$ . Set  $H := \prod_{i=1}^p H_i$  and endow it with the inner product  $\langle \cdot, \cdot \rangle := \sum_{i=1}^p \langle \cdot, \cdot \rangle_{H_i}$  and the associated norm  $\| \cdot \| := \sqrt{\langle \cdot, \cdot \rangle}$ . Consider the problem

$$(4.8) \quad \underset{x_i \in H_i}{\text{minimize}} \quad f(x_1, \dots, x_p) := h(x_1, \dots, x_p) + \sum_{i=1}^p g_i(x_i),$$

where  $h : H \rightarrow \mathbb{R}$  is continuously differentiable and each  $g_i : H_i \rightarrow \mathbb{R} \cup \{+\infty\}$  is a lower semi-continuous function. Moreover, we suppose that there is  $L \geq 0$  such that, for each  $(x_1, \dots, x_p) \in H$  and  $i \in \{1, \dots, p\}$ , the application

$$(4.9) \quad x \in H_i \mapsto h(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_p)$$

has a  $L$ -Lipschitz continuous gradient on bounded sets.

We shall present an algorithm that generates sequences converging to critical points of  $f$ . The sequences will be updated cyclically, meaning that, given  $(x_{1,k}, \dots, x_{p,k})$ , we start by updating the first variable  $x_{1,k}$  into  $x_{1,k+1}$ , and then we use  $(x_{1,k+1}, x_{2,k}, \dots, x_{p,k})$  to update the second variable, and so on. In order to have concise and clear notations, throughout this section we shall denote:

$$(4.10) \quad X_k := (x_{1,k}, \dots, x_{p,k}) \text{ and } X_{i,k} := (x_{1,k+1}, \dots, x_{i-1,k+1}, x_{i,k}, \dots, x_{p,k}).$$

Observe that  $X_{1,k} = X_k$  and that we can write  $X_{p+1,k} = X_{k+1}$ .

Let us now present the Alternate Forward-Backward (AFB) algorithm. As said before, it consists in doing a forward-backward step relatively to each variable, taking in account a possibly different metric. Then, for all  $i \in \{1, \dots, p\}$ , consider a sequence  $(A_{i,k})_{k \in \mathbb{N}} \subset S_{++}(H_i)$  which will model the metrics. Given a starting point  $X_0 \in H$ , the AFB algorithm generates a sequence  $(X_k)_{k \in \mathbb{N}}$  by taking, for all  $k \in \mathbb{N}$  and  $i \in \{1, \dots, p\}$ ,

$$(AFB) \quad x_{i,k+1} \in \text{prox}_{g_i^{A_{i,k}}} \left( x_{i,k} - A_{i,k}^{-1} \nabla_i h(X_{i,k}) \right).$$

Here  $\nabla_i h$  denotes the  $i$ -th component of  $\nabla h$  in  $H = \prod_{i=1}^p H_i$ . We shall consider some hypotheses on the operators  $A_{i,k}$ . For  $A \in S_{++}(H_i)$ , we set  $\alpha(A)$  as the infimum of the spectral values of  $A$ , being the best constant  $\|x\|_A^2 \geq \alpha(A) \|x\|^2$  for all  $x \in H$ . We define then  $\alpha_k := \min_{i=1..p} \alpha(A_{i,k})$  and  $\beta_k := \max_{i=1..p} \|A_{i,k}\|$ , which give bounds on the spectral values of  $(A_{i,k})_{i=1..p}$ . We make the following hypotheses on the parameters  $\alpha_k$  and  $\beta_k$ :

$$(HP) \quad \begin{array}{l} 1. \text{ There exists } \underline{\alpha} > 0 \text{ such that } \alpha_k \geq \underline{\alpha} > L \\ 2. \frac{1}{\beta_k} \notin \ell^1 \quad 3. \sup_{k \in \mathbb{N}} \frac{\beta_k}{\alpha_{k+1}} < +\infty. \end{array}$$

**Remark 4.1.1.** Here **HP**<sub>1</sub> is a bound on the spectral values by the Lipschitz constant of the gradient of  $h$ , in order to enforce the descent property of the sequence. For operators of the form  $\frac{1}{\lambda_{i,k}} id_{H_i}$ , we recover the classical bound  $L\lambda_{i,k} \leq L\bar{\lambda} < 1$ . In [107], the authors prove that, with an additional convexity assumption on the  $g_i$ 's, and boundedness of the parameters, one

can consider  $L\lambda_{i,k} \leq L\bar{\lambda} < 2$ . Item **HP**<sub>2</sub> states that the spectral values may diverge, but not too fast. Finally, **HP**<sub>3</sub> can be seen as an hypothesis on the variations of the extreme spectral values of the chosen operators. It clearly holds, for instance, if  $\beta_k$  is bounded. It is also sufficient to assume that the condition numbers

$$\kappa_{i,k} := \frac{\|A_{i,k}\|}{\alpha(A_{i,k})}$$

are bounded, with also  $\min\left\{\frac{\alpha_k}{\alpha_{k+1}}, \frac{\beta_k}{\beta_{k+1}}\right\}$  remaining bounded.

**Remark 4.1.2.** Even if  $\nabla h$  is globally Lipschitz continuous,  $L$  is not the Lipschitz constant of  $\nabla h$ , but a common Lipschitz constant for the functions defined in (4.9). As a consequence, the partial gradients  $\nabla_i h$  are  $\sqrt{p}L$ -Lipschitz continuous while  $\nabla h$  is  $pL$ -Lipschitz. This allows us to have a better bound in **HP**<sub>1</sub> which is of particular importance in the applications (see Section 4.2). In [72], the authors give a more precise analysis: at each substep  $X_{i,k}$  of the algorithm, they consider  $L_{i,k}$  as the Lipschitz constant of the gradient of  $x \in H_i \mapsto h(x_{1,k+1}, \dots, x_{i-1,k+1}, x, x_{i+1,k}, \dots, x_{p,k})$ . Then they take step sizes equal to  $\lambda_{i,k} = \frac{\epsilon_i}{L_{i,k}}$  where  $\epsilon_i < 1$  is a fixed and nonnegative constant. This approach can be related to the one in [106, 107]. However, they suppose *a priori* that the values  $L_{i,k}$  remain bounded. It would be interesting to know if it is possible to combine the two approaches (a variable Lipschitz constant and vanishing step sizes).

#### 4.1.2 The AFB Method with errors

In order to allow for approximate computation of the descent direction or the proximal mapping, we go further by considering an inexact AFB method. We introduce the sequences  $(r_{i,k})$  and  $(s_{i,k})$  for  $i \in \{1, \dots, p\}$ , which correspond respectively to errors arising at the explicit and implicit steps, relatively to the variable  $x_i$ . The *AFB method with Errors* is computed from an initial  $(x_{1,0}, \dots, x_{p,0}) \in H$  by

$$\begin{aligned} \text{(AFBE)} \quad y_{i,k+1} &\in \text{prox}_{g_i}^{A_{i,k}} \left( x_{i,k} - A_{i,k}^{-1} \nabla_i h(X_{i,k}) + r_{i,k} \right), \\ x_{i,k+1} &= y_{i,k+1} + s_{i,k+1}. \end{aligned}$$

We do specific hypothesis on the errors, in view to guarantee the convergence of the method. Observe in particular that we do not assume *a priori* that the errors converge to zero:

$$\begin{aligned} \text{(HE)} \quad &\text{There exists } \sigma \in [0, +\infty[, \rho \in ]0, 1] \text{ with } \frac{\sigma+1}{\rho} < \alpha L^{-1} \text{ such that} \\ &1. \|S_{i,k}\| \leq \frac{\sigma}{2} \|y_{i,k+1} - y_{i,k}\|, \text{ with } S_{i,k} \text{ defined from } (s_{i,k}) \text{ as in (4.10),} \\ &2. \|r_{i,k}\| \leq \frac{\sigma}{2} \|y_{i,k+1} - y_{i,k}\| + \mu_k, \text{ where } \mu_k \geq 0 \text{ with } \mu_k \in \ell^1, \\ &3. \langle r_{i,k} + s_{i,k}, y_{i,k+1} - y_{i,k} \rangle_{A_{i,k}} \leq \frac{1-\rho}{2} \|y_{i,k+1} - y_{i,k}\|_{A_{i,k}}^2. \end{aligned}$$

This AFB algorithm (with errors) is related to the abstract descent method studied in Section 3.2, as stated in the next proposition.

**Proposition 4.1.3.** Any bounded sequence  $Y_k = (y_{1,k}, \dots, y_{p,k})$  generated by the AFB algorithm with errors, which verify **HP** and **HE**, satisfies **H**<sub>1</sub>, **H**<sub>2</sub> and **H**<sub>3</sub>.

*Proof.* Since  $X_{i,k} = Y_{i,k} + S_{i,k}$ , we can rewrite the algorithm as

$$(4.11) \quad y_{i,k+1} \in \text{prox}_{g_i}^{A_{i,k}} (y_{i,k} - A_{i,k}^{-1} \nabla_i h(Y_{i,k} + S_{i,k}) + r_{i,k} + s_{i,k}).$$

We start by showing that **H**<sub>1</sub> is satisfied. Let  $i \in \{1, \dots, p\}$  be fixed; using the definition of the proximal operator  $\text{prox}_{g_i}^{A_{i,k}}$  in (4.11) and developing the squared norms gives

$$g_i(y_{i,k}) - g_i(y_{i,k+1}) \geq \frac{1}{2} \|y_{i,k+1} - y_{i,k}\|_{A_{i,k}}^2 + \langle y_{i,k+1} - y_{i,k}, \nabla_i h(Y_{i,k} + S_{i,k}) \rangle - \langle y_{i,k+1} - y_{i,k}, r_{i,k} + s_{i,k} \rangle_{A_{i,k}}.$$

Using **HE**<sub>3</sub> in the inequality above, together with the definition of  $\alpha_k$ , it results in

$$(4.12) \quad g_i(y_{i,k}) - g_i(y_{i,k+1}) \geq \frac{\rho\alpha_k}{2} \|y_{i,k+1} - y_{i,k}\|^2 + \langle y_{i,k+1} - y_{i,k}, \nabla_i h(Y_{i,k} + S_{i,k}) \rangle.$$

For fixed  $k \in \mathbb{N}$  and  $i \in \{1, \dots, p\}$ , introduce the function

$$\tilde{h}_{i,k} : y_i \in \mathbf{H}_i \mapsto (y_{1,k+1}, \dots, y_{i-1,k+1}, y_i, y_{i+1,k}, \dots, y_{p,k}) \in \mathbb{R},$$

which satisfies  $\tilde{h}_{i,k}(y_{i,k}) = h(Y_{i,k})$ ,  $\tilde{h}_{i,k}(y_{i,k+1}) = h(Y_{i+1,k})$  and  $\nabla \tilde{h}_{i,k}(y_{i,k}) = \nabla_i h(Y_{i,k})$ . Applying the descent lemma 2.1.13 to  $\tilde{h}_{i,k}$ , we obtain

$$(4.13) \quad h(Y_{i+1,k}) - h(Y_{i,k}) - \langle y_{i,k+1} - y_{i,k}, \nabla_i h(Y_{i,k}) \rangle \leq \frac{L}{2} \|y_{i,k+1} - y_{i,k}\|^2.$$

Then, combining (4.12) and (4.13) we get

$$(4.14) \quad \begin{aligned} & \frac{\rho\alpha_k - L}{2} \|y_{i,k+1} - y_{i,k}\|^2 + \langle y_{i,k+1} - y_{i,k}, \nabla_i h(Y_{i,k} + S_{i,k}) - \nabla_i h(Y_{i,k}) \rangle \\ & \leq g_i(y_{i,k}) - g_i(y_{i,k+1}) + h(Y_{i,k}) - h(Y_{i+1,k}), \end{aligned}$$

Using successively the Cauchy-Schwarz inequality, the Lipschitz property of  $\nabla_i h$  (see Remark 4.1.2) and **HE**<sub>1</sub>, one gets

$$-\frac{\sigma L}{2\sqrt{p}} \|y_{i,k+1} - y_{i,k}\|^2 \leq \langle y_{i,k+1} - y_{i,k}, \nabla_i h(Y_{i,k} + S_{i,k}) - \nabla_i h(Y_{i,k}) \rangle.$$

By inserting this estimation in (4.14), we deduce that

$$(4.15) \quad \frac{\rho\alpha_k - L(\frac{\sigma}{\sqrt{p}} + 1)}{2} \|y_{i,k+1} - y_{i,k}\|^2 \leq g_i(y_{i,k}) - g_i(y_{i,k+1}) + h(Y_{i,k}) - h(Y_{i+1,k}).$$

By summing (4.15) for  $i \in \{1, \dots, p\}$ , we obtain

$$\frac{\rho\alpha_k - L(\frac{\sigma}{\sqrt{p}} + 1)}{2} \sum_{i=1}^p \|y_{i,k+1} - y_{i,k}\|^2 \leq \sum_{i=1}^p g_i(y_{i,k}) - g_i(y_{i,k+1}) + h(Y_{i,k}) - h(Y_{i+1,k}),$$

which is exactly **H**<sub>1</sub> with  $a_k = \frac{\rho\alpha_k - L(\frac{\sigma}{\sqrt{p}} + 1)}{2}$ :

$$\frac{\rho\alpha_k - L(\frac{\sigma}{\sqrt{p}} + 1)}{2} \|Y_{k+1} - Y_k\|^2 \leq f(Y_k) - f(Y_{k+1}).$$

To prove **H**<sub>2</sub>, fix  $i \in \{1, \dots, p\}$  and use Fermat's rule for the Fréchet subdifferential in (4.11) to get:

$$(4.16) \quad 0 \in \partial^F g_i(y_{i,k+1}) + \{A_{i,k}(y_{i,k+1} - y_{i,k}) - A_{i,k}(r_{i,k} + s_{i,k}) + \nabla_i h(Y_{i,k} + S_{i,k})\}.$$

Define  $w_{i,k+1} := \nabla_i h(Y_k) - \nabla_i h(Y_{i,k} + S_{i,k}) - A_{i,k}(y_{i,k+1} - y_{i,k}) + A_{i,k}(r_{i,k} + s_{i,k})$ , which lies in  $\partial^F g_i(y_{i,k+1}) + \nabla_i h(Y_{k+1})$ , by (4.16). The triangle inequality gives

$$(4.17) \quad \|w_{i,k+1}\| \leq \beta_k (\|y_{i,k+1} - y_{i,k}\| + \|r_{i,k}\| + \|s_{i,k}\|) + \|\nabla_i h(Y_{i,k} + S_{i,k}) - \nabla_i h(Y_{k+1})\|.$$

Using the error estimations from **HE**, and the  $\sqrt{p}L$ -Lipschitz continuity of  $\nabla_i h$  in (4.17), we obtain:

$$(4.18) \quad \|w_{i,k+1}\| \leq (\beta_k(1 + \sigma) + \sqrt{p}L\sigma) \|y_{i,k+1} - y_{i,k}\| + \sqrt{p}L \|Y_{k+1} - Y_k\| + \beta_k \mu_k.$$

Define now  $W_{k+1} := (w_{1,k+1}, \dots, w_{p,k+1}) \in \partial^F f(Y_{k+1})$  (recall the definition of  $w_{i,k+1}$ ). Then, through the sum over  $i \in \{1, \dots, p\}$  of inequality (4.18), we deduce (using  $\sqrt{p} \leq p \leq p^2$ )

$$\|W_{k+1}\| \leq \sum_{i=1}^p \|w_{i,k+1}\| \leq p\beta_k\mu_k + p^2(\beta_k + L)(1 + \sigma)\|Y_{k+1} - Y_k\|.$$

Hence  $\mathbf{H}_2$  is verified with  $b_{k+1} = \frac{1}{p^2(1+\sigma)(\beta_k+L)}$  and  $\epsilon_{k+1} = \frac{\beta_k\mu_k}{p(1+\sigma)(\beta_k+L)}$ .

Now we just need to check that the hypotheses  $\mathbf{H}_3$  are satisfied with our hypotheses on  $\alpha_k$ ,  $\beta_k$  and  $\mu_k$ . Clearly  $\mathbf{H}_3(i)$  holds since we've supposed that  $\alpha_k \geq \underline{\alpha} > (\sigma\frac{\sqrt{p}}{p} + 1)\frac{L}{\rho}$ . Then  $\mathbf{H}_3(ii)$  asks that  $b_k \notin \ell^1$ , which is equivalent to  $\frac{1}{\beta_k+L} \notin \ell^1$  in our context. This holds since we've supposed that  $\frac{1}{\beta_k} \notin \ell^1$ . Hypothesis  $\mathbf{H}_3(iii)$  is satisfied because  $\frac{\beta_k}{\alpha_{k+1}}$  is supposed to be bounded. Finally,  $\mathbf{H}_3(iv)$  asks the summability of  $\frac{\beta_k\mu_k}{\beta_k+L}$  which is bounded by  $\mu_k \in \ell^1$ .  $\blacksquare$

Given this result, one could directly apply Theorem 3.2.2 to obtain convergence of the sequence  $(Y_k)_{k \in \mathbb{N}}$  to a critical point of  $f$ . However this result would suffer from some drawbacks. First, we are expecting that  $(X_k)_{k \in \mathbb{N}}$  converges to a critical point, not  $(Y_k)_{k \in \mathbb{N}}$ . So we should make the assumption that the errors  $S_k := X_k - Y_k$  tend to zero. Moreover we would suppose that  $(Y_k)$  is  $f$ -precompact, while we may only have an access to  $(X_k)$ . To handle this, we make the link between the asymptotic behaviour of  $(Y_k)$  and  $(X_k)$ :

**Proposition 4.1.4.** For any bounded sequence generated by the AFB method with errors satisfying **HP** and **HE**:

- i) If  $(Y_k)_{k \in \mathbb{N}}$  has finite length, then so does  $(X_k)_{k \in \mathbb{N}}$ .
- ii) If  $(f(Y_k))_{k \in \mathbb{N}}$  is bounded from below, then for all  $i \in \{1, \dots, p\}$ ,  $\|s_{i,k}\|$  and  $\|r_{i,k}\|$  lie in  $\ell^2$ . In particular,  $(Y_k)_{k \in \mathbb{N}}$  and  $(X_k)_{k \in \mathbb{N}}$  share the same limit points.
- iii)  $(Y_k)_{k \in \mathbb{N}}$  is precompact if and only if  $(f(Y_k))_{k \in \mathbb{N}}$  is bounded from below and  $(X_k)_{k \in \mathbb{N}}$  is precompact.

*Proof.* Item i) comes directly from **HE**<sub>1</sub>. To prove item ii), we use Proposition 4.1.3: from  $\mathbf{H}_1$  and  $\mathbf{H}_3(i)$ , we have

$$(4.19) \quad \underline{a}\|Y_{k+1} - Y_k\|^2 \leq f(Y_k) - f(Y_{k+1}),$$

whence  $(f(Y_k))$  is a decreasing sequence. Then we can sum (4.19) to obtain that

$$(4.20) \quad \underline{a} \sum_{k \in \mathbb{N}} \|Y_{k+1} - Y_k\|^2 \leq f(Y_0) - \inf_{k \in \mathbb{N}} f(Y_k) < +\infty.$$

Since we have  $\|r_{i,k}\| \leq \frac{\sigma}{2}\|y_{i,k+1} - y_{i,k}\| + \mu_k$  where  $\mu_k \in \ell^1$ , and  $\|y_{i,k+1} - y_{i,k}\| \leq \|Y_{k+1} - Y_k\|$  which is in  $\ell^2$ , we deduce that  $\|r_{i,k}\| \in \ell^2$ , and the same holds for  $\|s_{i,k}\|$ . So the errors converge to zero, and  $(X_k)$  and  $(Y_k)$  have the same limit points. Item iii) follows from item ii) and the following argument: suppose that we have a subsequence  $(Y_{n_k})$  converging to some  $Y_\infty = (y_{1,\infty}, \dots, y_{p,\infty})$  in  $H$ . Since  $f$  is lower semi-continuous and  $(f(Y_k))$  is decreasing, we have that  $\inf_{k \in \mathbb{N}} f(Y_k)$  is bounded from below by  $f(Y_\infty)$ .  $\blacksquare$

An other disadvantage of the direct application of Theorem 3.2.2 to Proposition 4.1.3 is that it asks the  $f$ -precompactness of  $(Y_k)_{k \in \mathbb{N}}$ . In some cases, precompactness of a sequence can be deduced using compact embeddings between Hilbert spaces. Sequences remaining in a sublevel set of an inf-compact function  $f$  are also precompact. However,  $f$ -precompactness is harder to obtain without further continuity assumption on  $f$ . Actually, both limit and  $f$ -limit points coincide whenever the parameters are bounded:

**Proposition 4.1.5.** If either  $\beta_k \leq \bar{\beta}$ , or  $f$  is continuous on its domain, then  $(Y_k)_{k \in \mathbb{N}}$  is  $f$ -precompact if and only if it is precompact.

*Proof.* Suppose that we have  $Y_{k_n}$  converging to  $Y_\infty$ , and show that  $f(Y_{k_n})$  converges also to  $f(Y_\infty)$ . Since  $f(Y_k)$  is decreasing and  $f$  is lower semi-continuous, we know that  $Y_\infty$  must lie in the domain of  $f$ . If  $f$  is continuous on its domain, the conclusion is immediate. On the other hand, suppose that  $\beta_k \leq \bar{\beta}$ . Since  $h$  is continuous, we only need to verify that  $\lim_{n \rightarrow +\infty} g_i(y_{i,k_n}) = g_i(y_{i,\infty})$  for each  $i \in \{1, \dots, p\}$ . By lower semi-continuity of  $g_i$ , we just have to prove that  $\limsup_{n \rightarrow +\infty} g_i(y_{i,k_n}) \leq g_i(y_{i,\infty})$ , following the ideas of [20].

Let  $n \in \mathbb{N}^*$  and  $k = k_n - 1$ ; by using the definition of the proximal operator, we have

$$\begin{aligned} & g_i(y_i^{k+1}) + \frac{1}{2} \|y_{i,k+1} - y_{i,k} + A_{i,k}^{-1} \nabla_i h(Y_{i,k} + S_{i,k}) - r_{i,k} - s_{i,k}\|_{A_{i,k}}^2 \\ & \leq g_i(y_i^\infty) + \frac{1}{2} \|y_i^\infty - y_{i,k} + A_{i,k}^{-1} \nabla_i h(Y_{i,k} + S_{i,k}) - r_{i,k} - s_{i,k}\|_{A_{i,k}}^2, \end{aligned}$$

and the latter implies (using Cauchy-Schwarz and  $\|A_{i,k}\| \leq \bar{\beta}$ ):

(4.21)

$$g_i(y_{i,k+1}) \leq g_i(y_{i,\infty}) + \frac{\bar{\beta}}{2} \|y_{i,\infty} - y_{i,k}\|^2 + \|y_{i,\infty} - y_{i,k+1}\| [\|\nabla_i h(Y_{i,k} + S_{i,k})\| + \bar{\beta} \|r_{i,k} + s_{i,k}\|].$$

Now recall that  $y_{i,k+1} = y_{i,k_n}$  tends to  $y_{i,\infty}$ , while  $r_{i,k} + s_{i,k}$  goes to zero (see Proposition 4.1.4). Observe also that  $\nabla_i h(Y_{i,k} + S_{i,k})$  is bounded since  $Y_{i,k+1} + S_{i,k}$  converges to  $Y_\infty$ . Moreover,  $\|y_{i,\infty} - y_{i,k}\|$  goes to zero since  $\|y_{i,\infty} - y_{i,k}\| \leq \|y_{i,\infty} - y_{i,k_n}\| + \|y_{i,k+1} - y_{i,k}\|$ , with  $y_{i,k_n} \rightarrow y_{i,\infty}$  and  $\|y_{i,k+1} - y_{i,k}\| \in \ell^2$  (see (4.20)). Passing to the upper limit in (4.21) leads finally to  $\limsup_{n \rightarrow +\infty} g_i(y_{i,k_n}) \leq g_i(y_{i,\infty})$ .  $\blacksquare$

As a direct consequence of Propositions 4.1.3, 4.1.4, 4.1.5, together with Theorem 3.2.2, we finally obtain our convergence result for the AFB algorithm with errors. It extends the results of [107] (when taking a cyclic permutation on the variables) in two directions: the functions  $g_i$  need not be continuous on their domain, or the step sizes can tend to 0.

**Theorem 4.1.6.** *Let  $f$  be a KL function. Let  $(Y_k)_{k \in \mathbb{N}}$  be a bounded precompact sequence generated by the AFB algorithm with errors, with **HP** and **HE** satisfied. Suppose that either  $\beta_k$  remains bounded, or that  $f$  is continuous on its domain. Hence, the sequence  $(X_k)_{k \in \mathbb{N}}$  has finite length, and converges toward a  $\partial^L$ -critical point of  $f$ .*

**Remark 4.1.7.** In the particular case where  $S_k \equiv 0$ , we know furthermore that the sequence  $(X_k)$  is convergent with respect to  $f$ . This is no longer true in general if  $f$  is not continuous and  $S_k \neq 0$ . As a simple counterexample, take  $f : x \in \mathbb{R} \mapsto |x|_0 \in \mathbb{R}$ , where  $|x|_0 = 0$  if  $x = 0$ ,  $|x|_0 = 1$  else. By taking as parameters  $A_k \equiv 2id$ ,  $r_k \equiv 0$ ,  $s_k = \frac{1}{k}$  and  $x_0 = 0$ , it is easy to see, after applying the AFB algorithm, that  $f(y_k) \equiv 0$  but  $f(x_k) \equiv 1$ .

An analog of the capture result in Theorem 3.2.3 can also be deduced:

**Theorem 4.1.8.** *Suppose that the KL property holds at a global minimum  $\bar{X}$  of  $f$ . Let  $(X_k)_{k \in \mathbb{N}}$  be a bounded sequence generated by the AFB algorithm with errors, satisfying **HP** and **HE** with  $\mu_k \equiv 0$ . Hence, there exist  $\delta > 0$  and  $\eta > 0$  such that, if  $X_0 \in \underline{\Gamma}_\eta(\bar{X}, \delta)$ , then  $(X_k)_{k \in \mathbb{N}}$  has finite length and converges to a global minimum of  $f$ .*

To prove this theorem, it suffices to use  $Y_0 = X_0$ , and to see at the end of the proof of Proposition 4.1.3 that  $\mu_k = 0$  iff  $\epsilon_k = 0$ , where  $\epsilon_k$  is the parameter involved in **H**<sub>3</sub>. Then, apply Theorem 3.2.3 together with Propositions 4.1.3 and 4.1.4.



### 4.1.3 Variable metric: Towards generalized newton methods

We focus here on the problem of minimizing a  $C^{1,1}$  function  $h : \mathbb{R}^N \rightarrow \mathbb{R}$  over a closed and nonempty set  $C \subset \mathbb{R}^N$ . The AFB algorithm reduces in this case to a projected-gradient method, and allow us to compute in the explicit step a descent direction governed by a chosen metric  $A_k$ . As an example, take  $h(x) = \frac{1}{2}\langle Ax, x \rangle - \langle b, x \rangle$  with  $A \in S_{++}(\mathbb{R}^N)$  being a symmetric definite positive matrix. In the unconstrained case, the Newton method (that is, taking  $A_k \equiv A$ ) is known to solve in a single step the problem. If we add a constraint  $C$ , it is easy to see that the Newton-projected method

$$(4.22) \quad x_{k+1} \in \text{proj}_C^{A_k} (x_k - A_k^{-1} \nabla h(x_k))$$

gives the minimum of  $h$  over  $C$  in one single step. For a general function  $h$ , (4.22) reduces to the minimisation over  $C$  of a quadratic model of  $h$ , as stressed in (4.5). One can see on this example that computing the proximal operator relatively to the metric  $A_n$ , used in the explicit step, (and not the ambient metric !) is of crucial importance in this method.

The spirit here is to use second-order information from  $h$  in order to improve the convergence of the method. In the unconstrained case, a popular choice of metric is given by Newton-like methods, where the metric at step  $k$  is induced by (an approximation of) the Hessian  $\nabla^2 h(x_k)$ . Since it is often impossible to know in advance whether or not the Hessian is uniformly elliptic at each  $x_k$ , a positive definite approximation has to be chosen.

We detail here a natural way to chose this positive definite  $A_k \sim \nabla^2 h(x_k)$  in closed loop, and show that this method remains in the setting of Theorem 4.1.6. Since it generalizes the *Levenberg-Marquardt* method used in the convex case (see [31]), we will refer to the *Generalized Levenberg-Marquardt* method for this way of designing  $A_k$ . One of the interesting aspect of the method is that such a matrix can be defined even if  $h$  is only  $C^{1,1}$  and not  $C^2$ , since the differentiability of  $\nabla h$  is not necessary in Theorem 4.1.6. Another interesting aspect is that the splitting approach led us to solve constrained minimization problems with a Newton-projected approach.

We set  $S_+(\mathbb{R}^N)$  as the closed convex cone of nonnegative  $N \times N$  matrices. Consider the generalized Hessian of  $h$ , by taking the generalized Jacobian of  $\nabla h$  in the sense of Clarke:

$$\partial^2 h(x) := \text{co} \left\{ \lim_{n \rightarrow +\infty} \nabla^2 h(x_n), \text{ where } \nabla h \text{ is differentiable at } x_n \text{ and } x_n \rightarrow x \right\}.$$

This set contains symmetric matrices bearing second-order information on  $h$ . Hence, the *Generalized Levenberg-Marquardt* method to compute  $A_k \in S_{++}(\mathbb{R}^N)$  from a given  $x_k \in \mathbb{R}^N$  is the following: for  $\varepsilon > 0$ ,

$$\begin{aligned} & \text{Take } H_k \in \partial^2 h(x_k), \\ & \text{Project } P_k := \text{proj}_{S_+(\mathbb{R}^N)}(H_k), \\ & \text{Regularize } A_k := P_k + \varepsilon I_N. \end{aligned}$$

A globalized version of the method can be considered by taking step sizes ensuring descent. Then, the following convergence result holds:

**Proposition 4.1.9.** Let  $f(x) := h(x) + \delta_C(x)$  be a KL function, where  $C \subset \mathbb{R}^N$  is closed and nonempty, and  $h$  is differentiable with a  $L$ -Lipschitz gradient. Let  $x_0 \in H$  and suppose that  $(x_k)$  is a bounded sequence generated by

$$x_{k+1} \in \text{proj}_C^{A_k} (x_k - \lambda_k A_k^{-1} \nabla h(x_k)),$$

where  $A_k$  is selected with the Generalized Levenberg-Marquardt process detailed above, and the stepsizes  $\lambda_k$  satisfy:

$$0 < \lambda_k \leq \bar{\lambda} < \frac{\varepsilon}{L}, \quad \lambda_k \notin \ell^1 \quad \text{and} \quad \sup_{k \in \mathbb{N}} \frac{\lambda_{k+1}}{\lambda_k} < +\infty.$$

Then, the sequence  $(x_k)_{k \in \mathbb{N}}$  has finite length and is converging to a critical point of  $f$ .

*Proof.* Start by observing that  $\text{proj}_C^{A_k} = \text{proj}_C^{\lambda_k^{-1}A_k}$ , so the algorithm falls in the setting of the AFB algorithm. According with the previous notations,  $\nabla h$  being  $L$ -Lipschitz continuous implies that the sequence  $(\|H_k\|)_{k \in \mathbb{N}}$  is bounded by  $L$ , and so  $(\|P_k\|)_{k \in \mathbb{N}}$  remains bounded by  $2L$ . To conclude through Theorem 4.1.6, we just need to check the hypotheses **HP** on the parameters  $\frac{1}{\lambda_k}A_k$ . We have here  $\alpha_k = \alpha(\frac{1}{\lambda_k}A_k) \geq \varepsilon\lambda_k^{-1} \geq \varepsilon\bar{\lambda}^{-1} > L$  and  $\beta_k = \|\frac{1}{\lambda_k}A_k\| \leq (2L + \varepsilon)\lambda_k^{-1}$ . Thus **HP**<sub>1</sub> is satisfied, while items **HP**<sub>2</sub> and **HP**<sub>3</sub> follow directly from the hypotheses made on  $(\lambda_k)_{k \in \mathbb{N}}$ . Since the indicator function  $\delta_C$  is continuous on its domain, the hypotheses of Theorem 4.1.6 are satisfied. ■

This extends, in a way, results from the convex setting to the nonconvex one, enforcing moreover the strong convergence (see [31, Theorem 7.1]).

A drawback of this method is that the Hessian increases the complexity of implementation since a matrix must be inverted in the explicit step. An alternative is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update scheme (see [65],[299]), using only first-order information to compute the inverse of the Hessian. On the other hand, the implicit step gains also in complexity since one must project onto a constraint relatively to a given metric, which is nontrivial even for simple constraints. For linear constraints, a particular second-order model of the Hessian can be taken in order to reduce the implicit step in a trivial orthogonal projection step (see [299, 167, 64]).

Newton-like methods are expected to have good convergence rates in exchange for a more expensive implementation. An interesting question is whether one can obtain convergence rates beyond the results in Subsection 3.3, by exploiting, not only the KL nature of the function, but also the specific properties of the matrices selected by the Generalized Levenberg-Marquardt process.

## 4.2 Some problems arising in the KL context

### 4.2.1 Nonconvex Feasibility

We present here a method to solve

$$\text{Find } x \in \bigcap_{i=1}^s C_i,$$

where each  $C_i$  is a closed nonempty subset of a Hilbert space  $H$ . This problem is called a *feasibility* problem. We suppose that each constraint  $C_i$  is well-known: we can easily access its elements, and the projection operator  $\text{proj}_{C_i}$  take nonempty values and can be computed. We emphasize the fact that the sets might be non-convex, which implies that  $\text{proj}_{C_i}$  can be set-valued. A natural approach to solve this problem consists in designing an iterative scheme involving projections onto each constraint  $C_i$ .

For instance, we can chose to successively project the current iterate onto one constraint, chosen in a cyclic way. This is the alternate projection method, designed by von Neumann for closed subspaces [252]. There is an abundant literature on this algorithm, see for instance [48, 49, 136] in the convex setting, and more recently [222, 19] for tame constraints. An other approach is the averaged projection algorithm, which works in two steps: first, project in parallel the current iterate onto each constraint, and then, define the new iterate as a convex combination of each projected point. The first convergence proof for this method is due to Auslender [34], see also [35, 135, 120, 49, 222]. There exists various other methods to solve the feasibility problem, let us mention for instance Dykstra's method [136, Chapter 9], or Haugazeau's method [190, 50, 112]. In this section, we will focus on relaxations and combinations of the averaged and alternate projection methods, invoking special instances of the Alternate Forward-Backward algorithm. In Theorem 4.2.8, we obtain a local convergence result.

We chose to approach this feasibility problem by introducing a penalization function. Consider for instance the following weighted sum of the distance functions to the constraints:

$$(4.23) \quad \underset{x \in \mathbb{H}}{\text{minimize}} \quad \frac{1}{2} \sum_{i=1}^s w_i \text{dist}^2(x, C_i) \quad \text{where } w_i \in ]0, 1[, \quad \sum_{i=1}^m w_i = 1.$$

If the functions  $x \mapsto \frac{1}{2} \text{dist}^2(x, C_i)$  are of class  $C^{1,1}$  on bounded sets (for instance if the sets are convex), we fall into the setting of the AFB method. Their gradient are  $x \mapsto x - \text{proj}_{C_i}(x)$ , which are 1-Lipschitz continuous, so the algorithm is written as follows:

$$(4.24) \quad x_{k+1} = (1 - \lambda_k)x_k + \lambda_k \left( \sum_{i=1}^s w_i \text{proj}_{C_i}(x_k) \right),$$

with  $(\lambda_k)_{k \in \mathbb{N}}$  in  $[\underline{\lambda}, \bar{\lambda}] \subset ]0, 1[$ . It is a relaxed version of the averaged projection algorithm of Auslender. Nevertheless, we cannot implement exactly the averaged projection method, without violating the constraint  $\lambda_k \ll 1$  due to the Lipschitz constant of  $x \mapsto x - \text{proj}_{C_i}(x)$ .

It might happen that the functions  $\text{dist}^2(\cdot, C_i)$  are not regular enough, in which case we say that the constraints are *hard* (in the previous case, we can talk about *soft* constraints). In that case, there is an other way to turn the original feasibility problem into a minimization problem solvable by the AFB method. It suffices to separate the variables, and solve

$$(4.25) \quad \underset{(x_1, \dots, x_s) \in \mathbb{H}^s}{\text{minimize}} \quad \sum_{i=1}^s \delta_{C_i}(x_i) + \frac{1}{4(s-1)} \sum_{1 \leq i, i' \leq s} \|x_i - x_{i'}\|^2.$$

The AFB method is written in that case as:

$$(4.26) \quad x_{i,k+1} \in \text{proj}_{C_i} \left( (1 - \lambda_{i,k})x_{i,k} + \lambda_{i,k} \left( \frac{x_{1,k+1} + \dots + x_{i-1,k+1} + x_{i+1,k} + \dots + x_{s,k}}{s-1} \right) \right),$$

with  $(\lambda_{i,k})_{k \in \mathbb{N}}$  in  $[\underline{\lambda}, \bar{\lambda}] \subset ]0, 1[$ . Here, we obtain an algorithm combining averaging and parallel projections.

In both cases, if the function to minimize satisfies the KL property, Theorem 3.2.3 guarantees the convergence to a solution, if the starting point is close enough to the solution set. The behaviour of the algorithms (4.24) and (4.26) is described in the following simple example.

**Example 4.2.1.** Consider two closed half-planes  $C_1, C_2$  in  $\mathbb{R}^2$ , with nonempty intersection. The algorithm (4.24) writes in that case

$$x_{k+1} = (1 - \lambda_k)x_k + \lambda_k \left( \frac{\text{proj}_{C_1}(x_k) + \text{proj}_{C_2}(x_k)}{2} \right).$$

If we take  $\lambda_k$  close to one, this scheme is very close to an averaged projection algorithm. It is illustrated in Figure 4.1, where we take  $\lambda_k \equiv 0.9$ .

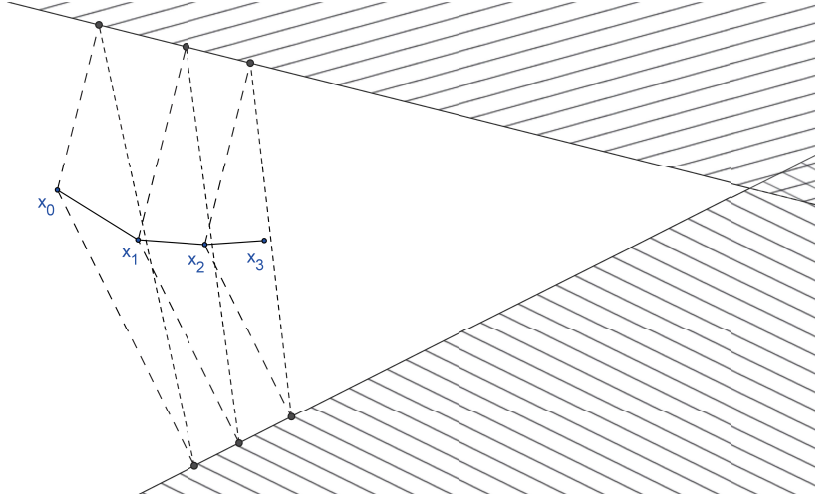


Figure 4.1: An relaxed averaged projection algorithm.

The algorithm (4.26) looks more to an alternating projection method. It generates two sequences  $(x_k)_{k \in \mathbb{N}}$  and  $(y_k)_{k \in \mathbb{N}}$ , defined by

$$\begin{aligned} x_{k+1} &= \text{proj}_{C_1}((1 - \lambda_{1,k})x_k + \lambda_{1,k}y_k), \\ y_{k+1} &= \text{proj}_{C_2}((1 - \lambda_{2,k})y_k + \lambda_{2,k}x_{k+1}). \end{aligned}$$

It is illustrated in Figure 4.1, where we also take  $\lambda_{1,k} = \lambda_{2,k} \equiv 0.9$ .

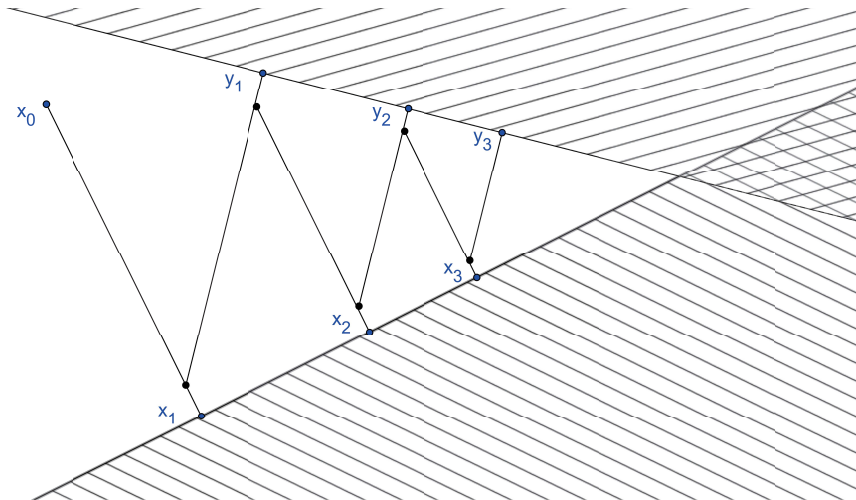


Figure 4.2: An relaxed alternate projection algorithm.

**Remark 4.2.2.** At first glance, it seems that the alternating projection method is slightly faster than the averaged projection method. This fact is observed in [222, Remark 7.5], under some regularity assumptions on the intersection.

**Remark 4.2.3.** In Example 4.2.1, we take  $m = 2$ , and observe that the algorithm (4.26) is very close to the classic alternating projection algorithm. Nevertheless, this is no more the case when  $m \geq 3$ , even if  $\lambda_{i,k} = 1$ . We do not know if it is possible to find a functional  $f$  for which the AFB method would lead to a ‘pure’ alternating projection algorithm. According to a recent result of Baillon-Combettes-Cominetti [40], we are tempted to conjecture that such a function does not exist.

**Remark 4.2.4.** When the distance functions  $\text{dist}^2(x, C_i)$  are not of class  $C^{1,1}$ , we said that we cannot use the averaged projection method (4.24). In fact, there exists an other penalization leading to an averaged projection-like method, and which is tractable even if the functions  $\text{dist}^2(x, C_i)$  are not regular. It suffices to consider

$$(4.27) \quad \underset{X, Y \in \mathbb{H}^s}{\text{minimize}} \quad \delta_C(X) + \delta_\Delta(Y) + \frac{1}{2} \|X - Y\|^2,$$

where  $C := \prod_{i=1}^s C_i$  and  $\Delta := \{(x, \dots, x) \in \mathbb{H}^s \mid x \in \mathbb{H}\}$ . Using the fact that

$$\text{proj}_C(X) = \prod_{i=1}^s \text{proj}_{C_i}(x_i) \quad \text{and} \quad \text{proj}_\Delta(X) = \frac{1}{s} \left( \sum_{i=1}^s x_i, \dots, \sum_{i=1}^s x_i \right),$$

the AFB method applied to (4.27) writes as

$$\begin{aligned} x_{i,k+1} &= \text{proj}_{C_i} \left( (1 - \lambda_k) x_{i,k} + \lambda_k y_k \right) \quad \text{for all } i \in \{1, \dots, s\}, \\ y_{k+1} &= (1 - \mu_k) y_k + \mu_k \sum_{i=1}^s \frac{1}{s} x_{i,k+1}. \end{aligned}$$

This algorithm can also be seen as a relaxation of the averaged projection algorithm.

We propose in the following a method to solve *mixed* feasibility problems, involving both soft and hard constraints. This will result in an algorithm combining the averaged projection algorithm and the alternate projection method.

#### 4.2.1.1 Mixed feasibility problem

Let  $s \geq 1$ ,  $p \geq 0$  and  $C_1, \dots, C_s, D_1, \dots, D_p$  be a family of weakly closed nonempty subsets of  $\mathbb{H}$ , having nonempty intersection. Consider the problem:

$$(4.28) \quad \text{Find } x \in S := C_1 \cap \dots \cap C_s \cap D_1 \cap \dots \cap D_p.$$

Here the  $C_i$ 's plays the role of the hard constraints (no regularity will be imposed), and we will always suppose that  $s \geq 1$ , taking if necessary  $C_1 = \mathbb{H}$ . The  $D_j$ 's are assumed to be *prox-regular*: a prox-regular set  $D \subset \mathbb{H}$  is a closed set admitting for each  $\bar{x} \in D$  a neighborhood  $U$  such that the projection  $\text{proj}_D : U \rightarrow D$  is well defined (single-valued) and strong-to-weak continuous. For instance convex sets, as well as manifolds of class  $C^2$ , are prox-regular. The prox-regular sets are interesting in our setting since the corresponding distance functions enjoy a good regularity property:

**Proposition 4.2.5.** [275] Let  $D \subset \mathbb{H}$  be a nonempty closed prox-regular set. Then, for all  $\bar{x} \in D$ , there exists  $\delta > 0$  such that:

- i) the projection  $\text{proj}_D(x)$  reduces to one element for all  $x \in B(\bar{x}, \delta)$ ,
- ii) the function  $h(x) := \frac{1}{2} \text{dist}^2(x, D)$  is of class  $C^{1,1}$  on  $B(\bar{x}, \delta)$ , with a 1-Lipschitz-continuous gradient given by  $\nabla h(x) = x - \text{proj}_D(x)$ .

To solve (4.28), we introduce a cost function, penalizing the prox-regular sets with their distance functions, and the hard constraints with their indicator functions:

$$(4.29) \quad f(x_1, \dots, x_s) = \sum_{i=1}^s \left[ \delta_{C_i}(x_i) + \frac{1}{2} \sum_{j=1}^p w_j \text{dist}^2(x_i, D_j) + \frac{1}{4} \sum_{i'=1}^s w \|x_i - x_{i'}\|^2 \right].$$

Here  $w, w_1, \dots, w_p$  are nonnegative weights satisfying

$$(s-1)w + \sum_{j=1}^p w_j = 1.$$

Note that  $(x_1, \dots, x_s)$  is a global minimum for  $f$  if and only if  $x_1 = \dots = x_s \in S$ , and that we recover (4.23) (respectively (4.25)) if there is only soft (respectively hard) constraints. This penalization function fits well with the setting of the previous chapter, since it is the sum of nonsmooth functions with separated variables (namely,  $\delta_{C_i}(x_i)$ ), and of a smooth function

$$(4.30) \quad h(x_1, \dots, x_s) := \sum_{i=1}^s \left[ \frac{1}{2} \sum_{j=1}^p w_j \text{dist}^2(x_i, D_j) + \frac{1}{4} \sum_{i'=1}^s w \|x_i - x_{i'}\|^2 \right].$$

Note nevertheless that, under the prox-regularity assumption, this function  $h$  is not differentiable on the whole space  $\mathbb{H}^s$ , but only in a neighborhood of the solution set:

**Proposition 4.2.6.** Let  $D_1, \dots, D_p$  be nonempty, weakly closed and prox-regular subsets of  $\mathbb{H}$ . Let  $s \geq 1$  and consider the function  $h$  defined in (4.30) with  $(s-1)w + \sum_{j=1}^p w_j = 1$ . Then, for all  $\bar{x} \in \bigcap_{j=1}^p D_j$ , there exists a neighborhood  $U$  of  $\bar{x}$  in  $\mathbb{H}$  such that  $h$  is continuously differentiable on  $U^s$ . Its partial gradients are

$$\forall i \in \{1, \dots, s\}, \forall (x_1, \dots, x_s) \in \mathbb{H}^s, \quad \nabla_i h(x_1, \dots, x_s) = x_i - \sum_{j=1}^p w_j \text{proj}_{D_j}(x_i) - \sum_{i' \neq i} w x_{i'},$$

and satisfy the following Lipschitz property:

$$\forall (x_1, \dots, x_s), (y_1, \dots, y_s) \in \mathbb{H}^s, \quad \|\nabla_i h(x_1, \dots, x_s) - \nabla_i h(y_1, \dots, y_s)\| \leq \sum_{i=1}^s \|x_i - y_i\|.$$

*Proof.* For each  $j \in \{1, \dots, p\}$ , let  $U_j$  be the neighborhood of  $\bar{x}$  given by (4.2.6) for the function  $x \in \mathbb{H} \mapsto \frac{1}{2} \text{dist}^2(x, D_j)$ . Thus, we can take  $U = \bigcap_{j=1}^p U_j$  on which  $h$  is continuously differentiable.

The computation of  $\nabla_i h$  is direct, coming directly from (4.2.6) and  $(s-1)w + \sum_{j=1}^p w_j = 1$ .

Now, let  $x := (x_1, \dots, x_s)$  and  $y := (y_1, \dots, y_s)$  be in  $U^s$ . Using the definition of  $\nabla_i h$  together with the triangular inequality and Proposition 4.2.6, one obtains

$$\begin{aligned} & \|\nabla_i h(x) - \nabla_i h(y)\| \\ &= \|w(s-1)(x_i - y_i) + \sum_{j=1}^p w_j [x_i - \text{proj}_{D_j}(x_i) - y_i + \text{proj}_{D_j}(y_i)] - \sum_{i' \neq i} w(y_{i'} - x_{i'})\| \\ &\leq w(s-1)\|x_i - y_i\| + \sum_{j=1}^p w_j \|x_i - \text{proj}_{D_j}(x_i) - y_i + \text{proj}_{D_j}(y_i)\| + \sum_{i' \neq i} w \|y_{i'} - x_{i'}\| \\ &\leq w(s-1)\|x_i - y_i\| + \sum_{j=1}^p w_j \|x_i - y_i\| + \sum_{i' \neq i} w \|y_{i'} - x_{i'}\|. \end{aligned}$$

Using again  $(s-1)w + \sum_{j=1}^p w_j = 1$  and  $w \leq 1$  we obtain:

$$\begin{aligned} \|\nabla_i h(x) - \nabla_i h(y)\| &\leq \|x_i - y_i\| + w \sum_{i' \neq i} \|y_{i'} - x_{i'}\| \\ &\leq \sum_{i=1}^s \|x_i - y_i\|. \end{aligned}$$

■

We can now define an *alternating average projection algorithm*, by applying the AFB method to (4.29). Each iteration of this algorithm consists in two steps: first, we compute a convex combination of the points lying in the hard constraints with some projections on the prox-regular sets. Next, we project onto a hard constraint, cyclically. More precisely, start with  $(x_{1,0}, \dots, x_{s,0}) \in \prod_{i=1}^s C_i$  and for all  $k \in \mathbb{N}$ ,  $i \in \{1, \dots, s\}$ , take  $\lambda_{i,k} \in [\underline{\lambda}, \bar{\lambda}] \subset ]0, 1[$  and apply

$$(4.31) \quad x_{i,k+1} = \text{proj}_{C_i} \left( (1 - \lambda_{i,k})x_{i,k} + \lambda_{i,k} \left[ \sum_{i' < i} w x_{i',k+1} + \sum_{j=1}^p w_j \text{proj}_{D_j}(x_{i,k}) + \sum_{i' > i} w x_{i',k} \right] \right)$$

The idea to obtain local convergence for this algorithm is to use Proposition 4.1.4 together with Theorem 3.2.3, under the hypothesis that  $f$  is KL and that the starting points are sufficiently close to an element of  $S$ . Nevertheless, the hypotheses of Proposition 4.1.4 are not directly satisfied here since in this case the gradient of  $h$  is not Lipschitz continuous on bounded sets but only in a neighborhood of any point of  $S$ . Then we shall verify that the sequence generated by this algorithm stays in an appropriate neighborhood of  $S$ . This is the purpose of the following technical lemma:

**Lemma 4.2.7.** Let  $\bar{x} \in S$  and  $U$  its neighborhood given by the previous proposition. We note  $\bar{z} = (\bar{x}, \dots, \bar{x}) \in S^s$ . Then, for any  $R > 0$  satisfying  $B(\bar{x}, R) \subset U$ , there exists  $r < R$  such that the stability hypothesis  $\mathbf{S}(\bar{z}, r, R)(i)$  holds for the algorithm given by (4.31), provided that  $(x_{1,0}, \dots, x_{s,0})$  is taken close enough to  $\bar{z}$ .

*Proof.* Let us start by defining a ‘good’ radius  $r < R$ . Given arbitrary  $r > 0$  and  $R_1 > 0$ , we define  $(R_2, \dots, R_s)$  by:

$$R_{i+1} = 4s(s+p)R_i + (2 + 4s(s+p))r, \text{ for } i \in \{2, \dots, s\}.$$

This progression being arithmetic-geometric, we can choose  $r$  and  $R_1$  small enough to have

$$r < R_1 \text{ and } R_s < \frac{R}{\sqrt{s}}.$$

By definition of these numbers, we easily see that  $R_1 < \dots < R_s$ . Moreover, we have this estimation which will be useful in the following:

$$(4.32) \quad \text{for all } 1 \leq i < s, \quad \frac{(R_{i+1} - 2r)^2}{8s(s+p)} \geq \frac{(R_{i+1} - 2r)^2}{16s(s+p)} = (r + R_i)^2.$$

Set  $\bar{X} := (\bar{x}, \dots, \bar{x})$  and, for all  $k \in \mathbb{N}$ , let  $X_k = (x_{1,k}, \dots, x_{s,k})$ . We assume that the initial point  $(x_{1,0}, \dots, x_{s,0})$  is close enough to  $\bar{X}$ , in the following sense (recall that  $f$  is continuous):

$$f(X_0) < \frac{(R_1 - 2r)^2}{32(s+p)} \text{ and } \forall i \in \{1, \dots, s\}, \|x_{0,i} - \bar{x}\| < r.$$

In view to prove  $\mathbf{S}(\bar{X}, R, r)(i)$ , we assume that  $X_0, \dots, X_k$  lies in  $B(\bar{X}, r)$ , and show that  $X_{k+1} \in B(\bar{X}, R)$ .

We start by giving a first estimation on  $\|y_{i,k+1} - x_{i,k}\|$ , where  $y_{i,k+1}$  is defined by:

$$y_{i,k+1} = (1 - \lambda_{i,k})x_{i,k} + \lambda_{i,k} \left( \sum_{j=1}^p w_j \text{proj}_{D_j}(x_{i,k}) + \sum_{i' < i} w x_{i',k+1} + \sum_{i' > i} w x_{i',k} \right).$$

Using the triangle inequality and the fact that  $\lambda_{i,k}$ ,  $w$  and  $w_j$  are strictly smaller than 1 for each  $i \in \{1, \dots, s\}$  and  $j \in \{1..p\}$ , one obtains:

$$\begin{aligned} \|y_{i,k+1} - x_{i,k}\|^2 &\leq \left\| \sum_{j=1}^p w_j(x_{i,k} - \text{proj}_{D_j}(x_{i,k})) + \sum_{i' < i} w(x_{i,k} - x_{i',k+1}) + \sum_{i' > i} w(x_{i,k} - x_{i',k}) \right\|^2 \\ &\leq (s+p) \left[ \sum_{j=1}^p w_j \|x_{i,k} - \text{proj}_{D_j}(x_{i,k})\|^2 + \sum_{i' < i} \|x_{i,k} - x_{i',k+1}\|^2 + \sum_{i' > i} w \|x_{i,k} - x_{i',k}\|^2 \right] \end{aligned}$$

Since  $\|x_{i,k} - \text{proj}_{D_j}(x_{i,k})\|^2 = \text{dist}^2(x_{i,k}, D_j)$ , we can write that

$$(4.33) \quad \|y_{i,k+1} - x_{i,k}\|^2 \leq 4(s+p)f(z^k) + (s+p) \sum_{i' < i} \|x_{i,k} - x_{i',k+1}\|^2.$$

We assumed that  $X_0, \dots, X_k \in B(\bar{X}, r)$ , where  $f$  is of class  $C^{1,1}$ . We can use the descent lemma 4.1.4, as in the proof of Proposition 4.1.3, to obtain that  $f(X_k) \leq f(X_0) < \frac{(R_1 - 2r)^2}{32(s+p)}$ . Recalling (4.33), we deduce

$$(4.34) \quad \|y_{i,k+1} - x_{i,k}\|^2 \leq \frac{(R_1 - 2r)^2}{8} + (s+p) \sum_{i' < i} \|x_{i,k} - x_{i',k+1}\|^2.$$

We shall prove now that  $\|x_{i,k+1} - \bar{x}\| \leq R_i$ , by using a recurrence on  $i \in \{1, \dots, s\}$ .

If  $i = 1$  then, by (4.34), we have  $\|y_{1,k+1} - x_{1,k}\|^2 \leq \frac{(R_1 - 2r)^2}{8} \leq \frac{(R_1 - 2r)^2}{4}$ , so

$$(4.35) \quad \|y_{1,k+1} - x_{1,k}\| \leq \frac{(R_1 - 2r)}{2}.$$

On the other hand,  $x_{1,k+1}$  being the projection of  $y_{1,k+1}$  onto  $C_1 \ni \bar{x}$ , we have  $\|x_1^{k+1} - \bar{x}\| \leq 2\|y_{1,k+1} - \bar{x}\|$ . This last inequality, combined with (4.35), gives:

$$\|x_{1,n+1} - \bar{x}\| \leq 2\|y_{1,k+1} - x_{1,k}\| + 2\|x_{1,k} - \bar{x}\| \leq (R_1 - 2r) + 2r = R_1.$$

Suppose now that  $\|x_{i',k+1} - \bar{x}\| \leq R_{i'}$  for any  $i' < i$ . In inequality (4.34) we have, using the recurrence hypothesis, that for all  $i' < i$ ,

$$\|x_{i,k} - x_{i',k+1}\| \leq \|x_{i,k} - \bar{x}\| + \|\bar{x} - x_{i',k+1}\| \leq r + R_{i'}.$$

Thus (recalling the monotonicity of the  $R_{i'}$ 's):

$$\sum_{i' < i} \|x_{i,k} - x_{i',k+1}\|^2 \leq \sum_{i' < i} (r + R_{i'})^2 \leq \sum_{i' < i} (r + R_{i-1})^2 \leq s(r + R_{i-1})^2$$

where  $s(r + R_{i-1})^2 \leq \frac{(R_i - 2r)^2}{8(s+p)}$  by definition of  $R_i$ . Combining this last inequality with (4.34) gives

$$\|y_{i,k+1} - x_{i,k}\|^2 \leq \frac{(R_1 - 2r)^2}{8} + \frac{(R_i - 2r)^2}{8},$$

where  $\frac{(R_1 - 2r)^2}{8} \leq \frac{(R_i - 2r)^2}{8}$  since  $r < R_1 < R_i$ .

Thus  $\|y_{i,k+1} - x_{i,k}\| \leq \frac{(R_i - 2r)}{2}$ , and using exactly the same arguments as for the case  $i = 1$  we can deduce that  $\|x_{i,k+1} - \bar{x}\| \leq R_i$ .

Hence, we obtain

$$\|X_{k+1} - \bar{X}\|^2 = \sum_{i=1}^s \|x_{i,k+1} - \bar{x}\|^2 \leq \sum_{i=1}^s R_i^2 \leq sR_s^2 < R^2,$$

which completes the proof. ■



We can now state our *capture* result:

**Theorem 4.2.8.** *Let  $C_1, \dots, C_s, D_1, \dots, D_p$  be a collection of nonempty weakly closed subsets of  $H$  with  $D_1, \dots, D_p$  supposed to be prox-regular. Suppose that  $f$ , defined by (4.29), satisfies the KL property at some  $\bar{x} \in S$ . Then if each  $x_{i,0}$  is close enough to  $\bar{x}$ , the sequence generated by (4.31) has finite length and converges to some  $(x_\infty, \dots, x_\infty) \in S^s$ .*

*Proof.* Let  $\bar{X} = (\bar{x}, \dots, \bar{x}) \in H^s$ . By Proposition 4.2.5 there is  $\delta > 0$  such that  $h$  is of class  $C^{1,1}$ , and  $f$  has the Kurdyka-Łojasiewicz property at  $\bar{x}$  both in a set  $\Gamma_\eta(\bar{x}, \delta) = \{x \in B(\bar{x}, \delta) : 0 < f(x) < \eta\}$ . Therefore, Lemma 4.2.7 tells us that if each starting point is close enough to  $\bar{x}$ , then  $\mathbf{S}(\bar{X}, \delta, r)(i)$  holds. We can even be sufficiently close to have  $\mathbf{S}(\bar{X}, \delta, r)(ii)$ . So the sequence remains in a ball where  $\nabla h$  is Lipschitz continuous, and it can be shown that  $\mathbf{H}_0, \mathbf{H}_1$  and  $\mathbf{H}_2$  are satisfied, following the lines of the proof of Proposition 4.1.3. Thus Proposition 3.2.4 allows us to conclude.  $\blacksquare$

#### 4.2.1.2 The KL property for the penalization function

The convergence result of Theorem 4.2.8 holds under the assumption that the cost function  $f$ , defined by (4.29), has the KL property. This property holds, for instance, if all the involved sets are defined in an o-minimal structure. We focus in this section on geometric sufficient conditions on the sets ensuring that  $f$  is KL.

**Definition 4.2.9.** Let  $F_1, \dots, F_m$  be a family of nonempty closed subsets of  $H$ . Its intersection  $S$  is said to be *strongly regular* at  $\bar{x} \in S$  if there exists  $\alpha > 0$  and  $\delta > 0$  such that

$$(4.36) \quad \sqrt{\sum_{i=1}^m \|x_i^*\|^2} \leq \alpha \sum_{i=1}^m \|x_i^*\|$$

holds, for all  $x_i \in F_i \cap B(\bar{x}, \delta)$  and  $x_i^* \in N_{F_i}^L(x_i)$ ,  $i \in \{1, \dots, m\}$ . If the intersection  $S$  is strongly regular at any of its points, we just say that it is strongly regular.

**Remark 4.2.10.** Note that this notion is *local*, which means that the strong regularity at  $\bar{x}$  is affected by the behaviour of the limiting normal cones  $N_{F_i}^L$  at  $x \sim \bar{x}$ . Observe also that, in finite dimension, we can equivalently replace the limiting normals by the Fréchet ones in Definition 4.2.9.

**Remark 4.2.11.** In its survey on regularity of collections of sets, Kruger [214] defines the strong regularity of  $F_1, \dots, F_m$  at  $\bar{x} \in F_1 \cap \dots \cap F_m$  as follows: there exists  $R > 0$  and  $\alpha > 0$  such that  $\forall x_i \in F_i \cap B(\bar{x}, R)$ ,  $\forall r \in ]0, R]$ ,  $\forall \varepsilon_i \in B(0, \alpha r)$ ,

$$\bigcap_{i=1}^m (F_i - x_i - \varepsilon_i) \cap B(0, r) \neq \emptyset.$$

Actually, these two notions of strong regularity are the same in finite dimensions spaces (see [214, Proposition 10, Corollary 2] and Remark 4.2.10).

In finite dimensions, the strong regularity also coincides with a well-known notion:

**Proposition 4.2.12.** Let  $F_1, \dots, F_m$  be a family of nonempty closed subsets of  $H$ , and assume that  $H$  has finite dimension. Let  $\bar{x} \in F_1 \cap \dots \cap F_m$ . Then, the strong regularity of the intersection at  $\bar{x}$  is equivalent to the following linear regularity:

$$(4.37) \quad \forall x_i^* \in N_{F_i}^L(\bar{x}), \sum_{i=1}^m x_i^* = 0 \Rightarrow x_1^* = \dots = x_m^* = 0.$$

*Proof.* Assume that the intersection is strongly regular at  $\bar{x}$ , *i.e.* there exists  $\alpha > 0$  and  $\delta > 0$  so that (4.36) holds. Let  $(x_1^*, \dots, x_m^*) \in \prod_{i=1}^m N_{F_i}^L(\bar{x})$  such that  $\sum_{i=1}^m x_i^* = 0$ . By definition of the limiting normal cone, there exists sequences  $(x_{i,k})_{k \in \mathbb{N}} \subset F_i$  and  $(x_{i,k}^*)_{k \in \mathbb{N}} \subset H$  such that  $x_{i,k}^* \in N_{F_i}^L(x_{i,k})$ , with

$$x_{i,k} \xrightarrow[k \rightarrow +\infty]{F_i} \bar{x} \text{ and } x_{i,k}^* \xrightarrow[k \rightarrow +\infty]{\|\cdot\|} x_i^*.$$

Because of the convergence of  $x_{i,k}$  to  $\bar{x}$  in  $F_i$ , we can assume, taking eventually a subsequence, that the sequence  $(x_{i,k})_{k \in \mathbb{N}}$  remains in  $F_i \cap B(\bar{x}, \delta)$ . Hence, we can invoke the strong regularity at  $\bar{x}$  to write, for all  $k \in \mathbb{N}$ :

$$\sqrt{\sum_{i=1}^m \|x_{i,k}^*\|^2} \leq \alpha \left\| \sum_{i=1}^m x_{i,k}^* \right\|.$$

Passing to the limit when  $k$  goes to  $+\infty$ , together with the fact that  $\sum_{i=1}^m x_i^* = 0$ , we deduce that  $\sum_{i=1}^m \|x_i^*\|^2 = 0$ , which proves the linear regularity.

On the reverse way, suppose that the linear regularity holds at  $\bar{x}$ , and assume, by contradiction, that there exists two converging sequences  $(x_{i,k})_{k \in \mathbb{N}} \subset F_i$  and  $(x_{i,k}^*)_{k \in \mathbb{N}} \subset H$  such that

$$x_{i,k}^* \in N_{F_i}^L(x_{i,k}) \text{ and } x_{i,k} \xrightarrow[k \rightarrow +\infty]{F_i} \bar{x}$$

and

$$\sqrt{\sum_{i=1}^m \|x_{i,k}^*\|^2} > k \left\| \sum_{i=1}^m x_{i,k}^* \right\|.$$

Adapting the proof in [222, Proposition 8.5], we define

$$y_{i,k}^* := \left( \sqrt{\sum_{i=1}^m \|x_{i,k}^*\|^2} \right)^{-1} x_{i,k}^* \in N_{F_i}^L(x_{i,k})$$

which satisfies

$$(4.38) \quad \sum_{i=1}^m \|y_{i,k}^*\|^2 = 1 \text{ and } \left\| \sum_{i=1}^m y_{i,k}^* \right\| \leq \frac{1}{k}.$$

Since the sequences  $(y_{i,k}^*)_{k \in \mathbb{N}}$  are bounded and that  $H$  is finite-dimensional, we can suppose that it converge to some  $y_i^*$  when  $k$  goes to  $+\infty$ . Moreover, the limiting normal cones have a closed graph in finite dimension, so  $y_i^* \in N_{F_i}^L(\bar{x})$  by construction. By taking the limit in (4.38), we obtain moreover

$$\sum_{i=1}^m \|y_i^*\|^2 = 1 \text{ and } \left\| \sum_{i=1}^m y_i^* \right\| \leq \frac{1}{k}.$$

This clearly in contradiction with the linear regularity assumption. ■

**Example 4.2.13.** In finite dimensions, we can derive from (4.37) the following simple characterization of the strong regularity between two sets:

$$N_{F_1}^L(\bar{x}) \cap -N_{F_2}^L(\bar{x}) = \{0\}.$$

In [222, Proposition 8.5], the authors consider a collection of closed prox-regular sets  $D_1, \dots, D_m$  in  $\mathbb{R}^n$  having a strong regular intersection, and derive the KL property of  $f$ . The same approach

is used in [20, Theorem 5.8], where the authors take a family of closed convex sets  $D_1, \dots, D_m$  in  $\mathbb{R}^n$  (one of them being compact), together with a general closed set  $C$ . We extend their approach to family of closed prox-regular sets in a Hilbert space, together with a general closed set.

**Proposition 4.2.14.** Let  $D_1, \dots, D_p$  be a collection of closed prox-regular subsets of  $H$ , and let  $C \subset H$  be closed and nonempty. Assume that the intersection  $S = C \cap D_1 \cap \dots \cap D_p \neq \emptyset$  is strongly regular at  $x^* \in S$ . Then, the function  $f(x) := \delta_C(x) + \frac{1}{2} \sum_{j=1}^p w_j \text{dist}^2(x, D_j)$ , has the Kurdyka-Łojasiewicz property at  $\bar{x}$  with a Łojasiewicz exponent  $\theta = \frac{1}{2}$ .

**Remark 4.2.15.** Note that in this setting, we know precisely from Theorem 4.2.8 that the convergence rate of the algorithm (4.31) is exponential.

**Remark 4.2.16.** Note that the strong regularity of an intersection of sets is equivalent to the *metric regularity* of an appropriate set-valued mapping (see the works of Kruger [213, Section 8], [214, Section 3.3] or [222, Section 3]). On the other hand, the Kurdyka-Łojasiewicz property of a function is also equivalent (under some assumptions) to the metric regularity on an (other) set-valued mapping. Thus, we can see –at least informally– that the result of Proposition 4.2.14 is not surprising.

*Proof.* We will find  $\delta > 0$  such that the KL inequality holds in  $\Gamma_{+\infty}(\bar{x}, \delta) = \{x \in B(\bar{x}, \delta) : 0 < f(x) < +\infty\}$ . The strong regularity of the intersection implies the existence of parameters  $\alpha > 0$  and  $\delta > 0$  such that

$$(4.39) \quad \sqrt{\|x^*\|^2 + \sum_{i=1}^m \|x_i^*\|^2} \leq \alpha \left\| x^* + \sum_{i=1}^p x_i^* \right\|$$

for each  $x_i^* \in N_{D_i}^L(x_i)$  with  $x_i \in D_i \cap B(\bar{x}, \delta)$ ,  $i = 1, \dots, p$ , and each  $x^* \in N_C^L(x)$  with  $x \in C \cap B(\bar{x}, \delta)$ . If necessary, shrink  $\delta$  to be in the context of Proposition 4.2.6 so that each projection  $\text{proj}_{D_i}$  for  $i = 1, \dots, p$ , is single-valued on  $B(\bar{x}, \delta)$ . Observe that

$$\partial^L f(x) = N_C^L(x) + \left\{ \sum_{i=1}^p w_i (x - \text{proj}_{D_i}(x)) \right\}.$$

Take  $x \in C \cap B(\bar{x}, \delta)$  and set  $x_i := \text{proj}_{D_i}(x) \in D_i \cap B(\bar{x}, \delta)$ . Note that  $x_i^* := w_i(x - x_i)$  lies in  $N_{D_i}^L(x_i)$ . We have

$$\|\partial^L f(x)\|_-^2 = \inf_{x^* \in N_C^L(x)} \left\| x^* + \sum_{i=1}^p x_i^* \right\|^2 \geq \frac{1}{\alpha^2} \left[ \inf_{x^* \in N_C^L(x)} \|x^*\|^2 + \sum_{i=1}^p \|x_i^*\|^2 \right]$$

by (4.39). Whence

$$\|\partial^L f(x)\|_-^2 \geq \frac{1}{\alpha^2} \sum_{i=1}^p \|x_i^*\|^2 = \frac{1}{\alpha^2} \sum_{i=1}^p w_i^2 \|x - x_i\|^2.$$

If we note  $\underline{w} := \min_{i=1..p} w_i > 0$ , then we obtain as desired

$$\|\partial^L f(x)\|_-^2 \geq \underline{w} \frac{2}{\alpha^2} f(x) = \underline{w} \frac{2}{\alpha^2} [f(x) - f(\bar{x})].$$

■

## 4.2.2 Inverse problems and sparsity

With the increase of data's size in signal or image processing, the research for sparse solutions gained in interest. The sparsity of a solution is generally expressed by means of the *counting function*<sup>1</sup>:

$$\forall x \in \mathbb{R}^N, \|x\|_0 := \#\{i \in \{1, \dots, N\} \mid x_i \neq 0\}.$$

When dealing with matrices, a natural analogue to the counting function is the *rank function*. These functions are lower semi-continuous and semi-algebraic, so they enter in the setting of the Alternate Forward-Backward algorithm. We present in this section some inverse problems involving sparsity, and provide different algorithms for solving them.

### 4.2.2.1 Compressed sensing

The search for sparse solutions of under-determined linear systems is an important problem in compressive sensing and appears naturally when studying signal denoising, deblurring or compression processes. Consider the following problem:

$$(P_0) \quad \min_{x \in \mathbb{R}^N} \|x\|_0 \text{ such that } Ax = b,$$

where  $b \in \mathbb{R}^M$  with  $M < N$  and  $A \in \mathbb{R}^{M \times N}$ . The interest of  $(P_0)$  in data compression is simple: suppose we have a measurement  $b := A\bar{x}$  of a signal  $\bar{x}$ , which we know to be sparse (meaning that  $\|\bar{x}\|_0$  is small). If  $\|\bar{x}\|_0 < \frac{M}{2}$ , and the columns of the matrix are supposed to be in *general position*, then  $\bar{x}$  is the only solution of problem  $(P_0)$  (see [149, 150]). This property holds in  $\mathbb{R}^{N \times M}$  with probability 1 and can be obtained by considering a matrix with random entries.

Solving  $(P_0)$  by means of combinatorial techniques is very difficult, because it is an NP-hard problem [249]. This is why this problem is generally handled with an iterative method. In order to solve  $(P_0)$ , which can be rewritten as

$$\min_{x \in \mathbb{R}^N} \|x\|_0 + \delta_{\{Ay=b\}}(x),$$

we could apply the proximal algorithm to the lower-semicontinuous function  $x \mapsto \alpha\|x\|_0 + \delta_{\{Ay=b\}}(x)$ . But the proximal operator associated to this sum is not easy to compute while the proximal operator associated to each summand is well known. Indeed,  $\text{prox}_{\alpha\|\cdot\|_0}$  is the *hard shrinkage operator*, that we will note  $\mathcal{H}_\alpha(x)$ , and is defined by:

$$\text{for all } i \in \{1, \dots, N\}, \quad (\mathcal{H}_\alpha(x))_i = \begin{cases} 0 & \text{if } |x_i| < \sqrt{2\alpha}, \\ \{0, x_i\} & \text{if } |x_i| = \sqrt{2\alpha}, \\ x_i & \text{if } |x_i| > \sqrt{2\alpha}. \end{cases}$$

On the other hand,  $\text{prox}_{\delta_{\{Ax=b\}}}(x)$  is the projection of  $x$  onto the linear set  $\{Ax = b\}$ . In other words,

$$\text{prox}_{\delta_{\{Ax=b\}}}(x) = \text{proj}_{\{Ax=b\}}(x) = (I_N - A^\dagger A)x + A^\dagger b,$$

where  $A^\dagger$  is the pseudo-inverse of  $A$ . Thus, it only remains to apply the splitting method seen in Section 4.1. Let us give some relaxations of  $(P_0)$ , which leads to different algorithms:

- Penalize quadratically the constraint, and consider for some  $\alpha > 0$ :

$$(P_0^{(1)}) \quad \underset{x \in \mathbb{R}^N}{\text{minimize}} \quad \alpha\|x\|_0 + \frac{1}{2}\|Ax - b\|^2.$$

<sup>1</sup>Sometimes called the  $\ell^0$  norm, which is not really an appropriate name.

The AFB algorithm becomes the *Hard-Shrunked Gradient algorithm*, defined as follows:

$$x_{k+1} \in \mathcal{H}_{\alpha\lambda_k}(x_k - \lambda_k A^*(Ax_k - b)),$$

where  $(\lambda_k)_{k \in \mathbb{N}}$  are step sizes lying in  $[\underline{\lambda}, \bar{\lambda}] \subset ]0, \|\|A^*A\|\|^{-1}[$ . This method appears in [20], and can be seen as an adaptation of the Thresholded Landweber algorithm [127], where the  $\ell^1$  norm is replaced by the counting function.

- Separate the variables by introducing the problem, for some  $\alpha > 0$ :

$$(P_0^{(2)}) \quad \underset{x \in \mathbb{R}^N}{\text{minimize}} \quad \alpha \|x\|_0 + \delta_{\{Ay=b\}}(y) + \frac{1}{2} \|x - y\|^2.$$

The AFB algorithm becomes the *Hard-Shrinkage Projection method*:

$$\begin{aligned} y_{k+1} &= \text{proj}_{\{Ax=b\}}(\mu_k x_k + (1 - \mu_k)y_k), \\ x_{k+1} &\in \mathcal{H}_{\alpha\lambda_k}(\lambda_k y_k + (1 - \lambda_k)x_{k+1}), \end{aligned}$$

with parameters  $(\lambda_k)_{k \in \mathbb{N}}, (\mu_k)_{k \in \mathbb{N}}$  lying in  $[\underline{\tau}, \bar{\tau}] \subset ]0, 1[$ . To our knowledge, this algorithm for solving  $(P_0)$  is new. See Section 4.2.2.3 for a practical implementation of this algorithm.

As a consequence of Theorem 4.1.6, together with the fact that the problem  $(P_0^{(1)})$  (resp.  $(P_0^{(2)})$ ) is semi-algebraic, we obtain that any bounded sequence generated by this algorithm converges. Moreover, Theorem 4.1.8 guarantees that the limit point is a solution of  $(P_0^{(1)})$  (resp.  $(P_0^{(2)})$ ), provided that the initial point is close enough to the solution.

#### 4.2.2.2 Sparse and Low-rank Matrix Decomposition

The problem of recovering the sparse and low-rank components of a matrix arises naturally in various areas, such as model selection in statistics or system identification in engineering (see [101] and references therein). Denote by  $\|X\|_0$  the number of nonzero components of a matrix  $X \in \mathcal{M}_{m,n}(\mathbb{R})$ . Given  $A \in \mathcal{M}_{m,n}(\mathbb{R})$ , the low-rank sparse matrix decomposition problem consists in finding  $X, Y$  in  $\mathcal{M}_{m,n}(\mathbb{R})$  such that  $A = X + Y$ , with  $X$  having a low rank, and  $Y$  being sparse:

$$(P_{rk,0}) \quad \underset{X+Y=A}{\text{minimize}} \quad \text{rank}(x) + \|Y\|_0.$$

An approach to solve this problem consists in doing a convex relaxation of the objective function: the counting norm and the rank functions are replaced by  $\ell^1$  and nuclear norms, respectively (see [283, 169, 322]). Nevertheless, the KL framework is well adapted to this nonconvex (but semialgebraic!) problem, and offers convergent numerical methods. Moreover, the AFB method is well suited for its structure in separated variables.

If we know some estimate on the desired rank (resp. sparsity) for  $X$  (resp.  $Y$ ), a tractable relaxation is:

$$\underset{X, Y \in \mathcal{M}_{m,n}(\mathbb{R})}{\text{minimize}} \quad \delta_{\{\text{rank}(\cdot) \leq r\}}(X) + \delta_{\{\|\cdot\|_0 \leq s\}}(Y) + \frac{1}{2} \|A - X - Y\|_F^2,$$

where  $\|\cdot\|_F$  denotes the Frobenius norm in  $\mathcal{M}_{m,n}(\mathbb{R})$ , and  $r, s \in \mathbb{N}$ . Applying the AFB method to this problem leads to an *Alternating Averaged Projected Method*:

$$\begin{aligned} X_{k+1} &\in \text{proj}_{\{\text{rank}(\cdot) \leq r\}}(\lambda_k(A - Y_k) + (1 - \lambda_k)X_k), \\ Y_{k+1} &\in \text{proj}_{\{\|\cdot\|_0 \leq s\}}(\mu_k(A - X_{k+1}) + (1 - \mu_k)Y_k), \end{aligned}$$

where  $(\lambda_k)_{k \in \mathbb{N}}, (\mu_k)_{k \in \mathbb{N}}$  are two sequences of parameters in  $[\underline{\tau}, \bar{\tau}] \subset ]0, 1[$ . This algorithm involves two projections at each step. To project onto  $\{\|\cdot\|_0 \leq s\}$ , one simply sets all the coefficients

to zero, except for the  $s$  largest ones (in absolute value). The projection of  $M \in \mathcal{M}_{m,n}(\mathbb{R})$  onto  $\{\text{rank} \cdot \leq r\}$  is given by Eckart-Young's Theorem [151]: writes the Singular Value Decomposition  $M = U\Sigma V$ , and put all the coefficients of the diagonal matrix  $\Sigma$  to zero, except for the  $s$  largest ones (in absolute value).

An other way to relax the problem ( $P_{rk,0}$ ) is the following: take some  $\alpha, \beta > 0$  and

$$\underset{X, Y \in \mathcal{M}_{m,n}(\mathbb{R})}{\text{minimize}} \quad \alpha \text{rank}(x) + \beta \|Y\|_0 + \frac{1}{2} \|A - X - Y\|_F^2.$$

The AFB algorithm applied to this problem leads to a similar algorithm:

$$\begin{aligned} X_{k+1} &\in \text{prox}_{\alpha \text{rank}} (\lambda_k(A - Y_k) + (1 - \lambda_k)X_k), \\ Y_{k+1} &\in \text{prox}_{\beta \|\cdot\|_0} (\mu_k(A - X_{k+1}) + (1 - \mu_k)Y_k), \end{aligned}$$

with  $(\lambda_k)_{k \in \mathbb{N}}, (\mu_k)_{k \in \mathbb{N}}$  in  $[\underline{\tau}, \bar{\tau}] \subset ]0, 1[$ . Here, two proximal operators are involved. The proximal operator of the counting norm for the matrices is exactly the same than for the vectors, the hard-shrinkage operator. Recently, Hiriart-Urruty and Le [194] gave an explicit description of the elements of  $\text{prox}_{\alpha \text{rank}}(M)$ . It suffices, for instance, to write the singular value decomposition  $M = U\Sigma V$ , and to apply the hard-shrinkage operator to  $\Sigma$ .

#### 4.2.2.3 Numerical results on data compression

Let  $\bar{x} \in \mathbb{R}^N$  be an original signal with sparsity  $s := \|\bar{x}\|_0$ , and  $b = A\bar{x}$  the compressed signal through  $A \in \mathbb{R}^{M \times N}$  generated with i.i.d. Gaussian entries<sup>2</sup>. The purpose is to recover  $\bar{x}$  from  $b$ , by solving

$$(P_0^{(2)}) \underset{x \in \mathbb{R}^N}{\text{minimize}} \quad \alpha \|x\|_0 + \delta_{\{Ay=b\}}(y) + \frac{1}{2} \|x - y\|^2$$

with the Hard Shrinkage Projection method. Rather than competing with the vast literature on compressive sensing, our principal aim here is to illustrate a new method from tame optimization to tackle problems involving sparsity. However, this clearly opens the door to new ideas from nonconvex optimization and maybe future works could focus on the enhancement of these algorithms, and their comparison to specific methods such as LASSO or Matching Pursuit.

Let us precise how we run the Hard Shrinkage Projection method. We generate randomly  $x_0 = y_0$  and run 10 steps of the algorithm with parameters  $\mu_k \equiv 10^{-3}$ ,  $\lambda_k \equiv 10^{-2}$  and  $\alpha = 10^5$ . Then, we do a restart at the current point for 10 more steps, with parameters  $\mu_k \equiv 10^{-3}$ ,  $\lambda_k \equiv \frac{1}{3}$  and  $\alpha = 130$ . Finally, we let the algorithm run with  $\mu_k \equiv 10^{-3}$ ,  $\lambda_k \equiv \frac{1}{3}$  and  $\alpha = 3$ . The choice of this initialization can be understood as follows: in a first time we use the hard shrinkage operator  $\mathcal{H}_{\alpha\lambda}$  with a high value for  $\alpha\lambda$ , which enforces  $x_k$  to keep a large number of zeroes. Then, we diminish  $\alpha\lambda$  progressively, which allows  $x_k$  to capture new nonzeros coefficients. All our numerical experimentations tend to show that a decreasing  $\alpha$  improves the convergence of the method.

---

<sup>2</sup> $A = (a_{i,j})$  with  $a_{i,j} \sim \mathcal{N}(0, \frac{1}{M})$ .

Let us start by evaluating the performance of this algorithm. For this, we generate random initial signals  $\bar{x} \in [0, 255]^{100}$ , that we try to recover for different values of  $s$  and  $M$ . We define  $\delta := \frac{M}{100}$  and  $\rho := \frac{s}{M}$ , being respectively the levels of compression and sparsity. We let  $\delta$  and  $\rho$  taking values between 0 and 1, and take for each couple  $(\delta, \rho)$  the values  $M = \lfloor 100\delta \rfloor$  and  $s = \lfloor 100\rho\delta \rfloor$ . For each couple of  $M$  and  $s$ , we run 10 instances of the Hard Shrinkage Projection algorithm with 100 steps, and take  $e(\delta, \rho)$  the average of the relative errors  $\frac{\|\bar{x} - x_{100}\|}{\|\bar{x}\|}$ . Figure 4.3 plots the values  $e(\delta, \rho)$  for  $(\delta, \rho) \in [0, 1]^2$ , in a gray scale: an exact recovery  $e(\delta, \rho) = 1$  corresponds to a black pixel, while  $e(\delta, \rho) = 0$  corresponds to a white pixel.

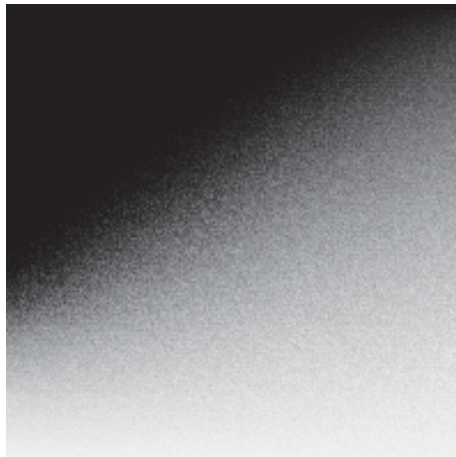


Figure 4.3: Values of  $e(\delta, \rho)$  for  $(\delta, \rho) \in [0, 1]^2$ .

We obtain a phase transition diagram: below a certain limit, the performance of the algorithm decreases dramatically. This behaviour is well-known for methods based on the  $\ell^1$  norm, see the works of Donoho [147] and Donoho-Stodden [148]. Observe nevertheless that our phase transition lies in a higher part of the diagram than the ones of [147, 148]. This tends to show that our method based on the direct minimization of the counting norm  $\|\cdot\|_0$  is more demanding in terms of the couple  $(\delta, \rho)$ .

We illustrate now the compression/decompression process by applying it to some picture  $X \in \mathbb{R}^{512 \times 512}$ . In general, images of every-day life are not sparse, so we cannot directly work with them. Instead, we will work in a wavelet space, applying to  $X$  a wavelet transform  $W$ . Thus,  $WX$  is almost sparse, in the sense that  $WX$  contains a lot of nearly zero coefficients. By taking  $\bar{X} = \mathcal{H}_\varepsilon(WX)$  for some  $\varepsilon \sim 0$ , we obtain a sparse matrix, such that  $W^{-1}\bar{X}$  is close to the original image (see Figure 4.4 below).



Figure 4.4: Top line: the original picture  $X$  and a representation of its wavelet transform  $WX$ . Bottom line: the matrix  $\bar{X}$  obtained after putting to zero 90% of the coefficients of  $WX$ , and the corresponding picture  $W^{-1}\bar{X}$ .

Once we have a sparse matrix  $\bar{X}$  to work with, we can compress it. For computation purposes, we do not compress directly  $\bar{X}$  through a huge matrix  $A$  of size  $\sim 512^4$ , but decompose  $\bar{X}$  into several vectors of size 200, that we can compress/decompress in parallel:  $\bar{X} = (\bar{x}_1, \dots, \bar{x}_{1311})$ . The decompression of these fragments  $\bar{x}_i$  is done by using exactly the same method as described above. Recall that the exact recovery is not guaranteed for each run of the Hard-Shrinkage Projection method. Hence, we keep in memory the sparsity of the original fragment  $s_i := \bar{x}_i$ , and restart eventually the algorithm until it converges to a vector having the same sparsity  $s_i$ .



Figure 4.5: Left: original image. Right: Image obtained after compression to 57% of the original size, and then decompression using 100 iterations of the Hard-Shrinkage Projection algorithm.



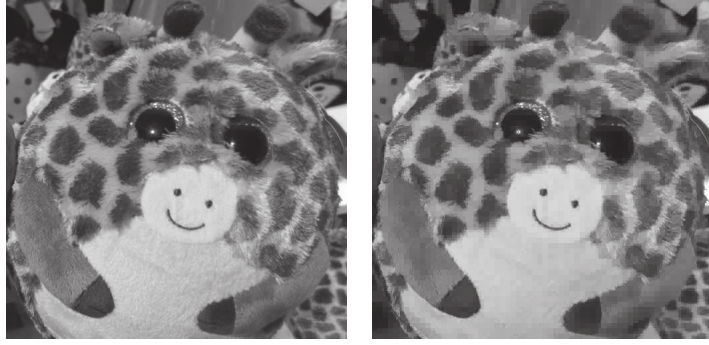


Figure 4.6: Left: original image. Right: Image obtained after compression to 60% of the original size, and then decompression using 100 iterations of the Hard-Shrinkage Projection algorithm.

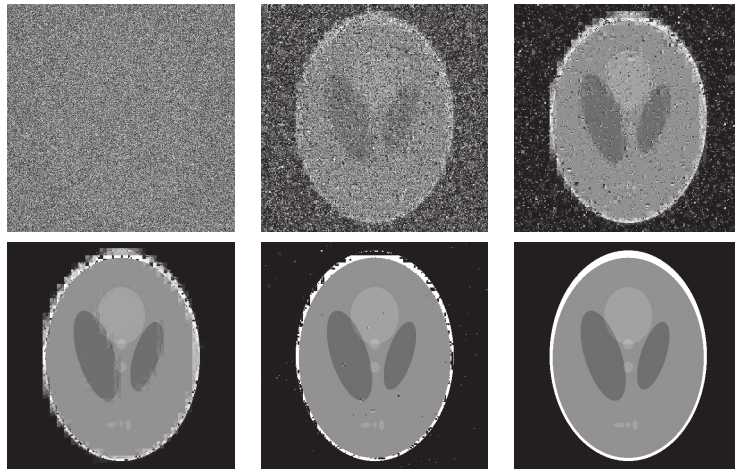


Figure 4.7: Reconstruction of the Shepp-Logan image, after compression to 60% of the original size. From left to right, top to bottom, reconstructed pictures after 0, 1, 5, 10, 25 and 32 iterations of the Hard-Shrinkage Projection algorithm.

### 4.3 Comments and perspectives

**Remark 4.3.1** (On the additive/relative errors). In Section 4.1, we presented a general splitting method, admitting *additive* errors. The convergence of the method is guaranteed under the set of assumptions **HE**, which is difficult to satisfy in practical. It would be of interest to design an algorithm admitting *relative* errors.

Consider for instance the gradient-like method

$$x_{k+1} = x_k - \lambda_k A_k^{-1} \nabla f(x_k), \quad \lambda_k > 0, \quad A_k \in S_{++}(\mathbb{H}),$$

which covers the gradient or Newton-like methods. At each iteration, the direction  $d_k := -A_k^{-1} \nabla f(x_k)$  must be computed. It can be obtained by solving, for instance, a quadratic program:

$$(4.40) \quad d_k = \operatorname{argmin}_{d \in \mathbb{H}} \frac{1}{2} \langle A_k d, d \rangle + \langle \nabla f(x_k), d \rangle.$$

The approach adopted for instance by Fliege and Svaiter [159] consists in solving (4.40) inexactly.

Define for this the optimal value

$$v_k := \inf_{d \in H} \frac{1}{2} \langle A_k d, d \rangle + \langle \nabla f(x_k), d \rangle = \frac{-1}{2} \|\nabla f(x_k)\|_{A_k^{-1}}^2 \leq 0.$$

Given a tolerance  $\sigma \in ]0, 1]$ , take  $x_{k+1} = x_k + \lambda_k d_k$ , where  $d_k$  is a descent direction such that

$$(4.41) \quad \frac{1}{2} \langle A_k d_k, d_k \rangle + \langle \nabla f(x_k), d_k \rangle \leq \sigma v_k.$$

The interest of this approach is that it allows an inexact computation of  $-A_k^{-1} \nabla f(x_k)$ , with an error estimation depending only on the parameter  $\sigma$ . It can be derived from (4.41) that such direction  $d_k$  satisfies

$$\langle \nabla f(x_k), d_k \rangle + \frac{\sigma}{2 \|A_k\|} \|\nabla f(x_k)\|^2 \leq 0 \text{ and } \|d_k\| \leq \frac{2}{\alpha(A)} \|\nabla f(x_k)\|,$$

which means that this method is gradient related, as defined in Example 3.2.1. As a consequence, this inexact method satisfies the hypotheses  $\mathbf{H}_1$  and  $\mathbf{H}_2$ , and enters into the setting of the convergence Theorem 3.2.2.

It would be of interest to take into account such relative errors for the proximal (or Forward-Backward) method. In that case, the analysis should rely on a perturbed proximal operator, for nonconvex functions. See the work of Rockafellar [287], or more recently the study of Salzo-Villa [292] (and the references therein), for examples of perturbed proximal methods.

**Remark 4.3.2** (Compressed sensing in optimal control). From the applications point of view, the *counting norm*  $\|\cdot\|_0$  evoked in Section 4.2.2 has a natural extension to an infinite-dimensional functional setting, namely, the measure of the support of a function  $u$  defined on some  $\Omega \subset \mathbb{R}^N$ . An interesting –but challenging– issue is to apply our algorithm to this extension in order to solve the problem of sparse-optimal control of partial differential equations. From the implementation point of view, it suffices to apply the one-dimensional hard shrinkage operator at each point. Nevertheless, the verification of the KL inequality for this function has not been established and will probably rely on sophisticated arguments concerning the geometry of Hilbert spaces. Then, there is the natural question whether this approach is more efficient than those using the  $L^1$  norm (see, for instance, [94]).

**Remark 4.3.3** (On numerical methods for solving semilinear PDE's). Let  $\Omega$  be an open bounded subset of  $\mathbb{R}^N$  of class  $C^2$ . We aim to study the following semilinear heat equation with Dirichlet boundary conditions:

$$\begin{cases} \dot{\mathbf{u}}(x, t) = \Delta \mathbf{u}(x, t) + g(x, \mathbf{u}(x, t)) & \text{for } (x, t) \in \Omega \times \mathbb{R}^+ \\ \mathbf{u}(x, t) = 0 & \text{for } (x, t) \in \partial\Omega \times \mathbb{R}^+, \end{cases}$$

where  $g : \Omega \times \mathbb{R} \rightarrow \mathbb{R}$  is a source term. As usual, we can use a variational approach for this parabolic PDE. Thus, consider trajectories  $u(t) = \mathbf{u}(\cdot, t)$  satisfying

$$(4.42) \quad \text{for all } t \in \mathbb{R}_+, \quad \dot{u}(t) = \Delta u(t) + g(\cdot, u(t)),$$

this equation being understood in  $H^{-1}(\Omega)$ . A natural problem when facing this equation evolution is the following: how to numerically implement the evolution of this dynamic, given an initial state  $u_0 \in H_0^1(\Omega)$ .

A popular approach consists in discretizing in time the equation (4.42). This leads to the following Euler scheme:

$$(4.43) \quad \forall k \in \mathbb{N}, \quad \frac{u_{k+1} - u_k}{\lambda_k} = \Delta u_{n_k} + g(\cdot, u_{n_k}) \quad \text{in } H^{-1}(\Omega),$$

where  $\lambda_k$  is some nonnegative stepsize, and  $n_k \in \{k, k+1\}$ . Here too, this equation shall be understood in a variational sense. When  $n_k = k$  (respectively  $n_k = k+1$ ), we talk about an explicit (resp. implicit) Euler scheme. The explicit Euler scheme is easy to implement, since  $u_{k+1}$  is computed through an explicit formula, but this method is not satisfactory since it suffers from a lack of stability, see for instance [200, Example 2.3]. On the contrary, the implicit Euler method enjoys a good stability property, see for instance the seminal work of Crandall-Liggett [119], or also Bénélian-Crandall-Pazy [57].

This is the approach adopted by Merlet and Pierre [240], which interpret the implicit Euler scheme as a ‘proximal’ method, applied to the underlying energy of the system:

$$\begin{aligned} \mathcal{E} : H_0^1(\Omega) &\longrightarrow \mathbb{R} \\ u &\longmapsto \frac{1}{2} \int_{\Omega} |\nabla u(x)|^2 dx - \int_{\Omega} \int_0^{u(x)} g(x, t) dt dx. \end{aligned}$$

Under some growth conditions on  $g$ , we are ensured that  $\mathcal{E}$  is of class  $C^2$  (see [209, Proposition I.17.6] and [97, Corollary 5.5.7]). Moreover, assuming  $g(x, t)$  to be analytic in  $t$ , uniformly with respect to  $x \in \Omega$ , a result of Haraux and Jendoubi [187] states<sup>3</sup> that  $\mathcal{E}$  inherits the KL property. Thus, the convergence of the implicit Euler method follows (see [240, Theorem 5.4]).

In the light of the work done in this chapter on splitting methods, it seems natural to adopt an explicit-implicit<sup>4</sup> Euler scheme (see [304, 121, 256, 184]):

$$(4.44) \quad \forall k \in \mathbb{N}, \quad \frac{u_{k+1} - u_k}{\lambda_k} = \Delta u_{k+1} + g(\cdot, u_k) \quad \text{in } H^{-1}(\Omega).$$

This method would be interesting, since it reduces the cost of computing  $u_{k+1}$ , (we only have a linear equation to solve), and we can hope that this method keeps some stability. Thus, it looks like it suffices to apply the Forward-Backward method to  $\mathcal{E}$ , to obtain the convergence of this explicit-implicit Euler scheme (4.44).

In fact, things are more complicated here. This is essentially due to the fact that the heat equation can be seen as a gradient flow for  $\mathcal{E}$  with respect to the  $L^2(\Omega)$  scalar product. In particular, the implicit Euler method corresponds to a proximal algorithm with respect to the  $L^2(\Omega)$  norm:

$$\begin{aligned} u_{k+1} \in H_0^1(\Omega) \quad \text{and} \quad \frac{u_{k+1} - u_k}{\lambda_k} &= \Delta u_{k+1} + g(\cdot, u_{k+1}) \quad \text{in } H^{-1}(\Omega) \\ \Leftrightarrow u_{k+1} &= \operatorname{argmin}_{u \in H_0^1(\Omega)} \mathcal{E}(u) + \frac{1}{2\lambda_k} \|u - u_k\|_{L^2}^2. \end{aligned}$$

Nevertheless, there is still some information to exploit there, like the decrease property:

$$\frac{1}{2\lambda_k} \|u - u_k\|_{L^2}^2 \leq \mathcal{E}(u_k) - \mathcal{E}(u_{k+1}).$$

When applying the Forward-Backward algorithm to  $\mathcal{E}$ , even a basic property like the descent property becomes complicated to obtain. To see this, define

$$\begin{aligned} \mathcal{J} : H_0^1(\Omega) &\longrightarrow \mathbb{R} & \text{and} & \quad \mathcal{G} : H_0^1(\Omega) &\longrightarrow \mathbb{R} \\ u &\longmapsto \frac{1}{2} \int_{\Omega} |\nabla u(x)|^2 dx & & & u &\longmapsto \int_{\Omega} \int_0^{u(x)} g(x, t) dt dx. \end{aligned}$$

<sup>3</sup>In fact, [187, Theorem 2.1] only states that  $\mathcal{E}$  verifies the KL inequality at the critical points having the regularity  $H^2(\Omega)$ . But an elliptic regularity argument ensures that all critical points of  $\mathcal{E}$  does have this regularity (see [97, Theorem 4.4.3] and [172, Theorem 8.12]).

<sup>4</sup>Sometimes written IMEX for short, see [200, 184].

The explicit-implicit Euler scheme (4.44) can then be rewritten as

$$u_{k+1} = \operatorname{argmin}_{u \in H_0^1(\Omega)} \mathcal{J}(u) + \frac{1}{2\lambda_k} \|u - u_k + g(\cdot, u_k)\|_{H^{-1}}^2.$$

As usual, we derive from the optimality condition above this first inequality (use the compact embedding  $L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$ ):

$$\frac{1}{2\lambda_k} \|u_{k+1} - u_k\|_{L^2}^2 + \langle g(\cdot, u_k), u_{k+1} - u_k \rangle_{H^{-1} \times H_0^1} \leq \mathcal{J}(u_k) - \mathcal{J}(u_{k+1}).$$

Assuming that the derivative  $D\mathcal{G} : H_0^1(\Omega) \rightarrow H^{-1}(\Omega)$  is  $L$ -Lipschitz continuous (use for instance [209, Proposition I.17.6]), we can use the descent lemma 2.1.13 to obtain:

$$\frac{1}{2\lambda_k} \|u_{k+1} - u_k\|_{L^2}^2 - \frac{L}{2} \|u_{k+1} - u_k\|_{H_0^1}^2 \leq \mathcal{E}(u_k) - \mathcal{E}(u_{k+1}).$$

We see that we cannot derive the descent property by means of usual methods, which is a clear obstacle. Thus, the study of this splitting method in this setting provides a good challenge for the future.

**Remark 4.3.4** (Going further). Although the set of hypotheses considered in Section 3.2 account for a wide variety of numerical methods described in this Chapter, there exist some algorithms that do not fit into this framework. An interesting task would be to extend the present convergence analysis to such methods:

- Acceleration schemes, like the ones studied in [251, 53, 54], have been proved to have a remarkable performance at least in terms of the values of the objective function. We point out a recent work [254] containing a first attempt in this direction.
- Dealing with on-line rules for updating certain parameters – such as (limited) minimization, Armijo, Goldstein, or Wolfe line search for the step sizes – may be quite useful. See for instance [89] for a gradient method with an inexact line-search satisfying  $\mathbf{H}_1$  and  $\mathbf{H}_2$ .
- Bregman distances provide a relevant alternative to Newton methods, as variable-metric schemes. A first attempt in this direction are the works of Quiroz-Oliveira [280] and Lageman [218], which study descent methods on manifolds.
- Gradient-type methods usually require Lipschitz-continuity of the gradient, at least in sublevel sets. The possibility to consider an asymptotic degeneration of this condition will broaden the scope of these methods.
- Finally, a challenging open problem is to obtain convergence results general primal-dual methods following a Lagrangian approach. Recent preprints give a first advance in this direction, providing convergence results for the ADMM method [225, 198, 316].



## Part II

# Dynamics for Vector optimization



Each individual always seeks to climb the hill of pleasure, to increase his ophelimity.

---

Vilfredo Pareto





## Chapter 5

# An ordered subdifferential for vector-valued functions

We study a general vector optimization problem associated to a locally Lipschitz function  $F : X \rightarrow Y$ , where  $X$  and  $Y$  are two Banach spaces. We consider a nonempty closed convex set  $C \subset X$ , which will model the constraints, and we equip  $Y$  with an order induced by a pointed closed convex cone  $K$ , with nonempty interior. In the following,  $\leq$  will denote the order induced by  $K$ , and  $\preceq$  the associated strict order induced by  $K \setminus \{0\}$ . Moreover, we will note  $<$  the strict order induced by the strict cone  $\text{int } K$ . We recall from Section 2.3.1 that this is equivalent to:

- $y_1 \leq y_2$  if and only if  $y_2 - y_1 \in K$ ,
- $y_1 \preceq y_2$  if and only if  $y_2 - y_1 \in K \setminus \{0\}$ ,
- $y_1 < y_2$  if and only if  $y_2 - y_1 \in \text{int } K$ .

Here we try to solve

$$(VOP) \quad \underset{x \in C}{\text{MIN}} \quad F(x),$$

and look for the set of its efficient points

$$\underset{x \in C}{\text{ARGMIN}} F = \{x \in X \mid [F \preceq F(x)] = \emptyset\},$$

or weak efficient points

$$\underset{x \in C}{\text{ARGMIN}_w} F = \{x \in X \mid [F < F(x)] = \emptyset\}.$$

From now, and until the end of this chapter (and in fact, until the end of Part II), we fix an element  $e \in \text{int } K$ , which gives us a base  $\Theta$  of  $K^+$  by taking:

$$(5.1) \quad \Theta := \{\theta \in K^+ \mid \langle \theta, e \rangle = 1\}.$$

As seen in Section 2.3,  $\Theta$  is weakly\* compact and convex, and the family of real-valued cost functions

$$\{f_\theta := \theta \circ F \mid \theta \in \Theta\}$$

controls the monotonicity of  $F$  (see Theorem 2.3.7). It also matters to consider  $\Xi$  the set of extreme points of  $\Theta$  in  $Y^*$ , and the corresponding subfamily of extreme cost functions

$$\{f_\xi := \xi \circ F \mid \xi \in \Xi\}.$$

The reader can keep in mind along its reading the essential case of multi-objective optimization. When  $(Y, K) = (\mathbb{R}^m, \mathbb{R}_+^m)$  and  $F = (f_1, \dots, f_m)$ , we recover from the definitions of efficiency the usual notions of Pareto and weak Pareto optimality (see Section 2.3). Moreover, if we take  $e = (1, \dots, 1)$ , the family of extreme cost functions is exactly  $\{f_1, \dots, f_m\}$ , while the cost functions  $\{f_\theta\}_{\theta \in \Delta_m}$  corresponds to all the convex combinations

$$\left\{ f_\theta = \sum_{i=1}^m \theta_i f_i \mid \sum_{i=1}^m \theta_i = 1, \theta_i \geq 0 \right\}.$$

In Section 5.1 we introduce the *ordered Clarke subdifferential* of  $F$  as a set-valued mapping  $\partial^c F : X \rightrightarrows X^*$ , defined through the Clarke subgradients of the cost functions:

$$\partial^c F(x) = \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \partial^c f_\theta(x).$$

In particular, in the multi-objective setting  $F = (f_1, \dots, f_m)$ , we show that the ordered Clarke subdifferential reduces to the convex hull of the subgradients of the  $f_i$ 's:

$$\partial^c F(x) = \text{co}\{\partial^c f_1(x), \dots, \partial^c f_m(x)\}.$$

We study in Section 5.1.2 some basic properties, such as its computation in the convex or smooth cases, in Proposition 5.1.8. In Propositions 5.1.11 and 5.1.13 we show some closure properties of the graph of  $\partial^c F : X \rightrightarrows X^*$ . This will be useful in the next chapters, when considering the asymptotic behaviour of a dynamic governed by  $\partial^c F$ . In Section 5.2 we study the descent directions provided by this ordered subdifferential, defined through the notion of *ordered Clarke derivative*

$$d^c F(x, d) := \sup_{\theta \in \Theta} d^c f_\theta(x, d) < 0.$$

In Section 5.2.1, after some definitions, we derive an Armijo's rule in Proposition 5.2.9. We also provide a Fermat's rule in Theorem 5.2.10, which relates the *critical points* satisfying

$$0 \in N_C(x) + \partial^c F(x)$$

to the weak Pareto points. In Section 5.2.2 we show that the ordered Clarke subdifferential  $\partial^c F(x)$  spans the normal cone to the sublevel set  $[F \leq F(x)]$ . Dually, the set of descent directions defined before is exactly the interior of the tangent cone to  $[F \leq F(x)]$ . In Section 5.2.3, we take  $X = H$  as a Hilbert space, and design one particular choice of descent direction, by considering the element of minimal norm in  $-\partial^c F(x)$  (or  $-N_C(x) - \partial^c F(x)$ , if there is a constraint). In a sense specified in Theorem 5.2.20, this direction is the *steepest* among the descent directions at  $x$ . That is why we call it the steepest descent direction at  $x$ , and note<sup>1</sup> it  $s(x)$ . This steepest descent vector field  $s : H \rightarrow H$  generalises the steepest descent vector field  $\nabla f : H \rightarrow H$  usually involved in the minimization of a smooth real-valued function  $f$ . This will naturally led us to study, in the next chapters, dynamics governed by  $s$ . The Figure 5.1 below sums quickly up the situation, for a bi-objective smooth function  $F = (f_1, f_2) : \mathbb{R}^2 \rightarrow (\mathbb{R}^2, \mathbb{R}_+^2)$ .

---

<sup>1</sup>Letter  $s$  stands for steepest.







## 5.1 An ordered Clarke subdifferential for vector-valued functions

### 5.1.1 Definition

We introduce what will be the central tools for our study of vector optimization problems.

**Definition 5.1.1.** For any  $x \in X$ , and  $d \in X$ , we define the *ordered (Clarke) directional derivative* of  $F$  at  $x$  in the direction  $d$  by

$$d_{\Theta}^C F(x, d) := \sup_{\theta \in \Theta} d^C f_{\theta}(x, d).$$

In general, if there is no ambiguity, we will omit the mention of the base  $\Theta$  and just write  $d^C F(x, d)$ .

We easily see that  $d^C F(x, \cdot)$  is sublinear, as the supremum of a family of sublinear functions. Moreover, since  $\Theta$  is bounded and  $F$  is locally Lipschitz, we also see that  $d^C F(x, \cdot)$  takes finite values and is Lipschitz continuous (see Proposition A.3.4).

**Remark 5.1.2.** Pay attention to the fact that  $d^C F(x, d)$  is a scalar, and must not be taken for the directional derivative  $DF(x, d) \in Y$ . When  $F$  is strictly Gateaux differentiable, we show in Proposition 5.1.8 that  $d^C F(x, d)$  and  $DF(x, d)$  are related through the support function of  $\Theta$ .

Like in the classic construction of the Clarke subdifferential, we derive from the ordered Clarke directional derivative above a dual object:

**Definition 5.1.3.** For any  $x \in X$ , we define the *ordered (Clarke) subdifferential* of  $F$  at  $x$  by

$$(5.2) \quad \partial_{\Theta}^C F(x) := \{x^* \in X^* \mid \forall d \in X, \langle x^*, d \rangle \leq d^C F(x, d)\}.$$

Again, if there is no ambiguity, we will omit the mention of the base  $\Theta$  and just write  $\partial^C F(x)$ .

It is a nonempty weakly\* compact convex set. So, by applying [5, Theorem 7.51] we see that the following *max formula* is satisfied for all  $x, d \in X$ :

$$(5.3) \quad d^C F(x, d) = \max_{x^* \in \partial^C F(x)} \langle x^*, d \rangle.$$

From this max formula we can deduce an explicit expression for  $\partial^C F(x)$ , involving the subgradients of the cost functions  $\{f_{\theta}\}_{\theta \in \Theta}$ :

**Proposition 5.1.4.** For all  $x \in X$ ,

$$\partial^C F(x) = \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \partial^C f_{\theta}(x).$$

*Proof.* From the sum rule for the Clarke subdifferential of scalar-valued functions, we know that for all  $\theta \in \Theta$ ,

$$d^C f_{\theta}(x, d) = \max_{x^* \in \partial^C f_{\theta}(x)} \langle x^*, d \rangle.$$

It follows that  $d^C F(x, d)$  is, by definition,

$$\max_{x^* \in \bigcup_{\theta \in \Theta} \partial^C f_{\theta}(x)} \langle x^*, d \rangle.$$

From the max formula (5.3) and Proposition 2.1.4 on support functions, the conclusion follows. ■

Next proposition shows that  $\partial^C F(x)$  is simpler in the polyhedral case: it only involves the subgradients of the extreme cost functions, corresponding to the extreme points of  $\Theta$ .

**Proposition 5.1.5.** (Polyhedral case) Suppose that  $K$  is polyhedral. Write  $\Theta = \text{co}\{\xi_1, \dots, \xi_m\}$  and let  $f_i := \xi_i \circ F$  for  $i \in \{1, \dots, m\}$ . Then

$$\partial^C F(x) = \text{co } \partial^C f_i(x) \text{ and } d^C F(x, d) = \max_{i \in \{1, \dots, m\}} d^C f_i(x, d).$$

**Example 5.1.6.** This Proposition applies in particular to the multicriteria optimization context: if  $(Y, K) = (\mathbb{R}^m, \mathbb{R}_+^m)$  and  $F(x) = (f_1(x), \dots, f_m(x))$ , then

$$d^C F(x; d) = \max_{i \in \{1, \dots, m\}} d^C f_i(x; d),$$

and  $\partial^C F(x)$  is the convex hull of the subdifferentials  $\partial^C f_i(x)$ . We see that we recover here the set studied by Cornet for smooth functions in [116], and revisited in [28]. In particular,  $\partial^C F(x)$  reduces to the classical Clarke subdifferential when  $m = 1$ .

Figure 5.2: The ordered Clarke subdifferential of  $F = (f_1, f_2)$  at  $x$  with respect to  $\mathbb{R}_+^2$  is the convex hull of  $\nabla f_1(x), \nabla f_2(x)$ .

*Proof of Proposition 5.1.5.* First, recall that, since it is assumed that  $\text{int } K$  is nonempty,  $K$  being polyhedral implies that  $Y$  is finite dimensional. It follows from Section 2.1.2 that  $K^+$  is also polyhedral, so we can indeed write  $\Theta = \text{co}\{\xi_1, \dots, \xi_m\}$ .

Let us start by proving that  $d^C F(x, d) = \max\{d^C f_i(x, d) \mid i = 1, \dots, m\}$ . Fix  $x$  and  $d$  in  $X$ , then it is clear from the definition of the  $f_i$ 's that

$$(5.4) \quad \max_{i \in \{1, \dots, m\}} d^C f_i(x, d) \leq \sup_{\theta \in \Theta} d^C f_\theta(x, d) = d^C F(x, d).$$

Let  $(\theta_n)_{n \in \mathbb{N}}$  be a sequence in  $\Theta$  such that  $d^C F(x, d) = \lim_{n \rightarrow +\infty} d^C f_{\theta_n}(x, d)$ . For each  $n \in \mathbb{N}$ , there exists  $(\lambda_1^n, \dots, \lambda_m^n)$  in the simplex unit  $\Delta_m$  such that  $\theta_n = \sum_{i=1}^m \lambda_i^n \xi_i$ . Hence, using the sum rule for the Clarke directional derivative :

$$(5.5) \quad d^C f_{\theta_n}(x, d) = d^C(\theta_n \circ F)(x, d) = d^C \left( \sum_{i=1}^m \lambda_i^n (\xi_i \circ F) \right) (x, d) \leq \sum_{i=1}^m \lambda_i^n d^C f_i(x, d).$$

Using the fact that  $\sum_{i=1}^m \lambda_i^n = 1$ , we see that the right member in (5.5) is bounded from above by  $\max\{d^C f_i(x, d) \mid i = 1, \dots, m\}$  for all  $n \in \mathbb{N}$ . By taking the limit when  $n \rightarrow +\infty$ , we conclude that

$$(5.6) \quad d^C F(x, d) = \max_{i \in \{1, \dots, m\}} d^C f_i(x, d)$$

Now, we use the max formula (5.3) for the ordered subdifferential to see that (5.6) means that the sets  $\partial^C F(x)$  and  $\bigcup_{i=1}^m \partial^C f_i(x)$  have the same support functions. Applying Theorem 2.1.4, we deduce that

$$\partial^C F(x) = \overline{\text{co}}^* \bigcup_{i=1}^m \partial^C f_i(x).$$

Since  $\partial^C f_i(x)$  is convex and compact for all  $i \in \{1, \dots, m\}$ , it follows from Proposition 2.1.3 that  $\text{co } \bigcup_{i=1}^m \partial^C f_i(x)$  is compact, hence closed. In other words,

$$\partial^C F(x) = \text{co } \bigcup_{i=1}^m \partial^C f_i(x).$$

■



In Section 5.2 we will study in details the descent directions induced by the directional derivative  $d^c F(x; \cdot)$ , exploiting the duality between  $d^c F(x; \cdot)$  and  $\partial^c F(x)$ . But first, we will focus on topological and geometrical properties of  $\partial^c F$  which will be needed later.

### 5.1.2 Properties of the ordered Clarke subdifferential

Let us start by showing that  $\partial^c F : X \rightrightarrows X^*$  is locally bounded, as a consequence of the local Lipschitz continuity of  $F$ .

**Proposition 5.1.7** (Local boundedness). For all  $\bar{x} \in X$ , there exists a neighbourhood  $U$  of  $\bar{x}$ , and  $M \geq 0$  such that

$$\forall x \in U, \forall x^* \in \partial^c F(x), \|x^*\| \leq M.$$

*Proof.* Let  $U$  be a neighborhood of  $\bar{x}$  on which  $F$  is Lipschitz continuous. Since  $\Theta$  is bounded, the family  $\{f_\theta = \theta \circ F \mid \theta \in \Theta\}$  is equi-Lipschitz continuous on  $U$ , for some Lipschitz constant  $M \geq 0$ . From Proposition 2.2.14 we have

$$\forall x \in U, \forall \theta \in \Theta, \forall x^* \in \partial^c f_\theta(x), \|x^*\| \leq M.$$

The conclusion follows the formula  $\partial^c F(x) = \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \partial^c f_\theta(x)$  proved in Proposition 5.1.4, and the weak\* lower semi-continuity of the norm in  $X^*$ . ■

We study now some hypotheses under which the ordered Clarke subdifferential has a simpler form. More exactly, we are interested on the cases for which  $\partial^c F(x)$  is exactly the reunion of the subgradients of the cost functions  $\{f_\theta\}_{\theta \in \Theta}$ . In other words, we look for conditions insuring that we can remove the closed convex hull in the expression

$$\partial^c F(x) = \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \partial^c f_\theta(x).$$

We will say that  $F : X \rightarrow Y$  is *positively Clarke regular* if,

$$\text{for all } \theta \in \Theta, f_\theta := \theta \circ F : X \rightarrow \mathbb{R} \text{ is Clarke regular, i.e. } \partial^B f_\theta = \partial^c f_\theta.$$

It is equivalent to ask the directional derivatives  $d^B f_\theta(x, \cdot)$  and  $d^c f_\theta(x, \cdot)$  to be equal. Recall also from Section 2.2.5 that, if  $X$  is reflexive, it is also equivalent to ask  $\partial^F f_\theta = \partial^c f_\theta$  for all  $\theta \in \Theta$ . It happens for instance if  $F$  is the sum of a convex and a strictly Gateaux differentiable functions.

**Proposition 5.1.8** (Ordered Clarke subdifferential in regular cases).

- i) If  $F$  is the sum of a convex function and a strictly Gateaux differentiable function at  $x \in X$ , then  $\partial^c F(x) = \bigcup_{\theta \in \Theta} \partial^c f_\theta(x)$  and  $d^c F(x; d) = \max_{\theta \in \Theta} d^c f_\theta(x; d)$  for all  $d \in X$ .
- ii) If  $F$  is positively Clarke regular, with  $X$  being a reflexive Banach space and  $Y$  finite dimensional, then the same conclusion than item i) holds.
- iii) When  $F$  is strictly Gateaux differentiable at  $x \in X$ , then

$$d^c F(x, d) = \sigma_\Theta(DF(x; d)) \text{ and } \partial^c F(x) = \Theta \circ DF(x) = D^* F(x; \Theta).$$

**Remark 5.1.9.** When  $F$  is smooth, according to item iii), taking an element in  $\partial^c F(x)$  is equivalent to take some  $\theta \in \Theta$  and compute  $\theta \circ DF(x)$ . For instance, in the multi-objective setting, it corresponds to the choice of some convex combination of the derivatives  $Df_i(x)$ . We will see later in Theorem 7.3.3 that it is also equivalent to chose a Radon probability  $\mu$  over  $\Theta$ , and compute  $\int_\Theta Df_\theta \, d\mu(\theta)$ .

*Proof.* We start with item iii). Considering that  $F$  is strictly Gateaux differentiable, we have for all  $\theta \in \Theta$  and  $d \in X$  that  $d^c f_\theta(x; d) = D(\theta \circ F)(x; d) = \theta \circ DF(x; d)$ . It follows that

$$d^c F(x; d) = \sup_{\theta \in \Theta} \theta \circ DF(x; d) = \sigma_\Theta(DF(x; d)).$$

In other words,  $d^c F(x; d) = \sigma_{\Theta \circ DF(x)}(d)$ , where  $\Theta \circ DF(x) := \{\theta \circ DF(x) \mid \theta \in \Theta\}$ , which is also equal to  $D^*F(x; \Theta)$ . From the definition of  $\partial^c F(x)$ , and the characterization of support functions in Theorem 2.1.4, we deduce that  $\partial^c F(x) = \overline{c\sigma^* \Theta \circ DF(x)}$ . Since  $\Theta$  is convex and weakly\* closed, item iii) follows.

We turn now on items i) and ii). The whole point is to prove that  $\bigcup_{\theta \in \Theta} \partial^c f_\theta(x)$  is convex and weakly\* closed. Note that, in both items i) and ii), the cost functions  $f_\theta$  verify  $\partial^F f_\theta = \partial^c f_\theta$ . Hence, it is enough to prove that  $\bigcup_{\theta \in \Theta} \partial^F f_\theta(x)$  is convex and weakly\* closed.

Let us prove first that it is convex. Consider  $x_1^*$  and  $x_2^*$  be in  $\bigcup_{\theta \in \Theta} \partial^F f_\theta(x)$ , and let  $x^* := \lambda x_1^* + (1 - \lambda)x_2^*$  for some  $\lambda \in [0, 1]$ . By definition, there exists  $\theta_1, \theta_2 \in \Theta$  such that  $x_1^* \in \partial^F f_{\theta_1}(x)$  and  $x_2^* \in \partial^F f_{\theta_2}(x)$ . Let  $\theta := \lambda \theta_1 + (1 - \lambda)\theta_2$ , from convexity of  $\Theta$  we know that  $\theta \in \Theta$ . We still have to verify that  $x^* \in \partial^F f_\theta(x)$ . Using  $f_\theta = \lambda f_{\theta_1} + (1 - \lambda)f_{\theta_2}$ , we obtain

$$\begin{aligned} & \liminf_{x' \rightarrow x} \frac{f_\theta(x') - f_\theta(x) - \langle x^*, x' - x \rangle}{\|x' - x\|} \\ &= \liminf_{x' \rightarrow x} \lambda \frac{f_{\theta_1}(x') - f_{\theta_1}(x) - \langle x_1^*, x' - x \rangle}{\|x' - x\|} + (1 - \lambda) \frac{f_{\theta_2}(x') - f_{\theta_2}(x) - \langle x_2^*, x' - x \rangle}{\|x' - x\|} \\ &\geq \lambda \liminf_{x' \rightarrow x} \frac{f_{\theta_1}(x') - f_{\theta_1}(x) - \langle x_1^*, x' - x \rangle}{\|x' - x\|} + (1 - \lambda) \liminf_{x' \rightarrow x} \frac{f_{\theta_2}(x') - f_{\theta_2}(x) - \langle x_2^*, x' - x \rangle}{\|x' - x\|} \\ &\geq 0. \end{aligned}$$

Now that the convexity of  $\bigcup_{\theta \in \Theta} \partial^c f_\theta(x)$  is proved, we study its weak\* closedness, distinguishing items i) and ii).

Suppose first that the hypothesis of item i) is satisfied, *i.e.* that  $F = G + H$ , where  $G$  is convex and  $H$  is strictly Gateaux differentiable at  $x$ . Let  $(x_\alpha^*)_{\alpha \in \mathcal{A}}$  be a net in  $\bigcup_{\theta \in \Theta} \partial^c f_\theta(x)$ , weakly\* converging to some  $x^* \in X^*$ . For all  $\alpha \in \mathcal{A}$ , there exists some  $\theta_\alpha \in \Theta$  such that

$$x_\alpha^* \in \partial^c f_{\theta_\alpha}(x) = \partial g_{\theta_\alpha}(x) + \theta_\alpha \circ DH(x),$$

where  $g_{\theta_\alpha} := \theta_\alpha \circ G$  is a convex function. We can assume, without loss of generality, that  $\theta_\alpha$  weakly\* converges to some  $\theta \in \Theta$ . Define  $z_\alpha^* := x_\alpha^* - \theta_\alpha \circ DH(x) \in \partial g_{\theta_\alpha}(x)$ , which is a net weakly\* convergent to  $z^* := x^* - \theta \circ DH(x)$ . Let us show that  $z^* \in \partial g_\theta(x)$ . Using  $z_\alpha^* \in \partial g_{\theta_\alpha}(x)$ , we have for all  $x' \in X$ :

$$(\theta \circ G)(x') - (\theta \circ G)(x) - \langle z^*, x' - x \rangle = \lim_{\alpha \in \mathcal{A}} (\theta_\alpha \circ G)(x') - (\theta_\alpha \circ G)(x) - \langle z_\alpha^*, x' - x \rangle \geq 0.$$

So  $z^* \in \partial g_\theta(x)$ , and it follows that  $x^* \in \partial^c f_\theta(x) \subset \bigcup_{\theta \in \Theta} \partial^c f_\theta(x)$ .

Suppose now that the hypotheses of item ii) are satisfied. Since  $X$  is supposed to be reflexive, and given that  $\bigcup_{\theta \in \Theta} \partial^c f_\theta(x)$  is convex, it is enough to verify that it is closed for the norm topology in  $X^*$ . Take then a sequence  $(x_n^*)_{n \in \mathbb{N}}$  in  $\bigcup_{\theta \in \Theta} \partial^c f_\theta(x)$  such that  $x_n^*$  tends to some  $x^* \in X^*$  for the norm topology. There exists a corresponding sequence  $\theta_n \in \Theta$  such that  $x_n^* \in \partial^c f_{\theta_n}(x)$ . Since  $\Theta$  is compact in  $Y$  (which is supposed to be finite-dimensional), we can assume, taking eventually a subsequence, that  $\theta_n$  converges to some  $\theta \in \Theta$ . We just need to show now that  $x^* \in \partial^c f_\theta(x)$ ,

which is equal to  $\partial^F f_\theta(x)$ , according to the positive Clarke regularity of  $F$  and the reflexivity of  $X$ . Start by writing, for all  $n \in \mathbb{N}$ :

$$\begin{aligned} & \liminf_{x' \rightarrow x} \frac{f_\theta(x') - f_\theta(x) - \langle x^*, x' - x \rangle}{\|x' - x\|} \\ = & \liminf_{x' \rightarrow x} \frac{(\theta_n \circ F)(x') - (\theta_n \circ F)(x) - \langle x_n^*, x' - x \rangle}{\|x' - x\|} + \frac{\langle \theta - \theta_n, F(x') - F(x) \rangle}{\|x' - x\|} + \frac{\langle x_n^* - x^*, x' - x \rangle}{\|x' - x\|} \\ \geq & \liminf_{x' \rightarrow x} \frac{\langle \theta - \theta_n, F(x') - F(x) \rangle}{\|x' - x\|} + \liminf_{x' \rightarrow x} \frac{\langle x_n^* - x^*, x' - x \rangle}{\|x' - x\|}, \end{aligned}$$

where we used  $x_n^* \in \partial^F f_{\theta_n}(x)$  in the last inequality. Using the Lipschitz property of  $F$  around  $x$ , together with the Cauchy-Schwarz inequality, we deduce that

$$(5.7) \quad \liminf_{x' \rightarrow x} \frac{f_\theta(x') - f_\theta(x) - \langle x^*, x' - x \rangle}{\|x' - x\|} \geq -\|\theta - \theta_n\|_Y \text{Lip}(F, x) - \|x_n^* - x^*\|.$$

Passing to the limit when  $n \rightarrow +\infty$  in (5.7), we finally obtain that  $x^* \in \partial^F f_\theta(x) = \partial^C f_\theta(x)$ . ■

**Example 5.1.10.** We give a simple example for which Proposition 5.1.8 fails. Take  $F = (f_1, f_0) : \mathbb{R}^2 \rightarrow (\mathbb{R}^2, \mathbb{R}_+^2)$ , defined by

$$\begin{aligned} f_1(x) &:= \|x\|_1 + \langle (2, 0), x \rangle, \\ f_0(x) &:= -\|x\|_1 + \langle (0, 2), x \rangle, \end{aligned}$$

where  $\|\cdot\|_1$  denotes the usual  $\ell^1$ -norm

$$\forall x = (x_1, x_2) \in \mathbb{R}^2, \quad \|x\|_1 := |x_1| + |x_2|.$$

We have in particular (we also note 0 for the origin in  $\mathbb{R}^2$ ):

$$\partial^C f_1(0) = [-1, 1] + (2, 0) \text{ and } \partial^C f_0(0) = [-1, 1] + (0, 2).$$

We consider now the convex combinations of  $f_1$  and  $f_0$ . For all  $\lambda \in [0, 1]$ , we note  $f_\lambda := \lambda f_1 + (1 - \lambda) f_0$ , that is

$$f_\lambda(x) = |2\lambda - 1| \|x\|_1 + \langle a_\lambda, x \rangle, \text{ with } a_\lambda := 2(\lambda, (1 - \lambda)).$$

The Clarke subdifferential of these functions is given by

$$\partial^C f_\lambda(0) = |2\lambda - 1|[-1, 1] + a_\lambda.$$

Thus, it is easy to see that

$$\forall \lambda \in [0, \frac{1}{2}], \partial^C f_\lambda(0) \subset \partial^C f_0(0) \text{ and } \forall \lambda \in [\frac{1}{2}, 1], \partial^C f_\lambda(0) \subset \partial^C f_1(0).$$

In other words (see Figure 5.3 below),

$$\bigcup_{\lambda \in [0, 1]} \partial^C f_\lambda(0) \subsetneq \partial^C F(0) = \text{co}\{\partial^C f_1(0), \partial^C f_0(0)\}.$$

This is essentially due to the lack of Clarke regularity of  $f_0$ .

Figure 5.3: A counterexample to Proposition 5.1.8.

We end now by studying the closed-graph properties of  $\partial^c F : X \rightrightarrows X^*$ . More generally, and because we shall want to take into account some constraints, we study the sum of  $\partial^c F$  and  $N_C$ , where  $N_C(x)$  denotes the normal cone to  $C$  at  $x$ .

**Proposition 5.1.11** (Bounded weak-strong outer semi-continuity, convex case). Suppose that  $F$  is convex, and let  $(x_\alpha, x_\alpha^*)_{\alpha \in \mathbb{N}}$  be a bounded net in  $X \times X^*$  such that

$$(5.8) \quad x_\alpha \xrightarrow[\alpha \in \mathcal{A}]{w} x, \quad x_\alpha^* \xrightarrow[\alpha \in \mathcal{A}]{\|\cdot\|} x^*, \quad \text{and } x_\alpha^* \in N_C(x_\alpha) + \partial^c F(x_\alpha).$$

Then  $x^* \in N_C(x) + \partial^c F(x)$  holds, provided that one of the two following hypotheses is satisfied :

- i)  $\dim Y < +\infty$  and  $F$  is Lipschitz continuous on bounded sets;
- ii) for all  $\alpha \in \mathcal{A}$ ,  $F(x) \leq F(x_\alpha)$ .

**Remark 5.1.12.** The case covered by item ii) will be of particularly interest when considering *descent* dynamics associated to  $N_C + \partial^c F$ , in Section 6.2.

*Proof.* Take nets as in (5.8), and use the convexity of  $F$  together with Proposition 5.1.8, to write  $\partial^c F(x_\alpha) = \bigcup_{\theta \in \Theta} \partial f_\theta(x_\alpha)$ . Then, for all  $\alpha \in \mathcal{A}$ , there exists  $\theta_\alpha \in \Theta$  such that

$$x_\alpha^* \in N_C(x_\alpha) + \partial^c f_{\theta_\alpha}(x_\alpha).$$

Since  $f_\theta$  is a scalar-valued convex continuous function, we can use the sum rule (Proposition 2.2.6) and equivalently say that  $x_\alpha^* \in \partial(\delta_C + f_\theta)(x_\alpha)$ , where  $\delta_C$  is the indicator<sup>2</sup> function of  $C$ . From the compactness of  $\Theta$ , we can assume that  $(\theta_\alpha)_{\alpha \in \mathcal{A}}$  weakly\* converges to some  $\theta \in \Theta$ . We write now, for all  $\alpha \in \mathcal{A}$ , that  $x_\alpha^* \in \partial(\delta_C + f_{\theta_\alpha})(x_\alpha)$  is equivalent to :

$$(5.9) \quad \forall x' \in C, \quad f_{\theta_\alpha}(x_\alpha) \leq f_{\theta_\alpha}(x') - \langle x_\alpha^*, x' - x_\alpha \rangle.$$

Look at the right member in (5.9). The first term  $f_{\theta_\alpha}(x') = \langle \theta_\alpha, F(x') \rangle$  converges to  $\langle \theta, F(x') \rangle = f_\theta(x')$  because  $\theta_\alpha \xrightarrow{w^*} \theta$ . The second term  $\langle x_\alpha^*, x' - x_\alpha \rangle$  converges to  $\langle x^*, x' - x \rangle$ , since  $x_\alpha^*$  strongly converges to  $x^*$  and  $x_\alpha$  is a bounded net weakly convergent to  $x$ . So, we have, passing to the limit in (5.9), that

$$\forall x' \in C, \quad \limsup_{\alpha \in \mathcal{A}} f_{\theta_\alpha}(x_\alpha) \leq f_\theta(x') - \langle x^*, x' - x \rangle.$$

So, all we need to prove now is

$$(5.10) \quad f_\theta(x) \leq \limsup_{\alpha \in \mathcal{A}} f_{\theta_\alpha}(x_\alpha),$$

since we would have

$$\forall x' \in C, \quad f_\theta(x) \leq f_\theta(x') - \langle x^*, x' - x \rangle,$$

and  $x^* \in \partial(\delta_C + f_\theta)(x) = N_C(x) + \partial f_\theta(x)$  would follow.

We prove now (5.10) using the hypotheses. Suppose first that  $\dim Y < +\infty$  and  $F$  is Lipschitz continuous on bounded sets. Using the convexity of  $f_\theta$ , and so, its weak lower-semicontinuity, we can write for all  $\alpha \in \mathcal{A}$  :

$$f_\theta(x) \leq \liminf_{\alpha \in \mathcal{A}} f_\theta(x_\alpha) \leq \limsup_{\alpha \in \mathcal{A}} (\theta_\alpha \circ F)(x_\alpha) + \limsup_{\alpha \in \mathcal{A}} \langle \theta - \theta_\alpha, F(x_\alpha) \rangle.$$

---

<sup>2</sup>The indicator function of  $C$  is a function taking as values  $\delta_C(x) = 0$  if  $x \in C$ ,  $+\infty$  otherwise.

In the last term, we have  $(\theta - \theta_\alpha)$  which converges to zero for the norm topology (recall  $\dim Y < +\infty$ ). Moreover,  $(x_\alpha)_{\alpha \in \mathcal{A}}$  is bounded in  $X$  and  $F$  is Lipschitz continuous on bounded sets, so  $F(x_\alpha)$  is also bounded. It follows that  $(\theta - \theta_\alpha)(F(x_\alpha))$  goes to zero, and (5.10) is proved.

Suppose now that  $F(x) \leq F(x_\alpha)$  is satisfied. In particular (see Proposition 2.3.7), for all  $\alpha \in \mathcal{A}$ ,

$$(\theta_\alpha \circ F)(x) \leq (\theta_\alpha \circ F)(x_\alpha).$$

Taking the limsup over  $\alpha \in \mathcal{A}$  on the inequality above, leads to

$$f_\theta(x) \leq \liminf_{\alpha \in \mathcal{A}} f_{\theta_\alpha}(x_\alpha)$$

and proves again (5.10). ■

We now study the strong-weak outer semi-continuity of  $\partial^c F$ . In the following, we will use of this set of hypotheses :

One of the three following properties is satisfied :

- H**
- 1)  $K$  is polyhedral,
  - 2)  $Y$  has finite dimension,  $X$  is reflexive and  $F$  is positively Clarke regular,
  - 3)  $F$  is the sum of a convex function and a strictly Gateaux differentiable function.

Observe that there is in these three hypotheses some trade-off between the regularity of the function  $F$  and the properties of  $X, Y, K$ .

**Proposition 5.1.13** (Strong-weak sequential outer semi-continuity, without constraint). Suppose that **H** holds, and let  $(x_\alpha, x_\alpha^*)_{n \in \mathbb{N}}$  be a net in  $X \times X^*$  such that

$$(5.11) \quad x_\alpha \xrightarrow[\alpha \in \mathcal{A}]{\|\cdot\|} x, \quad x_\alpha^* \xrightarrow[\alpha \in \mathcal{A}]{w^*} x^*, \quad \text{and } x_\alpha^* \in \partial^c F(x_\alpha).$$

Then  $x^* \in \partial^c F(x)$ .

*Proof.* Because of the hypothesis, we will prove this statement in two times: first we will consider that  $F$  is the sum of a convex and a strictly Gateaux differentiable function, and in a second time that either **H1**) or **H2**) holds.

Suppose then that  $F = G + H$ , with  $G$  being convex and  $H$  strictly Gateaux differentiable. Thanks to Proposition 5.1.8, and applying the sum rule for the convex subdifferential, we have for all  $\alpha \in \mathcal{A}$  that there exists  $\theta_\alpha$  satisfying  $x_\alpha^* \in \partial g_{\theta_\alpha}(x_\alpha) + \theta_\alpha \circ DH(x_\alpha)$ . Without loss of generality, we can assume that  $\theta_\alpha$  weakly\* converges to some  $\theta \in \Theta$ .

We know from Proposition 2.1.11 that the strict Gateaux differentiability of  $H$  in a neighbourhood of  $x$  implies that  $DH : U \rightarrow L(X, Y)$  is continuous,  $L(X, Y)$  being endowed with the topology of the pointwise convergence. In other words,

$$(5.12) \quad \text{for all } d \in X, \quad DH(x_\alpha; d) \xrightarrow[\alpha \in \mathcal{A}]{\|\cdot\|_Y} DH(x; d).$$

Hence, for all  $d \in X$ , we can use the boundedness and weak\* convergence of  $(\theta_\alpha)_{\alpha \in \mathcal{A}}$  to obtain :

$$(5.13) \quad \lim_{\alpha \in \mathcal{A}} \theta_\alpha \circ DH(x_\alpha)(d) = \lim_{\alpha \in \mathcal{A}} \langle \theta_\alpha, DH(x_\alpha; d) \rangle_{Y^* \times Y} = \langle \theta, DH(x; d) \rangle_{Y^* \times Y} = \theta \circ DH(x)(d).$$

The latter being true for any  $d \in X$ , it means that  $\theta_\alpha \circ DH(x_\alpha)$  converges weakly\* to  $\theta \circ DH(x)$ .

Write now  $\tilde{x}_\alpha := x_\alpha^* - \theta_\alpha \circ DH(x_\alpha) \in \partial g_{\theta_\alpha}(x_\alpha)$ . From what we just have shown,  $\tilde{x}_\alpha$  weakly\* converges to  $\tilde{x} := x^* - \theta \circ DH(x)$ . Use the convexity of  $g_{\theta_\alpha}$  to write, for all  $\alpha \in \mathcal{A}$ ,

$$(5.14) \quad \forall y \in X, \quad g_{\theta_\alpha}(y) - g_{\theta_\alpha}(x_\alpha) - \langle \tilde{x}_\alpha, y - x_\alpha \rangle \geq 0.$$

The first term  $g_{\theta_\alpha}(y) = \langle \theta_\alpha, G(y) \rangle$  tends to  $g_\theta(y)$ , using the weak\* convergence of  $(\theta_\alpha)_{\alpha \in \mathcal{A}}$ . The second term  $g_{\theta_\alpha}(x_\alpha)$  converges to  $g_\theta(x)$  because of the weak\* convergence of the bounded net  $(\theta_\alpha)_{\alpha \in \mathcal{A}}$ , and the strong convergence of  $F(x_\alpha)$  to  $F(x)$ . In the third term, we use strong-weak\* convergence of  $(x_\alpha, x_\alpha^*)$  to  $(x, x^*)$ , together with the fact that  $(x_\alpha^*)_{\alpha \in \mathcal{A}}$  is bounded, this being a consequence of the local boundedness of  $\partial^c F$  (see Proposition 5.1.7). We finally obtain

$$(5.15) \quad \forall y \in X, \quad g_\theta(y) - g_\theta(x) - \langle \tilde{x}, y - x \rangle \geq 0,$$

and it follows that  $\tilde{x} \in \partial g_\theta(x)$ . Hence,  $x^* = \tilde{x} + \theta \circ DH(x) \in \partial g_\theta(x) + \theta \circ DH(x) \subset \partial^c F(x)$ .

Now we suppose that either **H1**) or **H2**) holds. In both cases, the outer semi-continuity of  $\partial^c F$  will be a direct consequence of the upper semi-continuity of  $d^c F(\cdot, d)$ , that we prove first.

Let then  $d$  be in  $X$ , consider a converging net  $x_\alpha \xrightarrow[\alpha \in \mathcal{A}]{\|\cdot\|_X} x$ , and show that

$$(5.16) \quad \limsup_{\alpha \in \mathcal{A}} d^c F(x_\alpha, d) \leq d^c F(x, d).$$

We claim that there exists for all  $\alpha \in \mathcal{A}$  some  $\theta_\alpha \in \Theta$  such that  $d^c F(x_\alpha; d) = d^c f_{\theta_\alpha}(x_\alpha; d)$ . To see this when **H1**) holds, use Proposition 5.1.5, and if **H2**) holds, use Proposition 5.1.8. We assume that  $\theta_\alpha$  converges to some  $\theta \in \Theta$ , taking eventually a subnet. Consider a subnet  $(\alpha_\beta)_{\beta \in \mathcal{B}}$  for which the limsup is attained in (5.16), *i.e.*

$$(5.17) \quad \lim_{\beta \in \mathcal{B}} d^c f_{\theta_{\alpha_\beta}}(x_{\alpha_\beta}; d) = \limsup_{\alpha \in \mathcal{A}} d^c f_{\theta_\alpha}(x_\alpha; d) \quad (\text{and so } = \limsup_{\alpha \in \mathcal{A}} d^c F(x_\alpha; d)).$$

What we aim to show now is that

$$(5.18) \quad \limsup_{\beta \in \mathcal{B}} d^c f_\theta(x_{\alpha_\beta}; d) - d^c f_{\theta_{\alpha_\beta}}(x_{\alpha_\beta}; d) \geq 0.$$

Note  $c$  the limsup appearing in (5.18). By definition of Clarke's directional derivative,

$$(5.19) \quad c = \limsup_{\beta \in \mathcal{B}} \limsup_{\substack{\tilde{x}^{(\beta)} \rightarrow x_{\alpha_\beta} \\ t \downarrow 0}} \frac{f_\theta(\tilde{x}^{(\beta)} + td) - f_\theta(\tilde{x}^{(\beta)})}{t} - \limsup_{\substack{\tilde{x}^{(\beta)} \rightarrow x_{\alpha_\beta} \\ t \downarrow 0}} \frac{f_{\theta_{\alpha_\beta}}(\tilde{x}^{(\beta)} + td) - f_{\theta_{\alpha_\beta}}(\tilde{x}^{(\beta)})}{t}.$$

For all  $\beta \in \mathcal{B}$ , let  $\tilde{x}_\gamma^{(\beta)} \xrightarrow[\gamma \in \Gamma]{} x_{\alpha_\beta}$  and  $t_\gamma \xrightarrow[\gamma \in \Gamma]{} 0$  be two nets such that

$$\limsup_{\substack{\tilde{x}^{(\beta)} \rightarrow x_{\alpha_\beta} \\ t \downarrow 0}} \frac{f_{\theta_{\alpha_\beta}}(\tilde{x}^{(\beta)} + td) - f_{\theta_{\alpha_\beta}}(\tilde{x}^{(\beta)})}{t} = \lim_{\gamma \in \Gamma} \frac{f_{\theta_{\alpha_\beta}}(\tilde{x}_\gamma^{(\beta)} + t_\gamma d) - f_{\theta_{\alpha_\beta}}(\tilde{x}_\gamma^{(\beta)})}{t_\gamma}.$$

Injecting the latter in (5.19), we obtain

$$\begin{aligned} c &= \limsup_{\beta \in \mathcal{B}} \limsup_{\substack{\tilde{x}^{(\beta)} \rightarrow x_{\alpha_\beta} \\ t \downarrow 0}} \frac{f_\theta(\tilde{x}^{(\beta)} + td) - f_\theta(\tilde{x}^{(\beta)})}{t} - \lim_{\gamma \in \Gamma} \frac{f_{\theta_{\alpha_\beta}}(\tilde{x}_\gamma^{(\beta)} + t_\gamma d) - f_{\theta_{\alpha_\beta}}(\tilde{x}_\gamma^{(\beta)})}{t_\gamma} \\ &\geq \limsup_{\beta \in \mathcal{B}} \limsup_{\gamma \in \Gamma} \frac{f_\theta(\tilde{x}_\gamma^{(\beta)} + td) - f_\theta(\tilde{x}_\gamma^{(\beta)})}{t_\gamma} - \frac{f_{\theta_{\alpha_\beta}}(\tilde{x}_\gamma^{(\beta)} + t_\gamma d) - f_{\theta_{\alpha_\beta}}(\tilde{x}_\gamma^{(\beta)})}{t_\gamma} \\ &= \limsup_{\beta \in \mathcal{B}} \limsup_{\gamma \in \Gamma} \left\langle \theta - \theta_{\alpha_\beta}, \frac{F(\tilde{x}_\gamma^{(\beta)} + t_\gamma d) - F(\tilde{x}_\gamma^{(\beta)})}{t_\gamma} \right\rangle_{Y^* \times Y}. \end{aligned}$$

Now, use the fact that  $F$  is Lipschitz continuous around  $x$  to write

$$(5.20) \quad c \geq \limsup_{\beta \in \mathcal{B}} \limsup_{\gamma \in \Gamma} -\|\theta - \theta_{\alpha_\beta}\|_{\mathcal{Y}} \text{Lip}(F, x) \|d\|_{\mathcal{X}}.$$

Since  $\mathcal{Y}$  has finite dimension, the right member in (5.20) tends to zero, so  $c \geq 0$  and (5.18) is proved.

Now, use (5.17), (5.18), together with the upper semi-continuity of  $d^C f_\theta(\cdot, d)$  (see Section 2.2.4) to deduce

$$\begin{aligned} \limsup_{\alpha \in \mathcal{A}} d^C F(x_\alpha; d) &= \limsup_{\alpha \in \mathcal{A}} d^C f_{\theta_\alpha}(x_\alpha; d) = \lim_{\beta \in \mathcal{B}} d^C f_{\theta_{\alpha_\beta}}(x_{\alpha_\beta}; d) \\ &\leq \limsup_{\beta \in \mathcal{B}} d^C f_\theta(x_{\alpha_\beta}; d) \leq d^C f_\theta(x; d) \leq d^C F(x; d), \end{aligned}$$

which proves the upper semi-continuity of  $d^C F(\cdot; d)$ . As an immediate consequence, simply using its definition, we obtain the outer semi-continuity of  $\partial^C F$ .  $\blacksquare$

**Corollary 5.1.14** (Bounded strong-weak sequential outer semi-continuity). Suppose **H** holds, and let  $(x_\alpha, x_\alpha^*)_{\alpha \in \mathcal{A}}$  be a bounded net in  $\mathcal{X} \times \mathcal{X}^*$  such that

$$(5.21) \quad x_\alpha \xrightarrow{\|\cdot\|} x, \quad x_\alpha^* \xrightarrow{w^*} x^*, \quad \text{and } x_\alpha^* \in N_C(x_\alpha) + \partial^C F(x_\alpha).$$

Then  $x^* \in N_C(x) + \partial^C F(x)$ .

*Proof.* For all  $\alpha \in \mathcal{A}$ , we can write  $x_\alpha^* = \eta_\alpha^* + u_\alpha^*$ , where  $\eta_\alpha^* \in N_C(x_\alpha)$  and  $u_\alpha^* \in \partial^C F(x_\alpha)$ . Because of the strong convergence of  $x_\alpha$  to  $x$ , and using Proposition 5.1.7, we know that  $(u_\alpha^*)_{\alpha \in \mathcal{A}}$  is bounded. Hence, taking eventually a subnet, it weakly\* converges to some  $u^*$ , and we know from Proposition 5.1.13 that  $u^* \in \partial^C F(x)$ . Define now  $\eta^* := x^* - u^*$ , which is the weak\* limit of  $\eta_\alpha^*$ . From  $\eta_\alpha^* = x_\alpha^* - u_\alpha^*$ , we know that  $(\eta_\alpha^*)_{\alpha \in \mathcal{A}}$  is bounded. So, we deduce from the upper semi-continuity of  $N_C$  that  $\eta^* \in N_C(x)$  (see Proposition 2.2.4).  $\blacksquare$

## 5.2 Descent directions

### 5.2.1 Descent direction and Fermat's rule

Recall that  $C \subset \mathcal{X}$  is a nonempty closed convex set, which models the constraints.

**Definition 5.2.1** (Vector descent direction). We say that  $d \in \mathcal{X}$  is a *K-descent direction* for  $F$  at  $x \in C$  if  $d^C F(x, d) < 0$ . For convenience, we will just say that  $d$  is a descent direction. We say that it is an *admissible* descent direction if, moreover,  $d$  lies in the tangent cone  $T_C(x)$ .

Because of the positive homogeneity of  $d^C F(x, \cdot)$ , we see that the set of descent directions at  $x$  form a strict cone. It is important to note that the definition of decent direction is independent of the choice of  $\Theta$ :

**Proposition 5.2.2.** Let  $\Theta$  and  $\Theta'$  be two  $w^*$ -compact bases of  $K^+$ . Then for all  $x, d \in \mathcal{X}$ ,  $d^C_{\Theta'} F(x, d) < 0$  if and only if  $d^C_{\Theta} F(x, d) < 0$ .

*Proof.* Suppose by contradiction that there exists some  $d \in \mathcal{X}$  such that  $d^C_{\Theta} F(x, d) < 0$  and  $d^C_{\Theta'} F(x, d) \geq 0$ . Let  $\theta'_n$  be a sequence in  $\Theta'$  such that  $0 \leq d^C_{\Theta'} F(x, d) = \lim_{n \rightarrow +\infty} d^C f_{\theta'_n}(x, d)$ . Since  $\Theta' \subset K$  and  $\Theta$  is a base of  $K$ , there exists  $\lambda_n > 0$  and  $\theta_n \in \Theta$  for all  $n \in \mathbb{N}$  such that  $\theta'_n = \lambda_n \theta_n$ . In particular,  $f_{\theta'_n} = \lambda_n f_{\theta_n}$ .

Let us verify that the sequence  $(\lambda_n)_{n \in \mathbb{N}}$  is bounded from above. Indeed, suppose that  $\lambda_n$  tends to  $+\infty$  (taking eventually a subsequence). Since  $\lambda_n \theta_n$  lies in  $\Theta'$ , which is bounded, the

sequence  $(\lambda_n \|\theta_n\|)_{n \in \mathbb{N}}$  is bounded from above by some  $M > 0$ . Hence,  $\|\theta_n\| \leq \frac{M}{\lambda_n}$  tends to zero. But  $\Theta$  is  $w^*$ -closed, thus  $\|\cdot\|$ -closed, so  $0 \in \Theta$  would follow, which is impossible for a base.

Now we are assured that  $\lambda_n \leq \bar{\lambda}$  for all  $n \in \mathbb{N}$ , for  $\bar{\lambda} > 0$ . So

$$0 \leq d_{\Theta}^c F(x, d) = \lim_{n \rightarrow +\infty} d^c f_{\theta'_n}(x, d) = \lim_{n \rightarrow +\infty} \lambda_n d^c f_{\theta_n}(x, d) \leq \bar{\lambda} \lim_{n \rightarrow +\infty} d^c f_{\theta_n}(x, d) \leq \bar{\lambda} d_{\Theta}^c F(x, d) < 0,$$

which is a contradiction.  $\blacksquare$

**Example 5.2.3.** If  $F = (f_1, f_2) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  is smooth, the strict cone of descent directions at  $x \in \mathbb{R}^2$  consists in the vectors  $d$  satisfying

$$\langle \nabla f_1(x), d \rangle < 0 \text{ and } \langle \nabla f_2(x), d \rangle < 0.$$

In other words, it is the intersection of the interiors of the tangents cones to  $[f_1 \leq f_1(x)]$  and  $[f_2 \leq f_2(x)]$ . We will see in Theorem 5.2.12 that under some conditions, it is exactly the interior of the tangent cone to  $[F \leq F(x)] = [f_1 \leq f_1(x)] \cap [f_2 \leq f_2(x)]$ .

Figure 5.4: In dotted lines: the strict cone of descent directions for  $F = (f_1, f_2)$  at  $x$ .

We say that  $x \in C$  is a *critical point* of  $F$  if  $0 \in N_C(x) + \partial^c F(x)$ . According to the following result, this notion coincide with the one introduced by Smale [297]. In other words, critical points are exactly the points on which there is no admissible descent direction available.

**Theorem 5.2.4** (Pareto alternative). *Let  $x \in C$ . Then one and only one of the following statements is true :*

- $x$  is a critical point,
- there exists an admissible descent direction at  $x$ .

**Remark 5.2.5.** As a consequence of this Theorem and Proposition 5.2.2, the fact that  $x$  is critical does not depend on the choice of the base  $\Theta$ .

*Proof.* Suppose first that  $0 \in N_C(x) + \partial^c F(x)$ . Then there exists some  $x^* \in \partial^c F(x)$  such that  $-x^* \in N_C(x)$ . Since  $N_C(x)$  is defined as the polar cone of  $T_C(x)$ , for all  $d \in T_C(x)$  we have  $\langle -x^*, d \rangle \leq 0$ . From this we can deduce that

$$d^c F(x; d) = \sup_{x^* \in \partial^c F(x)} \langle x^*, d \rangle \geq 0 \text{ for all } d \in T_C(x).$$

As a consequence, there is no admissible descent direction at  $x$ .

Suppose now that  $0 \notin N_C(x) + \partial^c F(x)$ . Since  $\partial^c F(x)$  is convex and  $w^*$ -compact and  $N_C(x)$  is weakly\* closed convex, we have that  $N_C(x) + \partial^c F(x)$  is closed and convex in  $(X^*, w^*)$ . Using Hahn-Banach's separation theorem [325, Theorem 1.1.5], we obtain some  $d \in X$  and  $\alpha \in \mathbb{R}$  such that

$$(5.22) \quad \forall x^* \in N_C(x) + \partial^c F(x), \langle x^*, d \rangle \leq \alpha < 0.$$

Since  $0 \in N_C(x)$ , it follows directly that  $d^c F(x; d) = \sigma_{\partial^c F(x)}(d) \leq \alpha < 0$ . Moreover, using the weak\* compactness of  $\partial^c F(x)$  and the weak\* continuity of  $d$  as a linear functional over  $X^*$ , we can deduce that  $\sigma_{N_C(x) + \partial^c F(x)}(d) = \sigma_{\partial^c F(x)}(d) + \sigma_{N_C(x)}(d)$ . As a direct consequence of (5.22),  $\sigma_{N_C(x)}(d) \leq \alpha - \sigma_{\partial^c F(x)}(d)$ . Since  $\partial^c F(x)$  is compact,  $\sigma_{N_C(x)}(d) < +\infty$  follows. We can see, using  $(N_C(x))^* = T_C(x)$ , that this is equivalent to  $d \in T_C(x)$ .  $\blacksquare$



Figure 5.5: Illustration of the Pareto alternative in Theorem 5.2.4, for a bi-objective optimization problem. We take two points  $x_1$  and  $x_2$ , and we plotted the two corresponding strict cones of descent directions at  $x_1$  and  $x_2$  respectively. The thick dashed line represents the set of critical points. One can see that the closer  $x$  is to the set of Pareto points, the smaller is the strict cone of descent directions.

**Remark 5.2.6.** Note that in the smooth multi-objective case, this alternative theorem is exactly Gordan's alternative theorem [179],[77, Theorem 2.2.1].

We define now a stronger notion of descent direction, useful for designing algorithms. Recall that in all this Chapter,  $e$  denotes the element of  $\text{int } K$ , generating the base  $\Theta$  of  $K^+$ .

**Definition 5.2.7** (Armijo descent direction). Let  $d \in X$  be a descent direction at  $x \in C$ . We say that  $d$  is an *Armijo descent direction* at  $x$  if

$$(5.23) \quad \forall \beta \in ]0, 1[, \exists \varepsilon > 0 \text{ such that } \forall t \in ]0, \varepsilon[, F(x + td) \leq F(x) + \beta t d^c F(x, d)e.$$

We say that  $d$  is an admissible Armijo descent direction, if moreover  $x + td \in C$  for all  $t \in ]0, \varepsilon[$ .

Observe that admissible Armijo directions are exactly Armijo directions which lie in  $T_C^{ad}(x)$ , the admissible tangent cone to  $C$  at  $x$  (see Section 2.2.1). Note also that  $e$  being in  $\text{int } K$ , an Armijo descent direction  $d$  satisfies in particular

$$F(x + td) < F(x) \text{ for all } t \in ]0, \varepsilon[.$$

In a way, Armijo descent directions for  $F$  can be seen as vectors satisfying a kind of *uniform* Armijo condition for the family of cost functions  $\{f_\theta\}_{\theta \in \Theta}$ :

**Proposition 5.2.8.** Let  $x \in C$  and  $d \in X$  be a descent direction for  $F$  at  $x$ . Consider the following statements :

- i)  $\forall \beta \in ]0, 1[, \exists \varepsilon > 0$  s.t.  $\forall \theta \in \Theta, \forall t \in ]0, \varepsilon[, f_\theta(x + td) \leq f_\theta(x) + \beta t d^c f_\theta(x, d)$ .
- ii)  $\forall \beta \in ]0, 1[, \exists \varepsilon > 0$  s.t.  $\forall \theta \in \Theta, \forall t \in ]0, \varepsilon[, f_\theta(x + td) \leq f_\theta(x) + \beta t d^c F(x, d)$ .
- iii)  $d$  is an Armijo descent direction for  $F$  at  $x$ .
- iv)  $\exists \bar{\beta} \in ]0, 1[, \forall \beta \in ]0, \bar{\beta}[, \exists \varepsilon > 0$  s.t.  $\forall \theta \in \Theta, \forall t \in ]0, \varepsilon[, f_\theta(x + td) \leq f_\theta(x) + \beta t d^c f_\theta(x, d)$ .

Then, i)  $\Rightarrow$  ii)  $\Leftrightarrow$  iii)  $\Rightarrow$  iv).

*Proof.* From the definition of  $d^c F(x, d) = \sup_{\theta \in \Theta} d^c f_\theta(x, d)$ , we see that i) implies ii). In order to prove the equivalence between ii) and iii), use Proposition 2.3.7. Indeed, let  $\beta \in ]0, 1[$  and  $t > 0$ , then

$$\begin{aligned} F(x + td) &\leq F(x) + \beta t d^c F(x, d)e \\ \Leftrightarrow \forall \theta \in \Theta, f_\theta(x + td) &\leq f_\theta(x) + t \beta d^c F(x, d) \quad (\text{because } \theta(e) = 1). \end{aligned}$$

Assume now item ii), and define

$$\alpha := \frac{\inf_{\theta \in \Theta} d^c f_\theta(x, d)}{d^c F(x, d)}$$

which is bounded by Lipschitz property of  $F$  and boundedness of  $\Theta$ . By definition,  $\alpha \geq 1$ . Let then  $\beta \in ]0, \frac{1}{\alpha}[$ , and show that item iv) holds for this  $\beta$ . Since  $\alpha\beta \in ]0, 1[$ , we can apply item ii) to have some  $\varepsilon > 0$  such that for all  $t \in ]0, \varepsilon[$  and  $\theta \in \Theta$ ,

$$f_\theta(x + td) \leq f_\theta(x) + \alpha \beta t d^c F(x, d).$$

But, for all  $\theta \in \Theta$ , we have  $\alpha d^c F(x, d) \leq d^c f_\theta(x, d)$ , so the desired result follows with  $\bar{\beta} = \frac{1}{\alpha}$ . ■

We show now that, under assumptions, the Armijo descent directions coincide with the descent direction, and that they characterize the critical points of  $F$  (recall Theorem 5.2.4).

**Proposition 5.2.9.** Let  $x \in C$ , and suppose that either  $\dim Y < +\infty$ , or  $F$  is the sum of a convex function and a directionally derivable function.

- i) If  $d \in X$  is a descent direction at  $x$ , then it is an Armijo descent direction at  $x$ .
- ii) There exists an admissible descent direction at  $x$  if and only if there exists an admissible Armijo descent direction at  $x$ .

*Proof.* i) Let  $d \in X$  be a descent direction at  $x$ , that is,  $d^c F(x, d) < 0$ . We argue by contradiction, using Proposition 5.2.8: suppose that there exists  $\beta \in ]0, 1[$  such that  $\forall \varepsilon > 0$ , there exists  $t \in ]0, \varepsilon[$  and  $\theta \in \Theta$  satisfying

$$(5.24) \quad f_\theta(x + td) > f_\theta(x) + \beta t d^c F(x, d).$$

In other words, there exists a net  $(t_\alpha)_{\alpha \in \mathcal{A}}$  in  $\mathbb{R}_{++}$  converging to zero, and a net  $(\theta_\alpha)_{\alpha \in \mathcal{A}}$  in  $\Theta$  weakly\* converging to some  $\theta \in \Theta$  (use the compactness of  $\Theta$ ), such that for all  $\alpha \in \mathcal{A}$

$$(5.25) \quad \beta d^c F(x, d) < \frac{f_{\theta_\alpha}(x + t_\alpha d) - f_{\theta_\alpha}(x)}{t_\alpha}.$$

Suppose in a first time that  $\dim Y < +\infty$ , and write

$$(5.26) \quad \frac{f_{\theta_\alpha}(x + t_\alpha d) - f_{\theta_\alpha}(x)}{t_\alpha} = \frac{f_\theta(x + t_\alpha d) - f_\theta(x)}{t_\alpha} + \left\langle \theta_\alpha - \theta, \frac{F(x + t_\alpha d) - F(x)}{t_\alpha} \right\rangle.$$

Using the Lipschitz property of  $F$ , we see that the second term of the right member of (5.26) is bounded from above by  $\|\theta_\alpha - \theta\| \text{Lip}(F, x) \|d\|$  when  $t_\alpha$  is close enough to zero. So, using the norm-convergence of  $\theta_\alpha$  to  $\theta$ , we deduce that

$$\limsup_{\alpha \in \mathcal{A}} \frac{f_{\theta_\alpha}(x + t_\alpha d) - f_{\theta_\alpha}(x)}{t_\alpha} \leq \limsup_{\alpha \in \mathcal{A}} \frac{f_\theta(x + t_\alpha d) - f_\theta(x)}{t_\alpha} \leq d^c f_\theta(x, d) \leq d^c F(x, d).$$

It follows that  $\beta d^c F(x, d) \leq d^c F(x, d)$  which is in contradiction with the facts that  $d^c F(x, d) \leq 0$  and  $\beta \in ]0, 1[$ .

Suppose now that  $F = G + H$ , with  $G$  being convex and  $H$  being directionally derivable at  $x$ . Then, (5.25) rewrites as (we note  $g_\theta := \theta \circ G$  as usual)

$$(5.27) \quad \beta d^c F(x, d) < \frac{g_{\theta_\alpha}(x + t_\alpha d) - g_{\theta_\alpha}(x)}{t_\alpha} + \left\langle \theta_\alpha, \frac{H(x + t_\alpha d) - H(x)}{t_\alpha} \right\rangle, \quad \text{for all } \alpha \in \mathcal{A}.$$

Consider any fixed  $t > 0$ . Using the convergence of  $t_\alpha$  to zero, we obtain some  $\bar{\alpha} \in \mathcal{A}$  such that for all  $\alpha \succeq \bar{\alpha}$ ,  $t > t_\alpha$ . Write

$$(5.28) \quad x + t_\alpha d = \frac{t_\alpha}{t}(x + td) + \left(1 - \frac{t_\alpha}{t}\right)x, \quad \text{with } \frac{t_\alpha}{t} \in ]0, 1[ \text{ for all } \alpha \succeq \bar{\alpha}.$$

The convexity of  $G$  in (5.27) together with (5.28) leads to:

$$(5.29) \quad \beta d^c F(x, d) < \frac{g_{\theta_\alpha}(x + td) - g_{\theta_\alpha}(x)}{t} + \left\langle \theta_\alpha, \frac{H(x + t_\alpha d) - H(x)}{t_\alpha} \right\rangle, \quad \text{for all } \alpha \succeq \bar{\alpha}.$$

Since (5.29) holds for all  $\alpha \succeq \bar{\alpha}$ , we can take the liminf over  $\alpha \in \mathcal{A}$  :

$$(5.30) \quad \beta d^c F(x, d) \leq \liminf_{\alpha \in \mathcal{A}} \frac{g_{\theta_\alpha}(x + td) - g_{\theta_\alpha}(x)}{t} + \left\langle \theta_\alpha, \frac{H(x + t_\alpha d) - H(x)}{t_\alpha} \right\rangle.$$

Using the bounded weak\* convergence of  $(\theta_\alpha)_{\alpha \in \mathcal{A}}$  to  $\theta_\alpha$ , together with the directional derivability of  $H$  at  $x$ , we deduce from (5.30) that

$$(5.31) \quad \beta d^c F(x, d) \leq \frac{g_\theta(x + td) - g_\theta(x)}{t} + \langle \theta, DH(x, d) \rangle.$$

This being true for any fixed  $t > 0$ , we obtain by taking the limit when  $t \downarrow 0$ , together with (2.24):

$$(5.32) \quad \beta d^c F(x, d) \leq d^c(\theta \circ G)(x, d) + \langle \theta, DH(x, d) \rangle \leq d^c F(x, d).$$

As before, we see here a contradiction, and item i) is proved.

ii) From Definition 5.2.7 and  $T_C^{ad}(x) \subset T_C(x)$ , it is obvious that the admissible Armijo descent directions are admissible descent directions. We prove here the reverse statement. Let  $d \in X$  be an admissible descent direction at  $x$ , that is  $d \in T_C(x)$  and  $d^c F(x, d)$ . Since  $C$  is convex, we know from Proposition 2.2.2 that  $T_C(x) = \text{cl} T_C^{ad}(x)$ . Hence, there exists a sequence  $(d_n)_{n \in \mathbb{N}}$  in  $T_C^{ad}(x)$  converging in norm to  $d$ . Moreover,  $d^c F(x, \cdot)$  is continuous, so there must exist some  $N \in \mathbb{N}$  such that  $d^c F(x, d_N) < 0$ , which proves the claim. ■

As a consequence, we derive a natural Fermat's rule for vector optimization problems, giving a necessary condition for weak efficiency. In the convex setting, it becomes a necessary and sufficient condition, and we recover a known description of weak efficient points [139, Theorem 2.10].

**Theorem 5.2.10** (Fermat's rule). *Suppose that either  $\dim Y < +\infty$ , or  $F$  is the sum of a convex function and a directionally derivable function. Then any weak efficient point  $x$  is critical, i.e.  $0 \in N_C(x) + \partial^c F(x)$ .*

*If  $F$  is convex, then critical points coincide with weak efficient points, and we have*

$$(5.33) \quad \text{ARGMIN}_{x \in C} F = \bigcup_{\theta \in \Theta} \underset{x \in C}{\text{argmin}} f_\theta.$$

*If  $F$  is strictly convex, then critical points are efficient points.*

**Remark 5.2.11.** According to (5.33), a good approach for generating the set of Pareto points of a convex function  $F$ , is to take a sample among the cost functions  $\{f_\theta\}_{\theta \in \Theta}$ , and minimize them. In the multi-objective setting  $F = (f_1, \dots, f_m)$ , this is exactly the weighting method, which consists in the minimization of arbitrary (or chosen) convex combinations of the  $f_i$ 's. We discuss some drawbacks of this method later in Section 6.3.

*Proof.* Let  $x \in C$  be a weak efficient point. We argue by contradiction and suppose that  $x$  is not critical. By Theorem 5.2.4, there would exist an admissible descent direction at  $x$ . Because of the hypotheses, we can apply Proposition 5.2.9, so there exists  $d \in X$  an admissible Armijo direction. In particular, there exists  $\varepsilon > 0$  such that for all  $t \in ]0, \varepsilon[$ ,

$$(5.34) \quad x + td \in C \quad \text{and} \quad F(x + td) < F(x).$$

Take  $y := x + td$  for some  $t \in ]0, \varepsilon[$  and sees that it contradicts the weak efficiency of  $x$ .

Suppose now that  $F$  is convex, and show that critical points are weakly efficient. We know from Proposition 5.1.4 that in the convex case,

$$(5.35) \quad \partial^c F(x) = \bigcup_{\theta \in \Theta} \partial f_\theta(x).$$

If  $x$  is critical, there must exist  $\theta \in \Theta$  such that  $0 \in N_C(x) + \partial f_\theta(x)$ . Using the classic Fermat's rule for convex scalar functions, we see that it is equivalent for  $x$  to minimize  $f_\theta$  over  $C$ . If

we suppose now that  $x$  is not weakly efficient, it would exist  $x' \in C$  such that  $F(x') < F(x)$ . In particular (see Proposition 2.3.7) we would have  $f_\theta(x') < f_\theta(x)$  which would contradict  $x \in \operatorname{argmin}_{y \in C} f_\theta$ . As a consequence, critical points are weak minimizers.

Note that the equality (5.33) is proved by the same argument :  $x$  is weakly efficient if and only if it is a critical point, this being equivalent to say (see above) that there exists  $\theta \in \Theta$  such that  $x \in \operatorname{argmin}_{y \in C} f_\theta$ .

For the strictly convex case, argue exactly as for the convex case, just using the fact that a scalar-valued strictly convex function admits at most one minimizer. ■

### 5.2.2 Normals to level sets

There is an other geometric consequence of Proposition 5.2.9 : under appropriate assumptions, the subdifferential generates the Bouligand normal cone to the sublevel sets.

**Theorem 5.2.12** (Normals to sublevel sets). *Let  $x \in X$  be such that  $0 \notin \partial^c F(x)$ . Suppose that either  $F$  is weakly Clarke regular with  $\dim Y < +\infty$ , or  $F$  is the sum of a convex function and a strictly Gateaux differentiable function at  $x$ . Then*

- i)  $T_{[F \leq F(x)]}^B(x) = \{d \in X \mid d^c F(x, d) \leq 0\}$ ,
- ii)  $\operatorname{int} T_{[F \leq F(x)]}^B(x) = \{d \in X \mid d^c F(x, d) < 0\}$  (i.e. the descent directions at  $x$ ),
- iii)  $N_{[F \leq F(x)]}^B(x) = \mathbb{R}_+ \partial^c F(x)$ .

**Example 5.2.13.** Let  $F = (f_1, f_2) : \mathbb{R}^2 \rightarrow (\mathbb{R}^2, \mathbb{R}_+^2)$  be a smooth convex function and  $x \in \mathbb{R}^2$ . We plot in Figure 5.6 the sublevel set  $[F \leq F(x)]$  in thick continuous curves, and dotted lines represent respectively the normal and tangent cone to  $[F \leq F(x)]$  at  $x$ .

Figure 5.6: First-order analysis at the intersection of two sublevel sets.

*Proof of Theorem 5.2.12.* We start by showing the inclusion

$$(5.36) \quad T_{[F \leq F(x)]}^B(x) \subset \{d \in X \mid d^c F(x, d) \leq 0\}.$$

So let us consider  $d$  such that there exists  $t_n \downarrow 0$  and  $d_n \rightarrow d$  such that  $x + t_n d_n \in [F \leq F(x)]$  for all  $n \in \mathbb{N}$ . That is,  $f_\theta(x + t_n d_n) \leq f_\theta(x)$  for all  $\theta \in \Theta$ . After division by  $t_n > 0$  and taking the liminf on  $n \in \mathbb{N}$ , we obtain that  $d^B f_\theta(x; d) \leq 0$  for all  $\theta \in \Theta$ . In other words, because of the positive Clarke regularity of  $F$ ,  $d^c F(x, d) \leq 0$ .

We show now that

$$(5.37) \quad \{d \in X \mid d^c F(x, d) < 0\} \subset T_{[F \leq F(x)]}^B(x).$$

Consider  $d \in X$  such that  $d^c F(x, d) < 0$ . It is a descent direction at  $x$ , so it is an Armijo descent direction, according to Proposition 5.2.9 and the hypotheses made on  $F$ . In particular, there exists some  $\varepsilon > 0$  such that  $x + td \in [F \leq F(x)]$  for all  $t \in ]0, \varepsilon[$ . By definition of the Bouligand tangent cone, it follows that  $d \in T_{[F \leq F(x)]}^B(x)$  and (5.37) is proved.

In order to obtain item i), we will take the closure in (5.37). The left member is the strict sublevel set of the convex continuous function  $d^c F(x, \cdot) = d^B F(x, \cdot)$ . It is nonempty since it is assumed that  $0 \notin \partial^c F(x)$ , so we can use Theorem 5.2.4. Hence, we can apply Proposition A.1.4 which says that:

$$(5.38) \quad \operatorname{cl} \{d \in X \mid d^c F(x, d) < 0\} = \{d \in X \mid d^B F(x, d) \leq 0\},$$

$$(5.39) \quad \{d \in X \mid d^c F(x, d) < 0\} = \operatorname{int} \{d \in X \mid d^B F(x, d) \leq 0\}.$$

From (5.37), (5.38) and the closure of  $T_{[F \leq F(x)]}^B(x)$ , we deduce item i). Item ii) follows item i) and (5.39).

To prove item iii), start from item i) and write

$$\begin{aligned}
 (5.40) \quad T_{[F \leq F(x)]}^B(x) &= \{d \in X \mid d^c F(x, d) \leq 0\} \\
 &= \{d \in X \mid \langle x^*, d \rangle \leq 0 \ \forall x^* \in \partial^c F(x)\} \\
 &= \{d \in X \mid \langle x^*, d \rangle \leq 0 \ \forall x^* \in \mathbb{R}_+ \partial^c F(x)\} \\
 &= (\mathbb{R}_+ \partial^c F(x))^*.
 \end{aligned}$$

Since  $\partial^c F(x)$  is  $w^*$ -compact and does not contain the origin, Proposition 2.1.7 ensures that  $\mathbb{R}_+ \partial^c F(x)$  is a closed convex cone. By taking the polar on (5.40) we obtain the desired inequality.  $\blacksquare$

**Remark 5.2.14.** It must be underlined that, while we are minimizing a function  $F$  from  $X$  to  $Y$ , all the objects involved here (the ordered subdifferential, the sublevel sets) live in the decision space  $X$  or its dual. Of course, the order on  $Y$  determines these objects, but it is somehow implicit. This will be determinant in Chapter 7, where we construct an auxiliary function  $\mathcal{F} : X \rightarrow Z$ , for some well chosen ordered Banach space  $Z$ , which has the same sublevel set than  $F$ , and so, the same ordered subdifferential (see Theorem 7.3.3).

**Remark 5.2.15.** In a recent work [91], Cabot and Thibault give a general description of  $N_{[F \leq F(x)]}(x)$  in the multi-objective convex case, *without* the condition  $0 \notin \partial^c F(x)$ . When this condition is satisfied, they recover  $N_{[F \leq F(x)]}(x) = \mathbb{R}_+ \partial^c F(x)$ . It would be of interest to overcome the hypothesis  $0 \notin \partial^c F(x)$  in Theorem 5.2.12 by adapting the ideas in [91, Theorem 5.1] to the general vector case.

**Remark 5.2.16.** It is a known fact in scalar optimization (see [111, Theorem 2.4.7] or [263, Proposition 5.48]) that the sublevel sets of Clarke regular functions are themselves Clarke regular, which means that their Clarke and Bouligand tangent cones coincide. This result extends to our vectorial case when  $K$  is polyhedral: it is easy to exploit the finite number of objective functions to adapt the argument of [263, Proposition 5.48]. It is not clear whether it is true or not for nonpolyhedral cones. An other approach consists in writing  $[F \leq F(x)]$  as the infinite intersection of sublevel sets  $[f_\theta \leq f_\theta(x)]$ . Since the inclusions

$$T_{[F \leq F(x)]}^C(x) \subset T_{[F \leq F(x)]}^B(x) \subset \bigcap_{\theta \in \Theta} T_{[f_\theta \leq f_\theta(x)]}^B(x)$$

always hold, it is enough to prove, after using the Clarke regularity of  $\Gamma_\theta$ , that

$$\bigcap_{\theta \in \Theta} T_{[f_\theta \leq f_\theta(x)]}^C(x) \subset T_{[F \leq F(x)]}^C(x).$$

This kind of inclusion is known to be valid when a finite number of sets is involved, under appropriate conditions, see [288, Theorem 5]. But is is a nontrivial question to know whether it can be extended to an infinite number of sets.

### 5.2.3 The steepest descent direction in Hilbert spaces

Suppose now that  $X = H$  is a Hilbert space, endowed with its scalar product  $\langle \cdot, \cdot \rangle$ , and identify  $H$  with its dual  $H^*$ . We still suppose that  $Y$  is an arbitrary Banach space. We aim to introduce an analog of the usual steepest descent direction used in scalar optimization. Recall that we note  $C^0$  the element of minimal norm of a closed convex set<sup>3</sup>.

<sup>3</sup>In other words,  $C^0$  is the projection of zero on  $C$ .

**Definition 5.2.17.** For any  $x \in C$ , the unique element of minimal norm of the closed convex set  $-N_C(x) - \partial^c F(x)$  is called the *steepest descent direction* of  $F$  at  $x$ . It is denoted by

$$(5.41) \quad s(x) := \left( -N_C(x) - \partial^c F(x) \right)^0.$$

Note that, for any  $x \in C$ , the set  $-N_C(x) - \partial^c F(x)$  is a closed convex set, as being equal to the vectorial sum of two closed convex sets, one of them being weakly compact. Hence, it has a unique element of minimal norm, and  $s(x)$  is well defined.

This vector field clearly satisfies that  $x$  is a critical point if and only if  $s(x) = 0$ . Furthermore,  $s(x)$  is an admissible descent direction for  $F$  at any non critical  $x$  :

**Proposition 5.2.18** (Descent property). For all  $x \in C$  we have  $s(x) \in T_C(x)$  and

$$(5.42) \quad d^c F(x; s(x)) \leq -\|s(x)\|^2.$$

In particular,  $s(x)$  is an admissible descent direction for  $F$  at any  $x$ , as soon as  $x$  is not critical.

*Proof.* By definition,  $-s(x)$  is the projection of the origin onto the closed convex set  $N_C(x) + \partial^c F(x)$ . Hence, using the variational definition of the projection in a Hilbert space and  $0 \in N_C(x)$ , we obtain for any  $x^* \in \partial^c F(x)$  :

$$(5.43) \quad \langle 0 - (-s(x)), x^* - (-s(x)) \rangle \leq 0.$$

In other words,

$$(5.44) \quad \|s(x)\|^2 + \langle s(x), x^* \rangle \leq 0 \quad \text{for all } x^* \in \partial^c F(x),$$

and the desired inequality follows, by taking the sup over  $x^* \in \partial^c F(x)$ .

Suppose now that  $x$  is not critical. As noticed before, it implies that  $s(x) \neq 0$ , so it makes  $s(x)$  a descent direction. Verify now that  $s(x)$  is an *admissible* descent direction. By definition of  $s(x)$ , we can write  $s(x) = \left( -x^* - N_C(x) \right)^0$  for some  $x^* \in \partial^c F(x)$ . Since  $T_C(x)$  is the polar cone of  $N_C(x)$ , we can use Moreau's decomposition Theorem 2.1.10 and obtain,

$$\begin{aligned} (z - N_C(x))^0 &= z - \text{proj}_{N_C(x)} z \\ &= \text{proj}_{T_C(x)} z \end{aligned}$$

which shows that  $s(x) \in T_C(x)$ , and concludes the proof. ■

The result above has a simple geometrical interpretation. Take simply the multi-objective unconstrained problem, *i.e.*,  $C = \mathcal{H}$ ,  $(Y, K) = (\mathbb{R}^2, \mathbb{R}_+^2)$  with  $F = (f_1, f_2)$  being smooth. Then  $-s(x)$  is the orthogonal projection of the origin on the segment  $[\nabla f_1(x), \nabla f_2(x)]$ . By the classical result on the sum of the angles of a triangle, this forces the angles between  $-s(x)$  and  $\nabla f_i(x)$ ,  $i = 1, 2$ , to be acute and so  $\langle s(x), \nabla f_i(x) \rangle \leq 0$ .

Figure 5.7: The steepest descent direction associated to a smooth biobjective optimization problem.

**Remark 5.2.19.** In the multi-objective case  $F : \mathbb{H} \rightarrow (\mathbb{R}^m, \mathbb{R}_+^m)$ ,  $F(x) = (f_1(x), \dots, f_m(x))$ , computing the steepest descent direction at  $x$  requires to project the origin on the convex hull of the subdifferentials  $\partial^c f_i(x)$ :

$$-s(x) = \text{proj}_{\text{co}\partial^c f_i(x)}(0).$$

Suppose that we are in a non regular point  $x$ , such that the subdifferentials  $\partial^C f_i(x)$  are not reduced to one point. A naive approach to compute  $s(x)$  could be the following: take  $p_i(x) := \partial^C f_i(x)^0$  for each  $i \in \{1, \dots, m\}$ , and then compute  $-s(x)$  as the element of minimal norm of  $\text{co} \{p_i\}$ . But this approach doesn't work, because the operations of taking the convex hull and doing the projection do not commute. Example 5.1.10 provides a simple counter-example to this situation. In this example, such a direction would not be a descent direction because  $x$  is critical (recall the alternative Theorem 5.2.4).

Now that we have established that  $s(x)$  is a descent direction, one may wonder why it is called the *steepest* descent direction. Observe first that in the case of a smooth real-valued function  $f : H \rightarrow \mathbb{R}$ , the direction  $s(x)$  at  $x \in C$  is given by (use Moreau's theorem 2.1.10)

$$\begin{aligned} s(x) &= (-N_C(x) - \nabla f(x))^0, \\ &= \text{proj}_{T_C(x)}(-\nabla f(x)). \end{aligned}$$

It is known in that setting that the normalized vector  $\frac{s(x)}{\|s(x)\|}$  is the solution, if  $\nabla f(x) \neq 0$ , of the minimization problem

$$\min \{ \langle \nabla f(x), d \rangle \mid d \in T_C(x), \|d\| = 1 \},$$

whence the name of *steepest descent direction* for  $\text{proj}_{T_C(x)}(-\nabla f(x))$ . As shown below, this steepest descent property can be extended to vector-valued functions. Moreover, still in the case of a smooth real-valued function, it can be easily verified that

$$(5.45) \quad s(x) = \underset{d \in T_C(x)}{\text{argmin}} \frac{1}{2} \|d\|^2 + \langle \nabla f(x), d \rangle,$$

and this further characterization will also be generalized to the multiobjective case.

**Theorem 5.2.20** (Steepest descent property). *Let  $x \in C$  be such that  $s(x) \neq 0$ . Then  $s(x)$  can be formulated in the following equivalent forms:*

1.  $s(x) = \left( -N_C(x) - \partial^C F(x) \right)^0,$
2.  $\frac{s(x)}{\|s(x)\|^{\frac{p-2}{p-1}}} = \underset{d \in T_C(x)}{\text{argmin}} \frac{1}{p} \|d\|^p + d^C F(x, d), \quad \forall p \in [2, +\infty[,$
3.  $\frac{s(x)}{\|s(x)\|} = \underset{\substack{d \in T_C(x) \\ \|d\|=1}}{\text{argmin}} d^C F(x, d).$

**Remark 5.2.21.** The equivalence between formulations 1. and 3. of the steepest descent direction has been first obtained, in the multicriteria and smooth case, by Cornet in [116, Proposition 3.1]. In the same context, formulation 2. seems to be introduced for the first time by Fliege and Svaiter [159], for  $p = 2$

$$s(x) = \underset{d \in T_C(x)}{\text{argmin}} \left\{ \frac{1}{2} \|d\|^2 + d^C F(x, d) \right\},$$

but the equivalence between formulations 1. and 2. is seemingly new. The second formulation for  $p \neq 2$  is new, although it was stressed in [159] that  $\frac{1}{2} \|d\|^2$  could be replaced by any positive proper l.s.c strictly convex function which is dominated by the norm around the origin. The interest of considering  $p$  arbitrary large is that we can see -at least formally- the third formulation as the limit of the second when  $p \rightarrow +\infty$ :  $\frac{p-2}{p-1}$  tends to 1, while the function  $\frac{1}{p} \|\cdot\|^p$  is pointwise converging to the indicator function of the unit ball  $\delta_{\mathbb{B}}(\cdot)$ .

*Proof of Proposition 5.2.20.* Let us start by proving the formulation 2. Because of the powered norm term, and the fact that  $H$  is a Hilbert space,  $d \mapsto \frac{1}{p}\|d\|^p + d^{\mathcal{C}F}(x; d)$  is a coercive strictly convex function. Therefore, there exists a unique solution  $\bar{d}$  to the minimization problem (see for instance [21, Proposition 2.3.3])

$$(5.46) \quad \min_{d \in T_C(x)} \frac{1}{p} \|d\|^p + d^{\mathcal{C}F}(x; d).$$

Let us show that  $\bar{d} = \frac{s}{\|s\|^{\frac{p-2}{p-1}}}$ , where  $s = s(x)$ . We use a duality argument which relies on the equivalent formulation of (5.46) as the convex-concave saddle value problem

$$(5.47) \quad \min_{d \in T_C(x)} \max_{a \in \partial^{\mathcal{C}F}(x)} \frac{1}{p} \|d\|^p + \langle a, d \rangle.$$

It is associated with the convex-concave Lagrangian function

$$L(d, a) = \frac{1}{p} \|d\|^p + \langle a, d \rangle$$

defined on  $T_C(x) \times \partial^{\mathcal{C}F}(x)$ . Since  $L$  is convex and coercive with respect to the first variable, and  $\partial^{\mathcal{C}F}(x)$  is bounded, by the von Neumann's minimax theorem (see [21, Theorem 9.7.1]) there exists  $\bar{a} \in \partial^{\mathcal{C}F}(x)$  such that  $(\bar{d}, \bar{a})$  is a saddle point of (5.47), that is

$$(5.48) \quad \inf_{d \in T_C(x)} L(d, \bar{a}) = L(\bar{d}, \bar{a}) = \sup_{a \in \partial^{\mathcal{C}F}(x)} L(\bar{d}, a).$$

For any  $a \in \partial^{\mathcal{C}F}(x)$  let us define

$$(5.49) \quad d(a) := \operatorname{argmin}_{d \in T_C(x)} \frac{1}{p} \|d\|^p + \langle a, d \rangle.$$

Of course, by definition we have  $\bar{d} = d(\bar{a})$ . Using Fermat's rule for the above primal problem (5.49) together with a sum rule (recall 2.2.6) gives

$$(5.50) \quad 0 \in N_{T_C(x)}(d(a)) + d(a)\|d(a)\|^{p-2} + a,$$

where we used that the derivative of  $\frac{1}{p}\|\cdot\|^p$  at any  $x$  for  $p \geq 2$  exists and is  $x\|x\|^{p-2}$ . Using the variational characterization of the projection, it follows

$$(5.51) \quad d(a) = \operatorname{proj}_{T_C(x)} \left( \frac{-a}{\|d(a)\|^{p-2}} \right),$$

which can be rewritten, by Moreau's theorem, as

$$(5.52) \quad d(a) = \frac{1}{\|d(a)\|^{p-2}} (-a - N_C(x))^0.$$

Since  $\bar{d} = d(\bar{p})$ , we just need to prove that  $d(\bar{a}) = \frac{s}{\|s\|^{\frac{p-2}{p-1}}}$ . To identify  $\bar{a}$ , we use the dual formulation

$$(5.53) \quad \bar{a} = \operatorname{argmax}_{a \in \partial^{\mathcal{C}F}(x)} \min_{d \in T_C(x)} \left\{ \frac{1}{p} \|d\|^p + \langle a, d \rangle \right\},$$

which, by (5.49) and (5.51), can be rewritten as

$$(5.54) \quad \bar{a} = \operatorname{argmax}_{a \in \partial^{\mathcal{C}F}(x)} \frac{1}{p} \|d(a)\|^p - \|d(a)\|^{p-2} \left\langle \frac{-a}{\|d(a)\|^{p-2}}, \operatorname{proj}_{T_C(x)} \left( \frac{-a}{\|d(a)\|^{p-2}} \right) \right\rangle.$$



Using Moreau's theorem, we obtain

$$(5.55) \quad \bar{a} = \operatorname{argmax}_{a \in \partial^c F(x)} \frac{1}{p} \|d(a)\|^p - \|d(a)\|^{p-2} \left\| \operatorname{proj}_{T_C(x)} \left( \frac{-a}{\|d(a)\|^{p-2}} \right) \right\|^2,$$

which, by (5.51) and  $p \in [2, +\infty]$ , is equivalent to

$$(5.56) \quad \bar{a} = \operatorname{argmin}_{a \in \partial^c F(x)} \|d(a)\|^{p-1}.$$

From (5.52), we know that  $\|d(a)\|^{p-1} = \|(-a - N_C(x))^0\|$ . Therefore,  $s = (-\bar{a} - N_C(x))^0$  with  $\|d(\bar{a})\|^{p-1} = \|s\|$ . Using again (5.52), we obtain  $d(\bar{a}) = \frac{s}{\|s\|^{\frac{p-2}{p-1}}}$ , as expected.

Let us complete the proof by proving the third characterisation. As we said before, it relies on a limit argument. Define, for any  $p \geq 2$ , the functions  $F_p : d \in H \mapsto \frac{1}{p} \|d\|^p + d^c F(x; d) + \delta_{T_C(x)}(d)$ . It can be easily verified that the sequence  $(F_p)_{p \geq 2}$  epiconverges when  $p \rightarrow +\infty$  to

$$F : d \in H \mapsto \delta_{\{\|\cdot\| \leq 1\}}(d) + d^c F(x; d) + \delta_{T_C(x)}(d).$$

From item 1 and [21, Theorem 12.1.1] we can deduce that

$$(5.57) \quad \frac{s}{\|s\|} = \operatorname{argmin}_{\substack{v \in T_C(x) \\ \|v\| \leq 1}} d^c F(x; v),$$

where the inequality constraint  $\|v\| \leq 1$  can be replaced by  $\|v\| = 1$ , since  $\frac{s}{\|s\|}$  is a normalized vector. ■

**Example 5.2.22.** We work in the same setting that Example 2.3.8, and we assume to simplify that  $F : X \rightarrow Y$  is strictly Gateaux differentiable and  $C = H$ .

- If  $(Y, K) = (\mathbb{R}^m, \mathbb{R}_+^m)$ ,  $s(x)$  is the element of minimal norm in the convex hull of the gradients  $\nabla f_i(x)$ , and is characterized by  $\frac{s(x)}{\|s(x)\|} = \operatorname{argmin}_{\|d\|=1} \max_{i \in \{1, \dots, m\}} \langle \nabla f_i(x), d \rangle$ .
- If  $(Y, K) = (S^m(\mathbb{R}), S_+^m(\mathbb{R}))$ , then  $\frac{s(x)}{\|s(x)\|} = \operatorname{argmin}_{\|d\|=1} \lambda_{\max}(DF(x; d))$ , where  $\lambda_{\max}(M)$  is the greatest eigenvalue of  $M$ .
- If  $(Y, K) = (L^\infty(\Omega, \Sigma, \mu), L_+^\infty(\Omega, \Sigma, \mu))$ , then  $\frac{s(x)}{\|s(x)\|} = \operatorname{argmin}_{\|d\|=1} \operatorname{supess}(DF(x; d))$ .

### 5.3 Comments

We end here with a couple of remarks about this chapter.

**Remark 5.3.1** (The choice of the Clarke subdifferential). We chose to build the ordered subdifferential on the Clarke subdifferential for several reasons. The first one, and the most obvious, is that it enjoys very good duality properties with its Clarke directional derivatives, like the max formula (5.3). This allows us to characterize easily the ordered subdifferential  $\partial^c F(x)$  with its support function, namely, the ordered directional derivative  $d^c F(x; \cdot)$ .

An other reason, directly related to the first one, is that the Clarke directional derivative provides a good Armijo rule at descent directions (see Proposition 5.2.9). If we use for instance the Bouligand directional derivative for a function  $f : X \rightarrow \mathbb{R}$ , we would have a poorer Armijo rule (see for instance [223, Section 4]):

$$d^B f(x; d) < 0 \Rightarrow \forall \beta \in ]0, 1[, \forall T > 0, \text{ there exists } t_0 \in ]0, T[ \text{ and } \varepsilon > 0 \text{ such that}$$

$$f(x + td) < f(x) + \beta t d^B f(x; d), \quad \forall t \in [t_0 - \varepsilon, t_0 + \varepsilon].$$

While the Armijo rule has a clear importance for the study of descent algorithms with line-search, it has also a theoretical interest for us, since it is the key to the proof of Fermat's Theorem 5.2.10.

The last reason is that the convexity of the Clarke subdifferential is essential for us, since our main results rely on the separation of convex sets, see the proof of the alternative Theorem 5.2.4. Of course, one can try to go beyond convexity, using for instance the limiting Fréchet subdifferential. This is the approach adopted by Mordukhovich [245] and Bao, Mordukhovich [46], see also [61]. Such an approach would, of course, benefit from the fact that the limiting subdifferential is smaller. Thus, it would provide more accurate necessary conditions for weak optimality (see [205] for a discussion about criticalities). But it would ask for a completely different approach, less centered on the primal space and its directional derivatives.

This remark is also the occasion to say that, in most of the results of this chapter, we could have considered nonconvex constraints. If the constraint is a nonempty closed set  $\Omega \subset X$ , it suffices to consider the Clarke tangent and normal cones to  $\Omega$ . These cones are closed and convex, and enjoys a full duality by taking the polar. The only point for which more regularity is needed is Proposition 5.2.9, for which we would need the fact that the admissible tangent cone is dense in the Clarke tangent cone. Such a property is satisfied if  $\Omega$  is for instance prox-regular.

**Remark 5.3.2** (Relationship with generalized Jacobians, and other works). From the beginning of nonsmooth analysis, there have been attempts to provide good generalizations of the differential for nonsmooth vector-valued functions. For a locally Lipschitz continuous function  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , Clarke [111] defined the Clarke generalized Jacobian of  $F$  at  $x \in \mathbb{R}^n$  by

$$(5.58) \quad \mathcal{J}F(x) := \text{co}\left\{ \lim_{\substack{n \rightarrow +\infty \\ x_n \rightarrow x}} DF(x_n) \mid x_n \xrightarrow{n \rightarrow +\infty} x \text{ and } F \text{ is differentiable at } x_n \right\} \subset L(\mathbb{R}^n, \mathbb{R}^m).$$

This definition is based on Rademacher's theorem, which asserts that a locally Lipschitz function is Gateaux differentiable almost everywhere. When  $m = 1$ , it is equal to the usual Clarke subdifferential introduced in Section 2.2. This Jacobian satisfies the following chain rule (see [111, Proposition 2.6.4]):

$$\forall y^* \in Y^*, \quad \partial^C(y^* \circ F)(x) = y^* \circ \mathcal{J}F(x) := \{y^* \circ A \mid A \in \mathcal{J}F(x)\}.$$

Thus, if  $Y$  is ordered by a closed convex cone with nonempty interior, we can see that the Clarke generalized Jacobian is related to the ordered Clarke subdifferential by:

$$\partial^C F(x) = \Theta \circ \mathcal{J}F(x) := \{\theta \circ A \mid \theta \in \Theta, A \in \mathcal{J}F(x)\}.$$

The latter generalizes Proposition 5.1.8 iii), in finite dimensions. See the work of Thibault [305] for an extension to functions  $F : X \rightarrow Y$  between a separable Banach space  $X$  and a reflexive separable Banach space  $Y$ .

Considering a function  $F : X \rightarrow Y$  with values in an ordered Banach space, there has been in the 70's some attempts to build a generalized Jacobian which would respect the order structure of  $Y$ . In its seminal work, Valadier [310] considered convex continuous functions with values in an order complete Banach lattice<sup>4</sup> and introduced what we call<sup>5</sup> an ordered generalized Jacobian of  $F$  at  $x$ :

$$\mathcal{J}_{\leq} F(x) := \{A \in L(X, Y) \mid \forall x' \in X, 0 \leq F(x') - F(x) - A(x' - x)\}.$$

It is clear that  $\mathcal{J}_{\leq} F(x)$  enjoys the following Fermat's rule

$$0 \in \mathcal{J}_{\leq} F(x) \Leftrightarrow F(x) \leq F(x') \quad \forall x' \in X.$$

<sup>4</sup>A Banach lattice is an ordered Banach space such that, given two elements  $x, y$ , we can define their minimum and maximum. Order complete is an assumption which essentially ensures the existence of a supremum for bounded sets. See [5] for more details.

<sup>5</sup>We choose this denomination to underline the fact that it is an object living in  $L(X, Y)$ , in opposition with our ordered subdifferential living in  $X^*$ . In the mentioned papers, these objects are also called subdifferentials.

But, as we mentioned it in Section 2.3.3, this is in general pointless, since in a non totally ordered space, such global minimum does not always exist. It is easy to see that, in the context of this chapter,

$$\Theta \circ \mathcal{J}_{\leq} F(x) \subset \partial^c F(x),$$

but this inclusion is strict in general since  $\mathcal{J}_{\leq} F(x)$  might be empty. When  $X$  is a Banach space and  $Y$  a reflexive separable Banach space, Thibault [305] extended this notion to locally Lipschitz continuous functions, by taking<sup>6</sup>

$$\mathcal{J}_{\leq} F(x) := \text{cl}_{WOT} \text{coo} \left\{ \lim_{n \rightarrow +\infty} DF(x_n) \mid x_n \xrightarrow[n \rightarrow +\infty]{} x \text{ and } F \text{ is differentiable at } x_n \right\}.$$

It is an exercise to verify that, in the context of this chapter, we also have

$$\Theta \circ \mathcal{J}_{\leq} F(x) \subset \partial^c F(x).$$

The generalized Jacobian has been used in the analysis of vector-valued functions, in particular in the Lagrangian theory [110, 192, 111]. But it seems to us that these ordered Jacobians have not been a successful tool for the analysis of vector optimization problems. The fact that they characterize global minimum instead of Pareto/efficient points might be its major drawback.

**Remark 5.3.3** (Links with qualification conditions). We say that two sets  $C, D \subset X$  have a (*Bouligand*) *linearly regular* intersection at  $x \in C \cap D$ , if (see e.g. Example 4.2.13 and [214, 222])

$$N_C^B(x) \cap N_D^B(x) = \{0\}.$$

Recall from the alternative Theorem 5.2.4 that there exists an admissible descent direction at  $x$  if and only if  $0 \notin N_C(x) + \partial^c F(x)$ . It can be verified that this noncriticality of  $x$  is equivalent to

$$0 \notin \partial^c F(x) \text{ and } N_C(x) \cap -\mathbb{R}_+ \partial^c F(x) = \{0\}.$$

In the context of Theorem 5.2.12, we see that this is equivalent to

$$0 \notin \partial^c F(x) \text{ and } N_C(x) \cap -N_{[F \leq F(x)]}^B(x) = \{0\},$$

where this second property expresses that the intersection between the constraint  $C$  and the sublevel set  $[F \leq F(x)]$  must be linearly regular at  $x$ . We illustrate this situation in Figure 5.8. We take  $F = (f_1, f_2) : \mathbb{R}^2 \rightarrow (\mathbb{R}^2, \mathbb{R}_+^2)$ , and consider a linear constraint  $C$  (plotted in a continuous line) for which there exists no *admissible* descent directions at  $x \in C$ . This is because the intersection between  $C$  and  $[F \leq F(x)]$  is not regular : see how  $N_C(x) \cap -N_{[F \leq F(x)]}^B(x) \neq \{0\}$ .

Figure 5.8: First-order analysis at the intersection of two sublevel sets and a constraint.

This kind of regularity condition arises naturally when considering optimization problems with equality and inequality constraints. Consider for instance a finite family of functions

$$f, g_1, \dots, g_m, h_1, \dots, h_p : \mathbb{R}^n \rightarrow \mathbb{R},$$

---

<sup>6</sup>Here  $\text{cl}_{WOT}$  denotes the closure with respect with the *weak operator norm*, which is the topology of the pointwise convergence to respect with the weak topology of  $Y$ . In other words,  $A_\alpha \xrightarrow[\alpha \in \mathcal{A}]{WOT} A$  if and only if  $\langle y^*, A_\alpha(x) \rangle \xrightarrow[\alpha \in \mathcal{A}]{\mathbb{R}} \langle y^*, A(x) \rangle$  for all  $y^* \in Y^*$  and all  $x \in X$ . The “coo” is the *operator convex hull*, defined for a family  $\mathcal{F} \subset L(X, Y)$  as the set of finite sums  $\sum_{i=1}^N u_i \circ A_i$ , where the  $A_i$ ’s are elements of  $\mathcal{F}$ , and the  $u_i$ ’s are positive endomorphisms of  $Y$  (*i.e.*  $0 \leq y \Rightarrow 0 \leq u_i(y)$ ) such that  $\sum_{i=1}^N u_i = \text{id}_Y$ .

where  $f$  and  $G := (g_1, \dots, g_m)$  are locally Lipschitz continuous, and  $H := (h_1, \dots, h_p)$  is smooth. A classical approach to study

$$(P) \quad \min_{x \in \mathbb{R}^n} f(x) \text{ such that } g_1(x) \leq 0, \dots, g_m(x) \leq 0 \text{ and } H(x) = 0$$

consists in writing the corresponding Lagrangian

$$L(\mu, \theta, \lambda, x) := \mu f(x) + \sum_{i=1}^m \theta_i g_i(x) + \sum_{j=1}^p \lambda_j h_j(x), \text{ for } \theta = (\theta_i) \in \mathbb{R}^m \text{ and } \lambda = (\lambda_j) \in \mathbb{R}^p,$$

and study its critical points. For instance, if  $\bar{x} \in \mathbb{R}^n$  is a local solution for (P), there exists a triplet  $(\mu, \theta, \lambda) \in \mathbb{R} \times \mathbb{R}_+^m \times \mathbb{R}^p$ , called *Fritz-John coefficients*, such that  $0 \in \partial^c L(\mu, \theta, \lambda, \cdot)(\bar{x})$  (see [192, Theorem 2.1]). The whole point is to find *qualification conditions* on  $\bar{x}$  which ensures that  $(\mu, \theta, \lambda)$  are *Karush-Kuhn-Tucker coefficients*, i.e.  $\mu \neq 0$ . Such a standard assumption is the *Mangasarian-Fromovitz condition*, which demands that (assume that all the constraints are qualified<sup>7</sup> at  $\bar{x}$ ):

- i)  $DH(\bar{x})$  is surjective,
- ii) there exists  $d \in \text{Ker } DH(\bar{x})$  such that  $d^c g_i(\bar{x}; d) < 0, \forall i \in \{1, \dots, m\}$ .

According to Liusternik's theorem [77, Theorem 7.1.6], the Mangasaria-Fromovitz condition can be equivalently rewritten as:

- i)  $DH(\bar{x})$  is surjective,
- ii)  $[H = 0]$  and  $[G \leq 0]$  have a strongly regular intersection at  $\bar{x}$ ,

where  $[H = 0] := \{x \in \mathbb{R}^n \mid H(x) = 0\}$  and  $[G \leq 0] := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i \in \{1, \dots, m\}\}$ .

The discussion at the beginning of this remark says in particular that item ii) could be interpreted as the noncriticality of  $\bar{x}$  with respect to the multi-objective optimisation problem

$$(MOP) \quad \text{MIN}_{H(x)=0} G(x).$$

Of course, this would require the constraint  $[H = 0]$  to be convex, to fall into the setting of Theorems 5.2.4 and 5.2.12, which is false in general<sup>8</sup>. We will also see in Chapter 6, Proposition 6.1.17, that the condition i) will provide a good regularity property for the steepest descent vector field  $s$ .

<sup>7</sup>In other words, that  $g_i(\bar{x}) = 0$  for all  $i \in \{1, \dots, m\}$ .

<sup>8</sup>But most of the results of this chapter can be extended to more general constraints  $C$ , see Remark 5.3.1.



## Chapter 6

# The continuous steepest descent dynamic for vector optimization

Along all this chapter, we are considering a locally Lipschitz function  $F : H \rightarrow Y$ , where  $H$  is a Hilbert space identified with its dual, and  $Y$  is a Banach space ordered by a nonempty closed convex cone  $K$  having nonempty interior. As in the previous chapter, we equip the dual cone  $K^+$  with a fixed convex weakly\* compact base  $\Theta \subset Y^*$ . We refer to Section 2.3 for more details on ordered Banach spaces and bases of cones.

We recall from the introduction that Smale [297] studied the multi-objective optimisation problem

$$(MOP) \quad \text{MIN}_{x \in C} F(x), \text{ with } F(x) = (f_1(x), \dots, f_m(x)),$$

and defined the notion of *gradient process* for this problem. It is a differential equation

$$\dot{u}(t) = \psi(u(t))$$

where  $\psi : H \rightarrow H$  is a mapping which satisfies the following properties:

$$\begin{cases} \psi(u) \text{ is an admissible descent direction whenever } u \text{ is not a Pareto critical point,} \\ \psi(u) = 0 \text{ else.} \end{cases}$$

The interest of such a gradient process is twofold. First, the stationary points of the dynamic, those satisfying  $\dot{u}(t) = 0$ , are exactly the critical Pareto points for (MOP). This is a simple consequence of the Pareto alternative Theorem 5.2.4, which says that the critical Pareto points are exactly those for which no admissible descent direction can be found. Secondly, as long as  $u(t)$  is not a critical Pareto point, all the objective functions  $\{f_1, \dots, f_m\}$  are guaranteed to decrease.

The steepest descent vector field

$$\begin{aligned} s : H &\longrightarrow H \\ x &\longmapsto -(N_C(x) + \text{co } \partial^c f_i(x))^0, \end{aligned}$$

introduced in the previous chapter, induces such a gradient process for multi-objective optimization problems, according to Proposition 5.2.18. The idea of considering a continuous dynamic governed by this vector field goes back to Cornet [116], and has been revisited in the last years [295, 282, 248, 28].

This Chapter 6 is devoted to the study of the trajectories solutions of the *steepest descent* dynamical system

$$(SD) \quad \dot{u}(t) + \left( N_C(u(t)) + \partial^c F(u(t)) \right)^0 = 0,$$

where  $F : H \rightarrow Y$  is a locally Lipschitz function with values in an ordered Banach space  $(Y, K)$ . It can be written  $\dot{u}(t) = s(u(t))$ , where  $s$  denotes the steepest descent vector field introduced in Definition 5.2.17. This dynamic is a gradient process, in the sense of Smale, for the vector optimization problem

$$(VOP) \quad \underset{x \in C}{\text{MIN}} \quad F(x).$$

Indeed, if  $u : [0, +\infty[ \rightarrow H$  is a solution of (SD), its derivative  $\dot{u}(t)$  is, by definition, a descent direction for  $F$  at  $u(t)$ . Then, we expect that  $F(u(t))$  decreases along time, with respect to the order in  $Y$ . Moreover, the stationary points of  $u(\cdot)$  are critical points for (VOP), so we can also expect the trajectories solutions of (SD) to converge, when  $t \rightarrow +\infty$ , to a critical point of (VOP).

The Chapter is structured as follows. Section 6.1 is devoted to a first basic study of the steepest descent dynamic (SD). We start in Section 6.1.1 by defining properly what we mean by a solution of (SD). Indeed, the dynamic is governed by two nonsmooth operators, namely the normal cone to the constraint  $N_C$  and the ordered subdifferential  $\partial^c F$ . While it is clear that  $N_C(\cdot)$  is discontinuous, it also appears that  $\partial^c F : H \rightrightarrows H$  suffers from a lack of smoothness, even if  $F$  is regular (see Example 6.1.11). This is why we consider absolutely continuous solutions. In Section 6.1.2 we prove the announced decrease property of  $F$  along the trajectories, in Proposition 6.1.7. As it was already underlined in Chapter 5 and Section 2.3, this decrease property of  $F$  is essentially due to the simultaneous decrease of the family of cost functions

$$\{f_\theta = \theta \circ F : H \rightarrow \mathbb{R} \mid \theta \in \Theta \subset Y^*\}.$$

In sections 6.1.3 and 6.1.4 we illustrate the dynamic through some examples, and we compare (SD) to other dynamics, in particular the ones arising from scalarization methods. Section 6.1.5 consists in a discussion on the uniqueness of the trajectories. Indeed, since the steepest descent vector field is neither smooth or monotone, we cannot use classical results or techniques which ensure the uniqueness. We give in particular some geometrical sufficient conditions ensuring the uniqueness. The question of the uniqueness in general remains open.

Then comes Section 6.2, which is entirely devoted to the asymptotic behaviour of the solutions of (SD). Our main result is Theorem 6.2.6, which gives the convergence of the trajectories to solutions of (VOP) under some conditions. When  $F$  is convex, any bounded trajectory converges weakly to a weak Pareto point. If  $F$  is not convex but scalarly quasiconvex, we can still guarantee the convergence under additional hypotheses. The proof relies on the F ejer monotonicity of the trajectories, and Opial's Lemma. We also guarantee in Theorem 6.2.9 that the convergence is strong in the convex *symmetric* case.

Note finally that, in this Chapter, we assume the existence of solutions for (SD). This question will be treated in Chapter 7.

## 6.1 The steepest descent dynamic

### 6.1.1 Definitions

The dynamical system which is governed by  $u \mapsto s(u)$ , will be called the *steepest descent* dynamical system, (SD) for short. Its solution trajectories  $t \mapsto u(t)$  verify

$$(6.1) \quad (SD) \quad \dot{u}(t) + \left( N_C(u(t)) + \partial^c F(u(t)) \right)^0 = 0.$$

**Remark 6.1.1.** The vector fields that appear in Theorem 5.2.20 generate the same integral curves, with a different time scale. We could have chosen any of them to generate our dynamic.

In [297], it is assumed that the vector field  $\psi$  governing the gradient process is continuous, in a finite dimensional setting. In our context, the corresponding notions have been extended in order to cover dynamical systems governed by a discontinuous vector field on a general Hilbert space, to model a general preference relation on  $Y$ . In particular, instead of classical (continuously differentiable) solutions, we will consider strong solutions (absolutely continuous on bounded time intervals), the equality (6.1) being satisfied almost everywhere. We recall here the definition of absolutely continuous functions (see the monograph of Brezis [83, Appendix] for more details).

**Definition 6.1.2.** Given  $T \in \mathbb{R}_+$ , a function  $u : [0, T] \rightarrow \mathbf{H}$  is said to be absolutely continuous if one of the following equivalent properties holds:

i) there exists an integrable function  $g : [0, T] \rightarrow \mathbf{H}$  such that

$$u(t) = u(0) + \int_0^t g(s) ds \quad \forall t \in [0, T];$$

ii)  $u$  is continuous and its distributional derivative belongs to the Lebesgue space  $L^1([0, T]; \mathbf{H})$ ;

iii) for every  $\epsilon > 0$ , there exists  $\eta > 0$  such that for any finite family of intervals  $I_k = ]a_k, b_k[$ ,  $I_k \cap I_j = \emptyset$  for  $k \neq j$  and  $\sum_k |b_k - a_k| \leq \eta \implies \sum_k \|u(b_k) - u(a_k)\| \leq \epsilon$ .

We can now make precise the notion of solution for the steepest descent dynamic (SD).

**Definition 6.1.3.** We say that  $u : [0, +\infty[ \rightarrow C \subset \mathbf{H}$  is a *strong global solution* of (SD) if :

i)  $u(\cdot)$  is absolutely continuous on each interval finite interval  $[0, T]$ ,  $T \in ]0, +\infty[$ .

ii)  $\dot{u}(t) = s(u(t))$  for a.e.  $t \in ]0, +\infty[$ .

iii) there exists  $\eta : [0, +\infty[ \rightarrow \mathbf{H}$ ,  $v : [0, +\infty[ \rightarrow \mathbf{H}$ , such that for a.e.  $t \in ]0, +\infty[$  :

iii.a)  $\eta(t) \in N_C(u(t))$  and  $v(t) \in \partial^c F(u(t))$ ,

iii.b)  $\dot{u}(t) + \eta(t) + v(t) = 0$  for a.e.  $t \in ]0, +\infty[$ .

**Remark 6.1.4.** Observe that, by definition, the trajectory remains always in the constraint  $C$ . Indeed the nonvacuity of  $N_C(u(t))$  for a.e.  $t \geq 0$  means that  $u(t) \in C$  for a.e.  $t \geq 0$ . Since  $u$  is continuous and  $C$  closed, we deduce that  $u(t) \in C$  for all  $t \geq 0$ . It follows from the definition of the Fenchel tangent cone (recall Section 2.2.1) that  $\dot{u}(t) \in T_C(u(t))$  for a.e.  $t \in [0, T]$ .

## 6.1.2 Qualitative properties of the trajectories

We establish now the first qualitative properties of strong solutions of (SD). We take here for granted the existence of such strong global solutions. We will examine in details their existence in Chapter 7.

The main point of the following proposition is that the family of cost functions  $\{f_\theta\}_{\theta \in \Theta}$  is equicontinuous along the trajectories. This will be of great interest when studying the existence of such trajectories in Chapter 7.

**Proposition 6.1.5** (Equicontinuity of the cost functions). Let  $u : [0, +\infty[ \rightarrow C$  be a strong global solution of (SD).

i) *Lipschitz continuity:* The trajectory is Lipschitz continuous on any finite time interval  $[0, T]$ .



- ii) *Global Lipschitz continuity*: If we suppose that the trajectory is bounded, and that  $F$  is Lipschitz continuous on bounded sets, then the trajectory is Lipschitz continuous on the entire interval  $[0, +\infty[$ .
- iii) *Equicontinuity*: The family  $\{f_\theta \circ u : [0, T] \rightarrow \mathbb{R} \mid \theta \in \Theta\}$  is:
  - iii.a) uniformly Lipschitz continuous (*i.e.* Lipschitz continuous with the same Lipschitz constant), therefore equicontinuous,
  - iii.b) relatively compact in  $(C[0, T], \mathbb{R}), \|\cdot\|_\infty$ , the space of continuous real-valued functions on  $[0, T]$ , equipped with the uniform convergence topology.

**Remark 6.1.6.** If  $H$  has finite dimension, then the hypothesis of Lipschitz continuity on bounded sets which appears in item ii) can be removed, since it is equivalent to local Lipschitz continuity.

*Proof.* i) By Definition 6.1.3 of a strong global solution, we can write for a.e.  $t \in ]0, +\infty[$ ,

$$(6.2) \quad \dot{u}(t) + \eta(t) + v(t) = 0,$$

where  $\eta(t) \in N_C(u(t))$  and  $v(t) \in \partial^c F(u(t))$ . Taking the scalar product with  $\dot{u}(t)$  in (6.2) gives

$$(6.3) \quad \|\dot{u}(t)\|^2 = \langle \eta(t), -\dot{u}(t) \rangle + \langle v(t), -\dot{u}(t) \rangle.$$

Since  $u(t) \in C$ , we know from Proposition 2.2.11 that

$$(6.4) \quad \langle \eta, \dot{u}(t) \rangle = \frac{d}{dt} \delta_C \circ u(t) = 0 \quad \text{for all } \eta \in N_C(u(t)),$$

so (6.3) becomes  $\|\dot{u}(t)\|^2 = \langle v(t), -\dot{u}(t) \rangle$ . Using the Cauchy-Schwarz inequality, we obtain

$$(6.5) \quad \|\dot{u}(t)\| \leq \|v(t)\|.$$

From the compactness of  $[0, T]$  and the continuity of  $u$ , we know that  $u([0, T])$  is compact in  $H$ . Using this compactness and the local boundedness of  $\partial^c F$  (recall Proposition 5.1.7), we deduce that  $\partial^c F$  is uniformly bounded on  $u([0, T])$ . From  $v(t) \in \partial^c F(u(t))$  and (6.5), we deduce that  $\dot{u}$  is uniformly bounded on  $[0, T]$ , so  $u$  is Lipschitz continuous therein.

Item ii) follows exactly the same line, using the boundedness assumptions.

iii) For all  $\theta \in \Theta$ , and a.e.  $t \in [0, T]$ , we have  $\partial^c f_\theta(u(t)) \subset \partial^c F(x)$ . Since the ordered subdifferential  $\partial^c F(x)$  has been proved to be uniformly bounded on  $u([0, T])$ , there exists  $L \geq 0$  such that  $\partial^c f_\theta(u(t)) \subset L\mathbb{B}_H$ , for all  $\theta \in \Theta$  and a.e.  $t \in [0, T]$ . Therefore, each cost function  $f_\theta$  is  $L$ -Lipschitz continuous on a neighbourhood of  $u([0, T])$ . Added to the fact that  $u$  is Lipschitz continuous on the interval  $[0, T]$ , item iii.a) follows. To prove item iii.b), we want to use Ascoli's Theorem C.0.1. For this, consider for any  $t \in [0, T]$  the set

$$(6.6) \quad I_t := \{(f_\theta \circ u)(t) \mid \theta \in \Theta\}.$$

Using the fact that the base  $\Theta$  is bounded, and the definition  $f_\theta = \theta \circ F$ , it is clear that  $I_t$  is bounded in  $\mathbb{R}$ , hence relatively compact. Then Ascoli's Theorem applies, and item iii.b) follows. ■

Now we show that the cost functions are Lyapunov for the (SD) dynamic: they decrease along the trajectories. Even more, we see that the dynamic is a *descent* method, relatively to the order endowing  $Y$ .

**Proposition 6.1.7** (Descent property). Let  $u : [0, +\infty[ \rightarrow C$  be a strong global solution of (SD). Then, for each  $\theta \in \Theta$ , and for almost all  $t > 0$  :

$$(6.7) \quad \frac{d}{dt}(f_\theta \circ u)(t) \leq -\|\dot{u}(t)\|^2.$$

In particular,  $f_\theta$  is decreasing along the trajectories. It is also the case for  $F$  :

$$(6.8) \quad t_1 \leq t_2 \Rightarrow F(u(t_2)) \leq F(u(t_1)).$$

*Proof.* Let  $\theta \in \Theta$ . From Proposition 6.1.5, we know that  $f_\theta \circ u$  is locally Lipschitz on  $\mathbb{R}_+$ , thus differentiable almost everywhere. So, for a.e.  $t > 0$ ,  $f_\theta \circ u$  is differentiable and  $\dot{u}(t) = s(u(t))$  holds. For such  $t > 0$ , we have, using the chain rule for the Fréchet subdifferential (see Proposition 2.2.11):

$$(6.9) \quad \frac{d}{dt}(f_\theta \circ u)(t) = \langle x^*, \dot{u}(t) \rangle, \quad \forall x^* \in \partial^F f_\theta(u(t)).$$

Since  $\partial^F f_\theta(u(t)) \subset \partial^C f_\theta(u(t)) \subset \partial^C F(x)$ , (6.9) implies

$$(6.10) \quad \frac{d}{dt}(f_\theta \circ u)(t) \leq d^C F(u(t), \dot{u}(t)).$$

Using  $\dot{u}(t) = s(u(t))$  and its descent property (see Proposition 5.2.18), we obtain the desired inequality (6.7). The monotonic property of  $F \circ u$  (6.8) is a consequence of the monotony of  $f_\theta \circ u$ , and the characterization of the order in  $Y$  by the elements of the base  $\Theta$  (recall Proposition 2.3.7). ■

**Remark 6.1.8.** If we assume that, for a.e.  $t > 0$ ,  $\dot{u}(t) \neq 0$ , then we can easily improve the proof of (6.8) to obtain a *strict* monotonicity of  $F \circ u$ :

$$(6.11) \quad t_1 < t_2 \Rightarrow F(u(t_2)) < F(u(t_1)).$$

When we have a result of uniqueness for the trajectories, there is only two alternatives concerning the stationary points. Either the trajectory reaches a stationary point in finite time, and remains there, or the trajectory is never stationary. This question of uniqueness of trajectories is discussed in Section 6.1.5. In particular, Proposition 6.1.21 shows that, in the convex case, (6.11) holds until the trajectory reaches a weak Pareto point, in which case the trajectory stops there.

### 6.1.3 Some basic examples

We illustrate the (SD) dynamic through some examples. They suggest that its study is not a mere extension of some classic situation, because (SD) is governed by a vector field which is neither monotone, nor locally Lipschitz continuous. Without any further assumption, we cannot expect more than the Hölder continuity of this vector field (see Example 6.1.11 below, and Proposition 7.1.2 in Chapter 7).

For simplicity, we consider a strictly Gateaux differentiable bi-objective unconstrained case, *i.e.*

$$(6.12) \quad F = (f_1, f_2) : H \rightarrow \mathbb{R}^2, \text{ with } C = H \text{ and } K = \mathbb{R}_+^2.$$

The following elementary result provides an explicit description of the vector field in that case.

**Proposition 6.1.9.** Suppose we are in the setting of (6.12). Then, for any  $x \in H$  be such that  $\nabla f_1(x) \neq \nabla f_2(x)$ ,

$$(6.13) \quad \begin{cases} s(x) = -\lambda(x)\nabla f_1(x) - (1 - \lambda(x))\nabla f_2(x), \\ \lambda(x) = \text{proj}_{[0,1]} \left( \frac{\langle \nabla f_2(x) - \nabla f_1(x), \nabla f_2(x) \rangle}{\|\nabla f_2(x) - \nabla f_1(x)\|^2} \right). \end{cases}$$

*Proof.* By definition of  $s(x)$ , there exists  $\lambda(x) \in [0, 1]$  verifying

$$\begin{cases} s(x) = -\lambda(x)\nabla f_1(x) - (1 - \lambda(x))\nabla f_2(x), \\ \lambda(x) = \operatorname{argmin}_{\lambda \in [0,1]} \|\lambda\nabla f_1(x) + (1 - \lambda)\nabla f_2(x)\|^2. \end{cases}$$

Writing down the first-order optimality condition for  $\lambda(x)$ , we obtain, when  $\nabla f_1(x) \neq \nabla f_2(x)$ ,

$$\lambda(x) + N_{[0,1]}(\lambda(x)) \ni \frac{\langle \nabla f_2(x) - \nabla f_1(x), \nabla f_2(x) \rangle}{\|\nabla f_2(x) - \nabla f_1(x)\|^2}.$$

We conclude by using that the resolvent of the normal cone mapping is equal to the projection. ■

**Example 6.1.10.** Given  $a, b \in \mathbb{H}$ , consider the quadratic functions  $f_1(v) = \frac{1}{2}\|v - a\|^2$  and  $f_2(v) = \frac{1}{2}\|v - b\|^2$ . The set of Pareto points coincide with the set of weak Pareto points, and is equal to the segment  $[a, b]$ . The steepest descent vector field is given by  $s(x) = -(x - \operatorname{proj}_{[a,b]}x)$ . Trajectories are straight lines connecting the starting point and its projection on the Pareto set.

Figure 6.1 below shows some trajectories of the (SD) dynamic in the particular case  $\mathbb{H} = \mathbb{R}^2$ , and  $a = (1, 0)$ ,  $b = (-1, 0)$ . Figure 6.2 shows the Pareto front in the value space.

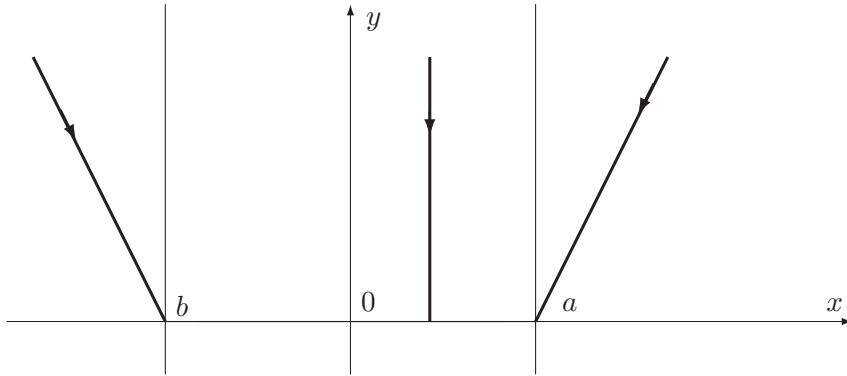


Figure 6.1: Trajectories associated to two distance functions in  $\mathbb{H} = \mathbb{R}^2$ .

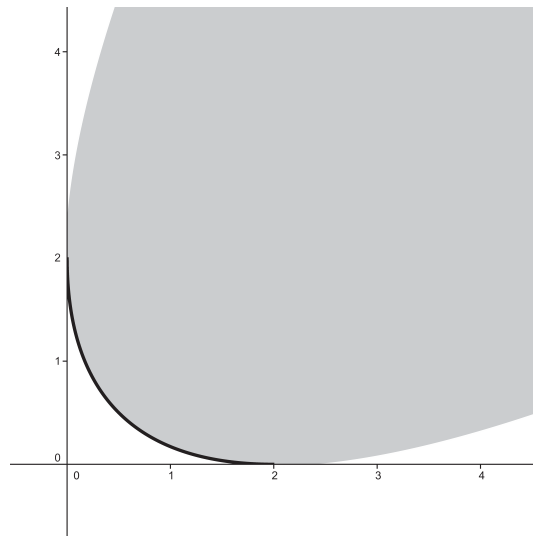


Figure 6.2: Pareto front in the value space  $Y = \mathbb{R}^2$ .  $F(X)$  is in grey, the Pareto front is the thick black curve.

**Example 6.1.11.** Given  $a \in H$ , consider the quadratic form  $f_1(v) = \frac{1}{2}\|v\|^2$  and the linear form  $f_2(v) = \langle a, v \rangle$ . The sets of Pareto and weak Pareto points both coincide with the half-line  $-\mathbb{R}_+a$ . The steepest descent vector field is deduced from Proposition 6.1.9:

$$(6.14) \quad s(x) = \begin{cases} -a & \text{if } \langle a, x \rangle \geq \|a\|^2, \\ -x & \text{if } \|x - \frac{a}{2}\|^2 \leq \|\frac{a}{2}\|^2, \\ -a + \frac{\langle a, a-x \rangle}{\|a-x\|^2}(a-x) & \text{else.} \end{cases}$$

The domain of this vector field appears to be split in three parts, where it behaves differently. Those three areas are a half-space supported by  $a$ , a ball centered in  $\frac{a}{2}$ , and the rest of the space. The fact that this domains splits in three can also be seen in Example 6.1.10. It is simply due to the fact that  $s(x)$  is the projection of the origin onto the segment  $[\nabla f_1(x), \nabla f_2(x)]$ . When doing this projection, three cases occurs. Either we project onto the extreme points of the segment, namely  $\nabla f_1(x)$  and  $\nabla f_2(x)$ , or we project onto the interior of  $[\nabla f_1(x), \nabla f_2(x)]$ .

Let us consider the particular case  $H = \mathbb{R}^2$ , with  $a = (1, 0)$ . Then  $f_1(x, y) = \frac{1}{2}(x^2 + y^2)$  and  $f_2(x, y) = x$ . The corresponding Pareto set is  $] -\infty, 0] \times \{0\}$ , and

$$(6.15) \quad s(x, y) = \begin{cases} -(1, 0) & \text{if } x \geq 1, \\ -(x, y) & \text{if } (x - \frac{1}{2})^2 + y^2 \leq \frac{1}{4}, \\ \frac{-1}{(x-1)^2+y^2}(y^2, y(1-x)) & \text{else.} \end{cases}$$

Figure 6.3 below shows some trajectories of the (SD) dynamic. One can see that both objective function decrease along the trajectories. Figure 6.4 shows the Pareto front in the value space.

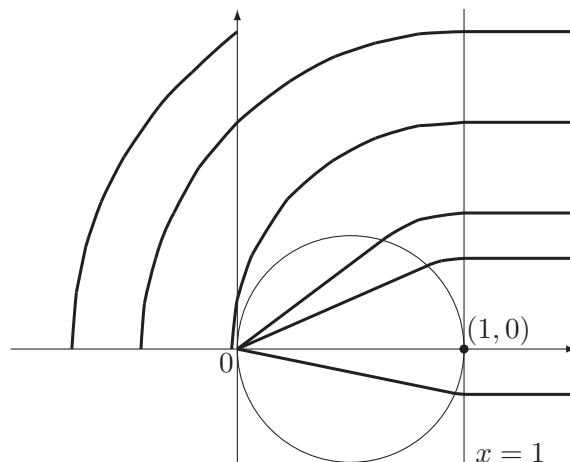


Figure 6.3: Trajectories in  $H = \mathbb{R}^2$  associated to a nondegenerated quadratic form and a linear form.

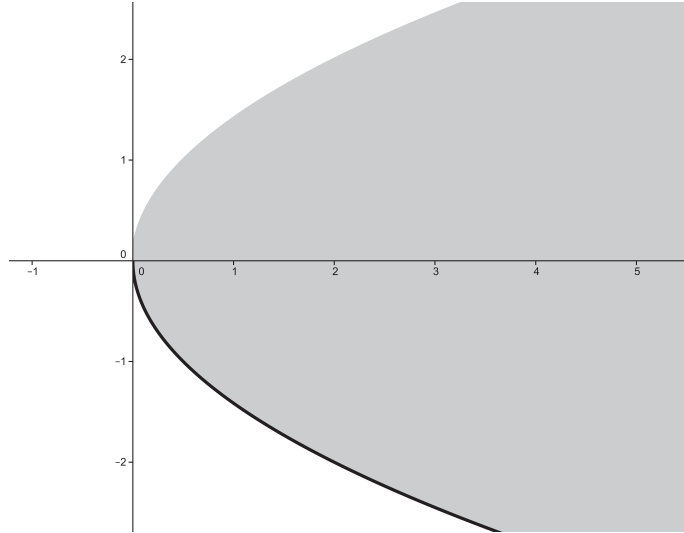


Figure 6.4: Pareto front in the value space  $Y = \mathbb{R}^2$ .  $F(X)$  is in grey, the Pareto front is the thick black curve.

This example is very interesting, because it has the property that the steepest descent vector field  $s(x) = -(\text{co}\{\nabla f_1(x), \nabla f_2(x)\})^0$  governing the dynamic may fail to be locally Lipschitz. This can be quite surprising at first sight, since the functions involved here are smooth. The lack of Lipschitz continuity occurs at the point  $(1, 0)$ , which is the point where the vector field “splits” into three parts, see (6.15). Figure 6.5 provides a simple example of parametrized vectors  $u_\alpha, v_\alpha$  that both converge to  $(1, 0)$ , when the angle parameter  $\alpha$  goes to zero, but such that

$$\|s(u_\alpha) - s(v_\alpha)\| = \sin(\alpha) \text{ and } \|u_\alpha - v_\alpha\| = \sin(\alpha) \tan(\alpha).$$

As a consequence,  $\frac{\|s(u_\alpha) - s(v_\alpha)\|}{\|u_\alpha - v_\alpha\|} \simeq \frac{\alpha}{\alpha^2}$  is unbounded when  $\alpha \rightarrow 0$ , and this breaks the Lipschitz continuity in the neighbourhood of  $(1, 0)$ .

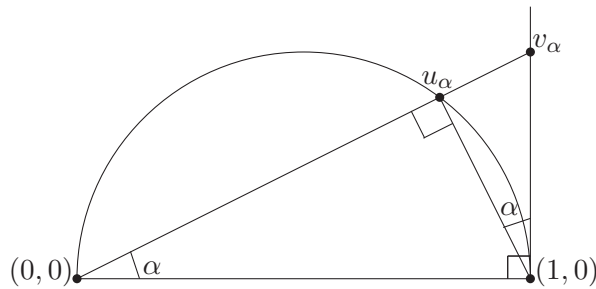


Figure 6.5: Lack of Lipschitz continuity for  $u \mapsto s(u)$ .

The point  $(1, 0)$  where fails the Lipschitz continuity of  $s$  is in fact the only one in the plane. To see this, go to Proposition 6.1.18, where we give a necessary condition for the failure of the Lipschitz continuity of  $s$ .

**Example 6.1.12.** Take  $H = \mathbb{R}^2$ , and consider the two quadratic degenerate forms  $f_1(x, y) = \frac{1}{2}x^2$  and  $f_2(x, y) = \frac{1}{2}y^2$ . Here the only Pareto point is the origin, and the set of weak Pareto points is and  $\mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R}$ . The multiobjective steepest descent vector field is, once computed:

$$s(x, y) = -\left(\frac{xy^2}{x^2 + y^2}, \frac{yx^2}{x^2 + y^2}\right) \text{ if } (x, y) \neq (0, 0), \quad s(0, 0) = (0, 0).$$

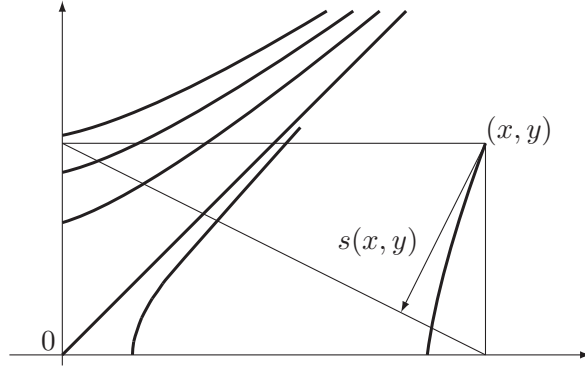


Figure 6.6: Trajectories in  $H = \mathbb{R}^2$  associated to two degenerated quadratic forms.



Figure 6.7: Pareto front in the value space  $Y = \mathbb{R}^2$ .  $F(X)$  is in grey, the Pareto front is the thick black curve.

Observe in Figure 6.6 that the trajectories tend to move away from each other. This reflects the fact that  $(x, y) \mapsto -s(x, y)$  is not a monotone operator. Indeed, for  $x > 0$ ,  $y > 0$ ,  $x \neq y$

$$\langle -s(x, y) + s(y, x), (x, y) - (y, x) \rangle = -2 \frac{xy(x-y)^2}{x^2 + y^2} < 0.$$

Nevertheless, in this example, it can be shown that  $(x, y) \mapsto -s(x, y)$  is *hypomonotone* on  $(\mathbb{R}^*)^2$ . This means that, locally, there exists some  $\alpha \geq 0$  such that  $-s + \alpha I$  is monotone. To see this, compute the Jacobian of  $-s$  at  $(x, y)$

$$D(-s)(x, y) = \frac{1}{(x^2 + y^2)^2} \begin{pmatrix} y^4 - x^2 y^2 & 2xy^3 \\ 2x^3 y & x^4 - x^2 y^2 \end{pmatrix},$$

and observe that  $D(-s)(x, y) + I$  is positive whenever  $(x, y) \neq (0, 0)$ . It would be interesting to find some conditions on  $F$  which would guarantee that  $-s$  is hypomonotone, since in that case the study of the dynamic would fall into a well-understood class of dynamics.

This example shows also that the operator  $s : H \rightarrow H$  does not derive from a potential. Indeed, we can see above that its differential  $Ds(x, y)$  is not symmetric.

### 6.1.4 Related dynamics

In classic optimization problems, when  $F = f : H \rightarrow \mathbb{R}$ , the (SD) dynamic reduces to

$$\dot{u}(t) + \left( N_C(u(t)) + \partial^c f(u(t)) \right)^0 = 0.$$

If moreover the function  $f$  is convex, this system is equivalent to

$$\dot{u}(t) + N_C(u(t)) + \partial^c f(u(t)) \ni 0,$$

because, in this case, the lazy solution property is automatically satisfied by the trajectories of the semigroup of contractions generated by the maximal monotone operator  $N_C + \partial f$ , see [83, Theorem 3.1]. In particular, our existence and asymptotic analysis for (SD) in Sections 2 and 3 extends the well-known results for the nonsmooth gradient flow, see [83].

This leads us to a natural question, which is the study of the relationship (or differences) between (SD) and the Vector Differential Inclusion ((VDI) for short)

$$(6.16) \quad (\text{VDI}) \quad \dot{u}(t) + N_C(u(t)) + \partial^c F(x) \ni 0.$$

For the sake of simplicity, we assume here that  $F$  is convex. It appears that (VDI) enjoys a weaker form of Proposition 6.1.7: we can only guarantee that *one* of the cost functions is decreasing at each instant.

**Proposition 6.1.13.** Let  $u : [0, +\infty[ \rightarrow C$  be a strong global solution of (VDI) in the sense of Definition 6.1.3 (except the lazy property). Suppose that  $F$  is convex. Then, for almost all  $t \geq 0$ , if  $u(t)$  is not critical, there exists some  $\theta \in \Theta$  (which depends on  $t$ ) such that

$$\frac{d}{dt}(f_\theta \circ u)(t) < 0.$$

*Proof.* Since  $u$  is a strong solution of (VDI), there exists  $\eta, p : [0, +\infty[ \rightarrow H$ , which satisfy for almost all  $t \geq 0$

$$\begin{aligned} \eta(t) &\in N_C(u(t)), \quad p(t) \in \partial^c F(u(t)), \\ \dot{u}(t) + \eta(t) + p(t) &= 0. \end{aligned}$$

Taking the scalar product of the above equation with  $\dot{u}(t)$ , we obtain

$$(6.17) \quad \|\dot{u}(t)\|^2 + \langle \eta(t), \dot{u}(t) \rangle + \langle p(t), \dot{u}(t) \rangle = 0.$$

By a similar argument than the one used for (6.4) in the proof of Proposition 6.1.5, we have

$$(6.18) \quad 0 = \frac{d}{dt} \delta_C(u(t)) = \langle \eta(t), \dot{u}(t) \rangle.$$

Hence, if we assume that  $u(t)$  is not critical, we have  $\dot{u}(t) \neq 0$ , and it follows that  $\langle p(t), \dot{u}(t) \rangle < 0$ . Since  $p(t) \in \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \partial^c f_\theta(u(t))$ , arguing by contradiction we easily see that there must exist some  $\theta \in \Theta$  (depending on  $t$ ) and some  $p_\theta \in \partial^c f_\theta(u(t))$  such that  $\langle p_\theta, \dot{u}(t) \rangle < 0$ . Using the chain rule of Proposition 2.2.11, together with the convexity of  $F$ , gives

$$(6.19) \quad \frac{d}{dt} f_\theta(u(t)) = \langle p_\theta, \dot{u}(t) \rangle < 0.$$

■

**Remark 6.1.14.** Proposition 6.1.13 tells us that, for any trajectory of (VDI), for almost all  $t > 0$ , at least one of the objective functions decreases. We will illustrate this on a few examples, and highlight the fact that, by contrast, for trajectories of (SD), they are all decreasing.

i) The scalarization approach consists in taking a constant merit function, say  $f_\theta$ . Clearly,  $\partial f_\theta(u) \subset \partial^c F(x)$ , and the corresponding trajectories are solutions of (VDI). According to Bruck's theorem [87], any orbit of the generalized gradient flow generated by  $\partial f_\theta$  converges to a minimizer of  $f_\theta$ , which, by Theorem 5.2.10, is a weak efficient point. This is the strategy underlying the weighting method, but it suffers from two drawbacks. First, it doesn't recover well the Pareto front in nonconvex cases (see Section 6.3). Secondly, there is no particular reason for this dynamic to improve all the cost functions. Take for instance in Example 6.1.10,  $\theta = (\lambda, (1 - \lambda)) \in \Delta_2$  for any  $\lambda \in ]0, 1]$ . When starting from  $(1, 0)$ , the trajectory goes straight to  $(1 - 2\lambda, 0)$  by decreasing  $f_1$  but increasing  $f_2$ .

ii) Consider the multiobjective setting  $F = (f_1, \dots, f_m)$ , and consider the steepest descent dynamic associated to the function  $f = \max_i f_i$ . This dynamic has some similarities with (SD), but it is different. As a supremum of a finite number of convex continuous functions,  $f$  is still convex continuous. The classical subdifferential rule for the supremum of convex functions (see for example [51, Theorem 18.5]) gives in our setting

$$(6.20) \quad \partial f(u) = \text{co} \{ \partial f_i(u) : i \in I(u) \}$$

where  $I(u) = \{i \in I : f_i(u) = f(u)\}$  is the set of the active indices at  $u$ . Clearly,  $\partial f(u) \subset \text{co} \partial f_i(u) = \partial^c F(u)$ . As a consequence, the trajectories of the steepest descent for  $f = \max_i f_i$  are also solutions of (VDI). But, in general, they fail to make decrease all the objective functions. Take for instance Example 6.1.10: when starting from some  $(x_0, y_0)$  with  $x_0 > y_0 > 1$ ,  $f_2$  is first decreasing along the trajectory, until the current point reaches the projection of  $(1, 0)$  on the line segment joining  $(x_0, y_0)$  to  $(-1, 0)$ , then it is increasing.

In the next proposition, we show that, among all the possible dynamics satisfying (VDI), the unique one making all the merit functions decrease with the estimation of Proposition 6.1.7, is the lazy one, *i.e.*, the steepest descent dynamic.

**Proposition 6.1.15.** Let  $u : [0, T[ \rightarrow H$  be a strong solution of (VDI), and suppose that  $F$  is convex. Then the two following statements are equivalent :

- i)  $u$  is a solution of (SD),
- ii)  $u$  is decreasing for all cost functions, more exactly,

$$\text{for a.e. } t \in [0, T] \text{ and for all } \theta \in \Theta, \quad \frac{d}{dt}(f_\theta \circ u)(t) \leq -\|\dot{u}(t)\|^2.$$

*Proof.* The implication i)  $\Rightarrow$  ii) has already been proved in Proposition 6.1.7. To prove the reverse implication, start by using the chain rule in Proposition 2.2.11, together with the convexity of  $F$ , to obtain for almost every  $t \in [0, T[$ :

$$\forall \theta \in \Theta, \forall x^* \in \partial^c f_\theta(u(t)), \quad \langle x^*, \dot{u}(t) \rangle \leq -\|\dot{u}(t)\|^2.$$

By definition of  $\partial^c F(u(t))$ , we obtain immediately

$$(6.21) \quad \forall x^* \in \partial^c F(u(t)), \quad \langle x^*, \dot{u}(t) \rangle \leq -\|\dot{u}(t)\|^2.$$

Moreover,  $u(t) \in C$  for all  $t \in [0, T[$  (recall Remark 6.1.4). It follows that  $\dot{u}(t) \in T_C(u(t))$  for a.e  $t \in [0, T[$ . Hence,

$$(6.22) \quad \forall \eta \in N_C(u(t)), \quad \langle \eta, \dot{u}(t) \rangle \leq 0.$$



By combining (6.21) and (6.22), we obtain for a.e.  $t \in [0, T[$

$$(6.23) \quad \forall x^* \in N_C(u(t)) + \partial^c F(u(t)), \langle x^*, \dot{u}(t) \rangle \leq -\|\dot{u}(t)\|^2,$$

which is equivalent to say that  $-\dot{u}(t)$  is the projection of the origin onto  $N_C(u(t)) + \partial^c F(u(t))$ . In other words,  $-\dot{u}(t) = (N_C(u(t)) + \partial^c F(u(t)))^0$ . ■

### 6.1.5 About the uniqueness of the trajectories

**Remark 6.1.16.** In the unconstrained multicriteria case, and for convex differentiable objective functions, illustrative examples of the (SD) dynamic were given in Section 6.1.3. In these elementary situations, we have been able to explicitly compute the vector field  $x \mapsto s(x)$ . We observed that it can be Lipschitz continuous (Examples 6.1.10 and 6.1.12) or only Hölder continuous (Example 6.1.11). This naturally raises the following question: in the unconstrained case, and for differentiable objective functions, what are the assumptions ensuring that the vector field  $x \mapsto s(x)$  is Lipschitz continuous? This is clearly a key property for the uniqueness of (SD), and we bring a first tribute to this question below.

The next result gives a sufficient condition for the local Lipschitz property of the vector field  $x \mapsto s(x)$ , and by extension to the uniqueness question:

**Proposition 6.1.17.** Suppose that  $H$  and  $Y$  are two Euclidean spaces, and  $C = H$ . Suppose that  $F$  is Gateaux differentiable with  $DF$  being Lipschitz continuous in a neighborhood of  $u \in H$ . If  $DF(u)$  is surjective, then the steepest descent vector field  $s : v \mapsto -(\partial^c F(v))^0$  is Lipschitz continuous in a neighborhood of  $u$ .

*Proof.* We have, for any  $v$  in the neighborhood of  $u$ ,

$$(6.24) \quad s(v) = -\theta(v) \circ DF(v) \text{ where } \theta(v) := \operatorname{argmin}_{\theta \in \Theta} \frac{1}{2} \|\theta \circ DF(v)\|^2.$$

Since  $\theta \circ DF(v)$  can also be written  $D^*F(v)(\theta)$ , we obtain after writing down the first-order optimality condition in (6.24)

$$(6.25) \quad 0 \in N_B(\theta(v)) + A(v)(\theta(v)),$$

where  $A(v) := DF(v) \circ D^*F(v)$  is a positive symmetric linear operator of  $Y$ . Note that  $A$  is Lipschitz continuous from  $H$  to  $L(Y)$  in a neighborhood of  $u$ , by assumption. Since we assume that  $DF(u)$  is surjective, then  $A(u)$  is in particular definite positive. Hence, by continuity of eigenvalues on matrix entries, we can suppose that there exists some  $\alpha > 0$  such that  $\langle A(v)x, x \rangle \geq \alpha \|x\|^2$  for all  $v$  in a neighborhood of  $u$ , and all  $x \in H$ .

The end of the proof is based on a standard argument for the study of stability for the solution of a strongly monotone variational inequality. Take two elements  $v_1, v_2$ , make the difference of the corresponding equations (6.25), and take the scalar product with  $\theta(v_2) - \theta(v_1)$ . By monotonicity of the normal cone mapping  $\theta \mapsto N_B(\theta)$ , we obtain

$$(6.26) \quad \langle A(v_2)\theta(v_2) - A(v_1)\theta(v_1), \theta(v_2) - \theta(v_1) \rangle \leq 0.$$

Let us rewrite (6.26) as

$$(6.27) \quad \langle A(v_2)(\theta(v_2) - \theta(v_1)), \theta(v_2) - \theta(v_1) \rangle \leq \langle (A(v_1) - A(v_2))\theta(v_1), \theta(v_2) - \theta(v_1) \rangle.$$

By the positive definite property of  $A(v)$ , and Cauchy-Schwarz inequality, we obtain

$$(6.28) \quad \alpha \|\theta(v_2) - \theta(v_1)\|^2 \leq \|(A(v_2) - A(v_1))\theta(v_1)\| \|\theta(v_2) - \theta(v_1)\|.$$

After simplification, and by using the Lipschitz continuous dependence of  $A(v)$  with respect to  $v$ , it follows that  $v \mapsto \theta(v)$  is locally Lipschitz continuous. Then combine this result with the local Lipschitz continuity of  $DF$  to conclude. ■

If we apply directly this result to the multiobjective case  $F = (f_1, \dots, f_m)$ , we obtain the Lipschitz continuity of  $s$  around  $u \in H$  under the condition that the gradients  $\nabla f_i(u)$  are linearly independent. In fact, in this context, one can modify slightly the proof above to obtain the same result under a slightly weaker assumption, which is the *affine independence* of the gradients. We say that a family  $\{x_1, \dots, x_m\}$  in  $H$  is affinely independent whether

$$(6.29) \quad \forall \Lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m, \sum_{i=1}^m \lambda_i = 1 \text{ and } \sum_{i=1}^m \lambda_i x_i = 0 \text{ implies } \Lambda = 0.$$

Equivalently, this means that the family  $\{x_1 - x_m, \dots, x_{m-1} - x_m\}$  is linearly independent.

**Proposition 6.1.18.** Suppose that  $H$  is an Euclidean space, with  $C = H$ , and that  $(Y, K) = (\mathbb{R}^m, \mathbb{R}_+^m)$ . Suppose that for all  $i \in \{1, \dots, m\}$ ,  $f_i : H \rightarrow \mathbb{R}$  is Gateaux differentiable with  $\nabla f_i$  being Lipschitz continuous in a neighborhood of  $u \in H$ . If  $\{\nabla f_i(u)\}_{i \in \{1, \dots, m\}}$  is affinely independent, then the steepest descent vector field  $s : v \mapsto -(\partial^c F(v))^0$  is Lipschitz continuous on a neighborhood of  $u$ .

**Remark 6.1.19.** This result is quite optimal, since in the case  $m = 2$  it ensures the Lipschitz continuity of  $s$  at  $u$  under the condition that  $\nabla f_1(u) \neq \nabla f_2(u)$ . Observe that in Example 6.1.11, this fails only at  $u = (-1, 0)$ , where the Lipschitz continuity is missing.

*Proof.* This proof follows the same lines than the previous one, but we will deal with additional affine terms. Write, for any  $v$  in the neighborhood of  $u$ ,

$$(6.30) \quad s(v) = - \sum_{i=1}^m \theta_i(v) \nabla f_i(v) \text{ where } \theta(v) := \operatorname{argmin}_{\theta \in \Delta_m} \frac{1}{2} \left\| \sum_{i=1}^m \theta_i(v) \nabla f_i(v) \right\|^2.$$

Equivalently, since  $\theta_m(v) = 1 - \sum_{i=1}^{m-1} \theta_i(v)$ , we can write

$$(6.31) \quad s(v) = -\nabla f_m(v) - \sum_{i=1}^{m-1} \theta_i(v) (\nabla f_i(v) - \nabla f_m(v))$$

where

$$(6.32) \quad \theta(v) := \operatorname{argmin}_{\theta \in \mathcal{T}^{m-1}} \frac{1}{2} \left\| \nabla f_m(v) + \sum_{i=1}^{m-1} \theta_i(v) (\nabla f_i(v) - \nabla f_m(v)) \right\|^2$$

with the notation  $\mathcal{T}^{m-1} := \{\theta \in \mathbb{R}_+^{m-1} \mid \sum_{i=1}^{m-1} \theta_i \leq 1\}$ .

Define  $G := (g_i)_{i \in \{1, \dots, m-1\}} : H \rightarrow \mathbb{R}^{m-1}$ , where  $g_i := f_i - f_m$ . Then, writing down the first-order optimality condition in (6.31) leads to

$$(6.33) \quad 0 \in N_{\mathcal{T}^{m-1}}(\theta(v)) + A(v)(\theta(v)) + B(v),$$

where  $A(v) := DG(v) \circ D^*G(v)$  and  $B(v) := DG(v)(\nabla f_N(v))$ . Note that  $A(u)$  is the Gram matrix corresponding to the linearly independent family  $\{\nabla g_i(u)\}_{i \in \{1, \dots, m-1\}} = \{\nabla f_i(u) - \nabla f_m(u)\}_{i \in \{1, \dots, m-1\}}$ , so  $A(u)$  is a definite positive symmetric linear operator on  $\mathbb{R}^{m-1}$ . Since  $A(v)$  and  $B(v)$  are both Lipschitz continuous with respect to  $v$  around  $u$ , we can use the same arguments than in the previous proof to conclude. ■

**Remark 6.1.20.** In the multi-objective case, if  $H = \mathbb{R}^n$  with  $n \geq m$ , then the necessary condition of Proposition 6.1.18 is equivalent to

$$\operatorname{rank} DF(u) \leq m - 1.$$

As already observed by Smale [297], this property is generically satisfied at the critical points of  $F$ . Thus, if we try to find a counterexample to the uniqueness of trajectories, we should look for a function  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  with  $n \sim m$ , or even  $n \ll m$ .

Since the uniqueness of trajectories are not guaranteed, one could fear some wild behaviour, such that a trajectory reaching a solution/stationary point and then escaping from it. Fortunately, it happens that weak Pareto points are stable stationary points.

**Proposition 6.1.21.** Let  $u$  be a strong global solution of (SD). If  $u(T)$  is a weak efficient point for some  $T > 0$ , then  $u(t) = u(T)$  for all  $t \geq T$ .

*Proof.* Let  $t \in [T, +\infty[$  be fixed. Since  $u(T)$  is a weak efficient point, there must exist a  $\theta \in \Theta$  such that

$$(6.34) \quad f_\theta(u(t)) \geq f_\theta(u(T)).$$

But, we know from Proposition 6.1.7 that  $f_\theta$  is decreasing along the trajectory, so this, combined with (6.34), means that

$$(6.35) \quad \forall s \in [T, t], f_\theta(u(s)) = f_\theta(u(T)).$$

As a consequence, we have  $\frac{d}{ds}(f_\theta \circ u)(s) = 0$  for a.e.  $s \in [T, t]$ . Using the energy estimation in Proposition REF, we deduce that  $\dot{u}(s) = 0$  for a.e.  $s \in [T, t]$ . Hence,  $u \equiv u(T)$  on  $[T, t]$ , this being true for all  $t \geq T$ . ■

Here, we just tackled the problem of uniqueness from the point of view of the regularity of  $s$ . Another interesting aspect concerns the monotonicity-like property of  $s$ . Indeed, in the Examples 6.1.11 and 6.1.12 of Section 6.1.3, there exists locally some positive constant  $\alpha$  such that  $x \mapsto -s(x) + \alpha x$  is monotone, a property which classically implies the uniqueness (see [83, Theorem 3.17]). Thus, it is an open question, that we do not address here, to know whether this property is satisfied by the (SD) system, at least under some general assumptions.

## 6.2 Asymptotic properties of the steepest descent dynamic

We start with a very simple result concerning the convergence of the values.

**Proposition 6.2.1.** Let  $u : [0, +\infty[ \rightarrow C$  be a strong global solution of (SD). Then :

- i) For all  $\theta \in \Theta$ ,  $f_\theta(u(t)) \downarrow \inf_{t \geq 0} f_\theta(u(t))$  when  $t$  goes to  $+\infty$ .
- ii) The sublevel sets  $([F \leq F(u(t))])_{t \geq 0}$  converge, in the Painlevé - Kuratowski sense, to  $\bigcap_{t \geq 0} [F \leq F(u(t))]$ , when  $t$  goes to  $+\infty$ .

*Proof.* Item i) is a direct consequence of the monotonicity of  $t \mapsto f_\theta(u(t))$ , see Proposition 6.1.7.

Now let us prove the convergence of  $[F \leq F(u(t))]$  in the Painlevé - Kuratowski sense, see [21, Remark 12.1.2] for a definition. For short, we will note  $\Gamma_F(u(t)) := [F \leq F(u(t))]$ . We will prove successively the following inclusions:

$$(6.36) \quad \limsup_{t \rightarrow +\infty} \Gamma_F(u(t)) \subset \liminf_{t \rightarrow +\infty} \Gamma_F(u(t)) \subset \bigcap_{t \geq 0} \Gamma_F(u(t)) \subset \limsup_{t \rightarrow +\infty} \Gamma_F(u(t)).$$

Start by taking  $x \in \limsup_{t \rightarrow +\infty} \Gamma_F(u(t))$ , then there exists some  $x_n \rightarrow x$ ,  $t_n \rightarrow +\infty$ , such that  $x_n \in \Gamma_F(u(t_n))$ . Given an arbitrary sequence  $\tilde{t}_k \rightarrow +\infty$ , let us show that there exists a corresponding sequence  $\tilde{x}_k$  converging to  $x$  such that  $\tilde{x}_k \in \Gamma_F(u(\tilde{t}_k))$ . For any  $k \in \mathbb{N}$ , there exists some  $n_k \in \mathbb{N}$  such that  $\tilde{t}_k \leq t_{n_k}$ , hence by Proposition 6.1.7 we deduce that  $x_{n_k} \in \Gamma_F(u(t_{n_k})) \subset \Gamma_F(u(\tilde{t}_k))$ . So it suffices to take  $\tilde{x}_k = x_{n_k}$  to deduce that  $x \in \liminf_{t \rightarrow +\infty} \Gamma_F(u(t))$ .

Consider now  $x \in \liminf_{t \rightarrow +\infty} \Gamma_F(u(t))$ , then there exists a sequence  $x_n \rightarrow x$  such that  $x_n \in \Gamma_F(u(t_n))$  for all  $n \in \mathbb{N}$ . Given some  $t \geq 0$ , and using again Proposition 6.1.7, we see that

$x_n \in \Gamma_F(u(n)) \subset \Gamma_F(u(t))$  for all  $n \geq t$ . Hence, by passing to the limit when  $n \rightarrow +\infty$ , along with the fact that  $\Gamma_F(u(t))$  is a closed set (due to the lower semi-continuity of the functions), we deduce that  $x \in \Gamma_F(u(t))$ . This being true for an arbitrary  $t \geq 0$ , we conclude that  $x \in \bigcap_{t \geq 0} \Gamma_F(u(t))$ .

Take now  $x \in \bigcap_{t \geq 0} \Gamma_F(u(t))$ . By defining  $t_n := n$  and  $x_n := x$ , we trivially have that  $x_n \in \Gamma_F(u(t_n))$  for all  $n \in \mathbb{N}$ , so  $x \in \limsup_{t \rightarrow +\infty} \Gamma_F(u(t))$  follows. ■

**Remark 6.2.2.** Note that this property of convergence for the sublevel sets holds more generally for any decreasing sequence of sets, in the sense of the inclusion.

We follow now with a general asymptotic property for the iterates which guarantees that the strong limit points of the trajectory are critical Pareto points. It is a direct consequence of the dissipative property of the dynamic, and the weak-strong outer semi-continuity of the operator  $x \rightrightarrows N_C(x) + \partial^c F(x)$ , studied in Section 5.1 (see Propositions 5.1.11 and 5.1.13). Recall that we introduced in Section 5.1 the following set of hypotheses (here  $H$  is a Hilbert space, thus reflexive):

One of the three following properties is satisfied:

- H**
- 1)  $K$  is polyhedral,
  - 2)  $Y$  has finite dimension, and  $F$  is positively Clarke regular,
  - 3)  $F$  is the sum of a convex function and a strictly Gateaux differentiable function.

**Proposition 6.2.3.** Suppose that one of the cost functions  $f_\theta$  remains bounded from below along the trajectory.

i) The trajectory has a finite energy:

$$(6.37) \quad \int_0^{+\infty} \|\dot{u}(t)\|^2 dt < +\infty.$$

ii) If **H** holds, then any strong limit point of the trajectory is Pareto critical.

iii) If the functions are convex, any weak limit point of the trajectory is a weak Pareto.

*Proof.* Let  $f_\theta$  be the cost function such that  $\inf_{t \geq 0} f_\theta(x(t)) > -\infty$ . From Proposition 6.1.7 and by integrating (6.7), we obtain

$$(6.38) \quad \int_0^{+\infty} \|\dot{u}(t)\|^2 dt \leq f_\theta(x(0)) - \inf_{t \geq 0} f_\theta(x(t)).$$

This proves item i).

Let  $u_\infty$  be a strong limit point of the trajectory, i.e.  $u(t_n) \rightarrow u_\infty$  strongly in  $H$  for some  $t_n \rightarrow +\infty$ . The finite energy property (6.38) implies that

$$(6.39) \quad \liminf_{t \rightarrow +\infty} \|\dot{u}(t)\| = 0.$$

Since relations (6.38) and (6.39) are satisfied for almost all  $t > 0$ , we can suppose, for some  $t_n \rightarrow +\infty$ , that:

$$(6.40) \quad \begin{aligned} u(t_n) &\rightarrow x_\infty \text{ strongly in } H, \\ \dot{u}(t_n) &\rightarrow 0 \text{ strongly in } H, \\ -\dot{u}(t_n) &\in N_C(u(t_n)) + \partial^c F(u(t_n)) \text{ for each } n \in \mathbb{N}. \end{aligned}$$

From Corollary 5.1.14 we know that  $u \rightrightarrows N_C(u) + \partial^c F(u)$  is sequentially upper semi-continuous, so point ii) follows. Point iii) is proved by the same argument, using the weak-strong closure instead of the strong-weak closure (cf. Proposition 5.1.11), and Fermat's rule (Theorem 5.2.10) for convex functions. ■

We are now going to prove a global convergence result for the iterates, under the assumption that the merit functions are quasi-convex. (Our argument is in the line of the proof of convergence of the steepest descent by Goudou and Munier [180] in the case of a single (quasi-convex) objective function.) Recall from Chapter 2, Section 2.3, that  $F$  is said to be scalarly quasiconvex whenever  $f_\theta$  is quasiconvex, for all  $\theta \in \Theta$ . It implies in particular that  $F$  is quasiconvex. In the following, we will make use of the following limit sets:

$$(6.41) \quad \text{the set of weak limit points } \Omega[u(t)] := \{u \in \mathbf{H} \mid \exists t_n \rightarrow +\infty \text{ s.t. } u(t_n) \xrightarrow{w} u\},$$

$$(6.42) \quad \text{the common sublevel limit set } \Gamma_F[u(t)] := \bigcap_{t \geq 0} [F \leq F(u(t))].$$

Observe that  $\Gamma_F[u(t)]$  is nothing but  $\{v \in \mathbf{H} \mid f_\theta(v) \leq f_\theta(u(t)) \forall t \geq 0, \forall \theta \in \Theta\}$ , and is closed convex in the quasiconvex setting.

**Proposition 6.2.4.** Let  $F$  be scalarly quasiconvex, then  $\Omega[x(t)] \subset C \cap \Gamma_F[x(t)]$ .

*Proof.* Suppose that  $u(t_n) \xrightarrow{w} u_\infty$  for some  $t_n \rightarrow +\infty$ . Since the trajectory remains in  $C$  which is closed convex, then  $u_\infty \in C$ . Moreover, recall that the quasiconvex functions are weakly lower semi-continuous, so for all  $\theta \in \Theta$ :

$$f_\theta(u_\infty) \leq \liminf_{n \rightarrow +\infty} f_\theta(u(t_n)) = \inf_{t \geq 0} f_\theta(u(t)).$$

Whence  $u_\infty \in \Gamma_F[u(t)]$ . ■

**Proposition 6.2.5.** (Fejer property)

Let  $F$  be scalarly quasiconvex, then for all  $z \in C \cap \Gamma_F[u(t)]$ ,  $t \mapsto \|u(t) - z\|$  is decreasing.

*Proof.* Let  $z \in C \cap \Gamma_F[u(t)]$ , and define  $h(t) := \frac{1}{2}\|u(t) - z\|^2$ . Since the trajectory  $u(\cdot)$  is absolutely continuous, so is  $h$ , and we can derive it for almost every  $t > 0$ :

$$\begin{aligned} h'(t) &= \langle \dot{u}(t), u(t) - z \rangle = \langle -s(u(t)), z - u(t) \rangle \\ &= \langle \eta(u(t)), z - u(t) \rangle + \langle p(u(t)), z - u(t) \rangle, \end{aligned}$$

where  $\eta(u(t)) \in N_C(u(t))$ ,  $p(u(t)) \in \partial^c F(u(t))$ . The fact that  $z \in C$ , implies immediately from the definition of  $N_C$  that  $\langle \eta(u(t)), z - u(t) \rangle \leq 0$ . Using the definition of the Clarke directional derivative, one obtains:

$$h'(t) \leq d^c F(u(t), z - u(t)).$$

Since  $z \in \Gamma_F[u(t)]$ , we know that  $f_\theta(z) \leq f_\theta(u(t))$  for all  $\theta \in \Theta$  and all  $t \geq 0$ . Let  $t > 0$  be fixed, and suppose on one hand that there is some  $\theta \in \Theta$  such that  $f_\theta(u(t)) = f_\theta(z)$ . Then the trajectory stops there. Indeed, by decrease property (Proposition 6.1.7) one has  $f_\theta(u(s)) = f_\theta(u(t))$  for all  $s \geq t$ . Using (6.7) we see that  $\dot{u}(s) = 0$  for a.e.  $s \geq t$ . Since  $u(\cdot)$  is absolutely continuous, we deduce that  $u(s) = u(t)$  for all  $s \geq t$ . In that case  $h$  is constant on  $[t; +\infty[$ . On the other hand, if  $f_\theta(z) < f_\theta(u(t))$  for all  $\theta \in \Theta$ , then the quasiconvexity and Proposition 2.2.16 tell us that  $d^c f_\theta(u(t), z - u(t)) \leq 0$ . In other words,  $d^c F(u(t), z - u(t)) \leq 0$  and so  $h'(t) \leq 0$ . Both cases lead to the fact that  $h$  is decreasing along time. ■

We can now state our main convergence result.

**Theorem 6.2.6.** Let  $F$  be scalarly quasiconvex. Then the trajectory is bounded if and only if  $C \cap \Gamma_F[u(t)] \neq \emptyset$ . In that case,

- i) the trajectory weakly converges to a point in  $C \cap \Gamma_F[u(t)]$ ,
- ii) the trajectory has finite energy :  $\int_0^{+\infty} \|\dot{u}(t)\|^2 dt < +\infty$ .

Under additional hypotheses, we have the stronger results :

- (iii) if  $F$  is convex, then the weak limit point is a weak efficient point.
- (iv) if the trajectory is precompact and **H** holds, then the trajectory converges strongly to a critical point.

The theorem is a direct corollary of Propositions 6.2.3, 6.2.4 and 6.2.5, and Opial's lemma (see [268, Lemma 5.2])

**Lemma 6.2.7** (Opial). Let  $S$  be a non empty subset of  $H$ , and  $x : [0, +\infty[ \rightarrow H$  a map. Assume that

- (i) for every  $z \in S$ ,  $\lim_{t \rightarrow +\infty} \|x(t) - z\|$  exists;
- (ii)  $\Omega[x(t)] \subset S$ .

Then  $x(t)$  weakly converges to some  $x_\infty \in S$ , when  $t \rightarrow +\infty$ .

*Proof of Theorem 6.2.6.* If the trajectory is bounded, then it admits a weakly convergent subsequence. In other words,  $\Omega[u(t)] \neq \emptyset$ , and with Proposition 6.2.4 it follows that  $C \cap \Gamma_F[u(t)] \neq \emptyset$ . If  $C \cap \Gamma_F[u(t)] \neq \emptyset$ , then using Opial's Lemma together with Propositions 6.2.4 and 6.2.5, gives the weak convergence of the trajectory. In that case one sees that the trajectory is in particular bounded. Items ii-iv) follow Proposition 6.2.3, using the fact that if the trajectory converges to  $u_\infty \in \Gamma_F[u(t)]$ , then each cost function  $f_\theta$  is minimized along the trajectory by  $f_\theta(u_\infty)$ . ■

**Remark 6.2.8.** a) Since each function  $t \mapsto f_\theta(u(t))$  is nonincreasing (see Proposition 6.1.7), a natural condition insuring that the trajectory remains bounded, is that one of the functions has bounded sublevel sets. b) Similarly, if one of the merit functions has relatively compact sublevel sets (inf-compactness property), then the trajectory is relatively compact, which is needed for the strong convergence in  $H$ .

As in the convex monocriteria case, strong convergence can be obtained by doing a symmetry hypothesis on the involved functions. We say that  $F$  is even if  $F(x) = F(-x)$  for all  $x \in H$ , and that  $C$  is symmetric whether  $C = -C$ .

**Theorem 6.2.9.** Suppose that  $F$  is convex and even, and that  $C$  is symmetric. Hence the trajectory strongly converges to a weak efficient point.

*Proof.* Let  $s \geq 0$  and define, for  $t \in [0, s]$  :

$$\gamma(t) := \|u(t)\|^2 - \|u(s)\|^2 - \frac{1}{2}\|u(t) - u(s)\|^2.$$

We will show that  $\gamma$  is decreasing and then derive a Cauchy property for the trajectory. Since  $u$ ,  $\gamma$  is differentiable almost everywhere on  $[0, s]$ , then for a.e  $t \in [0, s]$  :

$$\begin{aligned} (6.43) \quad \dot{\gamma}(t) &= 2\langle \dot{u}(t), u(t) \rangle - \langle \dot{u}(t), u(t) - u(s) \rangle = \langle -\dot{u}(t), -u(s) - u(t) \rangle \\ &= \langle \eta(u(t)), -u(s) - u(t) \rangle + \langle p(u(t)), -u(s) - u(t) \rangle \\ &\leq \langle \eta(u(t)), -u(s) - u(t) \rangle + d^c F(u(t), -u(s) - u(t)) \end{aligned}$$

where  $\eta(u(t)) \in N_C(u(t))$  and  $p(u(t)) \in \partial^c F(u(t))$ .

Using the symmetry of  $C$ , we see that  $u(s) \in C \Rightarrow -u(s) \in C$ , so combined with  $\eta(u(t)) \in N_C(u(t))$  we obtain

$$(6.44) \quad \langle \eta(u(t)), -u(s) - u(t) \rangle \leq 0.$$

Now, thanks to the decrease property (Proposition 6.1.7) and the symmetry, we have

$$(6.45) \quad \text{for all } \theta \in \Theta, f_\theta(-u(s)) = f_\theta(u(s)) \leq f_\theta(u(t)).$$

Since the merit functions are supposed to be convex, we can use Proposition 2.2.15 to deduce from (6.45) that  $d^c F(u(t), -u(s) - u(t)) \leq 0$ . By combining (6.43) with (6.44), we obtain  $\dot{\gamma}(t) \leq 0$  for a.e.  $t \in [0, s]$ , so  $\gamma(\cdot)$  is decreasing on  $[0, s]$ .

The decrease property of  $\gamma$  indicates in particular that  $\gamma(t) \geq \gamma(s) = 0$  for all  $t \in [0, s]$ . In other words :

$$(6.46) \quad \text{for all } 0 \leq t \leq s, \frac{1}{2} \|u(t) - u(s)\|^2 \leq \|u(t)\|^2 - \|u(s)\|^2.$$

A first observation on (6.46) is that  $0 \leq \|u(s)\|^2 \leq \|u(t)\|^2$ . Being decreasing and positive, we see that  $\|u(\cdot)\|$  converges in  $\mathbb{R}$ , and so it is Cauchy. As a consequence, by going back on (6.46), we see that the trajectory  $u(\cdot)$  is Cauchy, whence strongly convergent in  $H$ . We know that the limit point is weakly efficient using Theorem 6.2.6. ■

## 6.3 The discrete steepest descent method, and numerical results

### 6.3.1 Review of the existing algorithms, and links with (SD)

We discuss the discretization(s) in time of the continuous dynamic

$$(SD) \quad \dot{u}(t) + \left( N_C(u(t)) + \partial^c F(u(t)) \right)^0 = 0,$$

and relate it to the existing literature.

We start by assuming that  $F$  is of class  $C^1$ , and that there is no constraint ( $C = H$ ). In that case, a naive discretization of (SD) gives

$$(6.47) \quad x_{n+1} = x_n + \lambda_n s(x_n),$$

where  $\lambda_n$  is a real stepsize, and  $s(x_n)$  is the element of minimal norm of  $\partial^c F(x_n)$ . It can be useful to recall that the steepest descent direction has also the equivalent form (see Theorem 5.2.20 and Proposition 5.1.8):

$$(6.48) \quad s(x_n) = \operatorname{argmin}_{d \in X} \frac{1}{2} \|d\|^2 + \sigma_\theta(DF(x; d)).$$

Since the steepest descent direction enjoys an Armijo rule (see Propositions 5.2.18 and 5.2.9), it is natural to chose  $\lambda_n$  as:

$$\lambda_n := \max \left\{ \frac{1}{2^k} \mid k \in \mathbb{N}, F(x_n + \frac{1}{2^k} d) \leq F(x_n) + \beta \frac{1}{2^k} d^c F(x_n; d) \right\}.$$

As usual, if  $DF$  is  $L$ -Lipschitz continuous at  $x_n$ , then it is sufficient to take  $\lambda_n < \frac{1-\beta}{L}$ .

This algorithm (6.47) has been first introduced by Fliege and Svaiter [159], for the multi-objective case, and extended by Graña and Svaiter [176] to the general vector case (see also [122, 60]). According to these works, this discrete dynamic shares exactly the same behaviour with our (SD) dynamic. That is, like in Theorem 6.2.6, this algorithm weakly converges in the quasiconvex case, and the limit point is a weak efficient point in the convex case. Still like in Theorem 6.2.6, the convergence of the iterates  $(x_n)_{n \in \mathbb{N}}$  is conditioned by the boundedness hypothesis

$$\{x \in \mathbf{H} \mid F(x) \leq F(x_n), \forall n \in \mathbb{N}\} \neq \emptyset.$$

If we consider the presence of a convex constraint  $C$ , the discretization of (SD) becomes more delicate. Of course, we could directly write

$$x_{n+1} = x_n + \lambda_n s(x_n) = x_n + \lambda_n \left( N_C(x_n) + \partial^C F(x_n) \right)^0,$$

but this would face a serious drawback. Indeed, we know in the constrained case that the steepest descent direction  $s(x_n)$  is an *admissible* descent direction, *i.e.*  $s(x_n) \in T_C(x_n)$ . But while we perform an Armijo step, what we really need is  $s(x_n) \in T_C^{ad}(x_n)$ , that is,

$$s(x_n) \in \frac{C - x_n}{\mu}, \forall \mu \sim 0.$$

The problem is that  $T_C(x_n)$  can be strictly bigger than  $T_C^{ad}(x_n)$ , and that following such a direction could make us getting out from the constraint. Of course, there is some cases for which the tangent cone of admissible descent directions is closed (e.g. the locally polyhedral sets, see [239]), but an other method must be considered in the general case.

This problem already appears in the scalar case ( $f : \mathbf{H} \rightarrow \mathbb{R}$ ), for which we generally perform a gradient-projection step:

$$x_{n+1} = \text{proj}_C(x_n - \lambda_n \nabla f(x_n)).$$

Thus, a natural approach would be, in the vector case, to perform first a steepest descent step to decrease  $F$ , and then project onto the constraint  $C$ :

$$(6.49) \quad x_{n+1} = \text{proj}_C(x_n + \lambda_n d_n), \text{ where } d_n = (\partial^C F(x_n))^0.$$

But, unfortunately, this approach does not work anymore in the vector case, because we loose the descent property of  $F$ . The least we can say is that, at each step, one of the cost functions decreases (do the parallel with Section 6.1.4). But this is not true in general for all the cost functions, and hence for  $F$  (see a simple counter-example below).

**Proposition 6.3.1.** Let  $(x_n)_{n \in \mathbb{N}}$  be a sequence generated by the algorithm (6.49). Suppose that  $F$  is continuously differentiable, with a Lipschitz continuous derivative. Let  $L := \sup_{\theta \in \Theta} \text{Lip}(\nabla f_\theta; x_n)$ , and suppose that  $\lambda_n \in ]0, \frac{1}{L}[$ . Then, for all  $n \in \mathbb{N}$ , there exists  $\theta_n \in \Theta$  such that

$$(6.50) \quad \left( \frac{1}{2\lambda_n} - \frac{L}{2} \right) \|x_{n+1} - x_n\|^2 \leq f_{\theta_n}(x_n) - f_{\theta_n}(x_{n+1}).$$

*Proof.* By definition of the projection, and since  $x_n \in C$ , we have

$$(6.51) \quad \frac{1}{2} \|x_{n+1} - x_n - \lambda_n d_n\|^2 \leq \frac{1}{2} \|\lambda_n d_n\|^2,$$



which can be rewritten as

$$(6.52) \quad \frac{1}{2t_n} \|x_{n+1} - x_n - \lambda_n d_n\|^2 \leq \langle x_{n+1} - x_n, d_n \rangle.$$

Because of Proposition 5.1.8,  $d_n$  writes as  $\theta_n \circ DF(x_n) = \nabla f_{\theta_n}(x_n)$  for some  $\theta_n \in \Theta$ . The Lipschitz continuity of  $\nabla f_{\theta_n}$  gives

$$(6.53) \quad \langle x_{n+1} - x_n, -\nabla f_{\theta_n}(x_n) \rangle \leq f_{\theta_n}(x_n) - f_{\theta_n}(x_{n+1}) + \frac{L}{2} \|x_{n+1} - x_n\|^2.$$

We obtain from (6.52) and (6.53) that

$$(6.54) \quad \left( \frac{1}{2\lambda_n} - \frac{L}{2} \right) \|x_{n+1} - x_n\|^2 \leq f_{\theta_n}(x_n) - f_{\theta_n}(x_{n+1}).$$

■

The following example shows that we cannot expect an individual decrease property for all the functions.

**Example 6.3.2.** We build our counter-example in the bi-objective case. Consider two quadratic functions, distances to two points:  $f_1(x) = \frac{1}{2}\|x - a\|^2$  and  $\frac{1}{2}\|x - b\|^2$  for two distinct points  $a \neq b$ . They are smooth functions, with 1-Lipschitz gradient. For now we do not precise what is  $C$ .

Suppose that at our current iterate, we have a point  $x_n$  such that  $\|x_n - a\| = \|x_n - b\|$ . At this point,  $s(x_n)$  is exactly  $\frac{a+b}{2} - x_n$ . Then, whatever can be  $C$ , the algorithm reads, for some  $\lambda_n \in ]0, 1[$ ,

$$x_{n+1} = \text{proj}_C \left( (1 - \lambda_n)x_n + \lambda_n \frac{a + b}{2} \right).$$

Introduce now the constraint  $C$ : consider at  $x_n$  the level set of the function  $f_1$ , which is exactly the ball  $B(a, \|x_n - a\|)$ . Consider the hyperplane tangent to  $B(a, \|x_n - a\|)$  at  $x_n$ , at take  $C$  as the closed half space delimited by this hyperplane, which does *not* contain  $B(a, \|x_n - a\|)$ :

$$(6.55) \quad C = \{x \in H \mid \langle x - x_n, a - x_n \rangle \leq 0\}.$$

One can easily see that for any choice of  $\lambda_n \in ]0, 1[$ , the point  $x_{n+1}$  will fall outside  $B(a, \|x_n - a\|)$ .

Figure 6.8: Failure of the steepest descent-projection method for bi-objective optimization problems.

The problem here is that, from the point of view of  $f_1$ , we are performing an inexact projected-gradient method, where the gradient of  $f_1$  is replaced by a more general descent direction, and this cannot work.

An other approach to handle the constraint is to use an approximation of  $s(x_n)$  based on its variational form

$$s(x_n) = \underset{d \in T_C(x_n)}{\text{argmin}} \frac{1}{2} \|d\|^2 + \sigma_\theta(DF(x; d)).$$

Approximating  $T_C(x_n)$  by  $\frac{C - x_n}{\mu_n}$  for some  $\mu_n \in$ , we define

$$d_n = \underset{d \in \frac{C - x_n}{\mu_n}}{\text{argmin}} \frac{1}{2} \|d\|^2 + \sigma_\theta(DF(x; d)).$$

It is not a difficult exercise to verify that such  $d_n$  is an admissible Armijo direction, so it makes sense to use it to design an algorithm. This approach has been developed by Graña and Iusem [173], and followed by Fukuda and Graña [164, 165], Bello, Lucambio and Melo [63]. As for the unconstrained case, this algorithm shares the same asymptotic behaviour than (SD).

In the last fifteen years, there has been other algorithms proposed to manage the constraint. Fliege and Svaiter [159] propose a method to deal with inequality constraints. Recchioni [282], together with Miglierina, Molho and Recchioni [248], consider a box constraint that they penalize with a barrier method. Villacorta, Oliveira and Soubeyran [313] study also an interior point method, by adapting the trust-region methods to the multi-objective case. Bento, Ferreira, Oliveira [62] and then Bento and Cruz [59] adopted an approach based on Riemannian manifolds.

Suppose now that  $F$  is convex but nonsmooth, without constraint. A common way to discretize (SD) is to do it implicitly with respect to  $\partial^c F$ :

$$x_{n+1} + \lambda_n \partial^c F(x_{n+1})^0 = x_n.$$

But this nonlinear implicit equation in  $x_{n+1}$  seems to be too difficult to be solved. We can relax this equation by taking general elements in  $\partial^c F(x_{n+1})$ , instead of taking specifically its element of minimal norm (recall (VDI) in Section 6.1.4):

$$(6.56) \quad x_{n+1} + \lambda_n \partial^c F(x_{n+1}) \ni x_n.$$

Introducing a formal resolvent for  $\partial^c F$ , we see that the above equation can be equivalently rewritten as

$$x_{n+1} \in \text{PROX}_{\lambda_n F}(x_n), \quad \text{where } \text{PROX}_{\lambda F} := (Id + \lambda \partial^c F)^{-1}.$$

Since  $F$  is assumed to be convex, we have (Proposition 5.1.8) that  $\partial^c F(x) = \bigcup_{\theta \in \Theta} \partial^c f_\theta(x)$ , so we can deduce that

$$\text{PROX}_{\lambda_n F}(x_n) = \bigcup_{\theta \in \Theta} \text{prox}_{\lambda_n f_\theta}(x_n),$$

where (using that  $\langle \theta, e \rangle = 1$ )

$$\text{prox}_{\lambda_n f_\theta}(x_n) = \underset{x \in H}{\text{argmin}} \quad \theta \circ F(x) + \frac{1}{2\lambda_n} \|x - x_n\|^2 = \underset{x \in H}{\text{argmin}} \quad \langle \theta, F(x) + \frac{1}{2\lambda_n} \|x - x_n\|^2 e \rangle.$$

Using the classical representation of weak efficient points for convex functions (Theorem 5.2.10), we see that the algorithm in (6.56) is equivalent to

$$(6.57) \quad x_{n+1} \in \underset{x \in H}{\text{ARGMIN}_w} F(x) + \frac{1}{2\lambda_n} \|x - x_n\|^2 e.$$

Since the algorithm in (6.57) is based on the vector differential inclusion

$$(VDI) \quad \dot{u}(t) + \partial^c F(u(t)) \ni 0,$$

instead of the steepest descent one, there is no guarantee for  $F$  to decrease at each iteration. In their seminal paper, Bonnel, Iusem and Svaiter [73] propose a modified version of (6.57), enforcing the decrease property by adding a sublevel constraint:

$$x_{n+1} \in \underset{x \in [F \leq F(x_n)]}{\text{ARGMIN}_w} F(x) + \frac{1}{2\lambda_n} \|x - x_n\|^2 e.$$

This proximal method for vector optimization problems has also been studied by Ceng, Yao [99] and Ceng, Mordukhovich, Yao [98]. Villacorta and Oliveira [312] consider a modified version of

this algorithm, involving Bregman distances to handle a convex constraint  $C$ . Bento, Cruz and Soubeyran [61] propose a particular instance of this proximal algorithm, defined by

$$x_{n+1} \in \operatorname{argmin}_{x \in H} \sigma_{\Theta} \left( F(x) + \frac{1}{2\lambda_n} \|x - x_n\|^2 e + \delta_{[F \leq F(x_n)]}(x) e \right).$$

But a quick computation, involving Proposition A.3.6, shows that it is in fact just a proximal method associated to the scalar-valued function  $\sigma_{\Theta} \circ F$ , with a moving constraint  $[F \leq F(\cdot)]$ .

None of the aforementioned proximal methods can be seen as a direct discretization of the (SD) dynamic, because of the constraint  $[F \leq F(\cdot)]$ . It is, to our opinion, more related to the sweeping process

$$\dot{u}(t) + N_{[F \leq F(u(t))]}(u(t)) + \partial^C F(u(t)) \ni 0.$$

### 6.3.2 The effective computation of the steepest descent direction

Given a point  $x \in H$ , Theorem 5.2.20 provides a formula for the computation of the steepest descent at  $x$ :

$$\min_{d \in H} \frac{1}{2} \|d\|^2 + d^C F(x; d).$$

Following [159], this problem is equivalent to the following quadratic problem with convex constraint:

$$\min_{(d, \alpha) \in H \times \mathbb{R}} \frac{1}{2} \|d\|^2 + \alpha \text{ such that } d^C F(x; d) \leq \alpha.$$

If  $F$  is strictly Gateaux differentiable, it is also equivalent to the quadratic program with cone constraints

$$\min_{(d, \alpha) \in H \times \mathbb{R}} \frac{1}{2} \|d\|^2 + \alpha \text{ such that } DF(x; d) \leq \alpha e.$$

For instance, if  $(Y, K) = (\mathcal{S}^m, \mathcal{S}_+^m)$ , the space of symmetric matrices ordered by the cone of positive symmetric matrices, it reduces to a SDP program:

$$\min_{(d, \alpha) \in H \times \mathbb{R}} \frac{1}{2} \|d\|^2 + \alpha \text{ such that } DF(x; d) \leq \alpha Id.$$

In the multi-objective case, it reduces to a quadratic program with linear constraints

$$\min_{(d, \alpha) \in H \times \mathbb{R}} \frac{1}{2} \|d\|^2 + \alpha \text{ such that } \langle \nabla f_i(x), d \rangle \leq \alpha \text{ for all } i \in \{1, \dots, m\}.$$

Theorem 5.2.20 provides also a dual version of this problem, involving the linear forms  $\theta \in \Theta$ :

$$\max_{y^* \in Y^*} -\|D^* F(x; y^*)\|^2 \text{ such that } y^* \in \Theta,$$

which is equivalent to

$$\min_{\theta \in \Theta} \|\theta \circ DF(x)\|^2.$$

In its general form, it is a quadratic program, with linear and cone constraints since

$$\theta \in \Theta \Leftrightarrow \theta \in K^+ \text{ and } \langle \theta, e \rangle = 1.$$

In the multi-objective case  $F = (f_1, \dots, f_m)$ , this dual problem reduces to the projection onto the polyhedron  $\operatorname{co}\{\nabla f_1(x), \dots, \nabla f_m(x)\}$ . It is particularly easy to implement when  $m$  is low, for instance  $m = 2$  or  $3$ .

In the nonsmooth case, the computation of  $s(x)$  can be challenging. We propose a naive approach to compute it, in the case of a bi-criteria optimization problem  $F = (f, g) : \mathbb{R}^n \rightarrow$

$(\mathbb{R}^2, \mathbb{R}_+^2)$ . We assume that  $f$  is a smooth function, but that  $g$  is nonnecessarily differentiable. In that setting, we need to be able to project the origin onto  $\text{co}\{\nabla f(x), \partial^C g(x)\}$ . We aim to show that this problem can be solved quickly, provided that the projection onto  $\partial^C g(x)$  is easy. For instance, the subdifferential of the  $\ell^1$  norm falls into this setting.

The problem we need to solve here is

$$\min_{\lambda \in \mathbb{R}} \inf_{x^* \in \partial^C g(x)} \|(1 - \lambda)\nabla f(x) + \lambda x^*\|, \text{ such that } \lambda \in [0, 1].$$

This is equivalent to solve

$$\min_{\lambda \in [0, 1]} \|\text{proj}_{C_\lambda}(0)\|,$$

where  $C_\lambda := (1 - \lambda)\nabla f(x) + \lambda\partial^C g(x)$ . In other words, we need to minimize real-valued function

$$\begin{aligned} \phi : [0, 1] &\longrightarrow \mathbb{R} \\ \lambda &\longmapsto \|\text{proj}_{(1-\lambda)\nabla f(x) + \lambda\partial^C g(x)}(0)\|. \end{aligned}$$

The main trick is that this function is convex:

**Lemma 6.3.3.** The function  $\phi$  defined above is convex.

*Proof.* Let  $\lambda_1, \lambda_2 \in [0, 1]$  and  $\alpha \in [0, 1]$ . Define  $\lambda := \alpha\lambda_1 + (1 - \alpha)\lambda_2$ , and show that

$$\phi(\lambda) \leq \alpha\phi(\lambda_1) + (1 - \alpha)\phi(\lambda_2).$$

Using basic algebra on sets, we can write

$$(6.58) \quad \alpha C_{\lambda_1} + (1 - \alpha)C_{\lambda_2} = C_\lambda.$$

Note  $p_1 := \text{proj}_{C_{\lambda_1}}(0)$ ,  $p_2 := \text{proj}_{C_{\lambda_2}}(0)$ , and  $p := \alpha p_1 + (1 - \alpha)p_2$ . From (6.58), we have  $p \in C_\lambda$ . Thus, using the triangle inequality:

$$\phi(\lambda) \leq \|p\| \leq \alpha\|p_1\| + (1 - \alpha)\|p_2\| = \alpha\phi(\lambda_1) + (1 - \alpha)\phi(\lambda_2).$$

■

So, we can apply a golden section method to  $\phi$ . It finds (one of) its minimum  $\bar{\lambda}$  by dichotomy. This method only asks to compute the values of  $\phi$ , which is computationally equivalent to compute a projection on  $\partial^C g(x)$ , since

$$\text{proj}_{(1-\lambda)\nabla f(x) + \lambda\partial^C g(x)}(0) = \text{proj}_{\lambda\partial^C g(x)}(-(1 - \lambda)\nabla f(x)) + (1 - \lambda)\nabla f(x).$$

Once  $\bar{\lambda}$  is found, it suffices to take  $s(x) = \text{proj}_{(1-\bar{\lambda})\nabla f(x) + \bar{\lambda}\partial^C g(x)}(0)$ .

### 6.3.3 Generation of Pareto fronts

**Example 6.3.4.** Let  $F = (f_1, f_2) : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$  be defined for all  $x = (x_1, x_2)$  by

$$f_1(x) = \|x\|_1 \text{ and } f_2(x) = \frac{1}{x_1} + \|x\|_2^2 + 3e^{-100(x_1-0.3)^2} + 3e^{-100(x_1-0.6)^2},$$

where  $\|\cdot\|_1$  and  $\|\cdot\|_2$  denote respectively the usual  $\ell^1$  and  $\ell^2$ -norms on  $\mathbb{R}^2$ . This example is taken from [284], see also [248, Test 1].

We take a constraint  $C = [0.1, 1]^2$ , on which  $F$  is well-defined and locally Lipschitz continuous. Our purpose is to generate the Pareto front, that is, the minimal elements of  $F(C)$  in  $(\mathbb{R}^2, \leq)$ . Next figure shows an approximation of  $F(C)$ , and the Pareto front we try to recover. Due to the oscillations of  $f_2$  around the values  $x_1 = 0.3, 0.6$ , the image  $F(C)$  is not convex, and we see that the Pareto front is not connected.

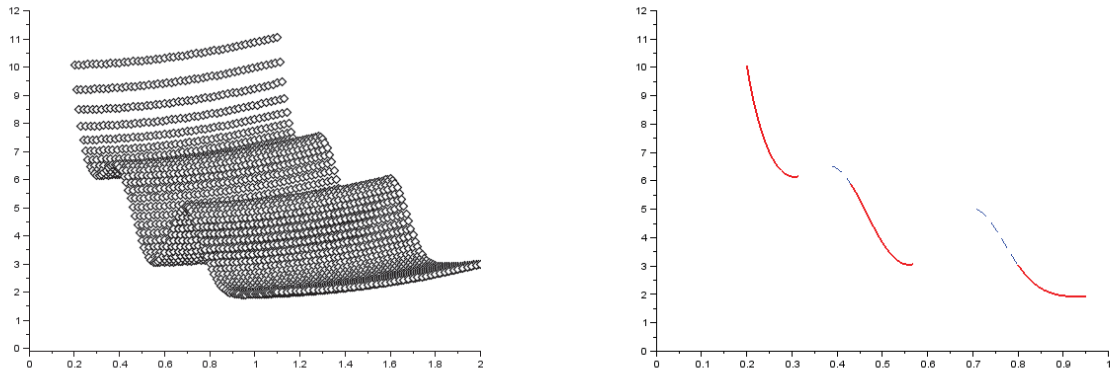


Figure 6.9: Left: approximation of  $F(C)$ . Right: the Pareto points are plotted in thick continuous red lines, and the local Pareto points are in thin dotted blue lines.

To compute the Pareto front, we compare two methods: the weighting method, and the steepest descent method. Let us give in detail our protocol. The weighting method consists in the choice of some  $\theta \in \Delta_2 = \{(\theta_1, 1 - \theta_1) \mid \theta_1 \in [0, 1]\}$ , and the minimization of the corresponding convex combination  $f_\theta = \theta_1 f_1 + (1 - \theta_1) f_2$ . Its minimization will be performed by a gradient projected algorithm

$$x_{k+1} = \text{proj}_C(x_k - \lambda \nabla f_\theta(x_k)), \text{ where } \lambda = 0.05.$$

The stopping criterion is  $\|\frac{x_{k+1} - x_k}{\lambda}\| < 0.05$ . In Figure 6.10, we show some samples of the weighting method. The starting point  $x_0$  and the combination  $\theta \in \Delta_2$  are either chosen randomly, or chosen on a uniformly distributed grid. We plot the value in  $Y = \mathbb{R}^2$  of the last iterate of each trajectory.

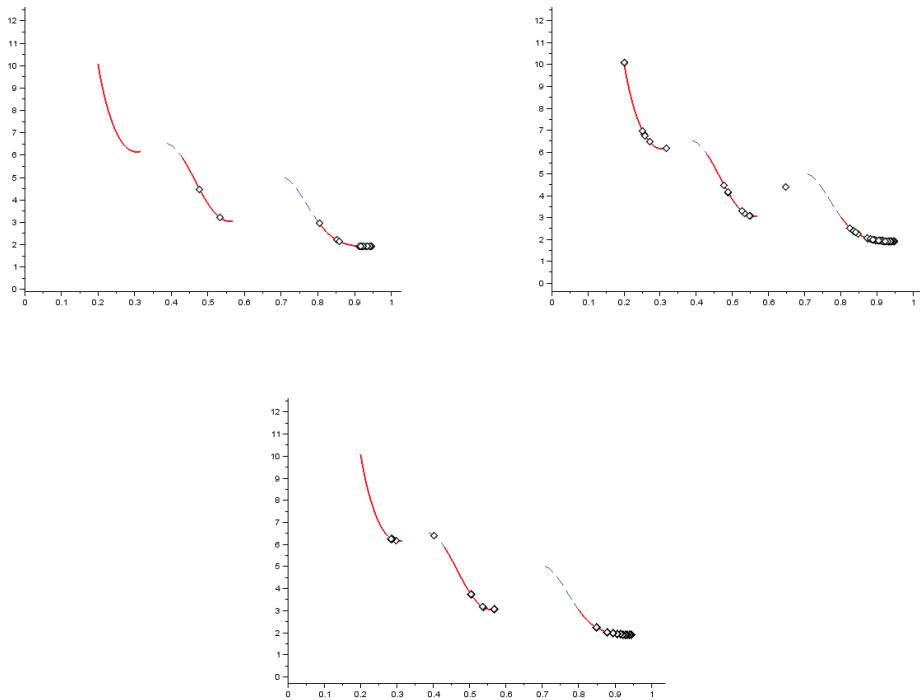


Figure 6.10: Top Left: 20 points with initial data chosen randomly. Top Right: 100 points with initial data chosen randomly. Bottom: 4000 points with initial data chosen on a uniform grid.

We observe that the points generated by the weighting method tends to be attracted in the zones where  $f_2$  has low values. We now pass to the steepest descent method. We use a basic discretized version of (SD), namely

$$x_{k+1} = x_k + \lambda s(x_k), \text{ where } \lambda = 0.05,$$

and the stopping criterion is again  $\| \frac{x_{k+1} - x_k}{\lambda} \| < 0.05$ . The results concerning the steepest descent method are given in Figure 6.11.

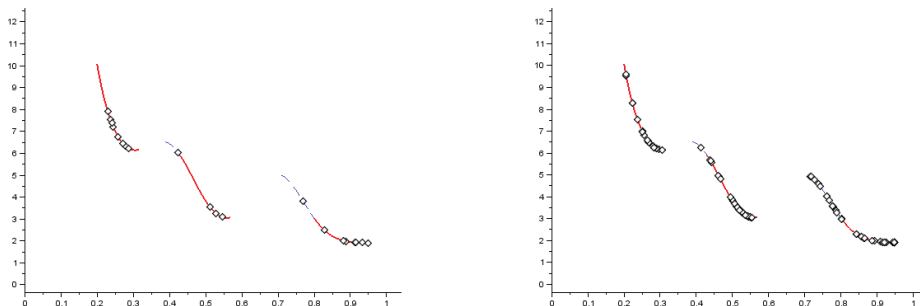


Figure 6.11: Left: 20 points with initial data chosen randomly. Right: 100 points with initial data chosen randomly.

We observe that the steepest descent approach tends to cover the Pareto front more uniformly.

**Example 6.3.5.** Let  $F = (f_1, f_2) : \mathbb{R}^n \rightarrow \mathbb{R}^2$  be defined for all  $x = (x_1, \dots, x_n)$  by

$$f_1(x) = \left( \sum_{i=1}^n x_i^2 - 10 \cos(2\pi x_i) + 10 \right)^{\frac{1}{4}} \quad \text{and} \quad \left( \sum_{i=1}^n (x_i - 1.5)^2 - 10 \cos(2\pi(x_i - 1.5)) + 10 \right)^{\frac{1}{4}},$$

and consider the constraint  $C = [-0.5, 2]^n$ . This example is adapted from [248, Test 5], see also [52, Section 6.3]. Here again, due to the oscillatory behaviour of  $f_1$  and  $f_2$ , the set of Pareto points is disconnected in  $\mathbb{R}^n$ , and the Pareto front is concave in  $\mathbb{R}^2$ .

We compare again the classic scalarization method with the steepest descent method. They are implemented exactly as we described it in the previous example, and we choose the initial data randomly. The results are presented in Figures 6.12 and 6.13. We observe that, with the increase of the dimension  $n$ , both methods present difficulties to recover the Pareto front. The trajectories tend to be attracted by some particular components of the Pareto front. An explanation could be the fact that the density of the Pareto set is not uniform (see [52]).

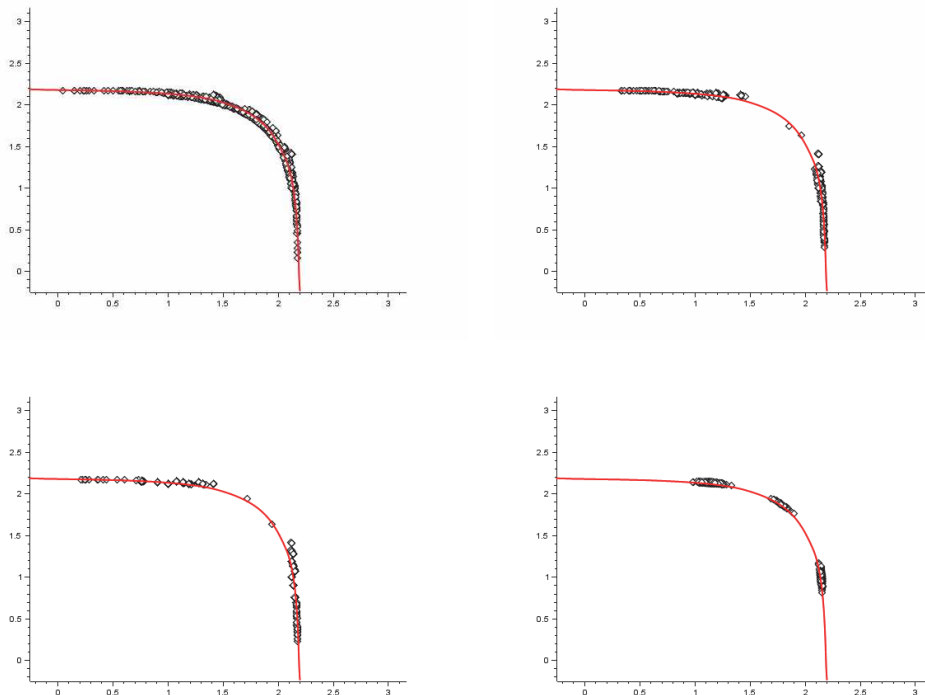


Figure 6.12: From left to right, and top to bottom: 1000 samples of the scalarization method, with  $n$  taking respectively the values 1, 2, 3, 50.

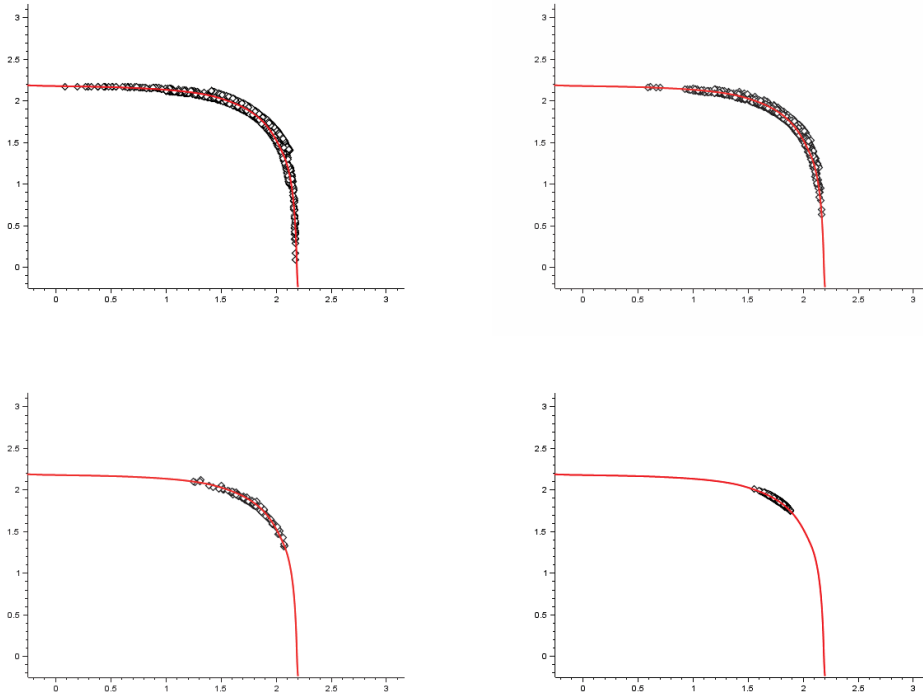


Figure 6.13: From left to right, and top to bottom: 1000 samples of the steepest descent method, with  $n$  taking respectively the values 1, 2, 3, 50.

## 6.4 Comments and perspectives

**Remark 6.4.1** (On the selection of an efficient point among others). Because the set of (weak) efficient points is rarely reduced to one point, we can aim to select one among them satisfying a desired property. For instance, we could look for the efficient point being the nearest from a desired state  $u_d$ . This problem of optimizing over the set of efficient points is sometimes called post-Pareto analysis, see [270, 58, 67, 74]. By analogy with the scalar optimization, we propose an approach based on a Tikhonov-like regularization. That is, we could add to the (SD) dynamic a vanishing term  $\varepsilon(t)(u(t) - u_d)$ , where  $\varepsilon(t) \xrightarrow[t \rightarrow +\infty]{} 0$ :

$$(6.59) \quad \dot{u}(t) + \partial^c F(u(t))^0 + \varepsilon(t)(u(t) - u_d) = 0.$$

More generally, if we want our solution to minimize some smooth potential  $g : \mathbb{H} \rightarrow \mathbb{R}$ , we could consider

$$\dot{u}(t) + \partial^c F(u(t))^0 + \varepsilon(t)\nabla g(u(t)) = 0,$$

which covers (6.59) by taking  $g = \frac{1}{2}\|\cdot - u_d\|^2$ .

This hierarchical dynamic is well-known in scalar optimization and is based on Tikhonov regularization ideas (see [23, 90, 124, 114, 24] for the continuous dynamic and [25, 267, 125] for the corresponding algorithms). When  $F = f : \mathbb{H} \rightarrow \mathbb{R}$ , it is proved in [285, 114] that the dynamic (6.59) converges to the closest element to  $u_d$  among  $\operatorname{argmin}_{x \in \mathbb{H}} f$ , provided that  $\varepsilon(t)$  tends to zero, but not too fast ( $\varepsilon(\cdot) \notin L^1([0, +\infty[, \mathbb{R}))$ ). We can hope to obtain similar results in our vector optimization problem, that is, finding the closest element to  $u_d$  among the weak efficient points.



**Remark 6.4.2** (On the convergence towards efficient points instead of weak efficient points). Theorem 6.2.6 only guarantees the convergence of the trajectories of (SD) towards weak efficient points. Of course, being a weak efficient point is a necessary condition to be an efficient point, but one might ask for a method ensuring the convergence to efficient points. A well-known method to turn weak efficient points into efficient points is to equip  $Y$  with a bigger cone than  $K$ . More exactly, consider a closed convex cone  $\tilde{K}$  such that  $K \setminus \{0\} \subset \text{int } \tilde{K}$ . It is easy, according to the definitions, to see that weak efficient points with respect to  $\tilde{K}$  are efficient points with respect to the original cone  $K$ . Thus, it suffices to consider a dynamic governed by the order induced by  $\tilde{K}$  instead of  $K$ .

As an example, consider the bi-objective smooth case  $F = (f_1, f_2) : H \longrightarrow (\mathbb{R}^2, \mathbb{R}_+^2)$ . Fix a parameter  $\varepsilon > 0$ , and consider the cone

$$\tilde{K}_\varepsilon := \{x = (x_1, x_2) \in \mathbb{R}^2 \mid (1 - \varepsilon)x_1 + \varepsilon x_2 \geq 0 \text{ and } \varepsilon x_1 + (1 - \varepsilon)x_2 \geq 0\}.$$

Its dual cone is

$$\tilde{K}_\varepsilon^+ = \{x = (x_1, x_2) \in \mathbb{R}^2 \mid \varepsilon x_2 \leq (1 - \varepsilon)x_1 \text{ and } \varepsilon x_1 \leq (1 - \varepsilon)x_2\},$$

whose base can be taken as

$$\Delta_{2,\varepsilon} := \{(\lambda, 1 - \lambda) \in \mathbb{R}^2 \mid \lambda \in [\varepsilon, 1 - \varepsilon]\}.$$

Figure 6.14: An enlargement for  $\mathbb{R}_+^2$ .

Clearly,  $\tilde{K}_\varepsilon$  is an enlargement of  $\mathbb{R}_+^2$ , such that  $\mathbb{R}_+^2 \setminus \{0\} \subset \text{int } \tilde{K}_\varepsilon$ . From a dual point of view, the basis  $\Delta_{2,\varepsilon}$  is a retract of the unit simplex  $\Delta_2$ . Since this new cone  $\tilde{K}_\varepsilon$  is still polyhedral, the dynamic induced by this order is the same, from the complexity point of view. Indeed, according to Proposition 5.1.5, the ordered subdifferential of  $F$  with respect to the basis  $\Delta_{2,\varepsilon}$  is

$$\partial^c F(x) = \text{co}\{\varepsilon \nabla f_1(x) + (1 - \varepsilon) \nabla f_2(x), \varepsilon \nabla f_2(x) + (1 - \varepsilon) \nabla f_1(x)\}.$$

This approach can be generalized to the general multi-objective setting  $F : H \longrightarrow (\mathbb{R}^m, \mathbb{R}_+^*)$ , by considering the reduced base

$$\Delta_{m,\varepsilon} := \{\theta = (\theta_i) \in \Delta_m \mid \theta_i \in [\varepsilon, 1 - (m - 1)\varepsilon]\}.$$

Note that this approach suffers some drawbacks. For instance, it guarantees the convergence to an efficient point, but not that *any* efficient point can be recovered. See also [164] for a short discussion on this topic.

**Remark 6.4.3** (The general differential inclusion and the parameter selection problem). Consider here that we are in the multi-objective setting  $F = (f_1, \dots, f_m) : H \longrightarrow \mathbb{R}^m$ . A very interesting feature of (SD) is that it selects itself, at (almost) each time, a convex combination  $\theta \in \Delta_m$ . Thus, we can see locally this dynamic as a steepest descent associated to the convex combination  $\sum_{i=1}^m \theta_i f_i$ . It is a clear improvement with respect to the usual weighting method, which asks to choose a fixed convex combination.

A parallel with this situation can be done with what we call the *parameter selection problem*. In optimization, and in particular in inverse problems, it happens that we often need to solve the problem

$$(6.60) \quad \min_{f(x)=0} R(x),$$

where  $R : H \rightarrow \mathbb{R}$  is a regularization term, and  $f : H \rightarrow \mathbb{R}$  is a data fidelity term. A common example is  $H = \mathbb{R}^n$ ,  $R(x) = \|x\|_1$  and  $f(x) = \frac{1}{2}\|Ax - b\|$ , for  $A \in \mathcal{M}^{n,p}(\mathbb{R})$  and  $b \in \mathbb{R}^p$ . In general  $R$  is not smooth, and managing both a nonsmooth function and a constraint can be difficult in practical. To overcome this difficulty, we can chose to relax the original problem into the following one

$$(6.61) \quad \min_{x \in H} R(x) + \alpha f(x),$$

where  $\alpha > 0$  is a parameter to choose. The whole problem is: how to find an  $\alpha$  so that the relaxed problem (6.61) will give us a satisfactory solution with respect to the original problem (6.60)? In general, finding this parameter is an expensive task which asks to try different values for  $\alpha$  until we are satisfied.

We feel that the situation here is similar to the multi-objective one, when one tries to find *the* good convex combination of the objective functions  $\{f_1, \dots, f_m\}$ . Our opinion is that a good approach could be to consider a dynamic which “naturally” let the parameter  $\alpha$  evolve. Note that (6.61) can be equivalently rewritten as

$$(6.62) \quad \min_{x \in H} \lambda R(x) + (1 - \lambda)f(x),$$

where  $\lambda = \frac{1}{1+\alpha} \in ]0, 1[$ . Thus, the dynamic we are looking for can be searched in the Vector Differential Inclusion discussed in Section 6.1.4:

$$(VDI) \quad \dot{u}(t) + \text{co}\{\partial^c R(x), \partial^c f(x)\} \ni 0.$$

One can try to consider the (SD) dynamic for  $F = (R, f)$ , but this might not be a good idea. Indeed, the (SD) dynamic minimizes simultaneously *too many* scalar functions, namely all the convex combinations of  $R$  and  $f$ . Thus, the trajectory will be quickly stucked in a Pareto critical point, which has nothing to do with the scalar optimization problem (6.62). Worst, the corresponding coefficient  $\lambda$  can tend to 0 or 1, which is too extreme in (6.62). In general, we look for a  $\lambda$  close to zero, but not too much.

A different approach could be to perform a steepest descent-like dynamic, but enforcing the parameter  $\lambda$  to remain in a desired zone, namely  $\lambda \in [a, b] \subset ]0, 1[$ . This is equivalent to consider the vector optimization problem for the function  $F = (R, f) : H \rightarrow \mathbb{R}^2$ ,  $\mathbb{R}^2$  being endowed with a bigger cone  $K$  than  $\mathbb{R}_+^2$ . As in the previous remark, consider the cone

$$K := \{x = (x_1, x_2) \in \mathbb{R}^2 \mid ax_1 + (1 - a)x_2 \geq 0 \text{ and } bx_1 + (1 - b)x_2 \geq 0\},$$

then its dual cone  $K^+$  admits the desired base

$$\Delta_{a,b} := \{(\lambda, (1 - \lambda)) \in \Delta_2 \mid \lambda \in [a, b]\}.$$

It would be interesting to look at how  $\lambda$  evolves with time, if it converges to a particular value, or if it tends to remain close to a particular value. Of course this is a simple suggestion, and we could imagine other dynamics for this problem, which would take more into account the structure of the original problem.

**Remark 6.4.4** (Links with cone-constrained problems). It would be interesting to investigate applications of the tools developed in the two previous chapters, to the class of optimization problems under cone constraint. They wrote as follows:

$$\min_{x \in H} h(x) \text{ under the constraint } G(x) \in -K,$$

where  $h : H \rightarrow \mathbb{R}$ ,  $G : H \rightarrow Y$  and  $K \subset Y$  is a closed convex cone with nonempty interior (we assume these functions to be smooth). This general problem covers for instance the

Nonlinear Complementary Problems, or the Semi Definite Programming. It arises for instance in mechanics [272], and is the subject of an active research, see [311, 294, 3, 2, 182] and the references therein. In the spirit of [159], we can propose an interior-point method for solving this cone-constrained problem, based on the steepest descent direction.

Choose an error parameter  $\varepsilon > 0$ , and note  $\Theta$  a weakly\* compact convex base of  $K^+$ . The condition

$$\sigma_{\Theta}(y) \leq -\varepsilon$$

is a barrier for the constraint  $-K$ , since Proposition A.3.5 asserts that

$$-K = \{y \in Y \mid \sigma_{\Theta}(y) \leq 0\}.$$

Thus, it seems natural to design an interior-point method as follows: Given a current iterate  $x_n \in H$ , test whether the condition  $\sigma_{\Theta}(G(x_n)) \leq -\varepsilon$  holds or no. If so, this means that  $G(x_n)$  lies in the interior of  $-K$ . In that case, we use a step of any classic descent method with respect to  $h$ , to obtain  $x_{n+1}$ . If  $\sigma_{\Theta}(G(x_n)) > -\varepsilon$ , it means that  $G(x_n)$  is too close to the boundary of  $-K$ . In that case, we make decrease both  $h$  and  $G$ , by computing a steepest descent step with respect to

$$\begin{aligned} F : H &\longrightarrow \mathbb{R} \times Y \\ x &\longmapsto F(x) := (h(x), G(x)), \end{aligned}$$

where  $\mathbb{R} \times Y$  is ordered by the cone  $\mathbb{R}_+ \times K$ . Its dual cone is  $\mathbb{R}_+ \times K^+$ , whose base can be taken as  $\{1\} \times \Theta$ . In other words, we propose to perform

$$x_{n+1} = x_n + \lambda_n s(x_n)$$

where

$$(6.63) \quad s(x_n) = \operatorname{argmin}_{d \in H} \frac{1}{2} \|d\|^2 + \langle \nabla h(x_n), d \rangle + \sigma_{\Theta}(DG(x_n; d)).$$

The main difficulty in this method would be the resolution of (6.63), which asks to solve a convex quadratic programming. Using the definition of the support function  $\sigma_{\Theta}$ , it can be rewritten as a saddle-point problem:

$$s(x_n) = \operatorname{argmin}_{d \in H} \sup_{\theta \in \Theta} \frac{1}{2} \|d\|^2 + \langle \nabla h(x_n) + D^*G(x_n; \theta), d \rangle.$$

Together with the explicit description of  $\sigma_{\Theta}$  that we know for classical cones (see Example 2.3.8), we also have from Proposition A.3.6 this useful characterization:

$$\forall y \in Y, \sigma_{\Theta}(y) = \inf\{t \in \mathbb{R} \mid y - te \in -K\},$$

where  $e \in \operatorname{int} K$  defines the base  $\Theta$  by  $\{y^* \in K^+ \mid \langle y^*, e \rangle = 1\}$ .

**Remark 6.4.5** (Vector optimization problems and the Kurdyka-Lojasiewicz property). In this chapter, we proved the convergence for the trajectories of (SD), assuming that  $F : X \rightarrow Y$  is convex. In the light of the results of Part I, it would be interesting to know whether a similar result can be obtained for vector-valued functions satisfying some analogue of the Kurdyka-Lojasiewicz inequality. We present briefly some ideas going in this direction. For simplicity, we assume that we are in the multiobjective setting, and that  $F(x) = (f_1(x), \dots, f_m(x))$  is of class  $C^1$ .

A natural hypothesis, when considering the finite family of functions  $\{f_1, \dots, f_m\}$ , is assuming that each of them is a KL function. From this, we aim to derive the finite length property for the trajectories of (SD). Consider then a strong global solution  $u : [0, +\infty[ \rightarrow H$  of (SD),

and assume that it has some strong limit point  $u_\infty$ . Consider  $h(t) = \sum_{i=1}^m \varphi_i(f_i(u(t)) - f_i(u_\infty))$ , where each  $\varphi_i$  is the desingularizing function corresponding to  $f_i$  around  $u_\infty$ . Its derivative is:

$$h'(t) = \sum_{i=1}^m \varphi_i'(f_i(u(t)) - f_i(u_\infty)) \langle \nabla f_i(u(t)), s(u(t)) \rangle.$$

Assume that the following angle condition holds:

$$(6.64) \quad \exists \varepsilon > 0 \text{ such that for a.e. } t \geq 0, \sum_{i=1}^m \frac{\langle -\nabla f_i(u(t)), s(u(t)) \rangle}{\|\nabla f_i(u(t))\| \|s(u(t))\|} \geq \varepsilon.$$

Then, by using successively the Kurdyka-Łojasiewicz inequality and this angle condition, we can deduce that

$$-h'(t) \geq \varepsilon \|\dot{u}(t)\|,$$

which is the key estimation for the proof of convergence in Theorems 3.1.8 and 3.2.2. Clearly, the angle hypothesis (6.64) is automatically satisfied when  $m = 1$ , since in that case  $s(u(t)) = -\nabla f_1(u(t))$ . More generally, (6.64) means that, at almost every  $t \geq 0$ , there exists some  $i(t)$  such that the cosine between  $\nabla f_{i(t)}(u(t))$  and  $s(u(t))$  is bounded from above by  $\varepsilon$ .

This angle hypothesis is quite difficult to verify in practice, so we propose an other approach by making a KL-like assumption directly on  $F$ . It is based on a generalized Kurdyka-Łojasiewicz inequality used in [44] to study general gradient-like systems. Suppose that we are given some locally Lipschitz Lyapunov function  $\mathcal{E} : \mathbb{H} \rightarrow \mathbb{R}$ , such that  $(\mathcal{E} \circ u)'(t) \leq 0$  for a.e.  $t \geq 0$ . Then, similarly to what has been done in Section 3.1.1, the finite length of  $u(\cdot)$  can be derived from the following generalized Kurdyka-Łojasiewicz inequality:

$$(6.65) \quad \forall x \sim \bar{x}, \forall x^* \in \partial^c \mathcal{E}(x), \varphi'(\mathcal{E}(x) - \mathcal{E}(\bar{x})) \langle x^*, -s(x) \rangle \geq \|s(x)\|.$$

We know from Proposition 6.1.7 that each function  $f_i$  is Lyapunov for the (SD) dynamic, as any convex combination of them. The max function  $\max_{i \in \{1, \dots, m\}} f_i$  is also Lyapunov for the dynamic.

It is an open question to know whether (6.65) is satisfied or no for one of the aforementioned Lyapunov functions.



## Chapter 7

# The continuous steepest descent: existence of trajectories

This chapter is devoted to the proof of the existence of trajectories for the dynamic (SD), introduced in Chapter 6. Its main result is:

**Theorem 7.0.1.** *Suppose that  $E$  is an Euclidean space, and let  $C \subset E$  be a nonempty closed convex set. Suppose that  $Y$  is a separable Banach space, ordered by a closed convex cone with nonempty interior. Let  $F : E \rightarrow Y$  be a locally Lipschitz continuous function, being convex and bounded from below. Then, for all  $u_0 \in E$ , there exists a strong global solution  $u : [0, +\infty[ \rightarrow E$  of*

$$(SD) \quad \dot{u}(t) + (N_C(u(t)) + \partial^c F(u(t)))^0 = 0$$

satisfying the Cauchy condition  $u(0) = u_0$ .

In Theorem 7.0.1, we make two structural hypotheses on the spaces  $E$  and  $Y$ . First, we suppose that  $E$  has finite-dimension. This is essentially because our existence result is based on the following Peano's Theorem (see [39, Theorem 2.8]), and not Cauchy-Lipschitz:

**Theorem 7.0.2 (Peano).** *Let  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be continuous. Then, for all  $x_0 \in H$  and  $t_0 \in \mathbb{R}$ , there exists some  $T > 0$  and  $x : [t_0, t_0 + T[ \rightarrow H$  of class  $C^1$ , such that*

$$(7.1) \quad \dot{x}(t) = \phi(x(t)) \quad \text{for all } t \in [t_0, t_0 + T[, \quad \text{with } x(t_0) = x_0.$$

The assumption of separability on  $Y$  relies on the necessity to use measurability results, like Castaing's measurable selection theorem, see Section B.1 in Appendix.

Now we discuss our strategy to prove this existence result. Our idea is to provide a constructive proof of existence, involving a regularization of the operator  $\partial^c F$  governing the (SD) dynamic. We aim to adapt to our situation the classical proof of the existence of strong solutions for evolution equations governed by subdifferentials of convex lower semicontinuous functions, see [83]. The idea is to approach the dynamic (SD) by a regularized one, parametrized by a real  $\lambda > 0$ :

$$(SD)_\lambda \quad \dot{u}_\lambda(t) + (N_C(u_\lambda(t)) + \partial^c F_\lambda(u_\lambda(t)))^0 = 0,$$

where  $F_\lambda$  must be defined as a smooth function such that  $\partial^c F_\lambda$  approximates  $\partial^c F$  when  $\lambda \downarrow 0$ . The problem  $(SD)_\lambda$  being easier because of the smoothness of  $F_\lambda$  (but not trivial, because of the noncontinuous operator  $N_C$ ), we should be able to prove the existence of approximated trajectories  $u_\lambda : [0, +\infty[ \rightarrow E$ . Then, it 'suffices' to let  $\lambda$  tend to zero and prove that the net  $(u_\lambda)_{\lambda > 0}$  tends to a trajectory satisfying (SD).

So we will proceed as follows. In Section 7.1, we focus on the smooth case and prove the existence of trajectories under this assumption. Note that the smooth multi-objective case has been addressed by Attouch and Goudou [28]. Our result derives from an abstract existence

result used in [28], which relies itself on Peano's Theorem. After that, we will need to define a 'good' approximation  $F_\lambda$  of  $F$ , and this will rely on the Moreau-Yosida approximation for convex real-valued functions, which is a widely used method in nonsmooth convex analysis (see [17, 33, 51, 83, 326, 268] for a detailed presentation).

Section 7.2 will be uniquely devoted to the convex (nonsmooth) multi-objective case. Indeed, in that setting, the function  $F$  is explicitly defined by a finite family of real-valued functions  $(f_1, \dots, f_m)$ . Then it suffices to define  $F_\lambda := (f_{1,\lambda}, \dots, f_{m,\lambda})$ , where  $f_{i,\lambda}$  is the Moreau-Yosida regularisation of  $f_i$  with index  $\lambda > 0$  (see Proposition 7.2.2 in Section 7.2 for more details). In that case  $(SD)_\lambda$  writes

$$(SD)_\lambda \quad \dot{u}_\lambda(t) + (N_C(u_\lambda(t)) + \text{co}\nabla f_{i,\lambda}(u_\lambda(t)))^0 = 0,$$

and we prove the existence of a solution of (SD) by passing to the limit in  $(SD)_\lambda$  when  $\lambda \rightarrow 0$ .

In section 7.3, we consider the general vector case, and we start by defining properly some function  $\mathcal{F}_\lambda$  satisfying the desired properties. As one can expect, this construction will rely on Moreau-Yosida approximations of the cost functions  $\{f_\theta\}_{\theta \in \Theta}$ . This being done, the analysis is almost the same than for the multi-criteria case.

Since the general existence result in Section 7.3 covers the multi-objective one in Section 7.2, one might question the pertinence of presenting in a separate way the results in section 7.2. We made this choice essentially because the definition of  $F_\lambda$  in the multicriteria case is direct and intuitive, which is not the case for the general vector case. Moreover, the proof in Section 7.3 faces some delicate questions about measurability which does not occur in the multicriteria case. With this presentation, the reader which is only interested in the multicriteria case can directly access an easier proof.

In all this chapter, it is assumed, if not specified, that  $E$  is an Euclidean space, that  $Y$  is a Banach space, and that  $F : E \rightarrow Y$  is locally Lipschitz continuous.

## 7.1 Existence in the smooth vector case

We start with a regularity result for the steepest descent vector field, when  $F$  is smooth. But before, we recall the notion of Hausdorff continuity for a set-valued mapping.

**Definition 7.1.1** (Hausdorff distance). Let  $A, B$  be two nonempty subsets of  $E$ . The *Hausdorff distance* between  $A$  and  $B$  is defined by

$$d_H(A, B) = \max\{\sup_{a \in A} d(a, B); \sup_{b \in B} d(b, A)\},$$

where  $d(a, B) := \inf_{b \in B} \|a - b\|$ .

We say that a set-valued mapping  $S : E \rightrightarrows E$  is Hausdorff continuous at  $x \in E$  whenever  $d_H(S(x_n), S(x)) \xrightarrow{n \rightarrow +\infty} 0$  for all sequence  $(x_n)_{n \in \mathbb{N}}$  converging to  $x$ . In an obvious way, we say that  $S$  is Lipschitz Hausdorff continuous on  $U \subset E$  if there exists some  $L \geq 0$  such that

$$\forall x, x' \in U, \quad d_H(S(x), S(x')) \leq L \|x - x'\|.$$

Of course, Lipschitz Hausdorff continuity on  $U$  entails Hausdorff continuity on  $U$ .

**Proposition 7.1.2.** Let  $F : E \rightarrow Y$  be a Fréchet differentiable function, with its derivative being Lipschitz continuous  $U \subset E$ . Then,

- i)  $x \mapsto \partial^c F(x)$  is Lipschitz Hausdorff continuous on  $U$ ,
- ii)  $x \mapsto \partial^c F(x)^0$  is Holder continuous on  $U$ ,

iii) there exists  $c > 0$  such that  $\forall x \in U, \forall x' \in \partial^c F(x), \|x'\| \leq c(1 + \|x\|)$ .

*Proof.* During all the proof, we note  $M > 0$  a bound for  $\Theta$  in  $Y^*$ . We start with item i). Take an arbitrary pair  $x_1, x_2 \in U$ . Consider any  $p_1 \in \partial^c F(x_1)$ , then there exists  $\theta_1 \in \Theta$  such that  $p_1 = \theta_1^* \circ DF(x_1)$ . Define  $p_2 := \theta_1^* \circ DF(x_2) \in \partial^c F(x_2)$ . It follows that

$$\begin{aligned} d(p_1, \partial^c F(x_2)) &= \inf_{p \in \partial^c F(x_2)} \|p_1, p\| \leq \|p_1 - p_2\| = \|\theta_1^* \circ (DF(x_1) - DF(x_2))\| \\ &\leq \|\theta_1^*\|_{Y^*} \|DF(x_1) - DF(x_2)\|_{L(E, Y)} \leq MLip(DF; U) \|x_1 - x_2\|. \end{aligned}$$

Since the latter is true for any  $p_1 \in \partial^c F(x_1)$ , it follows that

$$\sup_{p_1 \in \partial^c F(x_1)} d(p_1, \partial^c F(x_2)) \leq MLip(DF; U) \|x_1 - x_2\|.$$

By symmetry, we obtain

$$(7.2) \quad d_H(\partial^c F(x_1), \partial^c F(x_2)) \leq MLip(DF; U) \|x_1 - x_2\|.$$

We follow with item ii). Take two arbitrary  $x_1, x_2 \in U$ . For the sake of simplicity, we write  $s_1 := \partial^c F(x_1)^0$  and  $s_2 := \partial^c F(x_2)^0$ . Then :

$$(7.3) \quad \|s_1 - s_2\| = \|\text{proj}_{\partial^c F(x_1)}(0) - \text{proj}_{\partial^c F(x_2)}(0)\|.$$

According to [32, Proposition 5.1], we have the following estimation:

$$(7.4) \quad \|\text{proj}_{\partial^c F(x_1)}(0) - \text{proj}_{\partial^c F(x_2)}(0)\|^2 \leq (\|s_1\| + \|s_2\|) d_H(\partial^c F(x_1), \partial^c F(x_2)).$$

Then, combine (7.3) and (7.4) together with (7.2) to obtain

$$(7.5) \quad \|s_1 - s_2\|^2 \leq (\|s_1\| + \|s_2\|) MLip(DF; U) \|x_1 - x_2\|.$$

Since  $DF$  is Lipschitz continuous on  $U$ , it follows from the mean value theorem that  $F$  is also Lipschitz continuous on  $U$ . So, according to Proposition 5.1.7, we finally obtain

$$\|s_1 - s_2\|^2 \leq 2M^2 Lip(F; U) Lip(DF; U) \|x_1 - x_2\|.$$

We end the proof with item iii). Fix now some  $x_U \in U$ , and take any  $x \in U$  and  $x' \in \partial^c F(x)$ . Then, using the triangular inequality and the Lipschitz continuity of  $DF$  on  $U$ :

$$\begin{aligned} \|x'\| &\leq M \|DF(x)\| \leq M \|DF(x) - DF(x_U)\| + M \|DF(x_U)\| \\ &\leq MLip(DF; U) \|x' - x_U\| + M \|DF(x_U)\| \\ &\leq c(1 + \|x\|), \end{aligned}$$

where  $c := \max\{MLip(DF; U); MLip(DF; U) \|x_U\| + M \|DF(x_U)\|\}$ . ■

We can now state the main result of this section. It involves a space of bounded weak\* measurable functions, for which we refer to Appendix B.3 for its definitions and properties.

**Theorem 7.1.3.** *Let  $F : E \rightarrow Y$  be a Fréchet differentiable function, with a Lipschitz continuous gradient. Suppose that  $E$  is Euclidean. Then, for all  $u_0 \in C$ , there exists a strong global solution  $u : [0, +\infty[ \rightarrow E$  of*

$$(SD) \quad \dot{u}(t) + (N_C(u(t)) + \partial^c F(u(t)))^0 = 0$$

*such that  $u(0) = u_0$ . More precisely, there exists<sup>1</sup>  $\eta : [0, +\infty[ \rightarrow E$  and  $\theta : [0, +\infty[ \rightarrow Y^*$  such that*

---

<sup>1</sup>The reader might pay attention to the following fact: the theta  $\theta$  used here is a function with values in  $\Theta$ , whose generic elements are denoted by  $\theta$ , which is slightly more italic than  $\theta$ .



- i)  $\dot{u}(t) + \eta(t) + \theta(t) \circ DF(u(t)) = 0$  for a.e.  $t \in [0, +\infty[$ ,
- ii)  $\eta(t) \in N_C(u(t))$  and  $\theta(t) \in \Theta$  for a.e.  $t \in [0, +\infty[$ ,
- iii)  $\eta \in L^2([0, T], E)$  and  $\theta(\cdot) \circ DF(u(\cdot)) \in L^2([0, T], E)$  for each  $T > 0$ .

If we suppose moreover that  $Y$  is separable, then  $\theta \in L_{w^*}^\infty([0, +\infty[, Y^*)$ .

The proof is based on the following abstract existence result, which can be found in the work of Attouch and Goudou [28, Theorem 3.5] (see also the paper of Henry [191]).

**Theorem 7.1.4.** *Let  $S : E \rightrightarrows E$  be a Hausdorff continuous set-valued mapping, taking convex compact values, and satisfying the following growth condition :*

$$\exists c > 0 \text{ such that } \forall \bar{x} \in E, \forall x \in S(x), \|x\| \leq c(1 + \|\bar{x}\|).$$

Suppose that  $E$  is Euclidean. Then, for all  $u_0 \in C$ , there exists a strong global solution  $u : [0, +\infty[ \rightarrow E$  of

$$(7.6) \quad \dot{u}(t) + (N_C(u(t)) + S(u(t)))^0 = 0$$

such that  $u(0) = u_0$ . More precisely, there exists  $\eta, p : [0, +\infty[ \rightarrow E$  such that

- i)  $\dot{u}(t) + \eta(t) + p(t) = 0$  for a.e.  $t \in [0, +\infty[$ ,
- ii)  $\eta(t) \in N_C(u(t))$  and  $p(t) \in S(u(t))$  for a.e.  $t \in [0, +\infty[$ ,
- iii)  $\eta, p \in L^2([0, T], E)$  for each  $T > 0$ .

*Proof of Theorem 7.1.3.* Under our assumptions, and thanks to Proposition 7.1.2, we can apply Theorem 7.1.4 to obtain a strong global solution of (SD) satisfying items i-iii). The point now is to prove that  $\theta \in L_{w^*}^\infty([0, +\infty[, Y^*)$  when  $Y$  is separable. The fact that  $\theta$  is bounded on  $[0, +\infty[$  is clear since it takes its values in the bounded base  $\Theta$ . The whole point is to prove that it is weakly\* measurable.

Define the application

$$(7.7) \quad \begin{aligned} j : \mathbb{R}_+ \times \Theta &\longrightarrow \mathbb{R} \\ (t, \theta) &\longmapsto \|\theta \circ DF(u(t)) + \eta(t)\|_E. \end{aligned}$$

We verify first that  $j(t, \cdot)$  is  $w^*$ -continuous on  $\Theta$ , for any fixed  $t \geq 0$ . Consider for this any weakly-\* convergent net  $\theta_\alpha \xrightarrow[\alpha \in \mathcal{A}]{w^*} \theta$  in  $\Theta$ . For all  $d \in E$ ,  $DF(x; d) \in Y$  and  $\langle \theta_\alpha, DF(u(t); d) \rangle_{Y^* \times Y} \xrightarrow[\alpha \in \mathcal{A}]{\mathbb{R}} \langle \theta, DF(u(t); d) \rangle_{Y^* \times Y}$ . In other words,  $\theta_\alpha \circ DF(u(t))$  tends to  $\theta \circ DF(u(t))$  in  $(E^*, w^*)$ . But we supposed  $E$  to be finite dimensional, so this convergence holds in  $(E, \|\cdot\|)$ . Because of the strong continuity of the norm in  $E$ , it follows that  $j(t, \cdot)$  is  $w^*$ -continuous on  $\Theta$ . Now we verify that  $j(\cdot, \theta)$  is measurable for all  $\theta \in \Theta$ . We know that  $u(\cdot)$  is continuous, and so is  $DF$  by assumption, so it follows that  $t \mapsto \theta \circ DF(u(t))$  is continuous between  $\mathbb{R}_+$  and  $E$ , hence measurable. Moreover  $\eta$  is also measurable, because of item iii). Since the sum and the composition of measurable functions is measurable (see [5, Lemma 4.22]), we deduce that  $j(\cdot, \theta)$  is measurable.

We proved that  $j$  is a Caratheodory function, and because of the separability assumption on  $Y$ ,  $(\Theta, w^*)$  is a separable metrizable compact space (see Proposition C.0.2). So we can apply Castaing's measurable selection Theorem B.1.2, which asserts that the function  $\tilde{\theta} : \mathbb{R}_+ \rightarrow (\Theta, w^*)$  defined by

$$(7.8) \quad \tilde{\theta}(t) := \operatorname{argmin}_{\theta \in \Theta} \|\theta \circ DF(u(t)) + \eta(t)\|_E$$

is Borel measurable. Observe that, by definition of  $\theta$ , we have  $\theta(t) = \tilde{\theta}(t)$  for a.e.  $t \geq 0$ .

It remains only to verify that  $\tilde{\theta}$  is weakly-\* measurable. For this, consider any  $y \in Y$  and show that  $\tilde{\theta} \circ ev_y : \mathbb{R}_+ \rightarrow \mathbb{R}$  is measurable, where  $ev_y : Y^* \rightarrow \mathbb{R}$  is defined by  $ev_y(y^*) = \langle y^*, y \rangle_{Y^* \times Y}$ . We shown that  $\tilde{\theta}$  is Borel measurable between  $\mathbb{R}_+$  and  $(\Theta, w^*)$ , and we know that  $ev_y$  is continuous from  $(\Theta, w^*)$  to  $\mathbb{R}$ , so their composition is Borel measurable ([5, Lem 4.22]).  $\blacksquare$

## 7.2 Existence in the nonsmooth convex multiobjective case

Here we aim to prove a weaker version of Theorem 7.0.1, in the multi-objective case:

**Theorem 7.2.1.** *Let  $(f_i)_{i \in \{1, \dots, m\}} : E \rightarrow \mathbb{R}$  be a finite family of convex continuous functions being bounded from below, and  $C$  a nonempty closed convex subset of  $E$ . Suppose that  $E$  is an Euclidean space. Then, for all  $u_0 \in E$ , there exists a strong global solution  $u : [0, +\infty[ \rightarrow E$  of*

$$(SD) \quad \dot{u}(t) + (N_C(u(t)) + \text{co } \partial f_i(u(t)))^0 = 0$$

satisfying the Cauchy condition  $u(0) = u_0$ .

As we explained at the beginning of this chapter, the proof will heavily rely on a Moreau-Yosida regularization of the objective functions  $(f_i)_{i \in \{1, \dots, m\}}$ . We recall briefly the definition of the Moreau-Yosida approximation for a proper lower semi-continuous convex function  $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ , and summarize its main properties in the following statement, see in particular [83, Proposition 2.11], or more recently [268, Section 3.5.4]. Recall that we suppose in this Chapter 7 that  $E$  is a finite-dimensional Hilbert space.

**Proposition 7.2.2.** *Let  $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$  be a lower semi-continuous convex proper function. The Moreau-Yosida approximation of index  $\lambda > 0$  of  $f$  is the function  $f_\lambda : E \rightarrow \mathbb{R}$  which is defined for all  $x \in E$  by*

$$(7.9) \quad f_\lambda(x) = \inf_{d \in E} f(d) + \frac{1}{2\lambda} \|d - x\|^2.$$

i) The infimum in (7.9) is attained at a unique point  $\text{prox}_{\lambda f}(x) \in E$ , which satisfies

$$(7.10) \quad f_\lambda(x) = f(\text{prox}_{\lambda f}(x)) + \frac{1}{2\lambda} \|x - \text{prox}_{\lambda f}(x)\|^2;$$

$$(7.11) \quad \text{prox}_{\lambda f}(x) + \lambda \partial f(\text{prox}_{\lambda f}(x)) \ni x.$$

$\text{prox}_{\lambda f} : E \rightarrow E$  is a nonexpansive operator. It is the resolvent of index  $\lambda$  of  $\partial f : E \rightrightarrows E$ .

ii)  $f_\lambda$  is convex, and Gateaux differentiable. Its gradient at  $x \in E$  is equal to

$$(7.12) \quad \nabla f_\lambda(x) = \frac{1}{\lambda} (x - \text{prox}_{\lambda f}(x)).$$

The operator  $\nabla f_\lambda : E \rightarrow E$  is the Yosida approximation of index  $\lambda$  of the maximal monotone operator  $\partial f$ . It is Lipschitz continuous with Lipschitz constant  $\frac{1}{\lambda}$ .

iii) For any  $x \in \text{dom } \partial f$ ,  $\|\nabla f_\lambda(x)\| \leq \|\partial^0 f(x)\|$ , ( $\partial^0 f(x)$  is the element of minimal norm of  $\partial f(x)$ ).

iv) For any  $x \in E$ ,  $f_\lambda(x) \leq f(x)$ ,  $\inf_E f_\lambda = \inf_E f$  and  $f_\lambda(x) \uparrow f(x)$  as  $\lambda \downarrow 0$ .

For all  $i \in \{1, \dots, m\}$  and  $\lambda > 0$ , we will note  $f_{i,\lambda} := (f_i)_\lambda$  the Moreau-Yosida regularization of  $f_i$ , with index  $\lambda > 0$ . Define  $F_\lambda : E \rightarrow (\mathbb{R}^m, \mathbb{R}_+^m)$ , for  $\lambda > 0$ , by  $F_\lambda(x) := (f_{1,\lambda}(x), \dots, f_{m,\lambda}(x))$ . Proposition 7.2.2 asserts that  $F_\lambda$  is convex and continuously differentiable, with a  $\frac{1}{\lambda}$ -Lipschitz gradient. In that case, we can apply Theorem 7.1.3 to obtain, for all  $\lambda > 0$ , the existence of a trajectory  $u_\lambda : [0, +\infty[ \rightarrow E$  such that:

- $u_\lambda$  is absolutely continuous on  $[0, T]$  for all  $T > 0$ ,
- $\dot{u}_\lambda(t) + (N_C(u_\lambda(t)) + \text{co } \nabla f_{i,\lambda}(u_\lambda(t)))^0 = 0$  for a.e.  $t \geq 0$ , and  $u_\lambda(0) = u_0$ ,
- there exists  $\eta_\lambda : [0, +\infty[ \rightarrow E$  and  $\theta_\lambda = (\theta_{i,\lambda})_{i \in \{1, \dots, m\}} : [0, +\infty[ \rightarrow \mathbb{R}^m$  such that

- $\dot{u}_\lambda(t) + \eta_\lambda(t) + \sum_{i=1}^m \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)) = 0$  for a.e.  $t \geq 0$ ,
- $\eta_\lambda(t) \in N_C(u_\lambda(t))$  and  $\theta_\lambda(t) \in \Delta_m$  for a.e.  $t \geq 0$ ,
- $\eta_\lambda \in L^2([0, T], E)$  and  $\sum_{i=1}^m \theta_{i,\lambda}(\cdot) \nabla f_{i,\lambda}(u_\lambda(\cdot)) \in L^2([0, T], E)$  for all  $T > 0$ ,
- $\theta_\lambda \in L^\infty([0, +\infty[, \mathbb{R}^m)$ .

Now we turn on the proof of Theorem 7.2.1, by passing to the limit when  $\lambda$  tends to zero. We chose to divide the proof into three Lemmas. Basically, in the first lemma we pass to the limit when  $\lambda \downarrow 0$ , and in the two last lemmas we prove that the limiting trajectory is a solution of (SD). We will not recall that we assume the hypotheses of Theorem 7.2.1 to hold.

**Lemma 7.2.3.** When  $\lambda \downarrow 0$ , we have (taking eventually a subnet) for all  $T > 0$ :

- i)  $u_\lambda \xrightarrow{\|\cdot\|} u$  in  $C([0, T], E)$ ,
- ii)  $(\dot{u}_\lambda)_{\lambda>0}$  is bounded and  $\dot{u}_\lambda \xrightarrow{w} \dot{u}$  in  $L^2([0, +\infty[, E)$ ,
- iii)  $(\eta_\lambda)_{\lambda>0}$  is bounded and  $\eta_\lambda \xrightarrow{w} \eta$  in  $L^2([0, T], E)$ , with  $\eta(t) \in N_C(u(t))$  for a.e.  $t \geq 0$ ,
- iv)  $\forall i \in \{1, \dots, m\}$ ,  $\theta_{i,\lambda} \xrightarrow{w^*} \theta_i$  in  $L^\infty([0, T], \mathbb{R})$ , with  $\theta_i(t) \geq 0$  and  $\sum_{i=1}^m \theta_i(t) = 1$  for a.e.  $t \geq 0$ ,
- v)  $\forall i \in \{1, \dots, m\}$ ,  $f_{i,\lambda} \circ u_\lambda \xrightarrow{\|\cdot\|} f_i \circ u$  in  $C([0, T], R)$ .

*Proof.* Let us establish bounds for the net  $(u_\lambda)_{\lambda>0}$ , which are independent of  $\lambda$ . By a similar argument to that used in the asymptotic study (see Theorem 6.2.6), we obtain

$$\int_0^{+\infty} \|\dot{u}_\lambda(t)\|^2 dt \leq f_{i,\lambda}(u_0) - \inf_E f_{i,\lambda}.$$

Then notice that  $f_{i,\lambda}(u_0) \leq f_i(u_0)$ , and  $\inf_E f_{i,\lambda} = \inf_E f_i$ . Hence

$$\int_0^{+\infty} \|\dot{u}_\lambda(t)\|^2 dt \leq f_i(u_0) - \inf_E f_i,$$

and

$$(7.13) \quad \sup_{\lambda>0} \int_0^{+\infty} \|\dot{u}_\lambda(t)\|^2 dt < +\infty.$$

Since  $(\dot{u}_\lambda)_{\lambda>0}$  is bounded in  $L^2([0, +\infty[, E)$ , we can assume that it converges.

From

$$u_\lambda(t) = u_0 + \int_0^t \dot{u}_\lambda(\tau) d\tau,$$

and Cauchy-Schwarz inequality, we obtain

$$(7.14) \quad \|u_\lambda(t)\| \leq \|u_0\| + \sqrt{t} \left( \int_0^t \|\dot{u}_\lambda(\tau)\|^2 d\tau \right)^{\frac{1}{2}}.$$

Combining (7.13) with (7.14) we deduce that, for any  $T > 0$

$$(7.15) \quad \sup_\lambda \|u_\lambda\|_{L^\infty([0, T], E)} < +\infty.$$

By (7.13), (7.15), the generalized sequence  $(u_\lambda)$  is uniformly bounded and equi-continuous on  $[0, T]$ . Since  $E$  is finite dimensional, we deduce from Ascoli's theorem that, for any  $0 < T < +\infty$ , the generalized sequence  $(u_\lambda)$  is relatively compact for the uniform convergence topology on  $[0, T]$ . Thus, we can assume that  $(u_\lambda)_{\lambda>0}$  converges uniformly on bounded intervals to some  $u$ , and item i) is proved. For item ii), use the weak convergence in  $L^2([0, T], E)$  of  $(u_\lambda)_{\lambda>0}$  and  $(\dot{u}_\lambda)_{\lambda>0}$ , together with [83, Proposition A.6], to see that  $\dot{u}_\lambda \xrightarrow[\lambda \rightarrow 0]{w} \dot{u}$ .

Item iii) comes from the following observation: by definition of  $u_\lambda$ ,

$$(7.16) \quad \eta_\lambda(t) = -\dot{u}_\lambda(t) - \sum_{i=1}^m \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)),$$

where  $(\dot{u}_\lambda)_{\lambda>0}$  is bounded in  $L^2([0, +\infty[, E)$  and  $\theta_{i,\lambda}$  is bounded in  $L^\infty([0, +\infty[, \mathbb{R}^m)$ . Moreover, by Proposition 7.2.2, for any  $v \in E$ ,  $\lambda > 0$ , and  $i = 1, 2, \dots, m$

$$(7.17) \quad \|\nabla f_{i,\lambda}(v)\| \leq \|(\partial f_i)^0(v)\|.$$

Combining (7.15) with (7.17), and using that  $f_i$  is a convex continuous function whose subdifferential  $\partial f_i$  is bounded on bounded sets, we obtain, for any  $T > 0$

$$(7.18) \quad \sup_\lambda \|\nabla f_{i,\lambda}(u_\lambda)\|_{L^\infty([0, T]; E)} < +\infty.$$

We deduce that the net  $(\eta_\lambda)_{\lambda>0}$  remains bounded in  $L^2([0, T]; E)$  for any  $T > 0$ , and admits then a weak limit point  $\eta$ . From  $u_\lambda \xrightarrow[\lambda \downarrow 0]{C([0, T], E)} u$ ,  $\eta_\lambda \xrightarrow[\lambda \downarrow 0]{w} \eta$  and  $\eta_\lambda(t) \in N_C(u_\lambda(t))$ , and from the demi-closedness property of the extension to  $L^2(0, T; E)$  of the maximal monotone normal cone mapping  $N_C$ , we obtain  $\eta(t) \in N_C(u(t))$  for a.e.  $t \geq 0$ .

We now turn on item iv). For all  $i \in \{1, \dots, m\}$ ,  $(\theta_{i,\lambda})_{\lambda>0}$  is bounded in  $L^\infty([0, +\infty[, \mathbb{R})$ , so we can assume, taking eventually a subnet, that these nets weakly\* converge to some  $\theta_i \in L^\infty([0, +\infty[, \mathbb{R})$ . Since  $L_+^\infty([0, +\infty[, \mathbb{R})$  is a weakly\* closed cone (as the dual of  $L_+^1([0, +\infty[, \mathbb{R})$ ), we know that  $\theta_i(t) \geq 0$  for a.e.  $t \geq 0$ . We note  $\theta := (\theta_i)_{i \in \{1, \dots, m\}}$ , which is a weak\* limit of  $\theta_\lambda$  when  $\lambda \downarrow 0$ . Endowing  $L^\infty([0, +\infty[, \mathbb{R})$  with the weakly\* lower semi-continuous norm  $\|\phi\|_\infty := \sup_{t \geq 0} \sum_{i=1}^m |\phi_i(t)|$ , we deduce that, for a.e.  $t \geq 0$ ,  $\sum_{i=1}^m \theta_i(t) \leq 1$ . To prove that  $\sum_{i=1}^m \theta_i(t) = 1$ , consider any measurable  $I \subset [0, T]$ , and use the weak\* convergence of  $\theta_\lambda$  to  $\theta$  to show

$$\int_I \sum_{i=1}^m \theta_i(t) dt = |I|,$$

and the conclusion follows.

For item v), use Proposition 7.2.2 to obtain the pointwise convergence of  $(f_{i,\lambda})$  to  $f_i$ . Combining (7.17) with the local Lipschitz property of the objective functions  $(f_i)_{i \in \{1, \dots, m\}}$ , we obtain that the net  $(f_{i,\lambda})_{\lambda>0}$  is equi-Lipschitz continuous on bounded sets (recall that  $E$  has finite dimension). Hence, by Ascoli's Theorem,  $(f_{i,\lambda})$  uniformly converges to  $f_i$  on bounded sets. Combining this result with the uniform convergence of  $(u_\lambda)$  gives the result. ■

**Lemma 7.2.4.** For a.e.  $t \geq 0$ ,  $\dot{u}(t) + N_C(u(t)) + \text{co } \partial f_i(u(t)) \ni 0$ .

*Proof.* Our difficult task consists in passing to the limit in

$$(7.19) \quad \dot{u}_\lambda(t) + \eta_\lambda(t) + \sum_{i=1}^m \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)) = 0,$$

which contains a product of two weakly converging sequences  $(\theta_{i,\lambda})_{\lambda>0}$  and  $(\nabla f_{i,\lambda}(u_\lambda))_{\lambda>0}$ .

In order to circumvent this difficulty, we use a variational argument based on the convex differential inequality: for any  $\xi \in L^\infty([0, T], E)$ ,

$$\sum_{i=1}^m \theta_{i,\lambda}(t) f_{i,\lambda}(\xi(t)) \geq \sum_{i=1}^m \theta_{i,\lambda}(t) f_{i,\lambda}(u_\lambda(t)) + \left\langle \sum_{i=1}^m \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)), \xi(t) - u_\lambda(t) \right\rangle.$$

After integration on  $[0, T]$ , we obtain

$$\begin{aligned} \int_0^T \sum_{i=1}^m \theta_{i,\lambda}(t) f_{i,\lambda}(\xi(t)) dt &\geq \int_0^T \sum_{i=1}^m \theta_{i,\lambda}(t) f_{i,\lambda}(u_\lambda(t)) dt \\ &+ \int_0^T \left\langle \sum_{i=1}^m \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)), \xi(t) - u_\lambda(t) \right\rangle dt. \end{aligned}$$

By (7.19),  $\sum_{i=1}^m \theta_{i,\lambda}(t) \nabla f_{i,\lambda}(u_\lambda(t)) = -\dot{u}_\lambda(t) - \eta_\lambda(t)$ , so we can rewrite the inequality above by:

$$\begin{aligned} \int_0^T \sum_{i=1}^m \theta_{i,\lambda}(t) f_{i,\lambda}(\xi(t)) dt &\geq \int_0^T \sum_{i=1}^m \theta_{i,\lambda}(t) f_{i,\lambda}(u_\lambda(t)) dt \\ &+ \int_0^T \langle -\dot{u}_\lambda(t) - \eta_\lambda(t), \xi(t) - u_\lambda(t) \rangle dt. \end{aligned}$$

Since  $f_{i,\lambda}(\xi(t)) \leq f_i(\xi(t))$ , and  $\theta_{i,\lambda}(t) \geq 0$ , we obtain

(7.20)

$$\int_0^T \sum_{i=1}^q \theta_{i,\lambda}(t) f_i(\xi(t)) dt \geq \int_0^T \sum_{i=1}^q \theta_{i,\lambda}(t) f_{i,\lambda}(u_\lambda(t)) dt + \int_0^T \langle -\dot{u}_\lambda(t) - \eta_\lambda(t), \xi(t) - u_\lambda(t) \rangle dt.$$

For any  $\xi \in L^\infty([0, T]; E)$  and all  $i \in \{1, \dots, m\}$ , since  $f_i$  is Lipschitz continuous on bounded sets, we have  $f_i \circ \xi \in L^\infty([0, T], \mathbb{R})$ . Moreover  $\theta_{i,\lambda} \xrightarrow{\lambda \rightarrow 0} \theta_i$  in  $L^\infty([0, T], \mathbb{R})$ . Therefore, by passing to the limit on the left member of (7.20), we obtain

$$\lim_{\lambda \rightarrow 0} \int_0^T \sum_{i=1}^m \theta_{i,\lambda}(t) f_i(\xi(t)) dt = \int_0^T \sum_{i=1}^m \theta_i(t) f_i(\xi(t)) dt.$$

Let us now pass to the limit on the right-hand-side of (7.20). For the first term, we use Lemma 7.2.3. For the second term, we notice that  $\dot{u}_\lambda + \eta_\lambda$  converges weakly in  $L^2([0, T], E)$  to  $\dot{u} + \eta$ ,  $\xi - u_\lambda$  converges uniformly, and hence strongly in  $L^2([0, T], E)$  to  $\xi - u$ . We obtain

$$(7.21) \quad \int_0^T \sum_{i=1}^m \theta_i(t) f_i(\xi(t)) dt \geq \int_0^T \sum_{i=1}^m \theta_i(t) f_i(u(t)) dt + \int_0^T \langle -\dot{u}(t) - \eta(t), \xi(t) - u(t) \rangle dt.$$

Let us interpret this inequality in the duality pairing between the functional spaces  $L^\infty([0, T], E)$  and  $L^1([0, T], E) \subset (L^\infty([0, T], E))^*$ . For this, introduce  $\Phi$ , the integral functional on  $L^\infty([0, T], E)$

$$(7.22) \quad \Phi(\xi) = \int_0^T \sum_{i=1}^m \theta_i(t) f_i(\xi(t)) dt.$$

$\Phi : L^\infty([0, T], E) \rightarrow \mathbb{R}$  is convex and continuous on  $L^\infty([0, T], E)$ . Hence, (7.21) can be rewritten as

$$(7.23) \quad -\dot{u} - \eta \in \partial \Phi(u).$$

According to a duality theorem of Rockafellar for convex functional integrals that we extended in Appendix (see Corollary B.2.11), the inclusion (7.23) implies that for almost all  $t > 0$ ,

$$-\dot{u}(t) - \eta(t) = \sum_{i=1}^m \theta_i(t) v_i(t)$$

with measurable functions  $v_i \in L^\infty([0, T], \mathbb{E})$  such that

$$(7.24) \quad v_i(t) \in \partial f_i(u(t)) \quad \text{for almost all } t > 0.$$

In other words, we proved that

$$\dot{u}(t) + N_C(u(t)) + \text{co } \partial f_i(u(t)) \ni 0.$$

■

**Lemma 7.2.5.** For a.e.  $t \geq 0$ ,  $\dot{u}(t) + (N_C(u(t)) + \text{co } \partial f_i(u(t)))^0 = 0$ .

*Proof.* Since we already proved in 7.2.4 that  $u$  is a solution of the vector differential inclusion

$$\dot{u}(t) + N_C(u(t)) + \text{co } \partial f_i(u(t)) \ni 0,$$

we will use the criterion of Proposition 6.1.15 to verify that it is in fact a *lazy* solution. In other words, we need to prove that

$$(7.25) \quad \forall \theta = (\theta_i) \in \Delta_m, \quad \|\dot{u}(t)\|^2 + \sum_{i=1}^m \theta_i \frac{d}{dt}(f_i \circ u)(t) \leq 0.$$

We will use of course the fact that each  $u_\lambda$  is a lazy solution of  $(\text{SD})_\lambda$ , and satisfies (using again Proposition 6.1.15)

$$(7.26) \quad \forall \theta = (\theta_i) \in \Delta_m, \quad \|\dot{u}_\lambda(t)\|^2 + \sum_{i=1}^m \theta_i \frac{d}{dt}(f_{i,\lambda} \circ u_\lambda)(t) \leq 0.$$

So the whole point is to pass to the limit in (7.26), when  $\lambda \rightarrow 0$ , to obtain (7.25).

Take any  $\psi \in \mathcal{D}(]0, T[, \mathbb{R})$ , the space of infinitely differentiable functions with support unclued in  $]0, T[$ , which is dense in  $L^2([0, T], \mathbb{R})$  (see [83, Appendices]). Consider then the corresponding linear integrand

$$\begin{aligned} \Psi : L^2([0, T], \mathbb{E}) &\longrightarrow \mathbb{R} \\ w &\longmapsto \int_0^T \psi(t) \|w(t)\|^2 dt. \end{aligned}$$

It is easy to see that  $\Psi$  is strongly continuous, hence weakly lower semi-continuous. As a direct consequence of  $\dot{u}_\lambda \xrightarrow[\lambda \rightarrow 0]{w} \dot{u}$  in  $L^2([0, T], \mathbb{E})$ , we deduce that

$$(7.27) \quad \int_0^T \psi(t) \|\dot{u}(t)\|^2 dt \leq \liminf_{\lambda \rightarrow 0} \int_0^T \psi(t) \|\dot{u}_\lambda(t)\|^2 dt.$$

Now, use an integration by parts together with the fact that  $\psi(0) = \psi(T) = 0$  to write for all  $\theta \in \Delta_m$

$$(7.28) \quad \int_0^T \psi(t) \left( \sum_{i=1}^m \theta_i f_i \circ u \right)'(t) dt = - \int_0^T \psi'(t) \left( \sum_{i=1}^m \theta_i f_i(u(t)) \right) dt.$$

Because of Proposition 7.3.20,  $f_{i,\lambda} \circ u_\lambda$  converges uniformly to  $f_i \circ u$  in  $C([0, T], \mathbb{R})$ , hence from (7.28) we deduce

$$\int_0^T \psi(t) \left( \sum_{i=1}^m \theta_i f_i \circ u \right)'(t) dt = - \lim_{\lambda \rightarrow 0} \int_0^T \psi'(t) \left( \sum_{i=1}^m \theta_i f_{i,\lambda}(u_\lambda(t)) \right) dt.$$

Doing again an integration by parts, we obtain

$$(7.29) \quad \int_0^T \psi(t) \left( \sum_{i=1}^m \theta_i f_i \circ u \right)' (t) dt = \lim_{\lambda \rightarrow 0} \int_0^T \psi(t) \left( \sum_{i=1}^m \theta_i f_{i,\lambda} \circ u_\lambda \right)' (t) dt.$$

Combines now (7.27) and (7.29) to find

$$\int_0^T \psi(t) \left( \|\dot{u}(t)\|^2 + \sum_{i=1}^m \theta_i \frac{d}{dt} (f_i \circ u)(t) \right) dt \leq \liminf_{\lambda \rightarrow 0} \int_0^T \psi(t) \left( \|\dot{u}_\lambda(t)\|^2 + \sum_{i=1}^m \theta_i \frac{d}{dt} (f_{i,\lambda} \circ u_\lambda)(t) \right) dt.$$

If we suppose that  $\psi \in \mathcal{D}_+(]0, T[, \mathbb{R})$ , *i.e.* that  $\psi$  takes positive values, we can deduce from (7.26) that

$$\int_0^T \psi(t) \left( \|\dot{u}(t)\|^2 + \sum_{i=1}^m \theta_i \frac{d}{dt} (f_i \circ u)(t) \right) dt \leq 0.$$

By using the density of  $\mathcal{D}_+(]0, T[, \mathbb{R})$  in  $L_+^2(]0, T[, \mathbb{R})$ , which is a closed convex self-dual cone, we obtain that  $\|\dot{u}(\cdot)\|^2 + \sum_{i=1}^m \theta_i \frac{d}{dt} (f_i \circ u) \in -L_+^2(]0, T[, \mathbb{R})$ . This proves (7.25), and ends the proof with Proposition 6.1.15. ■

### 7.3 Existence in the nonsmooth convex vector case

When observing what we did in the multicriteria case, one might be tempted in the general case  $F : E \rightarrow Y$  to regularize the cost functions  $\{f_\theta\}_{\theta \in \Theta}$ . Indeed we saw in the previous chapters that these functions were the key to understand the behaviour of  $F$  with respect to the order in  $Y$ . Hence it would be natural, adapting the method of the previous section, to consider the following dynamic:

$$(SD)_\lambda \quad \dot{u}_\lambda(t) + \left( N_C(u_\lambda(t)) + \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \nabla f_{\theta,\lambda}(u_\lambda(t)) \right)^0 = 0,$$

where  $f_{\theta,\lambda}$  denotes the Moreau-Yosida regularisation of  $f_\theta$  with index  $\lambda > 0$ . Indeed it seems, at least at first sight, that  $\overline{\text{co}}^* \bigcup_{\theta \in \Theta} \nabla f_{\theta,\lambda}$  is a good approximation for  $\partial^C F = \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \partial^C f_\theta$ . But two basic observations must be made at this point.

First, if we reduce to the multi-objective case, this  $(SD)_\lambda$  dynamic is in principle different from the one used in the previous section, which involved the convex hull of the gradients  $\nabla f_{i,\lambda}(u_\lambda(t))$ . This is because the Moreau-Yosida regularization  $f_{\theta,\lambda}$  is not linear with respect to  $\theta$ . In other words, for general  $\theta \in \Delta^m$ , it happens that

$$\nabla f_{\theta,\lambda}(x) = \nabla \left( \sum_{i=1}^m \theta_i f_i \right)_\lambda (x) \neq \sum_{i=1}^m \theta_i \nabla f_{i,\lambda}(x).$$

The second concern with this formulation is that there is no explicit vector-valued function  $F_\lambda : E \rightarrow Y$  whose ordered subdifferential is given by  $\overline{\text{co}}^* \bigcup_{\theta \in \Theta} \nabla f_{\theta,\lambda}$ . And this is essential for who wants to pass to the limit when  $\lambda$  goes to zero.

The aim of the following subsection 7.3.1 is to give the good setting in which we will be able to exploit the cost functions  $f_\theta$ . In subsection 7.3.2 we define properly a Moreau-Yosida-like regularization of  $F$ , for which the ordered subdifferential will be as desired  $\overline{\text{co}}^* \bigcup_{\theta \in \Theta} \nabla f_{\theta,\lambda}$ . It

will rely on the appropriate choice of a function  $\mathcal{F}_\lambda : E \rightarrow Z$  for some well-chosen  $Z \neq Y$ . The essential point in this trick is that the ordered subdifferential essentially depends on the *initial* space. Subsection 7.3.3 contains the proof of our main existence theorem. It follows the lines of the proof in the multi-objective case.

### 7.3.1 An equivalent description of $\partial F$

Recall that  $\Theta$  is a  $w^*$ -compact convex subset of  $Y^*$ . We note  $C(\Theta)$  the vector space of (bounded) continuous functions from  $(\Theta, w^*)$  to  $(\mathbb{R}, |\cdot|)$ . It is a Banach space, once equipped with the uniform norm :

$$(7.30) \quad \forall \psi \in C(\Theta), \|\psi\|_{C(\Theta)} := \sup_{\theta \in \Theta} \psi(\theta).$$

This space is naturally ordered by the closed convex cone of positive continuous functions on  $\Theta$ :

$$C_+(\Theta) := \{\psi \in C(\Theta) \mid \psi(\theta) \geq 0 \text{ for all } \theta \in \Theta\}.$$

We will note  $\leq_{C(\Theta)}$  the order induced by  $C_+(\Theta)$ , and  $<_{C(\Theta)}$  the strict order induced by its interior cone, namely :

$$C_{++}(\Theta) := \text{int } C_+(\Theta) = \{\psi \in C(\Theta) \mid \psi(\theta) > 0 \text{ for all } \theta \in \Theta\} = \{\psi \in C(\Theta) \mid \inf_{\theta \in \Theta} \psi(\theta) > 0\}.$$

The topological dual space of  $C(\Theta)$  is noted  $\mathcal{M}^R(\Theta)$ , and can be identified with the Banach space of Radon measures on  $(\Theta, w^*)$ , equipped with the total variation norm<sup>2</sup>  $\|\cdot\|_{\mathcal{M}(\Theta)}$ . This space of Radon measures inherits an order induced by the dual cone of  $C_+(\Theta)$ , that we note  $\mathcal{M}_+^R(\Theta)$ , and which can be identified with the set of positive Radon measures on  $\Theta$ . Since  $C_+(\Theta)$  has a nonempty interior, we can find a weakly\* compact convex base for  $\mathcal{M}_+^R(\Theta)$ . In the following, we will always consider that the base of  $\mathcal{M}_+^R(\Theta)$  is

$$(7.31) \quad \mathcal{P}^R(\Theta) := \{\mu \in \mathcal{M}_+^R(\Theta) \mid \langle \mu, \mathbb{1} \rangle = 1\}$$

where  $\mathbb{1} \in C_{++}(\Theta) = \text{int } C_+(\Theta)$  is the constant application  $\mathbb{1} : \theta \in \Theta \mapsto 1 \in \mathbb{R}$ . In that case,  $\mathcal{P}^R(\Theta)$  can be identified with the set of Radon probabilities on  $\Theta$ . It is an easy exercise to verify that this space of Radon probabilities can be equivalently described by

$$(7.32) \quad \mathcal{P}^R(\Theta) = \{\mu \in \mathcal{M}^R(\Theta) \mid \langle \mu, \mathbb{1} \rangle = 1 \text{ and } \|\mu\|_{\mathcal{M}(\Theta)} \leq 1\}.$$

We will often use the weak\* topology in  $\mathcal{M}^R(\Theta)$ , relatively to the dual pair  $\langle C(\Theta), \mathcal{M}^R(\Theta) \rangle$ . Hence two weak\* topologies will appear in the following : the one in  $Y^*$  and the one in  $\mathcal{M}^R(\Theta)$ . For the sake of simplicity we will both note them  $w^*$ , instead of  $w_{Y^*}^*$  and  $w_{C(\Theta)}^*$ . In general it will be obvious to know which topology we are dealing with, because of the context, or the notations (we try as possible to use  $y^*, \theta$  for denoting the elements in  $Y^* \supset \Theta$ , and  $\mu$  for the elements of  $\mathcal{M}^R(\Theta) \supset \mathcal{P}^R(\Theta)$ ).

Our interest for  $C(\Theta)$  is that there is a natural linear embedding  $i : Y \hookrightarrow C(\Theta)$ . It is the composition of the canonical embedding  $Y \hookrightarrow (Y^*, w^*)^*$ , together with the embedding by restriction  $(Y^*, w^*)^* \hookrightarrow (\Theta, w^*)^*$  and the inclusion  $(\Theta, w^*)^* \subset C(\Theta)$ , that is :

$$\begin{aligned} i : Y &\hookrightarrow C(\Theta) \\ y &\mapsto i(y) : \Theta \longrightarrow \mathbb{R} \\ &\theta \longmapsto \langle \theta, y \rangle_{Y^* \times Y} \end{aligned}$$

This lead to the following definition :

---

<sup>2</sup>See [5, Section 10.10].



**Definition 7.3.1.** We define  $\mathcal{F} := i \circ F : E \rightarrow C(\Theta)$ , *i.e.* for all  $x \in E$ ,  $\mathcal{F}(x) \in C(\Theta)$  is defined by

$$(7.33) \quad \begin{aligned} \mathcal{F}(x) : \Theta &\longrightarrow \mathbb{R} \\ \theta &\longmapsto f_\theta(x). \end{aligned}$$

As we can guess from its simple definition,  $\mathcal{F}$  shares a lot of properties with  $F$ .

**Proposition 7.3.2.** We have the following :

- i) for all  $x_1, x_2 \in E$ ,  $F(x_1) \leq F(x_2) \Leftrightarrow \mathcal{F}(x_1) \leq_{C(\Theta)} \mathcal{F}(x_2)$ .
- ii) for all  $x_1, x_2 \in E$ ,  $F(x_1) < F(x_2) \Leftrightarrow \mathcal{F}(x_1) <_{C(\Theta)} \mathcal{F}(x_2)$ .
- iii) If  $F$  is locally Lipschitz at  $x$ , then so is  $\mathcal{F}$ .
- iv) If  $F$  is convex, then so is  $\mathcal{F}$ .
- v) If  $F$  is Gateaux differentiable at  $x$ , then so is  $\mathcal{F}$ , and

$$\text{for all } d \in E, \theta \in \Theta, D\mathcal{F}(x; d)(\theta) = D^*F(x; \theta)(d).$$

- vi) If  $F$  is bounded from below by  $m\mathbb{1}$  for some  $m \in \mathbb{R}$ , then  $\mathcal{F}$  is bounded from below by  $m\mathbb{1}$ .

*Proof.* Items i) and ii) essentially relies on Proposition A.3.5. Indeed,

$$\begin{aligned} F(x_1) \leq F(x_2) &\Leftrightarrow \sigma_\Theta(F(x_1) - F(x_2)) \leq 0 \\ &\Leftrightarrow \forall \theta \in \Theta, f_\theta(x_1) \leq f_\theta(x_2) \\ &\Leftrightarrow \mathcal{F}(x_1) \leq_{C(\Theta)} \mathcal{F}(x_2), \end{aligned}$$

and

$$\begin{aligned} F(x_1) < F(x_2) &\Leftrightarrow \sigma_\Theta(F(x_1) - F(x_2)) < 0 \\ &\Leftrightarrow \inf_{\theta \in \Theta} \langle \theta, F(x_2) - F(x_1) \rangle_{Y^* \times Y} \\ &\Leftrightarrow \inf_{\theta \in \Theta} (\mathcal{F}(x_2) - \mathcal{F}(x_1))(\theta) > 0 \\ &\Leftrightarrow \mathcal{F}(x_1) <_{C(\Theta)} \mathcal{F}(x_2). \end{aligned}$$

Items ii) to v) are a direct consequence of the definition of  $\mathcal{F}$  as the composition of  $F$  together with the continuous linear application  $i$ . Item vi) relies on the same argument than the one of item i). ■

An important consequence of Proposition 7.3.2 is that  $F$  and  $\mathcal{F}$  share the same sublevel sets, *i.e.*  $[F \leq F(x)] = [\mathcal{F} \leq_{C(\Theta)} \mathcal{F}(x)]$  for all  $x \in E$ . Hence, minimizing  $F$  or  $\mathcal{F}$  is the same :

$$(7.34) \quad \underset{x \in C}{\text{ARGMIN}} F(x) = \underset{x \in C}{\text{ARGMIN}} \mathcal{F}(x) \text{ and } \underset{x \in C}{\text{ARGMIN}}_w F(x) = \underset{x \in C}{\text{ARGMIN}}_w \mathcal{F}(x)$$

Moreover we saw in Proposition 5.2.12 that the ordered subdifferential of  $F$  is strongly related to the geometry of its sublevel sets in  $E$ . As we can expect from this observation, we will prove below the following important result:

**Theorem 7.3.3.** *If  $Y$  is separable, for all  $x \in E$ ,  $\partial^c F(x) = \partial^c \mathcal{F}(x)$ .*

Recall that from its definition, the ordered subdifferential  $\partial^c \mathcal{F}(x)$  depends on the choice of a base for  $\mathcal{M}_+^R(\Theta)$ . Here, and in all what follows, we fix the base of  $\mathcal{M}_+^R(\Theta)$  to be  $\mathcal{P}^R(\Theta)$ . In other words,

$$(7.35) \quad \partial^c \mathcal{F}(x) = \overline{\text{co}}^* \bigcup_{\mu \in \mathcal{P}^R(\Theta)} \partial^c f_\mu(x),$$

where

$$\begin{aligned} f_\mu &:= \mu \circ \mathcal{F} : \mathbb{E} \longrightarrow \mathbb{R} \\ x &\longmapsto \langle \mu, \mathcal{F}(x) \rangle = \int_{\Theta} f_\theta(x) \, d\mu(\theta). \end{aligned}$$

For a given  $\theta \in \Theta$ , we note  $\delta_\theta$  the Dirac probability at  $\theta$  defined by

$$(7.36) \quad \text{for all } \psi \in C(\Theta), \quad \langle \delta_\theta, \psi \rangle_{\mathcal{M}^R(\Theta) \times C(\Theta)} = \psi(\theta).$$

Then, we can see that  $\delta_\theta \circ \mathcal{F}$  is exactly  $f_\theta$ . This means that the family of functions  $\{f_\theta\}_{\theta \in \Theta}$  is included in the family  $\{f_\mu\}_{\mu \in \mathcal{P}^R(\Theta)}$ , which gives directly that  $\partial^c F(x) \subset \partial^c \mathcal{F}(x)$ . In fact, we have:

**Lemma 7.3.4.** Let  $(\Theta, w^*)$  be metrizable. Let  $\mu \in \mathcal{P}^R(\Theta)$ , and take the mean

$$\bar{\theta}_\mu := \int_{\Theta} \theta \, d\mu(\theta).$$

Then  $\bar{\theta}_\mu \in \Theta$  and  $f_\mu = f_{\bar{\theta}_\mu}$ .

*Proof.* Consider

$$\begin{array}{ccc} \phi : \Theta & \longrightarrow & Y^* & \text{and} & A : Y^* & \longrightarrow & \mathbb{R} \\ \theta & \longmapsto & \theta & & y^* & \longmapsto & \langle y^*, F(x) \rangle, \end{array}$$

where  $\phi$  and  $A$  are continuous. The metrizability assumption ensures, via Proposition B.2.3, that  $\bar{\theta}_\mu$  is well-defined. Use [5, Theorem 11.54] to verify that  $\bar{\theta}_\mu \in \Theta$ . It suffices to apply [5, Lemma 11.45] (see Proposition B.2.4) to obtain the result. ■

In other words, this Lemma with the above discussion say that the families  $\{f_\theta\}_{\theta \in \Theta}$  and  $\{f_\mu\}_{\mu \in \mathcal{P}^R(\Theta)}$  are equal. Since the metrizability of  $\Theta$  is ensured when  $Y$  is separable (see Proposition C.0.2) this proves Theorem 7.3.3.

### 7.3.2 A Moreau-Yosida approximation for convex vector-valued functions

Our purpose here is to propose an analogous Moreau-Yosida approximation for  $\mathcal{F}$ , so that we can approach equation (SD) by a sequence of approximate equations involving a smooth approximation of  $\mathcal{F}$ . Indeed, this case has been treated in Theorem 7.1.3, and we hope to pass to the limit on the corresponding solutions to obtain a solution of (SD).

Before defining what will be  $\mathcal{F}_\lambda$ , we need two technical results:

**Lemma 7.3.5.** If  $F$  is bounded from below, then for all  $x \in \mathbb{E}$ ,  $T > 0$ , the family

$$\{\text{prox}_{\lambda f_\theta}(x) \mid \theta \in \Theta, \lambda \in ]0, T[ \}$$

is bounded in  $\mathbb{E}$ .

*Proof.* By definition, for all  $\theta \in \Theta, \lambda \in ]0, T[, x \in E$ , we have

$$(7.37) \quad \frac{1}{2\lambda} \|\text{prox}_{\lambda f_\theta}(x) - x\|^2 \leq f_\theta(x) - f_\theta(\text{prox}_{\lambda f_\theta}(x)).$$

Use the fact that  $\Theta$  is bounded in  $Y^*$ , and that  $F$  is bounded from below together with Proposition 2.3.11 to conclude.  $\blacksquare$

Recall that, given  $\theta \in \Theta$  and  $\lambda > 0$ , we note  $f_{\theta,\lambda} := (\theta \circ F)_\lambda$ , *i.e.* the Moreau-Yosida transform of  $f_\theta = \theta \circ F$  with index  $\lambda$ .

**Proposition 7.3.6.** Let  $F : E \rightarrow Y$  be convex and bounded from below, where  $E$  is Euclidean.

Let  $(\theta_\alpha)_{\alpha \in \mathcal{A}} \subset \Theta$  such that  $\theta_\alpha \xrightarrow[\alpha \in \mathcal{A}]{w^*} \theta$  is  $\Theta$ , and  $(\lambda_\alpha) \subset ]0, +\infty[$  such that  $\lambda_\alpha \xrightarrow[\alpha \in \mathcal{A}]{\mathbb{R}} \lambda$  in  $]0, +\infty[$ . Then, for all  $x \in E$ ,

- i)  $\lim_{\alpha \in \mathcal{A}} \text{prox}_{\lambda_\alpha f_{\theta_\alpha}}(x) = \text{prox}_{\lambda f_\theta}(x)$  if  $\lambda \neq 0$ ,  $x$  if  $\lambda = 0$ .
- ii)  $\lim_{\alpha \in \mathcal{A}} f_{\theta_\alpha, \lambda_\alpha}(x) = f_{\theta, \lambda}(x)$  if  $\lambda \neq 0$ ,  $f_\theta(x)$  if  $\lambda = 0$ .

*Proof.* For the sake of notation, we write here  $p_\alpha := \text{prox}_{\lambda_\alpha f_{\theta_\alpha}}(x)$ , and  $p := \text{prox}_{\lambda f_\theta}(x)$  if  $\lambda \neq 0$ ,  $x$  if  $\lambda = 0$ . Hence item i) resumes to the proof of  $p_\alpha \xrightarrow[\alpha \in \mathcal{A}]{E} p$ . We know from Lemma 7.3.5 that  $(p_\alpha)_{\alpha \in \mathcal{A}}$  is bounded in  $E$ , which is finite dimensional. Then, it suffices to prove that any limit point of  $(p_\alpha)_{\alpha \in \mathcal{A}}$  is  $p$ . Take then any limit point  $\bar{p}$  of  $(p_\alpha)_{\alpha \in \mathcal{A}}$ , *i.e.* suppose that there exists a subnet satisfying  $p_\beta \xrightarrow[\beta \in \mathcal{A}]{E} \bar{p}$ .

Now, use the definition of  $p_\beta = \text{prox}_{\lambda_\beta(\theta_\beta \circ F)}(x)$  to write, for all  $x' \in E$  :

$$(7.38) \quad f_{\theta_\beta}(p_\beta) + \frac{1}{2\lambda_\beta} \|p_\beta - x\|^2 \leq f_{\theta_\beta}(x') + \frac{1}{2\lambda_\beta} \|x' - x\|^2.$$

We want to pass to the limit in (7.38) when  $\beta \in \mathcal{A}$ . Observe that  $p_\beta$  is supposed to strongly converge to  $\bar{p} \in E$ , that  $F$  is continuous from  $E$  to  $Y$ , and that  $\theta_\beta$  is bounded and  $w^*$ -converging to  $\theta$  in  $Y^*$ . It follows that for all  $x' \in E$ ,

$$(7.39) \quad \begin{aligned} f_{\theta_\beta}(p_\beta) &\xrightarrow[\beta \in \mathcal{A}]{\mathbb{R}} f_\theta(\bar{p}), \\ f_{\theta_\beta, \lambda_\beta}(x') &\xrightarrow[\beta \in \mathcal{A}]{\mathbb{R}} f_\theta(x'). \end{aligned}$$

To pass to the limit in the other terms of (7.38), we need to distinguish the cases  $\lambda = 0$  and  $\lambda \neq 0$ .

We first start with  $\lambda \neq 0$ . In that case, for all  $x' \in E$ ,

$$(7.40) \quad \begin{aligned} \frac{1}{2\lambda} \|p_\beta - x\| &\xrightarrow[\beta \in \mathcal{A}]{\mathbb{R}} \frac{1}{2\lambda} \|p - x\|, \\ \frac{1}{2\lambda} \|x' - x\| &\xrightarrow[\beta \in \mathcal{A}]{\mathbb{R}} \frac{1}{2\lambda} \|x' - x\|. \end{aligned}$$

Passing to the limit in (7.38) by using (7.39) and (7.40), we obtain

$$(7.41) \quad \text{for all } x' \in E, \quad f_\theta(\bar{p}) + \frac{1}{2\lambda} \|p - x\|^2 \leq f_\theta(x') + \frac{1}{2\lambda} \|x' - x\|^2.$$

The above inequality states that  $\bar{p} = \text{prox}_{\lambda f_\theta}(x)$ .

Now consider that  $\lambda = 0$ . In (7.38), take  $x' = x$  and pass to the limit using (7.39) to obtain

$$\limsup_{\beta \in \mathcal{A}} \frac{1}{2\lambda_\beta} \|p_\beta - x\|^2 \leq f_\theta(x) - f_\theta(\bar{p}).$$

We have  $\limsup_{\beta \in \mathcal{A}} \frac{1}{2\lambda_\beta} \|p_\beta - x\|^2 < +\infty$ , where  $\lambda_\beta \rightarrow 0$  and  $p_\beta \rightarrow \bar{p}$ , so necessarily,  $\bar{p} = x$ . Hence item i) is proved.

In view to prove item ii), write

$$(7.42) \quad f_{\theta_\alpha, \lambda_\alpha}(x) = f_{\theta_\alpha}(\text{prox}_{\lambda_\alpha f_{\theta_\alpha}}(x)) + \frac{1}{2\lambda_\alpha} \|\text{prox}_{\lambda_\alpha f_{\theta_\alpha}}(x) - x\|^2.$$

If  $\lambda \neq 0$ , it is clear that the right member of (7.42) converges toward  $f_\theta(p) + \frac{1}{2\lambda_\alpha} \|p - x\|^2$ , where  $p = \text{prox}_{\lambda f_\theta}(x)$ , according to item i). In that case, we obtain  $\lim_{\alpha \in \mathcal{A}} f_{\theta_\alpha, \lambda_\alpha}(x) = f_{\theta, \lambda}(x)$ . If  $\lambda = 0$ , then the first term of the right member of (7.42) tends to  $f_\theta(x)$  according to item i). For the second term, note that by definition of  $\text{prox}_{\lambda_\alpha f_{\theta_\alpha}}(x)$  we have

$$(7.43) \quad \frac{1}{2\lambda_\alpha} \|\text{prox}_{\lambda_\alpha f_{\theta_\alpha}}(x) - x\|^2 \leq f_{\theta_\alpha}(p_\alpha) - f_{\theta_\alpha}(x).$$

The right member of (7.43) goes to zero, since  $p_\alpha = \text{prox}_{\lambda_\alpha f_{\theta_\alpha}}(x)$  tends to  $x$ , according to item i). It follows that  $\frac{1}{2\lambda_\alpha} \|\text{prox}_{\lambda_\alpha f_{\theta_\alpha}}(x) - x\|^2$  goes also to zero in (7.42), which means that  $\lim_{\alpha \in \mathcal{A}} f_{\theta_\alpha, \lambda_\alpha}(x) = f_\theta(x)$ .  $\blacksquare$

For all  $x \in E$  and  $\lambda > 0$ , consider the application

$$\begin{aligned} \mathcal{F}_\lambda(x) : (\Theta, w^*) &\longrightarrow \mathbb{R} \\ \theta &\longmapsto f_{\theta, \lambda}(x), \end{aligned}$$

where  $f_{\lambda, \theta}$  is the Moreau-Yosida approximation of  $f_\theta = \theta \circ F$  with index  $\lambda$ . Using Proposition 7.3.6, we see that  $\mathcal{F}_\lambda(x) \in C(\Theta)$ . Hence, we can define the *Moreau-Yosida approximation of  $\mathcal{F}$* , with index  $\lambda > 0$ . It is the application

$$(7.44) \quad \begin{aligned} \mathcal{F}_\lambda : E &\longrightarrow C(\Theta) \\ x &\longmapsto \mathcal{F}_\lambda(x), \end{aligned}$$

where  $\mathcal{F}_\lambda(x)$  has been defined above. Its properties are gathered in the following theorem :

**Theorem 7.3.7.** *Let  $F : E \longrightarrow Y$  be convex and bounded from below, with  $E$  finite-dimensional. Then :*

- i) for all  $\lambda > 0$ ,  $\mathcal{F}_\lambda : E \longrightarrow C(\Theta)$  is convex.
- ii) for all  $\lambda > 0$ ,  $\mathcal{F}_\lambda$  is Gateaux differentiable. Its derivative  $D\mathcal{F}_\lambda : E \longrightarrow L(E, C(\Theta))$  satisfies the following properties:
  - ii.a)  $D\mathcal{F}_\lambda$  is  $\frac{1}{\lambda}$ -Lipschitz continuous,
  - ii.b) for all  $x \in E$ ,  $d \in E$ ,  $\theta \in \Theta$ ,  $D\mathcal{F}_\lambda(x; d)(\theta) = \langle \nabla f_{\theta, \lambda}(x), d \rangle$ ,
  - ii.c) for all  $x \in E$ ,  $\mu \in \mathcal{M}^R(\Theta)$ ,  $D^*\mathcal{F}_\lambda(x; \mu) = \int_\Theta \nabla f_{\theta, \lambda}(x) \, d\mu(\theta) \in E$ ,
  - ii.d) for all  $x \in E$ ,  $\partial^c \mathcal{F}_\lambda(x) = \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \nabla f_{\theta, \lambda}(x)$ .
- iii) for all  $\lambda > 0$ ,  $x \in E$ ,  $\|D\mathcal{F}_\lambda(x)\|_{L(E, C(\Theta))} \leq \sup_{\theta \in \Theta} \|\partial^0 f_\theta(x)\|$ , where  $\partial^0 f_\theta(x)$  denotes the element of minimal norm of  $\partial f_\theta(x)$ .

iv) for all  $\lambda > 0$ ,  $x \in E$ ,  $\mathcal{F}_\lambda(x) \leq_{C(\Theta)} \mathcal{F}(x)$ .

v) If  $F$  is bounded from below by  $m \in Y$  for some  $m \in \mathbb{R}$ , then for all  $\lambda > 0$ ,  $\mathcal{F}_\lambda$  is bounded from below by  $m\mathbf{1} \in C(\Theta)$ .

vi) for all  $x \in E$ ,  $\mathcal{F}_\lambda(x) \xrightarrow[\lambda \downarrow 0]{C(\Theta)} \mathcal{F}(x)$ .

Observe that in the monocriteria case, *i.e.* when  $(Y, K) = (\mathbb{R}, \mathbb{R}_+)$ , we recover the results of Proposition 7.2.2.

The rest of this Section 7.3.2 will be devoted to the proof of Theorem 7.3.7. For the sake of presentation, we will divide the result in individual Propositions. We start with item i) of Theorem 7.3.7, which is quite immediate.

**Proposition 7.3.8.** Let  $F : E \rightarrow Y$  be a convex function bounded from below, with  $E$  finite-dimensional. Then, for all  $\lambda > 0$ ,  $\mathcal{F}_\lambda : E \rightarrow C(\Theta)$  is convex.

*Proof.* Let  $x_1, x_2 \in E$ , and  $t \in [0, 1]$ . Using the definition of  $\leq_{C(\Theta)}$  and  $\mathcal{F}_\lambda$ , we have

$$\begin{aligned} & \mathcal{F}_\lambda(tx + (1-t)y) \leq_{C(\Theta)} t\mathcal{F}_\lambda(x) + (1-t)\mathcal{F}_\lambda(y) \\ \Leftrightarrow & \mathcal{F}_\lambda(tx + (1-t)y)(\theta) \leq t\mathcal{F}_\lambda(x)(\theta) + (1-t)\mathcal{F}_\lambda(y)(\theta) \quad \text{for all } \theta \in \Theta \\ \Leftrightarrow & f_{\theta,\lambda}(tx + (1-t)y) \leq tf_{\theta,\lambda}(x) + (1-t)f_{\theta,\lambda}(y) \quad \text{for all } \theta \in \Theta. \end{aligned}$$

Since the latter inequality always holds (see Proposition 7.2.2 ii)), the convexity of  $\mathcal{F}_\lambda$  is proved.  $\blacksquare$

Now we focus on one of the most important properties of the Moreau-Yosida approximation  $\mathcal{F}_\lambda$ , which is its *regularity*.

**Lemma 7.3.9.** Let  $F : E \rightarrow Y$  be a convex function bounded from below, with  $E$  finite-dimensional. Then, for all  $\lambda > 0$ , the family of functions

$$\{f_{\theta,\lambda} : E \rightarrow \mathbb{R} \mid \theta \in \Theta\}$$

is equi-Lipschitz continuous on bounded sets.

*Proof.* It is a direct consequence of the equi-Lipschitz continuity (see Proposition 7.2.2 ii)) of

$$\{\nabla f_{\theta,\lambda} : E \rightarrow E \mid \theta \in \Theta\}.$$

Indeed, for all  $x \in E$ , use the triangle inequality and the  $\frac{1}{\lambda}$ -Lipschitz continuity of  $f_{\theta,\lambda}$  to obtain

$$(7.45) \quad \begin{aligned} \|\nabla f_{\theta,\lambda}(x)\| & \leq \|\nabla f_{\theta,\lambda}(x) - \nabla f_{\theta,\lambda}(0)\| + \|\nabla f_{\theta,\lambda}(0)\| \\ & \leq \frac{1}{\lambda}\|x\| + \|\nabla f_{\theta,\lambda}(0)\|. \end{aligned}$$

We saw in Proposition 7.2.2 that  $\nabla f_{\theta,\lambda}(0) = \frac{-1}{\lambda}\text{prox}_{\lambda f_\theta}(0)$ , that we know to be uniformly bounded (with respect to  $\theta$ ) thanks to Lemma 7.3.5. Hence, it follows from (7.45) that  $\nabla f_{\theta,\lambda}$  sends bounded sets onto bounded sets, uniformly with respect to  $\theta \in \Theta$ . Using the mean value theorem, we conclude that  $f_{\theta,\lambda}$  is Lipschitz continuous on bounded sets, uniformly with respect to  $\theta \in \Theta$ .  $\blacksquare$

**Lemma 7.3.10.** Let  $F : E \rightarrow Y$  be a convex function bounded from below, with  $E$  finite-dimensional. Then, for all  $\lambda > 0$  and  $x \in E$ , the application

$$\begin{aligned} (\Theta, w^*) & \rightarrow E \\ \theta & \mapsto \nabla f_{\theta,\lambda}(x) \end{aligned}$$

is continuous.

*Proof.* Let  $(\theta_\alpha)_{\alpha \in \mathcal{A}}$  be a net in  $\Theta$ , weakly-\* convergent to some  $\theta \in \Theta$ . Let us show that  $\nabla f_{\theta_\alpha, \lambda}(x)$  converges to  $\nabla f_{\theta, \lambda}(x)$  in  $E$ .

Observe that, according to Proposition 7.2.2,  $\|\nabla f_{\theta_\alpha, \lambda}(x)\| \leq \|\partial^0 f_{\theta_\alpha}(x)\|$ . Given that  $\partial^0 f_{\theta_\alpha}(x) \in \partial F(x)$  for all  $\alpha \in \mathcal{A}$ , and using that  $\partial F(x)$  is bounded in  $E$ , it follows that the net  $(\nabla f_{\theta_\alpha, \lambda}(x))_{\alpha \in \mathcal{A}}$  is bounded in  $E$ . Since we assume that  $E$  is finite-dimensional, it is then sufficient to show that the only limit point of  $(\nabla f_{\theta_\alpha, \lambda}(x))_{\alpha \in \mathcal{A}}$  is  $\nabla f_{\theta, \lambda}(x)$ .

Suppose that, for some subnet,  $\nabla f_{\theta_\beta, \lambda}(x)$  converges when  $\beta \in \mathcal{A}$  to some  $\bar{x} \in E$ . Using the convexity of  $f_{\theta_\beta, \lambda}$ , we have for all  $\beta \in \mathcal{A}$  and  $x' \in E$ , that

$$(7.46) \quad f_{\theta_\beta, \lambda}(x') - f_{\theta_\beta, \lambda}(x) - \langle \nabla f_{\theta_\beta, \lambda}(x), x' - x \rangle \geq 0.$$

This can be rewritten, using the definition of  $\mathcal{F}_\lambda$ ,

$$(7.47) \quad \mathcal{F}_\lambda(x')(\theta_\beta) - \mathcal{F}_\lambda(x)(\theta_\beta) - \langle \nabla(\theta_\beta \circ F)_\lambda(x), x' - x \rangle \geq 0.$$

We pass to the limit in this last expression, using our hypothesis  $\nabla f_{\theta_\beta, \lambda}(x) \xrightarrow[\beta \in \mathcal{A}]{E} \bar{x}$ , and the continuity of  $\mathcal{F}_\lambda(x)$  on  $\Theta$  (recall Proposition 7.3.6) :

$$(7.48) \quad \text{for all } x' \in E, \quad f_{\theta, \lambda}(x') - f_{\theta, \lambda}(x) - \langle \bar{x}, x' - x \rangle.$$

Since  $f_{\theta, \lambda}$  is convex, the latter means that  $\bar{x}$  lies in the (Fenchel) subdifferential of  $f_{\theta, \lambda}$ . But this function being Gateaux differentiable (Proposition 7.2.2), this means that  $\bar{x} = \nabla f_{\theta, \lambda}(x)$ . ■

**Proposition 7.3.11.** Let  $F : E \rightarrow Y$  be a convex function bounded from below, with  $E$  finite-dimensional. Then, for all  $\lambda > 0$ ,  $\mathcal{F}_\lambda$  is Gateaux differentiable. Its derivative  $D\mathcal{F}_\lambda : E \rightarrow L(E, C(\Theta))$  is  $\frac{1}{\lambda}$ -Lipschitz continuous, and is defined by :

$$(7.49) \quad \text{for all } x \in E, d \in E, \theta \in \Theta, \quad D\mathcal{F}_\lambda(x; d)(\theta) = Df_{\theta, \lambda}(x; d) = \langle \nabla f_{\theta, \lambda}(x), d \rangle.$$

*Proof.* We define, for all  $x \in E$ , the application  $A_x : E \rightarrow C(\Theta)$ , where  $A_x(d) \in C(\Theta)$  is defined for all  $d \in E$  by  $A_x(d) : \theta \mapsto \langle \nabla f_{\theta, \lambda}(x), d \rangle$ . We shown in Lemma 7.3.10 that this last application is continuous on  $(\Theta, w^*)$ . Hence  $A_x$  is well-defined, and it is immediate to see that it is linear and continuous, *i.e.*  $A_x \in L(E, C(\Theta))$ . Let us verify now that  $A_x$  is the Gateaux derivative of  $\mathcal{F}_\lambda$  at  $x$ .

We know from Proposition 7.2.2 that, for all  $\lambda > 0$  and all  $\theta \in \Theta$ , that  $f_{\theta, \lambda}$  is convex and Gateaux differentiable at  $x$ , so

$$(7.50) \quad \text{for all } x' \in E, \quad f_{\theta, \lambda}(x') - f_{\theta, \lambda}(x) - \langle \nabla f_{\theta, \lambda}(x), x' - x \rangle \geq 0.$$

Moreover  $\nabla f_{\theta, \lambda}$  is  $\frac{1}{\lambda}$ -Lipschitz continuous on  $E$ , so we can deduce from the classic descent lemma (see [255, Proposition 3.2.12] or [268, Lemma 1.30]) that

$$(7.51) \quad \text{for all } x' \in E, \quad \frac{1}{\lambda} \|x' - x\|^2 \geq f_{\theta, \lambda}(x') - f_{\theta, \lambda}(x) - \langle \nabla f_{\theta, \lambda}(x), x' - x \rangle.$$

By combining (7.50) and (7.51), we obtain for all  $x' \in E$ ,  $x' \neq x$  :

$$(7.52) \quad \left| \frac{f_{\theta, \lambda}(x') - f_{\theta, \lambda}(x) - \langle \nabla f_{\theta, \lambda}(x), x' - x \rangle}{\|x' - x\|} \right| \leq \frac{1}{\lambda} \|x' - x\|.$$

Take now an arbitrary nonzero  $d \in E$ ,  $t \in \mathbb{R}_{++}$ . By replacing  $x'$  by  $x + td$  in (7.52), we obtain

$$(7.53) \quad \left| \frac{f_{\theta, \lambda}(x + td) - f_{\theta, \lambda}(x)}{|t|} - \langle \nabla f_{\theta, \lambda}(x), d \rangle \right| \leq \frac{1}{\lambda} t \|d\|^2,$$

which can be rewritten as

$$(7.54) \quad \left| \frac{\mathcal{F}_\lambda(x+td)(\theta) - \mathcal{F}_\lambda(x)(\theta)}{|t|} - A_x(d)(\theta) \right| \leq \frac{1}{\lambda} t \|d\|^2.$$

Taking the supremum over  $\theta \in \Theta$  in the above expression leads to

$$(7.55) \quad \left\| \frac{\mathcal{F}_\lambda(x+td) - \mathcal{F}_\lambda(x)}{|t|} - A_x(d) \right\|_{C(\Theta)} \leq \frac{1}{\lambda} t \|d\|^2.$$

If we take the limit when  $t$  goes to zero, we see that  $A_x(d)$  is the Gateaux derivative of  $\mathcal{F}_\lambda$  at  $x$  in the direction  $d$ .

Now we just need to check that  $D\mathcal{F}_\lambda$  is  $\frac{1}{\lambda}$ -Lipschitz continuous. Take two arbitrary  $x, x' \in E$ , then :

$$(7.56) \quad \begin{aligned} \|\!| D\mathcal{F}_\lambda(x') - D\mathcal{F}_\lambda(x) \|\!|_{L(E, C(\Theta))} &= \sup_{\|d\|_E=1} \|D\mathcal{F}_\lambda(x'; d) - D\mathcal{F}_\lambda(x; d)\|_{C(\Theta)} \\ &= \sup_{\|d\|_E=1} \sup_{\theta \in \Theta} |D\mathcal{F}_\lambda(x'; d)(\theta) - D\mathcal{F}_\lambda(x; d)(\theta)| \\ &= \sup_{\|d\|_E=1} \sup_{\theta \in \Theta} |\langle \nabla f_{\theta, \lambda}(x') - \nabla f_{\theta, \lambda}(x), d \rangle|. \end{aligned}$$

Using the Cauchy-Schwarz inequality in  $E$  and the  $\frac{1}{\lambda}$ -Lipschitz continuity of all  $\nabla f_{\theta, \lambda}(x)$ , for all  $\theta \in \Theta$ , we finally obtain

$$\begin{aligned} \|\!| D\mathcal{F}_\lambda(x') - D\mathcal{F}_\lambda(x) \|\!|_{L(E, C(\Theta))} &\leq \sup_{\|d\|_E=1} \sup_{\theta \in \Theta} \|\nabla f_{\theta, \lambda}(x') - \nabla f_{\theta, \lambda}(x)\|_E \|d\|_E \\ &\leq \sup_{\|d\|_E=1} \sup_{\theta \in \Theta} \frac{1}{\lambda} \|x' - x\|_E = \frac{1}{\lambda} \|x' - x\|_E. \end{aligned}$$

■

**Corollary 7.3.12.** Let  $F : E \rightarrow Y$  be a convex function bounded from below, with  $E$  finite-dimensional. Then, for all  $x \in E$ :

- i) for all  $\mu \in \mathcal{M}^R(\Theta)$ ,  $D^*\mathcal{F}_\lambda(x; \mu) = \int_\Theta \nabla f_{\theta, \lambda}(x) d\mu(\theta) \in E$ ,
- ii)  $\partial^c \mathcal{F}_\lambda(x) = \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \nabla f_{\theta, \lambda}(x)$ .

*Proof.* Let  $x \in E$ . Using Proposition 7.3.11, we know that for all  $d \in E$  and  $\mu \in \mathcal{M}^R(\Theta)$ , we have

$$\langle D^*\mathcal{F}_\lambda(x; \mu), d \rangle = \langle \mu, D\mathcal{F}_\lambda(x; d) \rangle = \int_\Theta \langle \nabla f_{\theta, \lambda}(x), d \rangle d\mu(\theta).$$

Moreover, because of Lemma 7.3.10 and Theorem B.2.2,  $\theta \mapsto \nabla f_{\theta, \lambda}(x)$  is Bochner integrable. It follows then that

$$D^*\mathcal{F}_\lambda(x; \mu) = \int_\Theta \nabla f_{\theta, \lambda}(x) d\mu(\theta).$$

Now we turn on the proof of item ii). Start by taking any  $\theta \in \Theta$ , and observe that

$$\nabla f_{\theta, \lambda}(x) = \int_\Theta \nabla f_{\tilde{\theta}, \lambda}(x) d\delta_\theta(\tilde{\theta}),$$

where  $\delta_\theta \in \mathcal{P}^R(\Theta)$  is the Dirac measure at  $\theta$ . Write now for all  $d \in E$ :

$$\langle \nabla f_{\theta, \lambda}(x), d \rangle = \langle D^*\mathcal{F}_\lambda(x; \delta_\theta), d \rangle = \langle \delta_\theta, D\mathcal{F}_\lambda(x; d) \rangle = \langle \delta_\theta \circ D\mathcal{F}_\lambda(x), d \rangle = \langle \nabla(\delta_\theta \circ \mathcal{F}_\lambda)(x), d \rangle.$$

We deduce that  $\nabla f_{\theta,\lambda}(x) = \nabla(\delta_\theta \circ \mathcal{F}_\lambda)(x)$ , so from the definition of the ordered subdifferential we obtain  $\nabla f_{\theta,\lambda}(x) \in \partial^c \mathcal{F}_\lambda(x)$ . Since this is true for any  $\theta \in \Theta$ , and because the ordered subdifferential  $\partial^c \mathcal{F}_\lambda(x)$  is convex and weakly\* closed, we can deduce

$$\partial^c \mathcal{F}_\lambda(x) \supset \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \nabla f_{\theta,\lambda}(x).$$

For the reverse inclusion, recall from Proposition 7.3.12 that  $\mathcal{F}_\lambda$  is smooth, so we have

$$\partial^c \mathcal{F}_\lambda(x) = \{D^* \mathcal{F}_\lambda(x; \mu) \mid \mu \in \mathcal{P}^R(\Theta)\}.$$

Then, take any  $\mu \in \mathcal{P}^R(\Theta)$ , and use item i) together with [5, Theorem 11.54] to obtain  $D^* \mathcal{F}_\lambda(x; \mu) \in \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \nabla f_{\theta,\lambda}(x)$ . It follows that

$$\partial^c \mathcal{F}_\lambda(x) \subset \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \nabla f_{\theta,\lambda}(x),$$

and the claim is proved. ■

We follow with item iii) of Theorem 7.3.7, which states a uniform boundedness of the derivatives, with respect to  $\lambda$ .

**Proposition 7.3.13.** Let  $F : E \rightarrow Y$  be a convex function bounded from below, with  $E$  finite-dimensional. Then, for all  $\lambda > 0$ ,  $x \in E$ ,

$$\|D\mathcal{F}_\lambda(x)\|_{L(E, C(\Theta))} \leq \sup_{\theta \in \Theta} \|\partial^0 f_\theta(x)\|,$$

where  $\partial^0 f_\theta(x)$  denotes the element of minimal norm of  $\partial f_\theta(x)$ .

*Proof.* The proof follows similar arguments as in the end of the proof of Proposition 7.3.11. Let  $x \in E$ , then

$$\begin{aligned} (7.57) \quad \|D\mathcal{F}_\lambda(x)\|_{L(E, C(\Theta))} &= \sup_{\|d\|_E=1} \|D\mathcal{F}_\lambda(x; d)\|_{C(\Theta)} \\ &= \sup_{\|d\|_E=1} \sup_{\theta \in \Theta} |D\mathcal{F}_\lambda(x; d)(\theta)| \\ &= \sup_{\|d\|_E=1} \sup_{\theta \in \Theta} |\langle \nabla f_{\theta,\lambda}(x), d \rangle|. \end{aligned}$$

Using the Cauchy-Schwarz inequality in  $E$ , and Proposition 7.2.2, the result follows. ■

We end now by studying the monotone properties of  $\mathcal{F}_\lambda$ , with respect to  $\lambda$ , to prove items iv), v) and vi) of Theorem 7.3.7.

**Proposition 7.3.14.** Let  $F : E \rightarrow Y$  be a convex function bounded from below, with  $E$  finite-dimensional. Then, for all  $\lambda > 0$ ,  $x \in E$ ,

$$m\mathbf{1} \leq_{C(\Theta)} \mathcal{F}_\lambda(x) \leq_{C(\Theta)} \mathcal{F}(x).$$

*Proof.* Start with the first inequality. By hypothesis, we have for all  $x \in E$  and all  $\theta \in \Theta$  that  $f_\theta(x) \geq m$  (see Proposition 2.3.11). Equivalently, for all  $\theta \in \Theta$ ,  $\inf_E f_\theta \geq m$ . From Proposition 7.2.2 we deduce that for all  $x \in E$  and all  $\theta \in \Theta$  that  $f_{\theta,\lambda}(x) \geq m$ . The latter means that  $m\mathbf{1} \leq_{C(\Theta)} \mathcal{F}_\lambda(x)$  for all  $x \in E$ . Hence, the first inequality is proved.

The second inequality comes from the fact that  $f_{\theta,\lambda}(x) \leq f_\theta(x)$ , for all  $x \in E, \theta \in \Theta$ , see Proposition 7.2.2. ■



To prove the convergence of  $\mathcal{F}_\lambda(x)$  to  $\mathcal{F}(x)$  in  $C(\Theta)$ , we will use the following technical lemmas:

**Lemma 7.3.15.** Let  $(K, \tau)$  be a compact topological vector space. Let  $\psi \in C(K)$  and  $(\psi_\alpha)_{\alpha \in \mathcal{A}} \subset C(K)$ . Suppose that

$$(7.58) \quad \text{for all } (k_\alpha)_{\alpha \in \mathcal{A}} \subset K, \text{ if } k_\alpha \xrightarrow[\alpha \in \mathcal{A}]{\tau} k \text{ then } \psi_\alpha(k_\alpha) \xrightarrow[\alpha \in \mathcal{A}]{\mathbb{R}} \psi(k).$$

Hence  $\psi_\alpha$  converges uniformly to  $\psi$  on  $K$ , i.e.  $\psi_\alpha \xrightarrow[\alpha \in \mathcal{A}]{C(K)} \psi$ .

*Proof.* Suppose by contradiction that  $\psi_\alpha$  does not converge uniformly to  $\psi$  on  $K$ . Then there exists  $\varepsilon > 0$  and a subnet  $(\psi_\beta)_{\beta \in \mathcal{A}}$  such that for all  $\beta \in \mathcal{A}$ ,  $\|\psi_\beta - \psi\| \geq \varepsilon$ . Since  $\psi_\beta - \psi$  is continuous by hypothesis, and  $K$  compact, by the extreme value theorem we obtain for all  $\beta \in \mathcal{A}$  some  $k_\beta \in K$  such that  $|\psi_\beta(k_\beta) - \psi(k_\beta)| \geq \varepsilon$ . But,  $(k_\beta)_{\beta \in \mathcal{A}}$  lies in the compact space  $K$ , so there exists a subnet  $(k_\gamma)_{\gamma \in \mathcal{A}}$  converging to some  $k \in K$ . By hypothesis, it satisfies

$$\psi_\gamma(k_\gamma) \xrightarrow[\gamma \in \mathcal{A}]{\mathbb{R}} \psi(k) \text{ and } \psi(k_\gamma) \xrightarrow[\gamma \in \mathcal{A}]{\mathbb{R}} \psi(k),$$

in other words  $|\psi_\gamma(k_\gamma) - \psi(k_\gamma)| \xrightarrow[\gamma \in \mathcal{A}]{} 0$  which is a contradiction.  $\blacksquare$

**Lemma 7.3.16.** Let  $F : E \rightarrow Y$  be a convex function bounded from below, with  $E$  finite-dimensional. For all  $x \in E$ , the following application is continuous :

$$\begin{aligned} \Psi : [0, 1] &\longrightarrow C(\Theta) \\ \lambda &\longmapsto \mathcal{F}_\lambda(x) \text{ if } \lambda \neq 0 \\ &\mathcal{F}(x) \text{ if } \lambda = 0. \end{aligned}$$

*Proof.* Let  $(\lambda_i)_{i \in I}$  be any net in  $[0, 1]$  converging to  $\lambda \in [0, 1]$ . If  $\lambda_i > 0$  for all  $i \in I$ , we could use Proposition 7.3.6 to obtain

$$(7.59) \quad \text{for all } (\theta_i)_{i \in I} \subset \Theta, \text{ if } \theta_i \xrightarrow[i \in I]{w^*} \theta \text{ then } \Psi(\lambda_i)(\theta_i) \xrightarrow[i \in I]{\mathbb{R}} \psi(\lambda)(\theta).$$

In the case  $\lambda \neq 0$ , it is clear that we can assume that  $\lambda_i \neq 0$ , taking eventually a subnet. If  $\lambda = 0$ , then two cases must be considered. Either there exists some  $i \in I$  such that for all  $j \succeq i$ ,  $\lambda_j = 0$ , in which case we have

$$\Psi(\lambda_j)(\theta_j) = \Psi(0)(\theta_j) = \langle \theta_j, F(x) \rangle_{Y^* \times Y} \xrightarrow[j \succeq i]{\mathbb{R}} \langle \theta, F(x) \rangle_{Y^* \times Y} = \psi(\lambda)(\theta).$$

Or it holds that for all  $i \in I$ , there exists some  $j \succeq i$  such that  $\lambda_j > 0$ . In that case  $J := \{j \in I \mid \lambda_j > 0\}$  induces two subnets  $(\lambda_j)_{j \in J}$  and  $(\theta_j)_{j \in J}$  for which we can apply Proposition 7.3.6.

Hence, we case assume in generality that (7.59) holds. Then, we can use Lemma 7.3.15 to obtain that  $\Psi(\lambda_\alpha) \xrightarrow[\alpha \in \mathcal{A}]{C(\Theta)} \Psi(\lambda)$ , which ends the proof.  $\blacksquare$

**Proposition 7.3.17.** Let  $F : E \rightarrow Y$  be a convex function bounded from below, with  $E$  finite-dimensional. Then, for all  $x \in E$ ,  $\lim_{\lambda \rightarrow 0} \mathcal{F}_\lambda(x) = \mathcal{F}(x)$  in  $C(\Theta)$ .

*Proof.* Just use the continuity at zero in the previous Lemma.  $\blacksquare$

**Remark 7.3.18.** Clearly, using Lemma 7.3.16 to prove Proposition 7.3.17 is kind of using a bazooka to kill a fly. Nevertheless we chose this way because we will need, later, to obtain some compactness for the family  $\{\mathcal{F}_\lambda(x)\}_\lambda$  in  $C(\Theta)$  (see Lemma Lemma 7.3.16).

If one wants to prove Proposition 7.3.17 directly, note that the net  $(\mathcal{F}_\lambda(x))_\lambda$  is monotone in  $C(\Theta)$  when  $\lambda \downarrow 0$ , so we can apply Dini's monotone convergence theorem(see [6, Theorem 9.04, p.70]<sup>3</sup>).

<sup>3</sup>Note that the Theorem there only deals with a sequence of functions, but the proof can easily be adapted to take into account our net  $(\mathcal{F}_\lambda(x))_{\lambda \in ]0, +\infty[}$ .

### 7.3.3 Proof of the existence

We are now ready to prove Theorem 7.0.1. We consider satisfied its hypotheses, *i.e.* we suppose that  $F : E \rightarrow Y$  is a convex locally Lipschitz function, with  $E$  Euclidean and  $Y$  a separable Banach space. Let  $u_0 \in E$  be fixed.

For all  $\lambda > 0$ ,  $\mathcal{F}_\lambda : E \rightarrow C(\Theta)$  is a locally Lipschitz function, whose derivative  $D\mathcal{F}_\lambda$  is globally  $\frac{1}{\lambda}$ -Lipschitz continuous. Hence, we can use Theorem 7.1.3 to obtain the existence of a trajectory  $u_\lambda : [0, +\infty[ \rightarrow E$  such that :

- $u_\lambda$  is absolutely continuous on  $[0, T]$  for all  $T > 0$ ,
- $\dot{u}_\lambda(t) + (N_C(u_\lambda(t)) + \partial^c \mathcal{F}_\lambda(u_\lambda(t)))^0 = 0$  for a.e.  $t \geq 0$ , and  $u_\lambda(0) = u_0$ ,
- there exists  $\eta_\lambda : [0, +\infty[ \rightarrow E$  and  $\mu_\lambda : [0, +\infty[ \rightarrow \mathcal{M}^R(\Theta)$  such that
  - $\dot{u}_\lambda(t) + \eta_\lambda(t) + \mu_\lambda(t) \circ D\mathcal{F}_\lambda(u_\lambda(t)) = 0$  for a.e.  $t \geq 0$ ,
  - $\eta_\lambda(t) \in N_C(u_\lambda(t))$  and  $\mu_\lambda(t) \in \mathcal{P}^R(\Theta)$  for a.e.  $t \geq 0$ ,
  - $\eta_\lambda \in L^2([0, T], E)$  and  $\mu_\lambda(\cdot) \circ D\mathcal{F}_\lambda(u_\lambda(\cdot)) \in L^2([0, T], E)$  for all  $T > 0$ ,
  - $\mu_\lambda \in L_{w^*}^\infty([0, +\infty[, \mathcal{M}^R(\Theta))$ .

Our intention is to pass to the limit when  $\lambda$  goes to zero, to obtain a trajectory solution of (SD).

During all this section, we use the following notations for the cost functions: for all  $\lambda > 0$  and each  $\mu \in \mathcal{P}^R(\Theta)$ , note

$$\begin{aligned} f_{\mu,\lambda} &:= \mu \circ \mathcal{F}_\lambda : E \rightarrow \mathbb{R} \\ x &\mapsto \mu \circ \mathcal{F}_\lambda(x) = \int_{\Theta} \mathcal{F}_\lambda(x) d\mu = \int_{\Theta} f_{\theta,\lambda}(x) d\mu(\theta). \end{aligned}$$

By Theorem 7.3.7,  $f_{\mu,\lambda}$  is a convex Fréchet differentiable function, whose gradient  $\nabla f_{\mu,\lambda}(x) = \mu \circ D\mathcal{F}_\lambda(x)$  is  $\frac{1}{\lambda}$ -Lipschitz continuous. Similarly, define

$$\begin{aligned} f_\mu &:= \mu \circ \mathcal{F} : E \rightarrow \mathbb{R} \\ x &\mapsto \mu \circ \mathcal{F}(x) = \int_{\Theta} \mathcal{F}(x) d\mu = \int_{\Theta} f_\theta(x) d\mu(\theta). \end{aligned}$$

It is also a convex function, such that  $(f_{\mu,\lambda})_{\lambda>0}$  is pointwise converging to  $f_\mu$  from below.

**Proposition 7.3.19.** Let  $F : E \rightarrow Y$  be convex and bounded from below, with  $E$  finite-dimensional and  $Y$  separable. Then, when  $\lambda \downarrow 0$ , we have (taking eventually a subnet) for all  $T > 0$ :

- i)  $u_\lambda \xrightarrow{\|\cdot\|} u$  in  $C([0, T], E)$ ,
- ii)  $(\dot{u}_\lambda)_{\lambda>0}$  is bounded and  $\dot{u}_\lambda \xrightarrow{w} \dot{u}$  in  $L^2([0, +\infty[, E)$ ,
- iii)  $(\eta_\lambda)_{\lambda>0}$  is bounded and  $\eta_\lambda \xrightarrow{w} \eta$  in  $L^2([0, T], E)$ , with  $\eta(t) \in N_C(u(t))$  for a.e.  $t \geq 0$ ,
- iv)  $(\mu_\lambda)_{\lambda>0}$  is bounded and  $\mu_\lambda \xrightarrow{w^*} \mu$  in  $L_{w^*}^\infty([0, T], \mathcal{M}^R(\Theta))$ , with  $\mu(t) \in \mathcal{P}^R(\Theta)$  for a.e.  $t \geq 0$ .

*Proof.* We start with energies estimations for the trajectory. For all  $\lambda > 0$  and  $\mu \in \mathcal{P}^R(\Theta)$ ,  $f_{\mu,\lambda}$  is differentiable so for a.e.  $t \geq 0$ ,

$$(7.60) \quad \frac{d}{dt}(f_{\mu,\lambda} \circ u_\lambda)(t) = \langle \nabla f_{\mu,\lambda}(u_\lambda(t)), \dot{u}_\lambda(t) \rangle = d^C f_{\mu,\lambda}(u_\lambda(t); \dot{u}_\lambda(t)).$$

Using  $\dot{u}_\lambda(t) = -(N_C(u_\lambda(t)) + \partial^C \mathcal{F}_\lambda(u_\lambda(t)))^0$ , we obtain from Proposition 6.1.7 that

$$(7.61) \quad \frac{d}{dt}(f_{\mu,\lambda} \circ u_\lambda)(t) \leq \|\dot{u}_\lambda(t)\|^2 \text{ for a.e. } t \geq 0,$$

and integrating (7.61) on  $[0, T]$  :

$$(7.62) \quad \int_0^T \|\dot{u}_\lambda(t)\|^2 dt \leq f_{\mu,\lambda}(u_0) - f_{\mu,\lambda}(u(T)).$$

Use now the fact that  $f_{\mu,\lambda} \leq f_\mu$  on  $E$ , and that  $F$  is bounded from below on  $E$  together with Theorem 7.3.7 v) to obtain:

$$(7.63) \quad \int_0^T \|\dot{u}_\lambda(t)\|^2 dt \leq f_\mu(u_0) - m.$$

Since the estimation in (7.63) is independent of  $\lambda$  and  $T$ , it follows that  $\dot{u}_\lambda \in L^2([0, +\infty[, E)$  with

$$(7.64) \quad \sup_{\lambda > 0} \|\dot{u}_\lambda\|_{L^2([0, +\infty[, E)} < +\infty.$$

Consider, for all  $\lambda > 0$  and all  $T > 0$ , the trajectory  $u_\lambda \in C([0, T], E)$ . To prove the existence of a converging subnet of this family, we will invoke Ascoli's Theorem. First recall that  $u_\lambda$  is absolutely continuous. So, for all  $t \geq 0$ ,

$$(7.65) \quad u_\lambda(t) = u_\lambda(0) + \int_0^t \dot{u}_\lambda(s) ds,$$

which implies, after using the Cauchy-Schwarz inequality in  $L^2([0, t], \mathbb{R})$ ,

$$(7.66) \quad \|u_\lambda(t)\| \leq \|u_\lambda(0)\| + \sqrt{t} \sqrt{\int_0^t \|\dot{u}_\lambda(s)\|^2 ds}.$$

As a direct consequence of (7.66) and (7.64), we have

$$(7.67) \quad \sup_{\lambda > 0} \|u_\lambda(t)\| \leq \|u_\lambda(0)\| + \sqrt{T} \sup_{\lambda > 0} \|\dot{u}_\lambda\|_{L^2([0, +\infty[, E)} < +\infty.$$

So  $\{u_\lambda\}_{\lambda > 0}$  is pointwise bounded, hence pointwise relatively compact in  $C([0, T], E)$  (recall that  $E$  has finite dimension). Now we verify that  $\{u_\lambda\}_{\lambda > 0}$  is equicontinuous in  $C([0, T], E)$ . For this, fix  $\lambda > 0$  and consider  $0 \leq s < t \leq T$ , and use the absolute continuity of  $u_\lambda$  to write

$$u_\lambda(t) - u_\lambda(s) = \int_s^t \dot{u}_\lambda(\tau) d\tau.$$

Using the Cauchy-Schwarz inequality as before, together with (7.64), we obtain

$$(7.68) \quad \|u_\lambda(t) - u_\lambda(s)\| \leq \int_s^t \|\dot{u}_\lambda(\tau)\| d\tau \leq \sqrt{t-s} \sqrt{\int_s^t \|\dot{u}_\lambda(\tau)\|^2 d\tau} \leq \sqrt{t-s} \sup_{\lambda > 0} \|\dot{u}_\lambda\|_{L^2([0, +\infty[, E)}.$$

The latter means that  $\{u_\lambda\}_{\lambda>0}$  is uniformly Hölder continuous, hence equicontinuous. So we can apply Ascoli's Theorem, and obtain that  $\{u_\lambda\}_{\lambda>0}$  is relatively compact in  $C([0, T], E)$ , so by taking eventually a subnet, we obtain that there exists some  $u : [0, +\infty[ \rightarrow E$  such that  $u_\lambda$  converges uniformly to  $u$  on  $[0, T]$  for each  $T > 0$ .

We prove now the weak convergence of  $\dot{u}_\lambda$  to  $\dot{u}$  in  $L^2([0, T], E)$ . We already now, from (7.64), that  $\{\dot{u}_\lambda\}_{\lambda>0}$  is relatively compact in  $L^2([0, T], E)$ . Let  $v \in L^2([0, T], E)$  be any limit point of  $\{\dot{u}_\lambda\}_{\lambda>0}$  when  $\lambda \downarrow 0$ , then use [83, Proposition A.6] to verify that  $v(t) = \dot{u}(t)$  for a.e.  $t \in ]0, T[$ .

We focus now on item iii). We have, for all  $\lambda > 0$  and a.e.  $t \geq 0$ ,

$$(7.69) \quad \|\eta_\lambda(t)\| = \|\dot{u}_\lambda(t) + \mu_\lambda(t) \circ D\mathcal{F}_\lambda(u_\lambda(t))\|.$$

Since  $[0, T]$  is compact and  $u$  continuous,  $u([0, T])$  is compact in  $E$ . But  $F$  is locally Lipschitz continuous, so there exists some  $\varepsilon > 0$  such that  $F$  is Lipschitz continuous on  $u([0, T]) + \varepsilon\mathbb{B}_E$ . Hence, using Proposition 5.1.7,

$$(7.70) \quad \exists M > 0 \text{ such that } \|x^*\| \leq M \ \forall x^* \in \partial^c F(x), \forall x \in u([0, T]) + \varepsilon\mathbb{B}.$$

Since  $u_\lambda$  converges uniformly to  $u$  on  $[0, T]$ , we can say that  $u_\lambda([0, T]) \subset u([0, T]) + \varepsilon\mathbb{B}$  for  $\lambda$  close enough to zero. From Theorem 7.3.7 iii) and (7.70), it follows that

$$(7.71) \quad \forall \lambda \sim 0, \forall t \in [0, T], \|\!|D\mathcal{F}_\lambda(u_\lambda(t))\|\!| \leq M,$$

and since  $\|\mu_\lambda(t)\|_{\mathcal{M}^R(\Theta)} = 1$ , we deduce

$$(7.72) \quad \forall \lambda \sim 0, \forall t \in [0, T], \|\mu_\lambda(t) \circ D\mathcal{F}_\lambda(u_\lambda(t))\| \leq M.$$

Combining (7.72), (7.64) and (7.69), we obtain that the net  $(\eta_\lambda)_\lambda$  is bounded in  $L^2([0, T], E)$ , hence weakly relatively compact. So, by taking eventually a subnet, we can say that  $\eta_\lambda$  weakly converges to some  $\eta \in L^2([0, T], E)$ . Since we know that  $u_\lambda$  strongly converges to  $u$  in  $L^2([0, T], E)$ , by using a classic argument involving the maximal monotonicity of  $N_C$  (see [83, Proposition 2.16]), it follows that  $\eta(t) \in N_C(u(t))$  for a.e.  $t \geq 0$ .

We end now the proof with item iv). We know that  $\|\mu_\lambda(t)\|_{\mathcal{M}^R(\Theta)} = 1$  for all  $\lambda > 0$  and a.e.  $t \geq 0$ , so  $\|\mu_\lambda\|_{L_{w^*}^\infty([0, T], \mathcal{M}^R(\Theta))} = 1$  for all  $\lambda > 0$ . This means that  $(\mu_\lambda)_{\lambda>0}$  is bounded in  $L_{w^*}^\infty([0, T], \mathcal{M}^R(\Theta))$ , hence relatively weakly\* compact. So, taking eventually a subnet,  $\mu_\lambda$  weakly\* converges to some  $\mu \in L_{w^*}^\infty([0, T], \mathcal{M}^R(\Theta))$ . It remains to verify that  $\mu(t) \in \mathcal{P}^R(\Theta)$  for a.e.  $t \geq 0$ .

Consider any measurable  $I \subset [0, T]$ , and define  $\mathbf{1}_I : [0, T] \rightarrow C(\Theta)$  as follows: if  $t \in I$ , then  $\mathbf{1}_I(t) = \mathbf{1} \in C(\Theta)$ , else  $\mathbf{1}_I(t) = 0$  *i.e.* the null function. This function is well-defined and measurable, and we easily see that it is Bochner integrable, *i.e.*  $\mathbf{1}_I \in L^1([0, T], C(\Theta))$ . From  $\mu_\lambda \xrightarrow[\lambda \rightarrow 0]{w^*} \mu$ , it follows<sup>4</sup>:

$$(7.73) \quad \langle \mu_\lambda, \mathbf{1}_I \rangle_{L_{w^*}^\infty \times L^1} \xrightarrow[\lambda \rightarrow 0]{\mathbb{R}} \langle \mu, \mathbf{1}_I \rangle_{L_{w^*}^\infty \times L^1}$$

On one hand, we know that for all  $\lambda > 0$ ,

$$\langle \mu_\lambda, \mathbf{1}_I \rangle_{L_{w^*}^\infty \times L^1} = \int_0^T \langle \mu_\lambda(t), \mathbf{1}_I(t) \rangle_{\mathcal{M}^R(\Theta) \times C(\Theta)} dt = \int_I \langle \mu_\lambda(t), \mathbf{1} \rangle_{\mathcal{M}^R(\Theta) \times C(\Theta)} dt = \int_I 1 dt = |I|.$$

Using (7.73), we deduce that  $\langle \mu, \mathbf{1}_I \rangle_{L_{w^*}^\infty \times L^1} = |I|$ . On the other hand,

$$(7.74) \quad \langle \mu, \mathbf{1}_I \rangle_{L_{w^*}^\infty \times L^1} = \int_I \langle \mu(t), \mathbf{1} \rangle_{\mathcal{M}^R(\Theta) \times C(\Theta)} dt,$$

<sup>4</sup> $\langle \cdot, \cdot \rangle_{L_{w^*}^\infty \times L^1}$  denotes the duality product between  $L_{w^*}^\infty([0, T], \mathcal{M}^R(\Theta))$  and  $L^1([0, T], C(\Theta))$ .

which means that

$$\int_I \langle \mu(t), \mathbb{1} \rangle_{\mathcal{M}^R(\Theta) \times C(\Theta)} dt = |I| \quad \text{for any measurable } I \subset [0, T].$$

The latter implies that

$$(7.75) \quad \langle \mu(t), \mathbb{1} \rangle_{\mathcal{M}^R(\Theta) \times C(\Theta)} = 1 \quad \text{for a.e. } t \in [0, T].$$

Recall now this alternative characterization of  $\mathcal{P}^R(\Theta)$  (see (7.32)):

$$\mathcal{P}^R(\Theta) = \{ \mu \in \mathcal{M}^R(\Theta) \mid \langle \mu, \mathbb{1} \rangle_{\mathcal{M}^R(\Theta) \times C(\Theta)} = 1 \text{ and } \|\mu\|_{\mathcal{M}^R(\Theta)} \leq 1 \}.$$

Since we already proved (7.75), it remains to verify that  $\|\mu(t)\|_{\mathcal{M}^R(\Theta)} \leq 1$  for a.e.  $t \in [0, T]$ . From the weak\* convergence of  $\mu_\lambda$  to  $\mu$ , and the weakly\* lower semi-continuity

$$(7.76) \quad \|\mu\|_{L_{w^*}^\infty} \leq \liminf_{\lambda \rightarrow 0} \|\mu_\lambda\|_{L_{w^*}^\infty}.$$

Moreover we know that  $\mu_\lambda(t) \in \mathcal{P}^R(\Theta)$  for all  $\lambda > 0$  and a.e.  $t \geq 0$ , which implies that  $\|\mu_\lambda\|_{L_{w^*}^\infty} = 1$  for all  $\lambda > 0$ . In the light of (7.76), we deduce that  $\|\mu\|_{L_{w^*}^\infty} \leq 1$ , which means that  $\|\mu(t)\|_{\mathcal{M}^R(\Theta)} \leq 1$  for a.e.  $t \in [0, T]$ , and ends the proof.  $\blacksquare$

**Proposition 7.3.20.** Let  $F : E \rightarrow Y$  be convex and bounded from below, with  $E$  finite-dimensional and  $Y$  separable. Then, when  $\lambda \downarrow 0$ , we have (taking eventually a subnet)  $\mathcal{F}_\lambda \circ u_\lambda \rightarrow \mathcal{F} \circ u$  in  $C([0, T], C(\Theta))$ .

*Proof.* We start by proving that  $\{\mathcal{F}_\lambda \circ u\}_{\lambda \in ]0, 1[}$  is relatively compact in  $C([0, T], C(\Theta))$ . For this we aim to use Ascoli's Theorem, for which we need to verify the hypotheses. Given a fixed  $t \in [0, T]$ , and using Lemma 7.3.16, we obtain the relative compactness of  $\{\mathcal{F}_\lambda(u(t))\}_{\lambda \in ]0, 1[}$  in  $C(\Theta)$ . Now we prove the equicontinuity of  $\{\mathcal{F}_\lambda \circ u\}_{\lambda \in ]0, 1[}$  in  $C([0, T], C(\Theta))$ . Take  $0 \leq s < t \leq T$ , and any  $\lambda > 0$ . Using the fact that  $u$  is absolutely continuous, and  $\mathcal{F}_\lambda$  is differentiable :

$$(7.77) \quad \|\mathcal{F}_\lambda(u(t)) - \mathcal{F}_\lambda(u(s))\| \leq \int_s^t \left\| \frac{d}{dt} \mathcal{F}_\lambda \circ u(\tau) \right\| d\tau \leq \int_s^t \|\mathbb{D}\mathcal{F}_\lambda(u_\lambda(\tau))\| \| \dot{u}(\tau) \| d\tau.$$

From the estimation we obtained in (7.71), and the Cauchy-Schwarz inequality in  $L^2([0, T], \mathbb{R})$ , we obtain

$$(7.78) \quad \|\mathcal{F}_\lambda(u(t)) - \mathcal{F}_\lambda(u(s))\| \leq M\sqrt{t-s} \|\dot{u}\|_{L^2([0, T], E)}.$$

We deduce from the latter that  $\{\mathcal{F}_\lambda \circ u\}_{\lambda \in ]0, 1[}$  is uniformly Hölder continuous, hence equicontinuous. So we can apply Ascoli's Theorem, and obtain that  $\{\mathcal{F}_\lambda \circ u\}_{\lambda \in ]0, 1[}$  is relatively compact in  $C([0, T], C(\Theta))$ .

Now we pass to the uniform convergence of  $\mathcal{F}_\lambda \circ u_\lambda$  to  $\mathcal{F} \circ u$  in  $C([0, T], C(\Theta))$ . For this, take any  $t \in [0, T]$ ,  $\lambda > 0$ , and use the triangular inequality to write

$$(7.79) \quad \|\mathcal{F}_\lambda \circ u_\lambda(t) - \mathcal{F} \circ u(t)\| \leq \|\mathcal{F}_\lambda \circ u_\lambda(t) - \mathcal{F}_\lambda \circ u(t)\| + \|\mathcal{F}_\lambda \circ u(t) - \mathcal{F} \circ u(t)\|.$$

On one hand, the first term of the right member in (7.79) can be bounded from above by using the mean value theorem:

$$(7.80) \quad \|\mathcal{F}_\lambda \circ u_\lambda(t) - \mathcal{F}_\lambda \circ u(t)\| \leq \sup_{c \in [u_\lambda(t), u(t)]} \|\mathbb{D}\mathcal{F}_\lambda(c)\| \|u_\lambda(t) - u(t)\|.$$

Using the same argument as in (7.71), we obtain some  $M > 0$  such that for all  $\lambda \sim 0$  and any  $t \in [0, T]$ ,

$$(7.81) \quad \|\mathcal{F}_\lambda \circ u_\lambda(t) - \mathcal{F}_\lambda \circ u(t)\| \leq M\|u_\lambda(t) - u(t)\|.$$

Since  $u_\lambda$  converges uniformly to  $u$ , we deduce from the latter that  $\mathcal{F}_\lambda \circ u_\lambda - \mathcal{F}_\lambda \circ u$  converges uniformly to zero. On the other hand, we know from Theorem 7.3.7 that  $\mathcal{F}_\lambda \circ u$  is pointwise convergent to  $\mathcal{F} \circ u$ . Moreover we shown in the first part of this proof that  $\{\mathcal{F}_\lambda \circ u\}_{\lambda \in ]0,1[}$  is relatively compact in  $C([0, T], C(\Theta))$ . These two properties entails the uniform convergence of  $\mathcal{F}_\lambda \circ u$  to  $\mathcal{F} \circ u$ , which means that the second term of the right member of (7.79) converges uniformly to zero when  $\lambda \rightarrow 0$ . So we proved the uniform convergence of  $\mathcal{F}_\lambda \circ u_\lambda$  to  $\mathcal{F} \circ u$  in  $C([0, T], C(\Theta))$ . ■

**Proposition 7.3.21.** For a.e.  $t \geq 0$ ,  $\dot{u}(t) + N_C(u(t)) + \partial^c F(u(t)) \ni 0$ .

*Proof.* Consider any  $w \in L^\infty([0, T], E)$  and  $\lambda > 0$ . Since  $\mathcal{F}_\lambda$  is convex and differentiable, we have for a.e.  $t \in [0, T]$

$$(7.82) \quad \mu_\lambda(t) \circ \mathcal{F}_\lambda(w(t)) - \mu_\lambda(t) \circ \mathcal{F}_\lambda(u_\lambda(t)) - \langle \nabla(\mu_\lambda(t) \circ \mathcal{F}_\lambda)(u_\lambda(t)), w(t) - u_\lambda(t) \rangle \geq 0.$$

Observe that by definition of  $\mu_\lambda$ , we have

$$(7.83) \quad \nabla(\mu_\lambda(t) \circ \mathcal{F}_\lambda)(u_\lambda(t)) = \mu_\lambda(t) \circ D\mathcal{F}_\lambda(u_\lambda(t)) = -\dot{u}_\lambda(t) - \eta_\lambda(t).$$

Using (7.83) together with the fact that  $\mathcal{F}_\lambda \leq \mathcal{F}$  on  $E$  (recall Theorem 7.3.7), we obtain from (7.82) that

$$(7.84) \quad \mu_\lambda(t) \circ \mathcal{F}(w(t)) - \mu_\lambda(t) \circ \mathcal{F}_\lambda(u_\lambda(t)) - \langle -\dot{u}_\lambda(t) - \eta_\lambda(t), w(t) - u_\lambda(t) \rangle \geq 0.$$

This expression is integrable on  $[0, T]$ , and we obtain after integration:

$$\begin{aligned} \int_0^T \langle \mu_\lambda(t), \mathcal{F}(w(t)) \rangle_{\mathcal{M}^R(\Theta) \times C(\Theta)} dt & - \int_0^T \langle \mu_\lambda(t), \mathcal{F}_\lambda(u_\lambda(t)) \rangle_{\mathcal{M}^R(\Theta) \times C(\Theta)} dt \\ & - \int_0^T \langle -\dot{u}_\lambda(t) - \eta_\lambda(t), w(t) - u_\lambda(t) \rangle_E dt \geq 0 \end{aligned}$$

In the latter, we recognize in the two first terms a duality product between  $L^1([0, T], C(\Theta))$  and  $L_{w^*}^\infty([0, T], \mathcal{M}^R(\Theta))$ . The last term involves  $u_\lambda \in C([0, T], E) \subset L^\infty([0, T], E)$  and  $\dot{u}_\lambda + \eta_\lambda \in L^2([0, T], E) \subset L^1([0, T], E)$ , which can be seen as a subspace of  $L^\infty([0, T], E)^*$ . Hence this last inequality can be rewritten as:

$$(7.85) \quad \langle \mu_\lambda, \mathcal{F} \circ w \rangle_{L_{w^*}^\infty \times L^1} - \langle \mu_\lambda, \mathcal{F}_\lambda \circ u_\lambda \rangle_{L_{w^*}^\infty \times L^1} - \langle -\dot{u}_\lambda - \eta_\lambda, w - u_\lambda \rangle_{L^1 \times L^\infty} \geq 0.$$

We can now pass to the limit in (7.85) when  $\lambda$  goes to zero, using Propositions 7.3.19 and 7.3.20 to obtain for all  $w \in L^\infty([0, T], E)$ :

$$(7.86) \quad \langle \mu, \mathcal{F} \circ w \rangle_{L_{w^*}^\infty \times L^1} - \langle \mu, \mathcal{F} \circ u \rangle_{L_{w^*}^\infty \times L^1} - \langle -\dot{u} - \eta, w - u \rangle_{L^1 \times L^\infty} \geq 0.$$

We want to interpret the equation above as a subdifferential inequality. Consider for this the integrand

$$\begin{aligned} \Phi : L^\infty([0, T], E) & \longrightarrow \mathbb{R} \\ w & \longmapsto \int_0^T \phi(t, w(t)) dt, \end{aligned}$$

where  $\phi$  is defined as follows

$$\begin{aligned} \phi : [0, T] \times E & \longrightarrow \mathbb{R} \\ (t, x) & \longmapsto \phi_t(x) := \langle \mu(t), \mathcal{F}(x) \rangle_{\mathcal{M}^R(\Theta) \times C(\Theta)}. \end{aligned}$$

In other words,  $\Phi(w) = \langle \mu, \mathcal{F} \circ w \rangle_{L^\infty_{w^*} \times L^1}$ . The convexity of  $\mathcal{F}$  implies the one of  $\Phi$ , and (7.86) expresses the fact that  $-\dot{u} - \eta \in \partial\Phi(u)$ . Using a celebrated result of Rockafellar ([286, Theorem 4]), we obtain for a.e.  $t \in [0, T]$  that

$$(7.87) \quad -\dot{u}(t) - \eta(t) \in \partial\phi_t(u(t)).$$

But for each  $t \in [0, T]$ ,  $\phi_t = \mu(t) \circ \mathcal{F}$ , so for all  $x \in E$ ,

$$(7.88) \quad \partial\phi_t(x) = \partial(\mu(t) \circ \mathcal{F})(x) \subset \partial^c\mathcal{F}(x).$$

Since  $\partial^c\mathcal{F} = \partial^cF$  from Theorem 7.3.3, we finally obtain from (7.87) and (7.88) that

$$-\dot{u}(t) - \eta(t) \in \partial^cF(u(t)) \text{ for a.e. } t \geq 0.$$

■

**Proposition 7.3.22.** For a.e.  $t \geq 0$ ,  $\dot{u}(t) + (N_C(u(t)) + \partial^cF(u(t)))^0 = 0$ .

*Proof.* To prove the lazyness of the trajectory  $u$ , we will use the one of each  $u_\lambda$  and pass to the limit when  $\lambda \downarrow 0$ . Indeed, we know for all  $\lambda > 0$  that for a.e.  $t \geq 0$ ,

$$(7.89) \quad \dot{u}_\lambda(t) + (N_C(u_\lambda(t)) + \partial^cF(u_\lambda(t)))^0 = 0.$$

Using the variational characterization of  $(N_C(u_\lambda(t)) + \partial^cF(u_\lambda(t)))^0$  as the projection of the origin onto  $N_C(u_\lambda(t)) + \partial^cF(u_\lambda(t))$ , together with the fact that  $0 \in N_C(u_\lambda(t))$ , we obtain for a.e.  $t \geq 0$

$$(7.90) \quad \forall x^* \in \partial^c\mathcal{F}_\lambda(u_\lambda(t)), \langle \dot{u}_\lambda(t), \dot{u}_\lambda(t) + x^* \rangle \leq 0,$$

which means that

$$(7.91) \quad \forall \mu \in \mathcal{P}^R(\Theta), \langle \dot{u}_\lambda(t), \dot{u}_\lambda(t) + \mu \circ D\mathcal{F}_\lambda(u_\lambda(t)) \rangle \leq 0,$$

or equivalently

$$(7.92) \quad \forall \mu \in \mathcal{P}^R(\Theta), \|\dot{u}_\lambda(t)\|^2 + \frac{d}{dt}(\mu \circ \mathcal{F}_\lambda \circ u_\lambda)(t) \leq 0.$$

Our aim is to pass to the limit in (7.92) to prove

$$(7.93) \quad \forall \mu \in \mathcal{P}^R(\Theta), \|\dot{u}(t)\|^2 + \frac{d}{dt}(\mu \circ \mathcal{F} \circ u)(t) \leq 0.$$

Indeed, this is a sufficient condition for the trajectory  $u$  to be lazy, recall Proposition 6.1.15.

Take any  $\psi \in \mathcal{D}(]0, T[, \mathbb{R})$ , the space of infinitely differentiable functions with support unclued in  $]0, T[$ , which is dense in  $L^2([0, T], \mathbb{R})$  (see [83, Appendices]). Consider then the corresponding linear integrand

$$\begin{aligned} \Psi : L^2([0, T], E) &\longrightarrow \mathbb{R} \\ w &\longmapsto \int_0^T \psi(t) \|w(t)\|^2 dt. \end{aligned}$$

It is easy to see that  $\Psi$  is strongly continuous, hence weakly lower semi-continuous. As direct consequence of  $\dot{u}_\lambda \xrightarrow[\lambda \rightarrow 0]{w} \dot{u}$  in  $L^2([0, T], E)$ , we deduce that

$$(7.94) \quad \int_0^T \psi(t) \|\dot{u}(t)\|^2 dt \leq \liminf_{\lambda \rightarrow 0} \int_0^T \psi(t) \|\dot{u}_\lambda(t)\|^2 dt.$$

Now, use an integration by parts together with the fact that  $\psi(0) = \psi(T) = 0$  to write for all  $\mu \in \mathcal{P}^R(\Theta)$

$$(7.95) \quad \int_0^T \psi(t)(\mu \circ \mathcal{F} \circ u)'(t) dt = - \int_0^T \psi'(t)(\mu \circ \mathcal{F} \circ u)(t) dt.$$

Because of Proposition 7.3.20,  $\mu \circ \mathcal{F}_\lambda \circ u_\lambda$  converges uniformly to  $\mu \circ \mathcal{F} \circ u$  in  $C([0, T], \mathbb{R})$ , hence from (7.95) we deduce

$$(7.96) \quad \int_0^T \psi(t)(\mu \circ \mathcal{F} \circ u)'(t) dt = - \lim_{\lambda \rightarrow 0} \int_0^T \psi'(t)(\mu \circ \mathcal{F}_\lambda \circ u_\lambda)(t) dt.$$

Doing again an integration by parts, we obtain

$$(7.97) \quad \int_0^T \psi(t)(\mu \circ \mathcal{F} \circ u)'(t) dt = \lim_{\lambda \rightarrow 0} \int_0^T \psi(t)(\mu \circ \mathcal{F}_\lambda \circ u_\lambda)'(t) dt.$$

Combines now (7.94) and (7.97) to find

$$\int_0^T \psi(t) \left( \|\dot{u}(t)\|^2 + \frac{d}{dt}(\mu \circ \mathcal{F} \circ u)(t) \right) dt \leq \liminf_{\lambda \rightarrow 0} \int_0^T \psi(t) \left( \|\dot{u}_\lambda(t)\|^2 + \frac{d}{dt}(\mu \circ \mathcal{F}_\lambda \circ u_\lambda)(t) \right) dt.$$

If we suppose that  $\psi \in \mathcal{D}_+(]0, T[, \mathbb{R})$ , *i.e.* that  $\psi$  takes positive values, we can deduce from (7.92) that

$$(7.98) \quad \int_0^T \psi(t) \left( \|\dot{u}(t)\|^2 + \frac{d}{dt}(\mu \circ \mathcal{F} \circ u)(t) \right) dt \leq 0.$$

By using the density of  $\mathcal{D}_+(]0, T[, \mathbb{R})$  in  $L^2_+(]0, T[, \mathbb{R})$ , which is a closed convex self-dual cone, we obtain that  $\|\dot{u}(\cdot)\|^2 + \frac{d}{dt}(\mu \circ \mathcal{F} \circ u) \in -L^2_+(]0, T[, \mathbb{R})$ . This proves (7.93), and ends the proof with Proposition 6.1.15. ■

## 7.4 Comments

**Remark 7.4.1** (On the necessity to work in general Banach spaces). When considering Chapters 5 and 6, it might seem superfluous to work with a function  $F$  taking its values in a general ordered Banach space  $Y$ . Indeed, we make the assumption that the cone  $K$  inducing the order in  $Y$  must have a nonempty interior, which reduces the “useful” spaces to work in to  $Y = L^\infty([0, 1], \mathbb{R}^p)$  or  $Y = C([0, 1])$ . When one sees the technicalities induced by this infinite-dimensional Banach setting, we could reasonably wonder if it is worth it.

In fact, even if  $Y = \mathbb{R}^m$ , as soon as  $K$  is not polyhedral, we cannot apply the techniques of Section 7.2 to prove the existence of trajectories. We need in the Section 7.3 to introduce an auxiliary Banach space of continuous functions on the base  $\Theta$ ,  $C(\Theta)$ , which is in general a nonreflexive Banach space.

**Remark 7.4.2** (Seeing  $\partial^c \mathcal{F}_\lambda$  as a Yosida approximation of  $\partial^c F$ ). In Section 6.3.1, we formally defined the resolvent of  $\partial^c F$  for a convex function  $F$  by

$$\forall x \in H, \text{PROX}_{\lambda F}(x) = (Id + \lambda \partial^c F)^{-1}(x) = \bigcup_{\theta \in \Theta} \text{prox}_{\lambda f_\theta}(x).$$

It is a known fact (see [51, Chapter 23]) that  $\text{prox}_{\lambda f_\theta}(x) = x - \lambda \nabla f_{\theta, \lambda}(x)$ . Then, we can write from Proposition 7.3.12 that

$$\partial^c \mathcal{F}_\lambda(x) = \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \nabla f_{\theta, \lambda}(x) = \frac{1}{\lambda} \left( x - \overline{\text{co}}^* \bigcup_{\theta \in \Theta} \text{prox}_{\lambda f_\theta}(x) \right).$$



In other words,

$$\partial^c \mathcal{F}_\lambda = \frac{1}{\lambda} (Id - \overline{\text{co}}^* \text{PROX}_{\lambda F}),$$

that we can interpret as the fact that  $\partial^c \mathcal{F}_\lambda$  is a formal Yosida transform of  $\partial^c F$ .

## Chapter 8

# A continuous inertial approach to multi-objective optimization

Let  $H$  be a Hilbert space, and consider the unconstrained multi-objective problem

$$(MOP) \quad \text{MIN}_{x \in H} F(x)$$

where  $F = (f_1, \dots, f_m) : H \rightarrow (\mathbb{R}^m, \mathbb{R}_+^m)$  is a continuously differentiable function. In Chapter 5, we introduced the *steepest descent* vector field for  $F$ :

$$(8.1) \quad \begin{aligned} s : H &\longrightarrow H \\ x &\longmapsto s(x) := -(\text{co}\nabla f_i(x))^0 \end{aligned}$$

where  $\text{co}\nabla f_i(u)^0$  denotes the element of minimal norm of the convex compact set  $\text{co}\nabla f_i(u)$ . The vector  $s(u)$  is called the *steepest descent direction* at  $x$ , and simply reduces to  $-\nabla f(u)$  if  $m = 1$ . It enjoys the following nice properties, which extends known facts about  $-\nabla f(u)$  in the mono-criteria case (recall Section 5.2):

- i)  $s(u) = 0$  if and only if  $u$  is Pareto critical.
- ii)  $s(u)$  is a *common* descent direction at  $u$  for all the objective functions. More exactly,

$$(8.2) \quad \forall i = \{1, \dots, m\}, \quad \langle \nabla f_i(u), s(u) \rangle \leq -\|s(u)\|^2.$$

- iii) It is the *steepest* common descent direction, in the sense that

$$(8.3) \quad \frac{s(u)}{\|s(u)\|} = \underset{\|d\|=1}{\text{argmin}} \max_{i \in \{1, \dots, m\}} \langle \nabla f_i(u), d \rangle \quad \text{whenever } s(u) \neq 0.$$

In Chapter 6, we introduced and studied the steepest descent dynamic (SD) to solve problem (MOP). It is a dynamic governed by the steepest descent vector field  $s$ , which reduces in this multi-objective case to

$$(SD) \quad \dot{u}(t) + (\text{co}\nabla f_i(u(t)))^0 = 0.$$

The main feature of (SD) is that, when the objective functions  $f_i$  are convex, all its bounded solutions weakly converge to weak Pareto point of (MOP).

When  $m = 1$ , this dynamic corresponds exactly to the classic steepest descent dynamic

$$\dot{u}(t) + \nabla f(u(t)) = 0.$$

It is known that its trajectories have poor convergence rates, when  $t \rightarrow +\infty$ . Thus, the (SD) dynamic for multi-objective optimization problems shares the same slow behaviour, which is

characteristic of first-order methods. This is a clear issue in the multi-objective setting. Indeed, while in scalar optimization we often look for one minimum, in multi-objective problems we aim to have a description of the set of Pareto efficient points. Hence, we need to perform in parallel a high number of trajectories, and it would be better that the trajectories converge quickly.

To circumvent this problem, we propose to introduce an inertial term in our dynamic. Inertial continuous dynamics for scalar optimization have been studied since Polyack [276], and Attouch-Goudou-Redont [29] introduced the so-called Heavy Ball with Friction dynamic:

$$(HBF) \quad \ddot{u}(t) + \gamma\dot{u}(t) + \nabla f(u(t)) = 0.$$

This system has a clear mechanical interpretation. The term  $\ddot{u}(t)$  is the acceleration of the physical point  $u(t)$ , on which acts the sum of two forces: the friction  $-\gamma\dot{u}(t)$  and the vector field  $\nabla f(u(t))$ . Just like a heavy ball sliding down the graph of  $f$ , due to the viscous friction effect, each trajectory tends to stabilize at a local minimum of  $f$ . This (HBF) dynamic has been extended by Cabot-Engler-Gaddat [92], which introduced a variable damping coefficient  $\gamma(\cdot)$ . This variable viscosity is of importance, in the light of the works of Su-Boyd-Candès [303] and Attouch-Chbani-Peypouquet-Redont [22]. Indeed they show that, by taking  $\gamma(t) = \frac{\alpha}{t}$ , we recover after discretization in time the inertial algorithms of Nesterov [251] and Chambolle-Dossal [100].

Our aim in this chapter is to provide a first study of an inertial gradient-based dynamical system for multi-objective optimization. To our knowledge, the combination of both inertial and multi-objective aspects has not been considered before. It is a second-order in time differential equation, which generalizes the (HBF) dynamic to the multi-objective setting, replacing the gradient of  $f$  by the ordered subdifferential of  $F$ . It is called the *Inertial Steepest Descent with Friction* dynamic, and is defined as follows:

$$(ISDF) \quad \ddot{u}(t) + \gamma\dot{u}(t) + \text{co}\nabla f_i(u(t))^0 = 0,$$

where the viscous damping coefficient  $\gamma$  is a fixed positive parameter. Clearly, when there is only one objective function, we recover the (HBF) dynamic.

Thus our program consists in studying the Inertial Steepest Descent with Friction dynamic, by combining the technics of the previous chapters and the ones in [29]. After introducing the Inertial Steepest Descent with Friction dynamic, we investigate in Section 8.1 the existence of solution trajectories for (ISDF) in finite dimensions. In Section 8.2, we study the properties of the trajectories generated by (ISDF). Under a convexity assumption on the objective functions, we show that the bounded trajectories converge to weak Pareto points of the problem. We recover in that case the convergence result concerning (HBF) in [29]. Of course, due to the effects of inertia, (ISDF) is not a descent dynamic, i.e. the values of the cost functions may not decrease over time. But we show that, with an appropriate choice of the initial velocity, the cost values are improved along the trajectory relative to the starting point (see Proposition 8.2.5). This is essential for one which wants to implement a numerical method with restarting, see [303] for instance. In Section 8.3, we illustrate a discretized version of this dynamic, and compare its convergence rate with the (SD) dynamic. The proper analysis of the corresponding algorithm is left for a future work, see Section 8.4.

## 8.1 Existence of trajectories

In this section, we question the existence of solutions for the Cauchy problem associated to (ISDF). Let  $t_0 \in \mathbb{R}$ ,  $T \in ]t_0, +\infty]$ , and  $(u_0, \dot{u}_0) \in \mathbb{H}^2$ . We say that  $u : [t_0, T[ \rightarrow \mathbb{H}$  is a solution

of (ISDF) if  $u$  is continuous on  $[t_0, T[$ , of class  $C^2$  on  $]t_0, T[$ , and satisfies

$$(IMOG) \quad \begin{cases} \ddot{u}(t) = -\gamma\dot{u}(t) + s(u(t)) & \text{for all } t \in ]t_0, T[, \\ u(t_0) = u_0, \dot{u}(t_0) = \dot{u}_0. \end{cases}$$

Here, contrary to Chapter 7, we consider solutions in the *classic* sense. This is because we restrict our analysis to the smooth case.

In view to apply an existence theorem for the dynamical system (IMOG), the key point is the regularity of the steepest descent vector field  $s$ . As a particular case of Proposition 7.1.2, we have:

**Proposition 8.1.1.** Suppose that the gradients  $\nabla f_i : H \rightarrow H$  are Lipschitz continuous on bounded sets. Then  $s$  is  $\frac{1}{2}$ -Holder continuous on bounded sets.

We will face for the (ISDF) dynamic the same regularity problem that with (SD). Thus, here also our existence proof will rely on Peano's existence result:

**Theorem 8.1.2.** (Peano) Let  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be continuous. Then, for all  $x_0 \in H$  and  $t_0 \in \mathbb{R}$ , there exists some  $T > 0$  and  $x : [t_0, t_0 + T[ \rightarrow H$  of class  $C^1$ , such that

$$(8.4) \quad \dot{x}(t) = \phi(x(t)) \quad \text{for all } t \in [t_0, t_0 + T[, \quad \text{with } x(t_0) = x_0.$$

The ingredients are now all gathered to get a first local existence result :

**Proposition 8.1.3.** Suppose that  $H$  has finite dimension, and that the gradients  $\nabla f_i : H \rightarrow H$  are Lipschitz continuous on bounded sets. Then, for all  $t_0 \in \mathbb{R}$ , and for all  $(u_0, \dot{u}_0) \in H \times H$ , there exists some  $T > 0$  and  $u : [t_0, t_0 + T[ \rightarrow H$  of class  $C^2$ , such that

$$(8.5) \quad \ddot{u}(t) = -\gamma\dot{u}(t) + s(u(t)) \quad \text{for all } t \in [t_0, t_0 + T[, \quad \text{with } u(t_0) = u_0, \dot{u}(t_0) = \dot{u}_0.$$

*Proof.* We just need to apply a change of variables in (ISDF) to get a first-order ODE. Let  $\phi : H^2 \rightarrow H^2$  be defined by

$$(8.6) \quad \phi(u, v) := (v, -\gamma v + s(u)).$$

Clearly, from Proposition 8.1.1,  $\phi$  is continuous on  $H^2$ . We can then apply Peano's Theorem at  $t_0$  and  $x_0 := (u_0, \dot{u}_0)$ , to get some  $x : [t_0, t_0 + T[ \rightarrow H^2$  of class  $C^1$  such that (8.4) holds. If we note  $x(t) = (u(t), v(t)) \in H \times H$ , (8.4) can be rewritten as :

$$(8.7) \quad \dot{u}(t) = v(t), \dot{v}(t) = -\gamma v(t) + s(u(t)) \quad \text{for all } t \in [t_0, t_0 + T[, \quad \text{with } u(t_0) = u_0, v(t_0) = \dot{u}_0.$$

Since  $x$  is of class  $C^1$ , we deduce that it is also the case for  $u$  and  $v$ . But from  $\dot{u}(t) = v(t)$ , we can see that  $u$  is of class  $C^2$  and satisfies (8.5). ■

**Remark 8.1.4.** For this result we use Peano's theorem, which asks the space to be finite dimensional. In fact, Peano's theorem can be stated in the Banach space setting, if one asks the vector field involved to be compact. We recall that an operator  $\phi : H \rightarrow H$  is said to be compact whenever it is continuous and maps bounded sets to relatively compact sets. Observe that if the gradients  $\nabla f_i$  are all compact operators, then  $s$  is also compact. Hence, one might want to apply Peano's result in this context. Nevertheless, by reducing (ISDF) to a first-order ODE, we do not deal directly with  $s$  but with  $(u, v) \mapsto (v, -\gamma v + s(u))$ . And it can be easily proved that if  $s$  is compact, then  $v \mapsto v$  is also compact, which would mean that  $H$  has finite dimension.

We can now state our main existence result. To get a global solution on  $[0, +\infty[$ , we do a stronger hypothesis on the gradients.

**Theorem 8.1.5.** *Suppose that  $\mathsf{H}$  has finite dimension, and that the gradients  $\nabla f_i$  are globally Lipschitz continuous. Then, for all  $t_0 \in \mathbb{R}$ ,  $(u_0, \dot{u}_0) \in \mathsf{H} \times \mathsf{H}$ , there exists  $u : [t_0, +\infty[ \rightarrow \mathsf{H}$  of class  $C^2$ , such that*

$$(8.8) \quad \ddot{u}(t) = -\gamma \dot{u}(t) + s(u(t)) \quad \text{for all } t \in [t_0, +\infty[, \text{ with } u(t_0) = u_0, \dot{u}(t_0) = \dot{u}_0.$$

*Proof of Theorem 8.1.5.* Proposition 8.1.3 provides us a local solution and, using Zorn's lemma, we can suppose that it is a maximal solution  $u : [t_0, T[ \rightarrow \mathsf{H}$ , with  $T \in [t_0, +\infty[$ . The whole point is to prove that  $T = +\infty$ . For this, we argue by contradiction by supposing that  $T < +\infty$ . We will show that the solution does not blow up in finite time, and extend it at  $T$  to obtain a contradiction.

Using the fact that the gradients are globally Lipschitz continuous, we obtain from Proposition 7.1.2 the following global growth property for  $s$  :

$$(8.9) \quad \exists c > 0 \text{ s.t. } \forall u \in \mathsf{H}, \quad \|s(u)\| \leq c(1 + \|u\|).$$

From this growth condition, we will obtain some energy estimates on the trajectory.

Let us show that  $\dot{u}$  and  $\ddot{u}$  lie in  $L^\infty([t_0, T], \mathsf{H})$ . For this, we consider as before

$$(8.10) \quad \phi : \mathsf{H}^2 \rightarrow \mathsf{H}^2, \quad (u, v) \mapsto \phi(u, v) = (v, -\gamma v + s(u)).$$

By defining  $x(t) := (u(t), \dot{u}(t))$  for all  $t \in [t_0, T[$ , we see that  $\dot{x}(t) = \phi(x(t))$  on  $[t_0, T[$ . Define  $h(t) := \|x(t) - x(t_0)\|$  on  $[t_0, T[$ , which is continuous on  $[t_0, T[$ . Equip  $\mathsf{H}^2$  with the scalar product inherited from  $\mathsf{H}$ , and note that  $h^2$  is derivable on  $[t_0, T[$ , so we can write for all  $t \in [t_0, T[$  :

$$(8.11) \quad \frac{d}{dt} \frac{1}{2} h^2(t) = \langle \dot{x}(t), x(t) - x(t_0) \rangle = \langle \phi(x(t)), x(t) - x(t_0) \rangle \leq \|\phi(x(t))\| h(t).$$

From the growth condition (8.9) we deduce an upper bound for  $\|\phi(x(t))\|$ . Indeed, for all  $x = (u, v) \in \mathsf{H}^2$ ,

$$\begin{aligned} \|\phi(x)\| &\leq (\|v\| + \|s(u) - \gamma v\|) && \text{using the equivalence between } \ell^1 \text{ and } \ell^2 \text{ norms} \\ &\leq (1 + \gamma)\|v\| + c(1 + \|u\|) && \text{using the triangle inequality with (8.9)} \\ &\leq c_2(1 + \|x\|) && \text{with } c_2 := \sqrt{2} \max\{c; 1 + \gamma\}. \end{aligned}$$

Using the triangle inequality with  $c_3 := c_2(1 + \|x(t_0)\|)$ , it follows for all  $t \in [t_0, T[$  that

$$(8.12) \quad \|\phi(x(t))\| \leq c_3(1 + h(t)).$$

Combining (8.11) and (8.12), we obtain

$$(8.13) \quad \frac{d}{dt} \frac{1}{2} h^2(t) \leq c_3 h(t)(1 + h(t)) \text{ for all } t \in [t_0, T[.$$

We will now conclude by using a Gronwall-type argument. Consider an arbitrary  $\varepsilon \in ]0, T - t_0[$ . After integration of (8.13) on  $[t_0, T - \varepsilon]$ , and using  $h(t_0) = 0$ , we obtain

$$(8.14) \quad \frac{1}{2} h^2(t) \leq \int_{t_0}^t c_3(1 + h(s))h(s) \, ds \text{ for all } t \in [t_0, T - \varepsilon].$$

Since  $h$  is continuous on  $[0, T - \varepsilon]$ , the function  $g : t \in [0, T - \varepsilon] \mapsto c_3(1 + h(t))$  is in  $L^1([0, T - \varepsilon], \mathbb{R})$ . Hence we can apply Lemma C.0.4 (we left it in the Appendix) to obtain

$$(8.15) \quad h(t) \leq \int_{t_0}^t c_3(1 + h(s)) \, ds \text{ for all } t \in [t_0, T - \varepsilon].$$

We easily obtain from (8.15) and  $T < +\infty$  that

$$(8.16) \quad h(t) \leq c_3 T + c_3 \int_{t_0}^t h(s) \, ds \text{ for all } t \in [t_0, T - \varepsilon],$$

so we can use the Gronwall-Bellman's Lemma (see Lemma C.0.3 in the Appendix), and obtain :

$$(8.17) \quad h(t) \leq c_3 T e^{c_3 t} \leq c_3 T e^{c_3 T} \text{ for all } t \in [t_0, T - \varepsilon].$$

Since the upper bound in (8.17) is independent of  $\varepsilon$  and  $t$ , we deduce that  $h \in L^\infty([0, T], \mathbb{R})$ .

From the definition of  $h$ , this means that  $u$  and  $\dot{u}$  lie in  $L^\infty([0, T], \mathbb{H})$ . Moreover, using the growth condition (8.9), it implies that  $s \circ u \in L^\infty([0, T], \mathbb{H})$ , so  $\ddot{u}(t) = (s(u(t)) - \gamma \dot{u}(t))$  lies also in  $L^\infty([0, T], \mathbb{H})$ . Now, since  $T$  is supposed finite, we can say that  $L^\infty([0, T], \mathbb{H}) \subset L^1([0, T], \mathbb{H})$ , so  $u$  can be continuously extended at  $T$  by

$$u(T) := u(0) + \int_0^T \dot{u}(t) \, dt,$$

and we can do the same for  $\dot{u}$ . Hence, we can apply Proposition 8.1.3 at  $t_0 = T$  with  $(u_0, \dot{u}_0) = (u(T), \dot{u}(T))$  to extend the solution  $u(\cdot)$ , which contradicts its maximality. ■

As already discussed in Section 6.1.5, the steepest descent vector field governing the dynamic is not Lipschitz continuous, neither monotone (even in the convex setting). So we cannot use methods from monotone operator theory, and the question of uniqueness of the trajectories remains open in the general context. Nevertheless, under some assumptions, we still can ensure the uniqueness.

**Corollary 8.1.6.** Let  $u$  be a trajectory solution of the Cauchy problem (8.8). Suppose that for all  $t \in [t_0, +\infty[$ , the family  $\{\nabla f_1(u(t)), \dots, \nabla f_m(u(t))\}$  is affinely independent. Then,  $u$  is the unique solution to (8.8).

*Proof.* It is proved in Proposition 6.1.17 that under these hypotheses, for all  $t \in [t_0, +\infty[$ , the steepest descent vector field is locally Lipschitz in the neighbourhood of  $u(t)$ . Hence, it suffices to apply the Cauchy-Lipschitz theorem to derive the uniqueness of  $u$ . ■

## 8.2 Qualitative study of the dynamic

Recall that, for a given function  $\phi : \mathbb{H} \rightarrow \mathbb{H}$  and a nonempty subset  $U \subset \mathbb{H}$ , we note  $Lip(\phi; U)$  the best Lipschitz constant of  $\phi$  over  $U$ , that is  $Lip(\phi; U) := \sup_{x \neq y \in U} \frac{\|\phi(x) - \phi(y)\|}{\|x - y\|}$ . We say that  $\phi$  is Lipschitz over  $U$  whenever  $Lip(\phi; U) < +\infty$ .

### 8.2.1 A dissipative system

We start our study of the dynamic by showing that it is a dissipative system. But before, we need the following chain rule:

**Lemma 8.2.1.** Let  $\phi : \mathbb{H} \rightarrow \mathbb{R}$  and  $u : I \rightarrow \mathbb{H}$ , where  $I$  is a nonempty open subset of  $\mathbb{R}$ . Suppose that  $\phi$  and  $u$  are of class  $C^{1,1}$  on  $I$ , and that  $Lip(\nabla \phi; u(I)) < +\infty$ . Then for a.e.  $t \in I$ ,

$$(8.18) \quad \frac{d^2}{dt^2}(\phi \circ u)(t) \leq Lip(\nabla \phi; u(I)) \|\dot{u}(t)\|^2 + \langle \nabla \phi(u(t)), \ddot{u}(t) \rangle.$$

*Proof.* By hypothesis,  $\dot{u}$  and  $\nabla\phi \circ u$  are locally Lipschitz continuous, hence differentiable almost everywhere. So, from  $\frac{d}{dt}(\phi \circ u)(t) = \langle \nabla\phi(u(t)), \dot{u}(t) \rangle$ , we have for a.e.  $t \in I$

$$(8.19) \quad \frac{d^2}{dt^2}(\phi \circ u)(t) = \left\langle \frac{d}{dt}(\nabla\phi \circ u)(t), \dot{u}(t) \right\rangle + \langle \nabla\phi \circ u(t), \ddot{u}(t) \rangle.$$

Moreover, we have in the second member above (using the Cauchy-Schwarz inequality and the Lipschitz property of  $\nabla\phi$ ) :

$$\begin{aligned} \left\langle \frac{d}{dt}(\nabla\phi \circ u)(t), \dot{u}(t) \right\rangle &= \lim_{h \rightarrow 0} \frac{1}{h} \langle \nabla\phi \circ u(t+h) - \nabla\phi \circ u(t), \dot{u}(t) \rangle \\ &\leq \lim_{h \rightarrow 0} \frac{1}{|h|} \|\nabla\phi \circ u(t+h) - \nabla\phi \circ u(t)\| \|\dot{u}(t)\| \\ &\leq \lim_{h \rightarrow 0} L \frac{1}{|h|} \|u(t+h) - u(t)\| \|\dot{u}(t)\| = L \|\dot{u}(t)\|^2 \end{aligned}$$

where  $L := Lip(\nabla\phi; u(I))$ . ■

Let us prove now the dissipativity of our dynamic :

**Proposition 8.2.2** (Dissipative property). Let  $u : [t_0, T[ \rightarrow \mathbf{H}$  be a solution of (ISDF). For all  $i \in \{1, \dots, m\}$ , define for all  $t \in [t_0, T[$  :

$$(8.20) \quad \mathcal{E}_i(t) := (f_i \circ u)(t) + \frac{1}{\gamma} (f_i \circ u)'(t) + \|\dot{u}(t)\|^2.$$

Then, for a.e.  $t \in [t_0, T[$ , if  $L_i := Lip(\nabla f_i; u([t_0, T[)) < +\infty$ , we have

$$(8.21) \quad \mathcal{E}'_i(t) \leq -\frac{1}{\gamma} \|\ddot{u}(t)\|^2 - \frac{1}{\gamma} (\gamma^2 - L_i) \|\dot{u}(t)\|^2$$

*Proof.* The dissipative property is a direct consequence of the variational characterisation of the projection of 0 over  $\text{co } \nabla f_i(u(t))$  in (ISDF). Indeed, for a.e.  $t \in [t_0, T[$ , we have  $-\ddot{u}(t) - \gamma\dot{u}(t) = \text{proj}_{\text{co}\{\nabla f_i(u(t))\}}(0)$ . It follows that

$$(8.22) \quad \langle \ddot{u}(t) + \gamma\dot{u}(t), \nabla f_i(u(t)) + \dot{u}(t) + \gamma\dot{u}(t) \rangle \leq 0,$$

which is equivalent, after distributing the terms and dividing by  $\gamma$ , to

$$(8.23) \quad \frac{1}{\gamma} \langle \nabla f_i(u(t)), \ddot{u}(t) \rangle + \frac{d}{dt} [(f_i \circ u) + \|\dot{u}\|^2] (t) \leq -\frac{1}{\gamma} \|\ddot{u}(t)\|^2 - \gamma \|\dot{u}(t)\|^2.$$

Use now Lemma 8.2.1 with  $Lip(\nabla f_i; u([t_0, T[)) < +\infty$  to obtain

$$\frac{1}{\gamma} \left( \frac{d^2}{dt^2} (f_i \circ u)(t) - L_i \|\dot{u}(t)\|^2 \right) + \frac{d}{dt} [(f_i \circ u) + \|\dot{u}\|^2] (t) \leq -\frac{1}{\gamma} \|\ddot{u}(t)\|^2 - \gamma \|\dot{u}(t)\|^2,$$

which ends the proof. ■

Proposition 8.2.2 suggests that we need an hypothesis on the parameters to ensure the dissipative property :

$$\mathbf{HP}_i \quad \gamma^2 > L_i \text{ where } L_i := Lip(\nabla f_i; u([t_0, T[)).$$

If  $\mathbf{HP}_i$  holds for all  $i \in \{1, \dots, q\}$  then we just write  $\mathbf{HP}$ . This hypothesis asks the friction parameter  $\gamma$  to be large enough, in order to limit the inertial effects, which induce oscillations (see Example 8.3.1). The hypothesis asks also implicitly the gradients  $\nabla f_i$  to be Lipschitz over

the trajectory (since  $\gamma \in \mathbb{R}$ ). Note that this last property holds whenever  $u(\cdot)$  is bounded, since the gradients are assumed to be Lipschitz continuous on bounded sets (see Corollary 8.2.4). See Section 8.4 for a discussion on this hypothesis.

As a direct consequence of the dissipative nature of the system, we obtain that the values  $(f_i(u(t)))$  are bounded from above by  $\max\{f_i(u_0); \mathcal{E}_i(t_0)\}$  :

**Corollary 8.2.3.** (Upper bound for the values) Let  $u : [t_0, T[ \rightarrow \mathbf{H}$  be a solution of (ISDF), such that  $\mathbf{HP}_i$  holds. Then, for all  $i \in \{1, \dots, m\}$  and  $t \geq t_0$ , we have the following upper bounds :

$$(8.24) \quad f_i(u(t)) \leq \mathcal{E}_i(t_0) + (f_i(u_0) - \mathcal{E}_i(t_0))e^{-\gamma(t-t_0)}.$$

*Proof.* It is a trivial consequence of the monotonicity property of  $\mathcal{E}_i$  obtained in Proposition 8.2.2. Indeed, we obtain for all  $t \in [t_0, +\infty[$  :

$$(8.25) \quad \frac{1}{\gamma}(f_i \circ u)'(t) \leq \mathcal{E}_i(t_0) - (f_i \circ u)(t).$$

The conclusion follows Gronwall's Lemma (see Lemma C.0.3), applied to  $t \mapsto (f_i \circ u)(t) - \mathcal{E}_i(t_0)$ . ■

This upper bound for the values has two interesting consequences. The first one is immediate, and gives a useful sufficient condition for the trajectory  $u(\cdot)$  to be bounded:

**Corollary 8.2.4.** Suppose that there exists  $i \in \{1, \dots, m\}$  such that  $f_i$  is coercive, and globally  $L_i$ -Lipschitz continuous, with  $\gamma^2 > L_i$ . Then any trajectory of (ISDF) is bounded.

The second consequence is that it tells us how to enforce the interesting property  $f_i(u(\cdot)) \leq f_i(u_0)$ . Indeed, we know that this dynamic is not a descent method for the functions because of the inertial effects which can create damped oscillations. But at least, one can choose appropriately the initial velocity so that each point on the trajectory is better than the initial one.

**Corollary 8.2.5.** Suppose that  $\mathbf{HP}_i$  holds for some  $i \in \{1, \dots, q\}$ . For all  $u_0 \in \mathbf{H}$ , if  $\dot{u}_0 \in \mathbf{H}$  is chosen to satisfy

$$(8.26) \quad \langle \nabla f_i(u_0), \dot{u}_0 \rangle \leq -\gamma \|\dot{u}_0\|^2,$$

then  $f_i(u(t)) \leq f_i(u_0)$  for all  $t \geq t_0$ . In particular, for all  $\lambda \in [0, \frac{1}{\gamma}]$ ,  $\dot{u}_0 = \lambda s(u_0)$  satisfies (8.26).

**Remark 8.2.6.** Observe that the set of vectors satisfying (8.26) recalls the notion of pseudo-gradient introduced by Miglierina [247].

*Proof of Corollary 8.2.5.* We saw in Corollary 8.2.3 that the conclusion holds whenever  $f_i(u_0) - \mathcal{E}_i(t_0) \geq 0$ . This condition, once rewritten, is exactly (8.26). Now, consider  $\dot{u}_0 = \lambda s(u_0)$  for some  $\lambda \in [0, \frac{1}{\gamma}]$ . We recall that this steepest descent direction satisfies for all  $i \in \{1, \dots, m\}$ , see (8.2),

$$(8.27) \quad \|s(u_0)\|^2 + \langle \nabla f_i(u_0), s(u_0) \rangle \leq 0.$$

So we deduce that (8.26) holds for  $\lambda s(u_0)$ . ■



### 8.2.2 Energy estimations

We will now use the dissipative property of the system to deduce energy estimations for a global solution of (ISDF).

**Proposition 8.2.7.** (Energy estimations) Let  $u : [t_0, +\infty[ \rightarrow \mathbb{H}$  be a bounded global solution of (ISDF) satisfying **HP**. Then,

- (i) For all  $i \in \{1, \dots, m\}$ ,  $\mathcal{E}_i(t) \downarrow \mathcal{E}_i^\infty \in \mathbb{R}$  whenever  $t \rightarrow +\infty$ .
- (ii)  $\dot{u} \in L^\infty([t_0, +\infty[, \mathbb{H}) \cap L^2([t_0, +\infty[, \mathbb{H})$  and  $\lim_{t \rightarrow +\infty} \|\dot{u}(t)\| = 0$ .
- (iii)  $\ddot{u} \in L^\infty([t_0, +\infty[, \mathbb{H}) \cap L^2([t_0, +\infty[, \mathbb{H})$  and  $\liminf_{t \rightarrow +\infty} \|\ddot{u}(t)\| = 0$ .
- (iv) For all  $i \in \{1, \dots, m\}$ ,  $(f_i \circ u)' \in L^\infty([t_0, +\infty[, \mathbb{R})$  and  $\lim_{t \rightarrow +\infty} (f_i \circ u)'(t) = 0$ .
- (v) For all  $i \in \{1, \dots, m\}$ ,  $(f_i \circ u) \in L^\infty([t_0, +\infty[, \mathbb{R})$  and  $\lim_{t \rightarrow +\infty} (f_i \circ u)(t) = \mathcal{E}_i^\infty$ .
- (vi) For all  $i \in \{1, \dots, m\}$ , there exists  $\theta_i \in L^\infty([t_0, +\infty[, \mathbb{R})$  such that for all  $t \in [t_0, T[$ ,

$$m\ddot{u}(t) + \gamma\dot{u}(t) + \sum_{i=1}^m \theta_i(t) \nabla f_i(u(t)) = 0 \text{ with } \theta(t) \in \Delta_m.$$

In particular, it follows that  $\sum_{i=1}^m \theta_i(\cdot) (f_i \circ u)' \in L^1([t_0, +\infty[, \mathbb{H})$ .

*Proof.* We start by proving that  $\dot{u} \in L^\infty([t_0, +\infty[, \mathbb{H})$ , from which the other results will follow easily. Take any  $i \in \{1, \dots, m\}$ , and define  $c := \inf_{t \geq t_0} f_i(u(t)) - \mathcal{E}_I(t_0)$  and  $M := \max_{i \in \{1, \dots, m\}} \sup_{t \geq t_0} \|\nabla f_i(u(t))\|$ .

Given that the gradients  $\nabla f_i$  are Lipschitz continuous on bounded sets, we deduce (using the mean value theorem) that the functions  $f_i$  are bounded on bounded sets. Since the trajectory is bounded, it follows that  $M$  and  $c$  are finite. In particular, it implies that  $\ddot{u} + \gamma\dot{u} \in L^\infty([t_0, +\infty[, \mathbb{H})$ , since, according to (ISDF), we have for a.e.  $t \geq t_0$  that  $-\ddot{u}(t) - \gamma\dot{u}(t) \in \text{co } \nabla f_i(u(t))$  which is bounded by  $M$ .

Using the monotonicity property of  $\mathcal{E}_i$  (see Proposition 8.2.2), we have for all  $t \geq t_0$ :

$$(8.28) \quad 0 \geq \mathcal{E}_i(t) - \mathcal{E}_i(t_0) \geq \|\dot{u}(t)\|^2 + \frac{1}{\gamma} (f_i \circ u)'(t) + c.$$

Using Cauchy-Schwarz inequality and the definition of  $M$ , one has

$$(8.29) \quad (f_i \circ u)'(t) = \langle \nabla f_i(u(t)), \dot{u}(t) \rangle \geq -\|\nabla f_i(u(t))\| \|\dot{u}(t)\| \geq -M \|\dot{u}(t)\|.$$

If we note  $b = \frac{1}{\gamma} M$ , we obtain

$$(8.30) \quad 0 \geq \|\dot{u}(t)\|^2 - b \|\dot{u}(t)\| + c.$$

If we consider now the real polynomial  $X^2 - bX + c$ , we can see that it takes negative values on a compact interval, independent of  $t$ . Since  $\|\dot{u}(t)\|$  lies therein, we conclude that  $\dot{u} \in L^\infty([t_0, +\infty[, \mathbb{H})$ .

We can now derive the other properties, and we start with (i). The decreasing property of the energies  $\mathcal{E}_i$  (see Proposition 8.2.2) ensures the existence of a limit  $\mathcal{E}_i^\infty$ , taking eventually the value  $-\infty$ . But now we can prove that for all  $i \in \{1, \dots, m\}$ ,  $\mathcal{E}_i^\infty \in \mathbb{R}$ . Indeed, using the same inequality as in (8.29),

$$(8.31) \quad \mathcal{E}_i^\infty = \lim_{t \rightarrow +\infty} \mathcal{E}_i(t) \geq \inf_{t \geq t_0} f_i(u(t)) - \frac{1}{\gamma} M \|\dot{u}\|_{L^\infty([t_0, +\infty[, \mathbb{H})} > -\infty.$$

We now prove (iii). Since  $\ddot{u} + \gamma\dot{u}$  and  $\dot{u}$  lie in  $L^\infty([t_0, +\infty[, H)$ , we directly obtain that  $\ddot{u} \in L^\infty([t_0, +\infty[, H)$ . For the  $L^2$  estimation, use Proposition 8.2.2 to obtain:

$$(8.32) \quad \frac{1}{\gamma} \int_{t_0}^{+\infty} \|\ddot{u}(t)\|^2 dt \leq \int_{t_0}^{+\infty} -\frac{d}{dt} \mathcal{E}_i(t) dt = \mathcal{E}_i(t_0) - \mathcal{E}_i^\infty.$$

It follows that  $\ddot{u} \in L^2([t_0, +\infty[, H)$ , and then,  $\liminf_{t \rightarrow +\infty} \|\ddot{u}(t)\| = 0$ .

Let us now to prove (ii). Using exactly the same argument as for  $\ddot{u}$ , one obtains  $\dot{u} \in L^2([t_0, +\infty[, H)$ . Moreover, we know that  $\dot{u}$  is Lipschitz continuous on  $[t_0, +\infty[$  (since  $\ddot{u} \in L^\infty([t_0, +\infty[, H)$ ), so it follows that  $\lim_{t \rightarrow +\infty} \|\dot{u}(t)\| = 0$ .

We continue with items (iv) and (v). From Cauchy-Schwarz inequality,  $|(f_i \circ u)'(t)| \leq M\|\dot{u}(t)\|$  for all  $t \geq t_0$ . As a direct consequence of (ii), we deduce  $(f_i \circ u)' \in L^\infty([t_0, +\infty[, H)$  and  $\lim_{t \rightarrow +\infty} (f_i \circ u)'(t) = 0$ . Then it follows directly from (i) that  $\lim_{t \rightarrow +\infty} (f_i \circ u)(t) = \mathcal{E}_i^\infty$ , and  $(f_i \circ u) \in L^\infty([t_0, +\infty[, H)$ .

We end the proof with item (vi). It is clear from the definition of (ISDF) that for all  $t \geq t_0$ , there exists  $\theta(t) = (\theta_i(t)) \in \Delta_m$  such that

$$\ddot{u}(t) + \gamma\dot{u}(t) + \sum_{i=1}^m \theta_i(t) \nabla f_i(u(t)) = 0.$$

To get  $\theta_i \in L^\infty([t_0, +\infty[, \mathbb{R})$ , the whole point is to verify that it can be taken measurable. For this, we write  $\theta(t)$  as a solution of the following optimality problem

$$\theta(t) \in \operatorname{argmin}_{\theta \in S^q} j(t, \theta), \quad \text{where } j(t, \theta) := \left\| \sum_{i=1}^m \theta_i \nabla f_i(u(t)) \right\|.$$

Since  $j$  is a Caratheodory function, we are guaranteed of the existence of a measurable selection  $\theta : t \mapsto \theta(t) \in \operatorname{argmin}_{\theta \in \Delta_m} j(t, \theta)$  (see B.1.2). Now we can write

$$\sum_{i=1}^m \theta_i(t) (f_i \circ u)'(t) = \sum_{i=1}^m \theta_i(t) \langle \nabla f_i(u(t)), \dot{u}(t) \rangle = \langle -\ddot{u}(t) - \gamma\dot{u}(t), \dot{u}(t) \rangle$$

where  $\dot{u}, \ddot{u} \in L^2([t_0, +\infty[, H)$ . So, using the Cauchy-Schwarz inequality and the measurability of  $\theta_i$ , we get directly that  $\sum_{i=1}^m \theta_i(\cdot) (f_i \circ u)' \in L^1([t_0, +\infty[, H)$ .  $\blacksquare$

### 8.2.3 Convergence of the trajectories

We present here the main result of this section. Under a convexity assumption, we show that the bounded trajectories of (ISDF) weakly converge to a solution. Recall that the **HP** hypothesis asks in particular the gradients  $\nabla f_i$  to be Lipschitz continuous in a neighborhood of the trajectory.

**Theorem 8.2.8.** *Suppose that the objective functions  $f_i$  are convex. Then any bounded trajectory of (ISDF)  $u : [t_0, +\infty[ \rightarrow H$  satisfying **HP** converges weakly to a weak Pareto optimum.*

We sketch here the main points of the proof. The convergence essentially relies on Opial's Lemma that we recall below (note  $\Omega[u(t)]$  the set of weak sequential cluster points of the trajectory) :

**Lemma 8.2.9** (Opial). Let  $S$  be a non empty subset of  $H$ , and  $u : [t_0, +\infty[ \rightarrow H$ . Assume that

- (i)  $\Omega[u(t)] \subset S$ ;
- (ii) for every  $z \in S$ ,  $\lim_{t \rightarrow +\infty} \|u(t) - z\|$  exists.

Then  $u(t)$  weakly converges to some element  $u^\infty \in S$ .

Like in Chapter 6, it is applied to the set

$$S := \{x \in H \mid f_i(x) \leq \lim_{t \rightarrow +\infty} f_i(u(t)) \text{ for all } i \in \{1, \dots, m\}\},$$

for which (i) is easy to obtain. The key point to prove the Féjer property (ii) is that  $h(t) := \frac{1}{2}\|u(t) - z\|^2$  satisfies a differential inequality. We will use for this the following result from [29, Lemma 4.2] or [30, Lemma 2.3] :

**Lemma 8.2.10.** Let  $h \in \mathcal{C}^1([t_0, +\infty[, \mathbb{R})$  be a positive function satisfying  $\ddot{h} + \gamma\dot{h} \leq g$  where  $\gamma > 0$  and  $g \in L^1([t_0, +\infty[, \mathbb{R})$ . Then  $\lim_{t \rightarrow +\infty} h(t)$  exists.

Once obtained the weak convergence of the trajectory, the characterisation of its limit point as a weak Pareto point is a direct consequence of the upper semi-continuity of  $u \rightrightarrows \text{co } \nabla f_i(u)$  (see Proposition 5.1.11) :

**Lemma 8.2.11.** Assume that the objective function  $\{f_1, \dots, f_m\}$  are convex. If  $u_n \xrightarrow{w} u_\infty$  and  $u_n^* \xrightarrow{n \rightarrow +\infty} 0$  with  $u_n^* \in \text{co } \nabla f_i(u_n)$ , then  $0 \in \text{co } \nabla f_i(u_\infty)$ .

*Proof of Theorem 8.2.8.* Since  $u(\cdot)$  is bounded, there exists some  $t_n \rightarrow +\infty$  such that  $u(t_n)$  converges weakly to some  $u_\infty$ . For all  $i \in \{1, \dots, m\}$ , since  $f_i$  is convex continuous, it is in particular weakly semi-continuous. Hence, using Proposition 8.2.7 we get

$$(8.33) \quad f_i(u_\infty) \leq \liminf_{n \rightarrow +\infty} f_i(u(t_n)) = \lim_{t \rightarrow +\infty} f_i(u(t)).$$

This proves that  $\Omega[u(t)] \subset S \neq \emptyset$ . To obtain convergence of the trajectory through Opial's Lemma, it remains to prove the Fejer property (ii). That is, given some  $z \in S$ , prove that  $\lim_{t \rightarrow +\infty} \|u(t) - z\|$  exists.

Define  $h(t) := \frac{1}{2}\|u(t) - z\|^2$  for all  $t \geq 0$ . Since  $\dot{u}$  is absolutely continuous, then  $h$  is twice differentiable for a.e.  $t \in [0, +\infty[$ , and

$$(8.34) \quad \dot{h}(t) = \langle \dot{u}(t), u(t) - z \rangle,$$

$$(8.35) \quad \ddot{h}(t) = \langle \ddot{u}(t), u(t) - z \rangle + \|\dot{u}(t)\|^2.$$

A linear combination of (8.34) and (8.35) gives

$$(8.36) \quad \ddot{h}(t) + \gamma\dot{h}(t) = \|\dot{u}(t)\|^2 + \langle -\ddot{u}(t) - \gamma\dot{u}(t), z - u(t) \rangle.$$

Let  $\theta_i(t) \in \Delta_m$  be such that  $-\ddot{u}(t) - \gamma\dot{u}(t) = \sum_{i=1}^m \theta_i(t) \nabla f_i(u(t))$ , then we can rewrite

$$(8.37) \quad \ddot{h}(t) + \gamma\dot{h}(t) = \|\dot{u}(t)\|^2 + \sum_{i=1}^m \theta_i(t) \langle \nabla f_i(u(t)), z - u(t) \rangle.$$

For any  $i \in \{1, \dots, m\}$ , we use the monotone property of  $\mathcal{E}_i$  and  $z \in S$  (recall that  $\mathcal{E}_i^\infty = \lim_{t \rightarrow +\infty} f_i(u(t))$ ) together with the convexity of  $f_i$ , to obtain for all  $t \in [0, +\infty[$  :

$$(8.38) \quad \begin{aligned} \mathcal{E}_i(t) &= f_i(u(t)) + \frac{1}{\gamma}(f_i \circ u)'(t) + \|\dot{u}(t)\|^2 \\ &\geq \mathcal{E}_i^\infty \geq f_i(z) \geq f_i(u(t)) + \langle \nabla f_i(u(t)), z - u(t) \rangle. \end{aligned}$$

Thus, it follows from (8.37) and (8.38) that

$$(8.39) \quad \ddot{h}(t) + \gamma\dot{h}(t) \leq 2\|\dot{u}(t)\|^2 + \frac{1}{\gamma} \sum_{i=1}^m \theta_i(t) (f_i \circ u)'(t),$$

where the right member of (8.39) lies in  $L^1([t_0, +\infty[, \mathbb{H})$  (see Proposition 8.2.7).

Thus, hypothesis of Lemma 8.2.10 is satisfied, and  $\lim_{t \rightarrow +\infty} h(t)$  exists. It follows from Opial's Lemma that  $u(t)$  weakly converges to some  $u_\infty \in S$ . It remains to prove that  $u_\infty$  is a weak Pareto. In (ISDF), we have  $-\ddot{u}(t) - \gamma \dot{u}(t) \in \text{co } \nabla f_i(u(t))$ , where (see Proposition 8.2.7)

$$u_n \xrightarrow[n \rightarrow +\infty]{w} u_\infty \text{ and } \liminf_{t \rightarrow +\infty} \|\ddot{u}(t) + \gamma \dot{u}(t)\| = 0.$$

Thus, we can apply Lemma 8.2.11 to obtain  $0 \in \text{co } \nabla f_i(u_\infty)$ . Following Theorem 5.2.10, this is equivalent for  $u_\infty$  to be a weak Pareto point. ■

**Remark 8.2.12.** If the objective functions are not convex, we still can say something on the limits points: each weak limit point of a bounded trajectory of (ISDF) is a critical Pareto point (see Proposition 8.2.7 and 8.2.11).

### 8.3 Examples and numerical results

**Example 8.3.1.** Consider two quadratic functions from  $\mathbb{R}^2$  to  $\mathbb{R}$ , defined by  $f_1(x, y) = \frac{1}{2}(x + 1)^2 + \frac{1}{2}y^2$  and  $f_2(x, y) = \frac{1}{2}(x - 1)^2 + \frac{1}{2}y^2$ . The corresponding Pareto set is  $[-1, +1] \times \{0\}$  and the steepest descent vector field is given by :

$$(8.40) \quad s(x, y) = \begin{cases} -(x - 1, y) & \text{if } x > 1, \\ -(0, y) & \text{if } -1 \leq x \leq 1, \\ -(x + 1, y) & \text{if } x < -1. \end{cases}$$

Figure 8.1 shows some trajectories of the (ISDF) dynamic, with the steepest descent vector field plotted in background. We took the Cauchy data  $(u_0, \dot{u}_0)$  randomly. Here the trajectories are computed exactly, since in this simple example (ISDF) can be solved explicitly. We can observe the following: all the trajectories converge to a Pareto point, the dynamic is clearly not a descent method, and can be highly oscillating whenever the friction parameter is too close to zero.

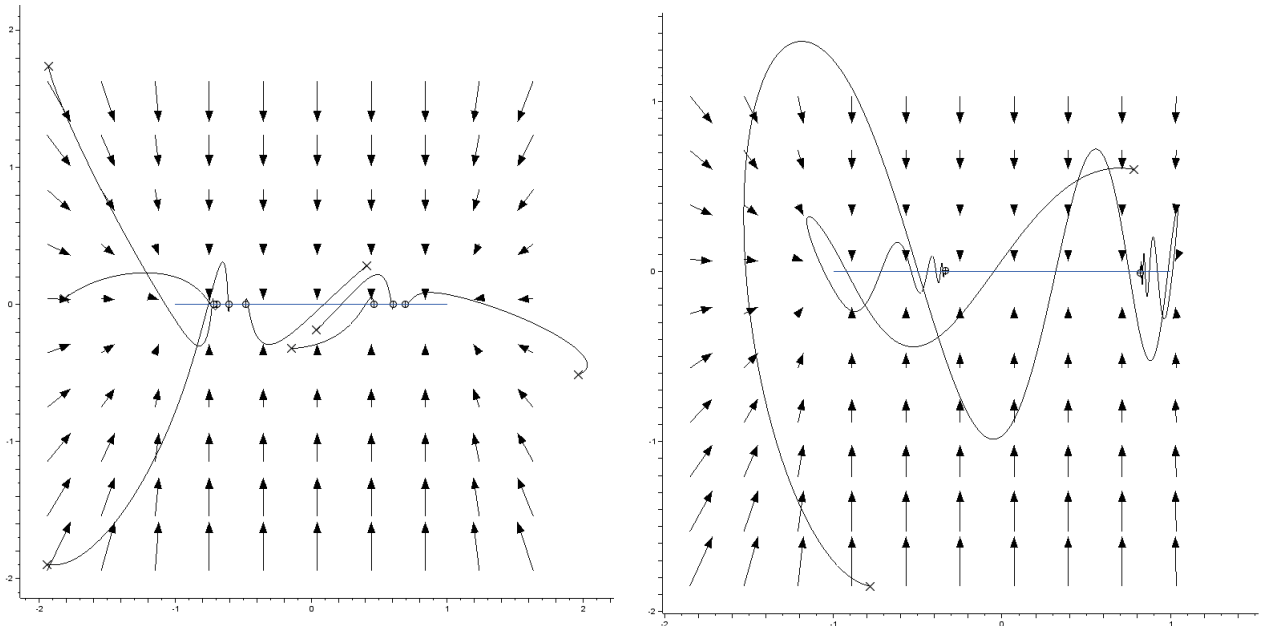


Figure 8.1: Friction parameter  $\gamma = 1$  (left) and  $\gamma = 0.1$  (right). For each trajectory, the initial point is indicated by the symbol  $\times$ , and the limit point by  $\oplus$ .

**Example 8.3.2.** Let  $f_1(x, y) = \frac{1}{2}(x^2 + y^2)$  and  $f_2(x, y) = x$ , that we already seen in Example 8.3.2. The corresponding Pareto set is  $] -\infty, 0] \times \{0\}$ , plotted in blue in Figure 8.2. Once computed, we see that the steepest descent vector field is defined according to three areas of the plane (these areas are delimited by red lines in Figure 8.2) :

$$(8.41) \quad s(x, y) = \begin{cases} -(1, 0) & \text{if } x \geq 1, \\ -(x, y) & \text{if } (x - \frac{1}{2})^2 + y^2 \leq \frac{1}{4}, \\ \frac{-1}{(x-1)^2+y^2}(y^2, y(1-x)) & \text{else.} \end{cases}$$

In this case we plot the trajectories using an explicit discretization in time of (IMOG):

$$\begin{aligned} \frac{u_{n+1} - 2u_n + u_{n-1}}{\tau^2} + \gamma \frac{u_{n+1} - u_n}{\tau} - s(u_n) &= 0 \\ \Leftrightarrow u_{n+1} &= u_n + \frac{1}{1+\tau\gamma}(u_n - u_{n-1}) + \frac{\tau^2}{1+\tau\gamma}s(u_n). \end{aligned}$$

We take  $\gamma = 1$ ,  $\tau = 0.05$ , and consider a randomly chosen initial data. More exactly, generate randomly  $(u_0, \dot{u}_0)$ , and take  $u_1 = u_0 + \tau \dot{u}_0$ .

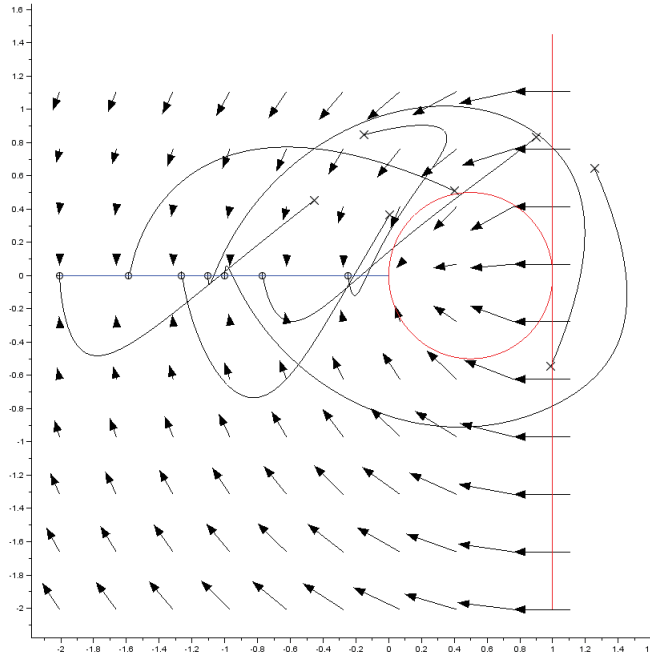


Figure 8.2: For each trajectory, the initial point is indicated by the symbol  $\times$ , and the limit point by  $\oplus$ .

**Remark 8.3.3** (Towards Fast Iterative Methods for Multi-Objective optimization). The study of the Inertial Steepest Descent with Friction dynamic should be a first step towards fast numerical methods to solve multi-objective optimization problems. Future works should start by considering the (ISDF) dynamic with a variable damping coefficient,

$$\ddot{u}(t) + \gamma(t)\dot{u}(t) + \text{co}\nabla f_i(u(t))^0,$$

where  $\gamma(t)$  should be able to tend to zero. In the mono-criteria case, this dynamic has been studied only very recently [92, 303, 22]. The analysis of this multi-objective generalization might be very delicate, because of the assumption we need on the damping parameter  $\gamma^2 > L$ .

The idea is to develop, in a second time, fast converging Nesterov-like algorithms for multi-objective problems. They would write as follows:

$$\begin{aligned} y_k &= x_k + \lambda s(x_k), \\ x_{k+1} &= y_k + \beta_n(y_k - y_{k-1}) \end{aligned}$$

where  $\lambda > 0$  is a stepsize, and  $\beta_n \in [0, 1[$  is an inertial parameter converging to 1. To guarantee the fast convergence of the method,  $\beta_n$  shall be finely tuned, look for instance in the works of Nesterov [251] and Beck-Teboulle [53].

**Example 8.3.4.** Let  $F = (f_1, f_2) : \mathbb{R}^n \rightarrow \mathbb{R}^2$  be the function defined for all  $x = (x_1, \dots, x_n)$  by:

$$f_1(x) = \left( \sum_{i=1}^n x_i^2 - 10 \cos(2\pi x_i) + 10 \right)^{\frac{1}{4}} \quad \text{and} \quad \left( \sum_{i=1}^n (x_i - 1.5)^2 - 10 \cos(2\pi(x_i - 1.5)) + 10 \right)^{\frac{1}{4}}.$$

It is the function considered in Example 6.3.5, in which we removed the square roots, so that  $F$  can be smooth. We compare here three numerical methods:

- The first order steepest descent, used in Examples 6.3.4 and 6.3.5.
- The Euler discretization of the (ISDF) dynamic, used in Example 8.3.2.
- A FISTA-like version of the steepest descent, defined as follows:

$$\begin{aligned} y_k &= x_{k-1} + \tau s(x_k) \\ x_k &= y_k + \frac{t_{k-1}-1}{t_k}(y_k - y_{k-1}), \quad \text{where } t_k = \frac{1 + \sqrt{1 + 4t_{k-1}^2}}{2}. \end{aligned}$$

We take  $n = 10$ ,  $\tau = 0.5$ ,  $\gamma = 1$ , and run in parallel these methods for 1000 iterations. The result is presented in Figure 8.3, in which we measure on  $k \in \{1, \dots, 1000\}$  the relative decay  $\max \left\{ \frac{f_i(x_k) - f_i(x_{1000})}{f_i(x_0) - f_i(x_{1000})}, i \in \{1, 2\} \right\}$ .

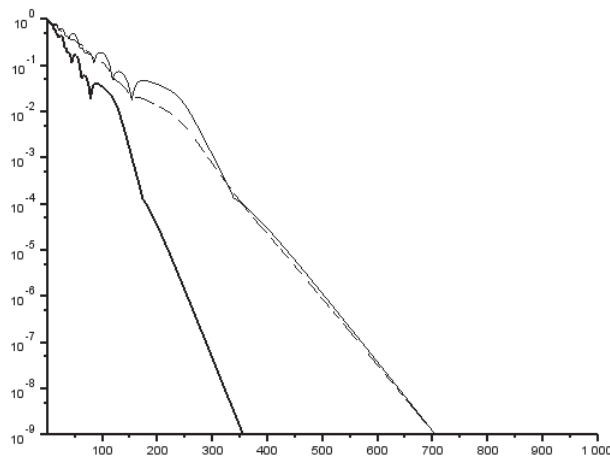


Figure 8.3: Convergence rates for the steepest descent method (dotted curve), the discretized (ISDF) (continuous curve) and its improvement using FISTA-like decreasing friction (thick curve).

## 8.4 Comments and perspectives

**Remark 8.4.1** (On the hypothesis on the friction parameter). To ensure the convergence of the trajectories of (ISDF), we assume in Section 8.2 the relation  $\gamma^2 > L$ , where  $\gamma$  is the damping parameter and  $L$  is a Lipschitz constant for the gradients of the objective functions  $\{f_1, \dots, f_m\}$ . When we reduce to the mono-objective case ( $m = 1$ ), we recover the (HBF) dynamic

$$(HBF) \quad \ddot{u}(t) + \gamma\dot{u}(t) + \nabla f(u(t)) = 0$$

for which such assumption is not needed to guarantee the convergence of the trajectories, see [29]. Thus, one might think that this hypothesis **HP** is not necessary, and is just a consequence of a ‘bad’ choice of Lyapunov function in Proposition 8.2.2. In addition, the analysis of (HBF) relies on the Lyapunov function

$$\mathcal{E}(t) = f(u(t)) + \|\dot{u}(t)\|^2,$$

while our energy functions involve in addition the derivatives of the  $f_i$ ’s:

$$\mathcal{E}_i(t) = f_i(u(t)) + \frac{1}{\gamma} \langle \nabla f_i(u(t)), \dot{u}(t) \rangle + \|\dot{u}(t)\|^2.$$

Thus, the analysis of (ISDF) and (HBF) are of different nature.

Recently, Attouch and Maingé [30] studied the following extension of (HBF):

$$\ddot{u}(t) + \gamma\dot{u}(t) + \nabla f(u(t)) + Au(t) = 0,$$

where  $f : \mathbb{H} \rightarrow \mathbb{R}$  is a convex differentiable function, and  $A : \mathbb{H} \rightarrow \mathbb{H}$  is a  $\beta$ -cocoercive<sup>1</sup> maximal monotone operator. The solutions of this dynamic are proved to weakly converge, when  $t \rightarrow +\infty$ , under the assumption  $\gamma^2 > \frac{1}{\beta}$ . As underlined by the authors, this assumption only involves the cocoercivity of the *nonpotential part* of the operator  $\nabla f + A$  which governs the dynamic. Thus, the parallel can be done with our dynamic (ISDF), which is governed by an operator  $s : \mathbb{H} \rightarrow \mathbb{H}$  which does not derive from a potential (see Example 6.1.12). Of course,  $-s$  is not monotone, and it is not clear whether it is  $\frac{1}{L}$ -cocoercive, with  $L := \max_{i \in \{1, \dots, m\}} \text{Lip}(\nabla f_i, \mathbb{H})$ .

But the similarities between those dynamics is striking.

**Remark 8.4.2** (Some things left behind). In this chapter, we did not take into account a constraint  $C$  for (MOP). This is because the analysis of a nonsmooth operator  $N_C$  is really difficult once we introduce inertia into the dynamic. From a mechanical point of view, in this second-order dynamic (ISDF), when a trajectory reaches the boundary of the constraint, there is a shock, which is hard to manage. Nevertheless, we still can penalize the constraint with a smooth barrier function  $\phi$ , and solve the relaxed problem

$$(MOP) \quad \underset{x \in \mathbb{H}}{\text{MIN}} \quad (f_1(x) + \phi(x), \dots, f_m(x) + \phi(x)).$$

The corresponding (ISDF) dynamic would be

$$(ISDF) \quad \ddot{u}(t) + \gamma\dot{u}(t) + (\nabla\phi(u(t)) + \text{co } \nabla f_i(u(t)))^0.$$

For more details on this approach, see [10].

An other missing point in this chapter, is the quantitative analysis of the trajectories around their limit point. We think that it will be worth it to study the convergence rate of this dynamic whenever we will be able to let the friction parameter to tend to zero.

<sup>1</sup>An operator  $A : \mathbb{H} \rightarrow \mathbb{H}$  is said to be  $\beta$ -cocoercive if, for all  $x, y \in \mathbb{H}$ ,  $\langle Ax - Ay, x - y \rangle \geq \beta \|Ax - Ay\|^2$  (see [51, Definition 4.4]). For instance, for a convex Fréchet differentiable function  $f : \mathbb{H} \rightarrow \mathbb{R}$ , its gradient  $\nabla f : \mathbb{H} \rightarrow \mathbb{H}$  is  $\beta$ -cocoercive if and only if it is  $\frac{1}{\beta}$ -Lipschitz continuous. It is the Baillon-Haddad theorem, see [51, Corollary 18.16].

# Appendices





# Appendix A

## Variational analysis tools in Banach spaces

In this chapter, we give in detail the proofs of some results exposed in Chapter 2. This chapter is structured exactly as the introducing Chapter, with three sections corresponding to the ones of Chapter 2.

### A.1 Basic topology and differential calculus

**Proposition A.1.1.** In a Hausdorff locally convex topological vector space  $(X, \tau)$ , let  $A \subset X$  be a compact set not containing the origin. Then  $\mathbb{R}_+A$  is closed.

*Proof.* Let  $(t_\alpha a_\alpha)_{\alpha \in A} \subset \mathbb{R}_+A$  be a net, converging to  $x \in X$ . Let us show that  $x \in \mathbb{R}_+A$ . By compactness of  $A$ , and considering eventually a subnet, one can assume that the net  $(a_\alpha)_{\alpha \in A}$  converges to some  $a \in A$ . Let  $(p_i)_{i \in I}$  be a family of semi-norms generating the topology of  $\tau$ . Since it is assumed that  $0 \notin A$ , and that  $(X, \tau)$  is Hausdorff, there exists some  $i \in I$  such that  $p_i(a) \neq 0$ .

We have  $t_\alpha a_\alpha \rightarrow x$  in  $X$  and  $p_i(a_\alpha) \rightarrow p_i(a) \neq 0$  in  $\mathbb{R}$ , so  $\frac{t_\alpha a_\alpha}{p_i(a_\alpha)}$  converges to  $\frac{x}{p_i(a)}$  in  $X$ . Now we can write  $t_\alpha = p_i\left(\frac{t_\alpha a_\alpha}{p_i(a_\alpha)}\right)$ , which tends to  $p_i\left(\frac{x}{p_i(a)}\right)$ . Define  $t := p_i\left(\frac{x}{p_i(a)}\right) \geq 0$ , and use again the Hausdorff property to conclude that  $x = ta$ . ■

**Proposition A.1.2.** Let  $(X, \tau)$  be a locally convex topological vector space, and  $K \subset X$  a polyhedral cone with a nonempty interior. Then  $X$  has finite dimension.

*Proof.* Since  $K$  has finite interior, there exists a nonempty open set  $U \subset K$ . For any  $x \in U$ ,  $U - x$  is a neighborhood of the origin which implies that  $X = \mathbb{R}_+(U - x)$ . Since  $\mathbb{R}_+(U - x) \subset \text{span } U$ , we have  $X = \text{span } U$ . But  $U \subset K$ , so  $X = \text{span } K$ .

By definition of a polyhedral cone, there exists  $m \in \mathbb{N}$  and some  $\{a_1, \dots, a_m\} \subset X$  such that  $K = \mathbb{R}_+\text{co}\{a_1, \dots, a_m\}$ . It follows that  $X = \text{span } K = \text{span } \{a_1, \dots, a_m\}$ , and proves the claim. ■

**Proposition A.1.3** ([325, Theorem 1.1.2]). Let  $C$  be a convex set in a topological vector space  $(X, \tau)$ . If  $\text{int}_\tau C \neq \emptyset$ , then

$$\text{cl}_\tau \text{int}_\tau C = \text{cl}_\tau C \text{ and } \text{int}_\tau \text{cl}_\tau C = \text{int}_\tau C.$$

For a function  $f : X \rightarrow \mathbb{R}$ , we use the following notation to denote its sublevel sets:

$$[f \leq \lambda] := \{x \in X \mid f(x) \leq \lambda\}, [f < \lambda] := \{x \in X \mid f(x) < \lambda\}.$$

**Proposition A.1.4.** Let  $(X, \tau)$  be a topological vector space, and  $f : X \rightarrow \mathbb{R}$  be a convex continuous function. If there exists  $\bar{x} \in X$  such that  $f(\bar{x}) < 0$ , then

$$\text{cl}_\tau[f < 0] = [f \leq 0] \text{ and } \text{int}_\tau[f \leq 0] = [f < 0].$$

*Proof.* This proof is inspired by the one of Hiriart-Urruty and Lemaréchal in [195, Proposition VI.1.33], where this result is stated with  $X = \mathbb{R}^n$ .

Let  $x \in \text{cl}_\tau[f < 0]$ , i.e. there exists a net  $(x_\alpha)_{\alpha \in \mathcal{A}}$  in  $[f < 0]$  such that  $x_\alpha \xrightarrow[\alpha \in \mathcal{A}]{\tau} x$ . Since  $f(x_\alpha) < 0$  for all  $\alpha \in \mathcal{A}$ , and using the lower semi-continuity of  $f$ , we obtain

$$f(x) \leq \liminf_{\alpha \in \mathcal{A}} f(x_\alpha) \leq 0,$$

that is,  $x \in [f \leq 0]$ .

Take now any  $x \in [f \leq 0]$ , and define for all  $n \in \mathbb{N}$ :

$$x_n := \frac{1}{n}\bar{x} + \left(1 - \frac{1}{n}\right)x.$$

We can rewrite the latter as

$$x_n - x = \frac{1}{n}(\bar{x} - x),$$

where the right member goes to zero when  $n \rightarrow +\infty$ . Hence,  $x_n$  converges to  $x$ . Moreover, using the definition of  $x_n$  together with the convexity of  $f$  and the hypothesis  $f(\bar{x}) < 0$  gives

$$f(x_n) \leq \frac{1}{n}f(\bar{x}) + \left(1 - \frac{1}{n}\right)f(x) \leq \frac{1}{n}f(\bar{x}) < 0.$$

So,  $x_n \in [f < 0]$  and it follows that  $x \in \text{cl}_\tau[f < 0]$ . This proves item i)

For item ii), we apply Proposition A.1.3 to  $C = [f < 0]$ , which is convex since  $f$  is convex. For this we need to verify that  $[f < 0]$  has a nonempty  $\tau$ -interior. But  $f$  being upper semi-continuous implies that  $[f \geq 0]$  is  $\tau$ -closed, and its complement is exactly  $[f < 0]$ . Hence,  $[f < 0]$  is  $\tau$ -open and nonempty since we assume that  $\bar{x} \in [f < 0]$ . So Proposition A.1.3 applies, and we obtain

$$\text{int}_\tau \text{cl}_\tau[f < 0] = \text{int}_\tau[f < 0].$$

Since  $[f < 0]$  is open and satisfies item i), we deduce that

$$\text{int}_\tau[f \leq 0] = [f < 0].$$

■

Now we prove the results concerning the differentiability of a function  $F : X \rightarrow Y$ , where  $X$  and  $Y$  are two Banach spaces. First, we will need some technical results.

**Proposition A.1.5.** If  $F : X \rightarrow Y$  is Gateaux differentiable at  $\bar{x} \in X$ , then for all  $d \in X$ , the application  $\phi : t \in \mathbb{R} \mapsto F(\bar{x} + td) \in Y$  is continuous and derivable at zero. In particular,  $\phi'(0) = DF(\bar{x}; d)$ .

*Proof.* Let us show that  $\phi$  is right derivable at zero, using the Gateaux differentiability of  $F$  at  $\bar{x}$ :

$$\lim_{h \rightarrow 0^+} \frac{\phi(h) - \phi(0)}{h} = \lim_{h \rightarrow 0^+} \frac{F(\bar{x} + hd) - F(\bar{x})}{h} = DF(\bar{x}; d).$$

Similarly, it is left derivable at zero:

$$\lim_{h \rightarrow 0^-} \frac{\phi(h) - \phi(0)}{h} = \lim_{h \rightarrow 0^-} \frac{F(\bar{x} + hd) - F(\bar{x})}{h} = \lim_{h \rightarrow 0^+} \frac{F(\bar{x} - hd) - F(\bar{x})}{-h} = -DF(\bar{x}; -d).$$

Since  $-DF(\bar{x}; -d) = DF(\bar{x}; d)$  by linearity of  $DF(\bar{x})$ , we have that  $\phi$  is derivable at zero, and its derivative is  $\phi'(0) = DF(\bar{x}; d)$ . In particular,  $\phi$  is continuous at zero. ■

We immediately derive the following corollary, just by observing that the derivability at  $\bar{t} \in [0, 1]$  of  $t \mapsto F(\bar{x} + td)$  is exactly the same than the derivability at zero of  $t \mapsto F(\bar{x} + \bar{t}d + td)$ .

**Corollary A.1.6.** If  $F : X \rightarrow Y$  is Gateaux differentiable on a segment  $[\bar{x}_1, \bar{x}_2] \subset X$ , then for all  $d \in X$ , the application  $\phi : t \in \mathbb{R} \mapsto F((1-t)\bar{x}_2 + t\bar{x}_1) \in Y$  is continuous and derivable on  $[0, 1]$ . In particular,  $\phi'(t) = DF((1-t)\bar{x}_2 + t\bar{x}_1; \bar{x}_2 - \bar{x}_1)$  for all  $t \in [0, 1]$ .

Applying the fundamental theorem of calculus to  $\phi : [0, 1] \rightarrow Y$ , ones obtain

**Corollary A.1.7.** If  $F : X \rightarrow Y$  is Gateaux differentiable on a segment  $[\bar{x}_1, \bar{x}_2] \subset X$ , then

$$F(\bar{x}_2) - F(\bar{x}_1) = \int_0^1 DF((1-t)\bar{x}_2 + t\bar{x}_1; \bar{x}_2 - \bar{x}_1) dt.$$

Now we can prove Proposition 2.1.11, that we recall here:

**Proposition A.1.8.** Let  $F : X \rightarrow Y$  be Gateaux differentiable on an open set  $U \subset X$ . Then, for all  $\bar{x} \in U$ , the following is equivalent:

- i)  $F$  is strictly Gateaux differentiable at  $\bar{x}$ ,
- ii)  $DF : U \rightarrow L(X, Y)$  is *pointwise continuous* at  $\bar{x}$ , i.e. continuous with respect to the norm topology of  $X$  and the pointwise<sup>1</sup> topology of  $L(X, Y)$ . In other words, for all converging net  $(x_\alpha)_{\alpha \in \mathcal{A}} \subset U$  converging to  $\bar{x}$ , we have  $DF(x_\alpha; d) \xrightarrow[\alpha \in \mathcal{A}]{\|\cdot\|_Y} DF(\bar{x}; d)$  for all  $d \in X$ .

*Proof.* Suppose first that  $F$  is strictly Gateaux differentiable at  $\bar{x}$  and fix some  $d \in X$ . Using the strict Gateaux differentiability at  $\bar{x}$ , there exists some  $\delta > 0$  such that

$$(A.1) \quad \forall x \in \bar{x} + \delta\mathbb{B}, \forall t \in ]0, \delta[, \left\| \frac{F(x + td) - F(x)}{t} - DF(\bar{x}; d) \right\| \leq \frac{\varepsilon}{2}.$$

Given this  $\delta$ , for any  $x \in \bar{x} + \delta\mathbb{B}$  we can assume (taking eventually a smaller  $\delta$ ) that  $F$  is Gateaux differentiable at  $x$ . Then, for all  $x \in \bar{x} + \delta\mathbb{B}$  there exists some  $t_x \in ]0, \delta[$  such that

$$(A.2) \quad \left\| \frac{F(x + t_x d) - F(x)}{t_x} - DF(x; d) \right\| \leq \frac{\varepsilon}{2}.$$

Combining (A.1) and (A.2) together with the triangle inequality, we proved that

$$\exists \delta > 0, \forall x \in \bar{x} + \delta\mathbb{B}, \quad \|DF(\bar{x}; d) - DF(x; d)\| \leq \varepsilon.$$

Item ii) being proved, suppose now the reverse, that is item i) holds. To prove that  $F$  is strictly Gateaux differentiable at  $\bar{x}$ , fix some  $d \in X$ , and  $\varepsilon > 0$ . Using the continuity of  $DF$  at  $\bar{x}$  in the sense of item ii), we obtain the existence of  $\delta > 0$  such that  $B(\bar{x}, \delta) \subset U$ , and which satisfies

$$(A.3) \quad \text{for all } x \in B(\bar{x}, \delta), \quad \|DF(x; d) - DF(\bar{x}; d)\| \leq \varepsilon.$$

Take arbitrary  $x \in B(\bar{x}, \frac{\delta}{2})$  and  $t \in ]0, \frac{\delta\|d\|}{2}[$ . These bounds ensure that  $x + td$  remains in  $B(\bar{x}, \delta)$ , where  $F$  is Gateaux differentiable. Since  $[x; x + td] \subset B(\bar{x}, \delta)$ , we can use Corollary A.1.7 which gives us

$$F(x + td) - F(x) = \int_0^1 DF(x + \tau td, td) d\tau,$$

<sup>1</sup>The pointwise topology is the locally convex topology  $\tau_{pw}$  on  $L(X, Y)$  defined by the family of seminorms  $\{p_x\}_{x \in X}$ , where  $p_x : A \in L(X, Y) \mapsto \|Ax\|_Y$ . This topology is also called the weak operator topology

or equivalently

$$(A.4) \quad F(x + td) - F(x) - tDF(\bar{x}; d) = \int_0^1 t(DF(x + \tau td, d) - DF(\bar{x}; d)) \, d\tau,$$

But for all  $\tau \in [0, 1]$ ,  $x + \tau td$  lies in  $[x, x + td] \subset B(\bar{x}, d)$  where (A.3) applies. We deduce immediately from (A.4) that

$$\|F(x + td) - F(x) - tDF(\bar{x}; d)\| \leq \varepsilon t.$$

Hence, we proved that for all  $x \in B(\bar{x}, \frac{\delta}{2})$  and  $t \in ]0, \frac{\delta \|d\|}{2}[$ ,

$$\left\| \frac{F(x + td) - F(x)}{t} - DF(\bar{x}; d) \right\| \leq \varepsilon,$$

which proves the strict Gateaux differentiability of  $F$  at  $\bar{x}$ . ■

In a similar fashion, we prove Proposition 2.1.12:

**Proposition A.1.9.** Let  $F : X \rightarrow Y$  be Frechet differentiable on an open set  $U \subset X$ . Then, for all  $\bar{x} \in U$ , the following is equivalent:

- i)  $F$  is strictly Frechet differentiable at  $\bar{x}$ ,
- ii)  $DF : U \rightarrow L(X, Y)$  is *strongly continuous* at  $\bar{x}$ , i.e. continuous with respect to the norm topology of  $X$  and the usual operator norm topology of  $L(X, Y)$ .

We will use for this proof the fact that the Frechet differentiability (and strict Frechet differentiability) of  $F$  at  $\bar{x}$  is equivalent to

$$(A.5) \quad \lim_{t \downarrow 0} \sup_{\|d\| \leq 1} \frac{\|F(\bar{x} + td) - F(\bar{x}) - tDF(\bar{x}; d)\|}{t} = 0,$$

$$(A.6) \quad \lim_{\substack{x \rightarrow \bar{x} \\ t \downarrow 0}} \sup_{\|d\| \leq 1} \frac{\|F(x + td) - F(x) - tDF(\bar{x}; d)\|}{t} = 0.$$

*Proof.* Suppose first that  $F$  is strictly Frechet differentiable at  $\bar{x}$ . Using the strict Frechet differentiability at  $\bar{x}$  and (A.6), there exists some  $\delta > 0$  such that

$$(A.7) \quad \forall x \in B(\bar{x}, \delta), \forall t \in ]0, \delta[, \sup_{\|d\| \leq 1} \left\| \frac{F(x + td) - F(x)}{t} - DF(\bar{x}; d) \right\| \leq \frac{\varepsilon}{2}.$$

Given this  $\delta$ , for any  $x \in B(\bar{x}, \delta)$  we can assume (taking eventually a smaller  $\delta$ ) that  $F$  is Frechet differentiable at  $x$ . Then, for all  $x \in B(\bar{x}, \delta)$  we use (A.5) to obtain some  $t_x \in ]0, \delta[$  such that

$$(A.8) \quad \sup_{\|d\| \leq 1} \left\| \frac{F(x + t_x d) - F(x)}{t_x} - DF(x; d) \right\| \leq \frac{\varepsilon}{2}.$$

Combining (A.7) and (A.8) together with the triangle inequality, we proved that

$$\exists \delta > 0, \forall x \in B(\bar{x}, \delta), \sup_{\|d\| \leq 1} \|DF(\bar{x}; d) - DF(x; d)\| \leq \varepsilon.$$

In other words, using the definition of the operator norm in  $L(X, Y)$ ,

$$\exists \delta > 0, \forall x \in B(\bar{x}, \delta), \quad \|DF(\bar{x}) - DF(x)\| \leq \varepsilon.$$

Item ii) being proved, suppose now the reverse, that is item i) holds. To prove that  $F$  is strictly Frechet differentiable at  $\bar{x}$ , fix some  $\varepsilon > 0$ . Using the continuity of  $DF$  at  $\bar{x}$  in the sense of item ii), we obtain the existence of  $\delta > 0$  such that  $B(\bar{x}, \delta) \subset U$ , and which satisfies

$$(A.9) \quad \text{for all } x \in B(\bar{x}, \delta), \quad \|DF(x) - DF(\bar{x})\| \leq \varepsilon,$$

or, equivalently,

$$(A.10) \quad \text{for all } x \in B(\bar{x}, \delta), \quad \sup_{\|d\| \leq 1} \|DF(x; d) - DF(\bar{x}; d)\| \leq \varepsilon.$$

Take arbitrary  $x \in B(\bar{x}, \frac{\delta}{2})$  and  $t \in ]0, \frac{\delta}{2}[$ . For all  $d \in X$  such that  $\|d\| \leq 1$ , these bounds ensure that  $x + td$  remains in  $B(\bar{x}, \delta)$ , where  $F$  is Frechet differentiable. Since  $[x; x + td] \subset B(\bar{x}, \delta)$ , we can use Corollary A.1.7 (recall that Frechet differentiability entails Gateaux differentiability) which gives us

$$F(x + td) - F(x) = \int_0^1 DF(x + \tau td, td) \, d\tau,$$

or equivalently

$$(A.11) \quad F(x + td) - F(x) - tDF(\bar{x}; d) = \int_0^1 t(DF(x + \tau td, d) - DF(\bar{x}; d)) \, d\tau,$$

But for all  $\tau \in [0, 1]$ ,  $x + \tau td$  lies in  $[x, x + td] \subset B(\bar{x}, \delta)$  where (A.10) applies. We deduce immediately from (A.11) that

$$\sup_{\|d\| \leq 1} \|F(x + td) - F(x) - tDF(\bar{x}; d)\| \leq \varepsilon t.$$

Hence, we proved that for all  $x \in B(\bar{x}, \frac{\delta}{2})$  and  $t \in ]0, \frac{\delta\|d\|}{2}[$ ,

$$\sup_{\|d\| \leq 1} \left\| \frac{F(x + td) - F(x)}{t} - DF(\bar{x}; d) \right\| \leq \varepsilon,$$

which proves the strict Gateaux differentiability of  $F$  at  $\bar{x}$ . ■

**Remark A.1.10.** The similarity between the proofs of Propositions A.1.8 and A.1.9 is simply due to the fact that and Frechet differentiability are just two particular instances of *bornological* differentiability, for which this equivalence extends.

## A.2 Nonsmooth analysis for extended-real-valued functions

**Proposition A.2.1.** Let  $x \in \Omega \subset X$ . Then  $T_\Omega^{ad}(x) = T_\Omega^r(x)$  if and only if  $\Omega$  is radial at  $x$ . In particular  $T_\Omega^{ad}(x) = T_\Omega^r(x)$  for all  $x \in \Omega$  when  $\Omega$  is convex.

*Proof.* Suppose that  $T_\Omega^{ad}(x) = T_\Omega^r(x)$  and take  $y \in \Omega$ . We have that  $y = x + (y - x) \in \Omega$ , so  $y - x \in T_\Omega^r(x) = T_\Omega^{ad}(x)$ . So there exists some  $\delta > 0$  such that for all  $t \in [0, \delta]$ ,  $x + td \in \Omega$ . In other words,  $[x, y] \cap B(x, \delta) \subset \Omega$ . Conversely, suppose that  $\Omega$  is radial at  $x$ , and consider  $d \in T_\Omega^r(x)$ . So we have  $x + \lambda d \in \Omega$  for some  $\lambda > 0$ . By hypothesis, there exists a  $\delta > 0$  such that  $[x, x + \lambda d] \cap B(x, \delta) \subset \Omega$ . In other words, for all  $t \in [0, \lambda\delta]$ , we have  $x + td \in \Omega$ , whence  $d \in T_\Omega^{ad}(x)$ . Now suppose that  $\Omega$  is convex. In particular it is radial, and the equality  $T_\Omega^{ad}(x) = T_\Omega^r(x)$  holds. Moreover,  $\text{co}T_\Omega^r(x) = \text{co}\mathbb{R}_+(\Omega - x) = \mathbb{R}_+\text{co}(\Omega - x) = \mathbb{R}_+(\Omega - x) = T_\Omega^r(x)$  which proves that  $T_\Omega^r(x)$  is convex. ■

**Proposition A.2.2.** Let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  and  $x \in \text{dom } f$ . Consider the following sets:

- i)  $\{x^* \in X^* \mid \forall y \in X, f(y) - f(x) - \langle x^*, y - x \rangle \geq 0\}$ ,
- ii)  $\{x^* \in X^* \mid x \text{ is a global minimum of } y \mapsto f(y) - \langle x^*, y \rangle\}$ ,
- iii)  $\{x^* \in X^* \mid (x^*, -1) \in N_{\text{epi } f}(x, f(x))\}$ ,
- iv)  $\{x^* \in X^* \mid \forall d \in X, \langle x^*, d \rangle \leq d^D f(x; d)\}$ .

Hence the first three sets always coincide, and we call it the *convex subdifferential* of  $f$  at  $x$ , noted  $\partial f(x)$ . Moreover, the fourth set always contains  $\partial f(x)$ , with equality when  $f$  is convex.

*Proof.* Equivalence between the sets 1 and 2 is immediate.

Now prove equivalence between sets 1 and 3. Suppose that  $(x^*, -1) \in N_{\text{epi } f}(x, f(x))$ , then by definition,  $\forall (d, \lambda) \in T_{\text{epi } f}(x, f(x))$ ,  $\langle x^*, d \rangle \leq \lambda$ . Take now any  $y \in X$ , and pose  $d := y - x$ ,  $\lambda := f(y) - f(x)$ . One easily sees that  $(x, f(x)) + (d, \lambda) \in \text{epi } f$ , that is  $(d, \lambda) \in T_{\text{epi } f}^r(x, f(x)) \subset T_{\text{epi } f}(x, f(x))$ . So we deduce  $\langle x^*, y - x \rangle \leq f(y) - f(x)$  which is what we wanted. Now suppose that for all  $y \in X$  we have  $f(y) - f(x) - \langle x^*, y - x \rangle \geq 0$ . Take now any  $(d, \lambda) \in T_{\text{epi } f}(x, f(x))$ , then by definition there exists sequences  $d_n \rightarrow d$ ,  $\lambda_n \rightarrow \lambda$  and  $t_n > 0$  such that  $f(x + t_n d_n) \leq f(x) + t_n \lambda_n$ . Then

$$(A.12) \quad \langle x^*, d \rangle = \lim_{n \rightarrow +\infty} \frac{\langle x^*, x + t_n d_n - x \rangle}{t_n} \leq \lim_{n \rightarrow +\infty} \frac{f(x + t_n d_n) - f(x)}{t_n} \leq \lim_{n \rightarrow +\infty} \lambda_n = \lambda.$$

Now we check that in the general case, any  $x^* \in \partial f(x)$  is a linear minorant for  $d^D f(x; \cdot)$ . Indeed for any  $d \in X$ ,

$$(A.13) \quad \langle x^*, d \rangle = \liminf_{t \downarrow 0} \frac{\langle x^*, x + td - x \rangle}{t} \leq \liminf_{t \downarrow 0} \frac{f(x + td) - f(x)}{t} = d^D f(x; d).$$

Suppose now that  $f$  is convex, and take any linear minorant  $x^*$  of  $d^D f(x; \cdot)$ . Then for any  $y \in X$ ,

$$(A.14) \quad \langle x^*, y - x \rangle \leq \liminf_{t \downarrow 0} \frac{f(x + t(y - x)) - f(x)}{t} \leq \liminf_{t \downarrow 0} \frac{tf(y) + (1 - t)f(x) - f(x)}{t} = f(y) - f(x). \quad \blacksquare$$

**Proposition A.2.3.** Let  $x \in \Omega \subset X$ . Then:

- i)  $\text{cl } T_{\Omega}^{ad}(x) \subset T_{\Omega}^B(x) \subset \text{cl } T_{\Omega}^r(x) = T_{\Omega}(x)$ .
- ii) If  $\Omega$  is radial at  $x$ , then the above inclusions become equalities.

*Proof.* We start with item i), and prove that  $\text{cl } T_{\Omega}^{ad}(x) \subset T_{\Omega}^B(x)$ . Let  $d = \lim_{n \rightarrow +\infty} d_n$  with  $d_n \in T_{\Omega}^{ad}(x)$ , that is there exists some  $\delta_n > 0$  such that for all  $t \in [0, \delta_n[$ , one has  $x + td_n \in \Omega$ . We can select a sequence satisfying  $t_n \downarrow 0$  and  $t_n < \delta_n$ , so that  $x + t_n d_n \in \Omega$  for all  $n \in \mathbb{N}$ . Whence  $d \in T_{\Omega}^B(x)$ . Now we show that  $T_{\Omega}^B(x) \subset \text{cl } T_{\Omega}^r(x)$ . Let  $d \in T_{\Omega}^B(x)$ , then there exists  $t_n \downarrow 0$  and  $d_n \rightarrow d$  such that  $x + t_n d_n \in \Omega$  for all  $n \in \mathbb{N}$ . Clearly,  $d_n \in T_{\Omega}^r(x)$ , which proves the claim. The inequalities concerning the weak Bouligand tangent cone follows exactly the same proof, replacing strong by weak topology. Item ii) is a direct consequences of Proposition A.2.1. \blacksquare

**Proposition A.2.4.** Let  $x \in \Omega \subset X$ , then  $T_{\Omega}^B(x)$  is closed.

*Proof.* Let  $d_n$  be a sequence in  $T_\Omega^B(x)$  strongly converging to some  $d \in X$ . Then there exists for all  $n \in \mathbb{N}$  a sequence  $d_n^k \xrightarrow[k \rightarrow +\infty]{} d_n$  and  $t_n^k \xrightarrow[k \rightarrow +\infty]{} 0$  such that for all  $k \in \mathbb{N}$  we have  $x + t_n^k d_n^k \in \Omega$ . So for all  $n \in \mathbb{N}$ , there exists some  $k_n \in \mathbb{N}$  such that for all  $k \geq k_n$  we have

$$(A.15) \quad \|d_n^k - d_n\| \leq \frac{1}{n} \text{ and } t_n^k \leq \frac{1}{n}.$$

By setting  $\tilde{d}_n := d_n^{k_n}$  and  $\tilde{t}_n$  one can see that  $\tilde{d}_n \rightarrow d$ ,  $\tilde{t}_n \rightarrow 0$  with  $x + \tilde{t}_n \tilde{d}_n \in \Omega$ . This achieves the proof.  $\blacksquare$

**Proposition A.2.5.** Let  $u : I \subset \mathbb{R} \rightarrow X$  and  $f \circ u : I \rightarrow \mathbb{R} \cup \{+\infty\}$ , where  $I$  is an open interval of  $\mathbb{R}$ . Suppose that both  $u$  and  $f \circ u$  are derivable at  $t \in I$ , and that  $u(t) \in \text{dom } \partial^B f$ . Then,

$$(A.16) \quad (f \circ u)'(t) = \langle x^*, \dot{u}(t) \rangle, \quad \forall x^* \in \partial^B f(u(t)).$$

*Proof.* Let  $x^*$  be in  $\partial^B f(u(t))$ , then  $\langle x^*, \dot{u}(t) \rangle \leq d^B f(u(t), \dot{u}(t))$  where

$$\begin{aligned} d^B f(u(t), \dot{u}(t)) &= \liminf_{h \downarrow 0, d' \rightarrow \dot{u}(t)} \frac{f(u(t) + hd') - f(u(t))}{h} \\ &\leq \liminf_{h \downarrow 0, \epsilon \rightarrow 0} \frac{f(u(t) + h \frac{u(t+\epsilon) - u(t)}{\epsilon}) - f(u(t))}{h} \end{aligned}$$

Taking  $\epsilon = h$  we obtain  $\langle x^*, \dot{u}(t) \rangle \leq (f \circ u)'(t)$ . Now we write  $\langle x^*, -\dot{u}(t) \rangle \leq d^B f(u(t), -\dot{u}(t))$ , and using the same method with  $\epsilon = -h$  one obtains  $\langle x^*, -\dot{u}(t) \rangle \leq -(f \circ u)'(t)$  which ends the proof.  $\blacksquare$

**Proposition A.2.6.** Let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be a quasi-convex function such that  $f(x_2) \leq f(x_1)$ . Then  $\langle x^*, x_2 - x_1 \rangle \leq 0$  for all  $x^* \in \partial^B f(x_1) \supset \partial^F f(x_1)$ .

*Proof.* Introducing the convex sublevel set  $\Omega = [f \leq f(x_1)]$ , we see that  $x_1$  and  $x_2$  lie therein. Hence  $tx_2 + (1-t)x_1$  lies in this set for all  $t \in [0, 1]$ , and:

$$(A.17) \quad f(x_1 + t(x_2 - x_1)) - f(x_1) \leq 0 \text{ for any } t \in [0, 1].$$

After dividing by  $t$  and passing to the limit when  $t \downarrow 0$ , we obtain  $d^D f(x_1; x_2 - x_1) \leq 0$ . Conclusion follows  $d^B f(x_1; \cdot) \leq d^D f(x_1; \cdot)$ .  $\blacksquare$

**Proposition A.2.7.** Let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be a quasi-convex function such that  $f(x_2) < f(x_1)$ . Then  $\langle x^*, x_2 - x_1 \rangle \leq 0$  for all  $x^* \in \partial^L f(x_1)$ . If  $f$  is locally Lipschitz then  $\partial^L$  can be replaced by  $\partial^C$ .

*Proof.* We start by proving our statement for the limiting subdifferential. Let  $x^* \in \partial^L f(x_1)$ , then it is the weak limit of a sequence  $(x_n^*)_{n \in \mathbb{N}}$  such that  $x_n^* \in \partial^F f(x_n)$  with  $x_n \xrightarrow[n \rightarrow +\infty]{\|\cdot\|} x_1$  and  $f(x_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{R}} f(x_1)$ . Since we have the strict inequality  $f(x_2) < f(x_1)$ , we can assume that  $f(x_2) < f(x_n)$  for all  $n \in \mathbb{N}$ . In particular, Proposition 2.2.15 applies, and  $\langle x_n^*, x_2 - x_n \rangle \leq 0$ . The conclusion follows directly by passing to the limit on  $n$ .

It is easy to extend this property to the Clarke subdifferential when  $f$  is locally Lipschitz continuous and  $X$  reflexive, by using  $\partial^C f(x_1) = \overline{\text{co}} \partial^L f(x_1)$ . Otherwise, we have to prove that  $d^C f(x_1; x_2 - x_1) \leq 0$ .

If  $x' \in X$  is taken arbitrary close to  $x_1$ , and because of the continuity of  $f$  together with  $f(x_2) < f(x_1)$ , we can assume that  $f(x_2) \leq f(x')$ . Using that  $f$  is quasi-convex together with  $x', x_2 \in [f \leq f(x_1)]$ , we obtain that

$$\forall t \in [0, 1], f(x' + t(x_2 - x')) - f(x') \leq 0.$$



Hence, for all  $t \in [0, 1]$  and  $x'$  close to  $x_1$ , use the inequality above and the Lipschitz continuity at  $x_1$  to obtain:

$$\begin{aligned} f(x' + t(x_2 - x_1)) - f(x') &= f(x' + t(x_2 - x_1)) - f(x' + t(x_2 - x')) + f(x' + t(x_2 - x')) - f(x') \\ &\leq t \text{Lip}(f, x_1) \|x' - x\|. \end{aligned}$$

As a consequence,

$$d^c f(x_1; x_2 - x_1) = \limsup_{\substack{x' \rightarrow x_1 \\ t \downarrow 0}} \frac{f(x' + t(x_2 - x_1)) - f(x')}{t} \leq \limsup_{\substack{x' \rightarrow x_1 \\ t \downarrow 0}} \text{Lip}(f, x_1) \|x' - x\| = 0.$$

■

### A.3 Optimization for vector-valued functions

Here are some results concerning the bases of a cone  $K$  in a locally convex topological vector space  $(X, \tau)$ . Let us fix some definitions in that setting. We say that a set  $A \subset K$  is  $\tau$ -*bounded* if for any neighbourhood of zero  $V$ , there exists some  $\lambda > 0$  such that  $A \subset \lambda V$ . This simply corresponds to the usual norm-boundedness in the following cases:  $X$  is a normed space equipped with its norm or weak topology, or  $X$  is the topological dual of a Banach space, equipped with the norm, weak or weak star topology. It is a consequence of the uniform boundedness principle [5, Theorems 6.14–6.15], see also Mackey's theorem [5, Theorem 6.20], [215, XX.11.7]. We will also consider the *strong topology* associated to  $(X, \tau)$ , that we note  $\beta(X, \tau)$ . It is the topology induced by the  $w^*$ -bounded sets of  $(X, \tau)^*$ , see [5, Chapter 5.19]. Here again, as expected, the strong topology reduces to the usual norm topology when  $X$  is a normed space equipped with its norm or weak topology, or  $X$  is the topological dual of a Banach space, equipped with the norm, weak or weak star topology.

We present some necessary and sufficient conditions on a cone  $K$  for the existence of such a base. We say that a cone  $K$  is *pointed* if  $K \cap -K = \{0\}$ .

#### Proposition A.3.1.

- i) If  $K$  admits a  $\tau$ -closed and  $\tau$ -bounded base, then  $K$  is  $\tau$ -closed.
- ii) If  $K$  admits a convex base, then  $K$  is pointed and convex.

As a consequence, if  $K$  admits a  $\tau$ -closed convex and  $\tau$ -bounded base, then  $K$  is  $\tau$ -closed, convex, pointed, with a nonempty strong interior.

*Proof.* Item ii) can be found in [206, Lemma 1.14]. Both items i) and ii) can be found in [139, Proposition I.1.7], where they are presented as a result of Jameson [207]. ■

Given a strict cone  $S \subset X$ , we also introduce its *strict dual cone*  $S^{++} \subset (X, \tau)^*$  as

$$(A.18) \quad S^{++} := \{x^* \in X^* \mid \langle x^*, x \rangle > 0 \text{ for all } x \in S\}.$$

Analogously, we define  $K^{++}$  the strict dual cone of a cone  $K$ , as the strict dual cone of  $K \setminus \{0\}$ . Do not mistake  $K^{++}$  with the bidual  $(K^+)^+$ . The strict dual cone of  $K$  is sometimes called the *quasi-interior* of  $K^+$ , since  $K^{++}$  is exactly the quasi-interior of  $K^+$  with respect to the weak\* topology of  $X^*$  when  $K$  is closed convex (see [80, Proposition 2.1.1]). It is easy to see that if  $K^{++} \neq \emptyset$  then  $K$  is pointed, and in fact the reverse statement is true in the normed separable spaces:

**Theorem A.3.2** (Krein-Rutman, [206, Theorem 3.38]). *Let  $K$  be a convex closed cone. If  $K_{++} \neq \emptyset$ , then  $K$  is pointed. The reverse is true if  $X$  is a separable normed vector space.*

Now we will see that the existence of a base having good properties is directly connected with the properties of  $K^+$ . We mention in this result the algebraic strict dual cone  $K'^{++}$  which is defined exactly as the strict dual cone  $K^{++}$ , but with elements in the algebraic dual space  $X'$  instead of the topological one  $X^*$ . Despite the fact that it won't be useful in the sequel, it helps to have a better comprehension of what happen here.

**Theorem A.3.3.** *Let  $K$  be a convex cone in a locally convex Hausdorff topological vector space  $(X, \tau)$ , and  $\Theta \subset X$ .*

- i)  $\Theta$  is a convex base of  $K \Leftrightarrow \exists e' \in K'^{++}$  such that  $\Theta = \{x \in K \mid \langle e', x \rangle = 1\}$ .
- ii)  $\Theta$  is a convex base of  $K$  with  $0 \notin \text{cl}_\tau \Theta \Leftrightarrow \exists e^* \in K^{++}$  such that  $\Theta = \{x \in K \mid \langle e^*, x \rangle = 1\}$ .
- iii)  $\Theta$  is a  $\tau$ -bounded convex base of  $K$  with  $0 \notin \tau \text{cl} \Theta \Leftrightarrow \exists e^* \in \beta \text{int} K^+$  such that  $\Theta = \{x \in K \mid \langle e^*, x \rangle = 1\}$ .

We see that deeper in the 'interiors' of  $K^+$  we can take  $e^*$  ( $\beta \text{int} K^+ \subset K^{++} \subset K'^{++}$ ), better is the base we construct for  $K$ . Moreover the Theorem gives us an easy way to construct such bases, it suffices to pick an appropriate  $e^* \in K^{++}$ .

*Proof.* Proof of item i) is based on Zorn's Lemma (see [206, Lemma 3.3]), but can also be proved without the axiom of choice if  $K$  spans  $X$  (see [206, Lemma 1.28]). Item iii) is proved in [207, Theorem 3.8.4], and makes use of the duality provided by the Alaoglu-Bourbaki theorem. Note that Jameson assume the cone  $K$  to be pointed<sup>2</sup> but this is not necessary since both members of the equivalence imply pointedness. The direct implication of item ii) is proved in [206, Corollary 3.19], but without mentioning  $0 \notin \text{cl}_\tau \Theta$  in the hypotheses<sup>3</sup>. It seems that the author made a mistake there. Given a convex base he proves that  $0$  do not lie in the *algebraic closure* of  $\Theta$ , and says that this algebraic closure equals the topological one  $\text{cl}_\tau B$  invoking [206, Lemma 1.32]. But this result asks the interior of  $\Theta$  to be nonempty, which is not guaranteed here. So adding  $0 \notin \text{cl}_\tau \Theta$  in the left member saves the day. The reverse implication is immediate, since it is clear that such  $\Theta$  is convex (see also item i)), and  $0 \notin \text{cl}_\tau \Theta$  follows the continuity of  $e^*$ . ■

As a corollary, we can prove Theorem 2.3.4:

*Proof of Theorem 2.3.4.* Suppose that  $K$  is a closed convex cone in a Banach space  $X$ , and apply item iii) of Theorem A.3.3 to the weakly\* closed convex cone  $K^+$  in  $(X^*, w^*)$ .

Suppose in a first time that we have some weakly\* compact convex base  $\Theta$  of  $K^+$ . We have in particular that  $\Theta$  is  $w^*$ -bounded. Moreover a base satisfies by definition  $0 \notin \Theta$ , and this  $\Theta$  is weakly\* closed, so  $0 \notin \text{cl}_{w^*} \Theta$ . Then Theorem A.3.3 applies, and using the fact that  $(X^*, w^*)^* = X$ , we obtain that  $\beta \text{int}(K^+)^+ \subset X$  is nonempty. Since  $K$  is assumed to be closed and convex, we have  $(K^+)^+ = K$ , and as said at the beginning of this section, the strong topology coincides here with the norm topology. In other words,  $\text{int} K \neq \emptyset$ , and  $\Theta = \{x^* \in K^+ \mid \langle x^*, e \rangle = 1\}$  for some  $e \in \text{int} K$ .

Suppose now that  $\text{int} K \neq \emptyset$ . Take some  $e \in \text{int} K$  and consider  $\Theta = \{x^* \in K^+ \mid \langle x^*, e \rangle = 1\}$ . Applying Theorem A.3.3, we obtain that this is a  $w^*$ -bounded convex base of  $K^+$ . Moreover,  $K^+$  is weakly\* closed, so by definition  $\Theta$  is also weakly\* closed. Since bounded weakly\* closed sets are exactly the weakly\* compact sets in  $(X^*, w^*)$ , the desired property is proved. ■

<sup>2</sup>For Jameson, 'cone' means pointed convex cone in our setting.

<sup>3</sup>Which is strange since it would prove the equivalence between a purely algebraic property with a topological one.

Given a base  $\Theta$  of  $K \subset (X, \tau)$ , we study now its support function

$$\begin{aligned} \sigma_{\Theta} : X^* &\longrightarrow \mathbb{R} \cup \{+\infty\} \\ x^* &\longmapsto \sup_{\theta \in \Theta} \langle x^*, \theta \rangle. \end{aligned}$$

**Proposition A.3.4.** Let  $\Theta$  be a  $\tau$ -bounded set of  $K$ . Then  $\sigma_{\Theta} : X^* \rightarrow \mathbb{R} \cup \{+\infty\}$  takes finite values, and is continuous for the strong topology on  $X^*$ .

*Proof.* Because of Mackey's theorem, the  $\tau$ -boundedness of  $\Theta$  implies its boundedness for the weak topology on  $X$ . As a consequence, for all  $x^* \in X^*$ ,  $x^*(\Theta)$  is a bounded set in  $\mathbb{R}$ . It is why  $\sigma_{\Theta}$  take finite values on  $X^*$ .

Let us turn now on continuity. From its definition, we know that  $\sigma_{\Theta}$  is lower semi-continuous for the weak\* topology, which is coarser than the strong topology  $\beta(X^*, w^*)$ . It follows that  $\sigma_{\Theta}$  is  $\beta$ -lower semi-continuous. Let us show now that it is also strongly upper semi-continuous. Take for this an arbitrary net  $(x_{\alpha}^*)_{\alpha \in \mathcal{A}}$  converging to some  $x^* \in X^*$  for the strong topology. In other words, we assume for all  $\tau$ -bounded sets  $A \subset X$  that  $\sup_{x \in A} |\langle x^* - x_{\alpha}^*, x \rangle|$  tends to zero when  $\alpha \in \mathcal{A}$ . In particular it holds for  $\Theta$  and we deduce that

$$(A.19) \quad |\sigma_{\Theta}(x^* - x_{\alpha}^*)| \leq \sup_{x \in \Theta} |\langle x^* - x_{\alpha}^*, x \rangle| \xrightarrow[\alpha \in \mathcal{A}]{\mathbb{R}} 0.$$

But from the sublinearity of  $\sigma_{\Theta}$ , we have

$$(A.20) \quad \sigma_{\Theta}(x_{\alpha}^*) \leq \sigma_{\Theta}(x^* - x_{\alpha}^*) + \sigma_{\Theta}(x^*) \text{ for all } \alpha \in \mathcal{A}.$$

So take the limsup over  $\alpha \in \mathcal{A}$  in (A.20), together with the limit obtained in (A.19), to deduce

$$\limsup_{\alpha \in \mathcal{A}} \sigma_{\Theta}(x_{\alpha}^*) \leq \sigma_{\Theta}(x^*)$$

and conclude. ■

Recall that the polar cone  $K^*$  and the dual cone  $K^+$  are just the negative of each other:  $K^* = -K^+$ .

**Proposition A.3.5.** Let  $K \subset (X, \tau)$  be a cone.

- i) If  $\Theta$  is a base of  $K$ , then  $K^* = [\sigma_{\Theta} \leq 0]$ .
- ii) If moreover  $\Theta$  is  $\tau$ -bounded and  $0 \notin \tau\text{-cl co}\Theta$ , then  ${}_{\beta}\text{int } K^* = [\sigma_{\Theta} < 0]$  is nonempty.

*Proof.* Item i) is immediate from  $K = \mathbb{R}_+\Theta$ :

$$x^* \in K^* \Leftrightarrow \forall x \in K, \langle x^*, x \rangle \leq 0 \Leftrightarrow \forall x \in \Theta, \langle x^*, x \rangle \leq 0 \Leftrightarrow \sigma_{\Theta}(x^*) \leq 0.$$

For item ii), take the strong interiors in item i), i.e.

$${}_{\beta}\text{int } K^* = {}_{\beta}\text{int } [\sigma_{\Theta} \leq 0].$$

We will now apply Proposition A.1.4 to conclude. We already know that  $\sigma_{\Theta}$  is convex, and is strongly continuous by Proposition A.3.4. We just need to verify that  $[\sigma_{\Theta} < 0]$  is nonempty. For this, consider the hypothesis  $0 \notin \tau\text{-cl co}\Theta$ , and apply Hanh-Banach separation theorem to obtain some  $x^* \in X^*$  and  $\alpha \in \mathbb{R}$  such that  $\sigma_{\Theta}(x^*) \leq \alpha < 0$ . ■

We can see that Theorem 2.3.7 is a direct consequence of Proposition A.3.5, applied to  $K^+ \subset (X^*, w^*)$ .

**Proposition A.3.6.** Let  $K$  be a convex cone having a convex base  $\Theta$ . Let  $e' \in K'^{++}$  be such that  $\Theta = \{x \in K \mid \langle e', x \rangle = 1\}$ . Then :

- i) for all  $x^* \in X^*$ ,  $\sigma_{\Theta}(x^*) = \inf_{t \in \mathbb{R}} \{x^* \leq_{X^*} te'\}$  where  $\leq_{X^*}$  denotes the dual order on  $X^*$  induced by  $K^+$ .
- ii) for all  $x^* \in X^*$  and  $\lambda \in \mathbb{R}$ ,  $\sigma_{\Theta}(x^* + \lambda e') = \sigma_{\Theta}(x^*) + \lambda$ .

*Proof.* Let  $t \in \mathbb{R}$ , then using the definition of  $e'$  and Proposition A.3.5, the item i) follows

$$(A.21) \quad x^* \leq_{X^*} te' \Leftrightarrow \forall \theta \in \Theta, \langle x^*, \theta \rangle \leq \langle te', \theta \rangle \Leftrightarrow \forall \theta \in \Theta, \langle x^*, \theta \rangle \leq t \Leftrightarrow \sigma_{\Theta}(x^*) \leq t.$$

Item ii) is a direct consequence of item i):

$$\begin{aligned} \sigma_{\Theta}(x^* + \lambda e') &= \inf_{t \in \mathbb{R}} \{x^* + \lambda e' \leq_{X^*} te'\} = \inf_{t \in \mathbb{R}} \{x^* \leq_{X^*} (t - \lambda)e'\} \\ &= \inf_{t \in \mathbb{R}} \{x^* \leq_{X^*} (t)e'\} - \lambda = \sigma_{\Theta}(x^*) - \lambda. \end{aligned}$$

■



# Appendix B

## Measurability and $L^p$ spaces

We present here some definitions and results which are necessary in Chapter 7. For the basic definitions in measure theory, we refer to [5].

### B.1 Measurability

Let  $(X, \tau_X)$  and  $(Y, \tau_Y)$  be two topological spaces. Let  $\Sigma_X$  be the Borel  $\sigma$ -algebra of  $(X, \tau_X)$ , i.e. the  $\sigma$ -algebra generated by the open sets of  $(X, \tau_X)$ . We say that  $f : X \rightarrow Y$  is *Borel measurable*<sup>1</sup> if  $f^{-1}(V) \in \Sigma_X$  for all open sets  $V$  of  $(Y, \tau_Y)$ . For instance, continuous functions are Borel measurable.

**Definition B.1.1.** Let  $T > 0$  and  $(X, \tau)$  a topological space. We say that  $f : [0, T] \times X \rightarrow \mathbb{R}$  is a Caratheodory function whenever

- i) for all  $x \in X$ ,  $f_x := f(\cdot, x) : [0, T] \rightarrow \mathbb{R}$  is Borel measurable,
- ii) for all  $t \in [0, T]$ ,  $f_t := f(t, \cdot) : X \rightarrow \mathbb{R}$  is continuous.

If  $S : X \rightrightarrows Y$  is a set-valued mapping, we say that it admits a *Borel measurable selection* whenever it exists a Borel measurable function  $s : X \rightarrow Y$  such that  $s(x) \in S(x)$  for all  $x \in X$ .

**Theorem B.1.2** (Castaing's measurable selection). *Let  $(X, \tau)$  be a metrizable separable compact topological space, and  $f : [0, T] \times X \rightarrow \mathbb{R}$  a Caratheodory function. Then the set-valued mapping*

$$\begin{aligned} [0, T] &\rightrightarrows X \\ t &\mapsto \operatorname{argmin}_{x \in X} f(t, x) \end{aligned}$$

*has nonempty compact values, and admits a Borel measurable selection.*

The proof of this result can be found for instance in [5, Theorem 18.19], or in the book of Castaing and Valadier [95] (Lemma III.39 p.36 and the application below).

### B.2 Strong measurability and $L^p$ spaces

Let  $(\mathcal{T}, \tau)$  be a compact topological space, and  $(Z, \|\cdot\|)$  be a Banach space. We equip  $\mathcal{T}$  with a measure  $\mu$  on its Borel sets, so that  $(\mathcal{T}, \Sigma_{\mathcal{T}}, \mu)$  is a finite measure space. We say that  $f : \mathcal{T} \rightarrow Z$  is a  $(\mu)$ -simple function whenever  $f(\mathcal{T})$  is finite, say  $f(\mathcal{T}) = \{z_1, \dots, z_N\}$ , and  $f^{-1}(z_i)$  is Borel measurable for each  $i \in \{1, \dots, N\}$ .

<sup>1</sup>See [5, Definition 4.21 and Corollary 4.24]

**Definition B.2.1.** We say that  $f : \mathcal{T} \rightarrow Z$  is *strongly measurable* (or *Bochner measurable*) if there exists a sequence  $(f_n)_{n \in \mathbb{N}}$  of simple functions such that for  $\mu$ -a.e.  $t \in \mathcal{T}$ ,  $f_n(t) \xrightarrow[n \rightarrow +\infty]{Z} f(t)$ .

Strong measurability is related to Borel measurability by the following Pettis's result<sup>2</sup>.

**Theorem B.2.2** (Pettis's measurability Theorem, [5, Lemma 11.37]). *A function  $f : \mathcal{T} \rightarrow Z$  is strongly measurable if and only if it is Borel measurable and is essentially separably valued<sup>3</sup>.*

In particular, we see that strong and Borel measurability coincide if  $Z$  is a separable Banach space. An easy example of strongly measurable function are the continuous functions:

**Proposition B.2.3.** Let  $f : \mathcal{T} \rightarrow Z$  be continuous, and assume that  $(\mathcal{T}, \tau)$  is metrizable. Then  $f$  is strongly measurable.

*Proof.* It suffices to exploit the fact that  $f$  is continuous on a compact metric space, hence uniformly continuous, to build a sequence of simple functions pointwise converging to  $f$ . Let  $d$  be the metric inducing the topology of  $\mathcal{T}$ . Using the uniform continuity of  $f$ , we can say that for all  $n \in \mathbb{N}$ , there exists  $\alpha_n > 0$  such that

$$(B.1) \quad \forall t, s \in \mathcal{T}, d(t, s) \leq \alpha_n \Rightarrow \|f(t) - f(s)\| \leq \frac{1}{n}.$$

Since a compact metric space is totally bounded [5, Theorem 3.28], we can cover  $\mathcal{T}$  with a finite number of open balls of radius  $\frac{\alpha_n}{2}$ . Using a finite number of boolean operations on these balls, we obtain a disjoint covering of  $\mathcal{T}$

$$\mathcal{T} \subset \bigsqcup_{k=1}^{K_n} B_k^n,$$

where the  $B_k^n$  are Borel sets, with diameter lower than  $\frac{\alpha_n}{2}$ . For all  $k \in \{1, \dots, K_n\}$ , define

$$\chi_{B_k^n} : \mathcal{T} \rightarrow \mathbb{R} \\ t \mapsto \begin{cases} 1 & \text{if } t \in B_k^n, \\ 0 & \text{if } t \notin B_k^n, \end{cases}$$

and pick some  $t_k^n \in B_k^n$ . Define then  $f_n := \sum_{k=1}^{K_n} f(t_k^n) \chi_{B_k^n}$ , which is clearly a simple function. Thus, given any  $\varepsilon > 0$ , we can take  $N \in \mathbb{N}$  satisfying  $\frac{1}{N} < \varepsilon$ , so that  $\|f(t) - f_n(t)\| \leq \varepsilon$  for all  $n \geq N$  and  $t \in \mathcal{T}$ . ■

According to [5, Lemma 11.39] or [137, Theorem II.2.2], for any strongly measurable function  $f : \mathcal{T} \rightarrow Z$ , the function  $\|f\| : \mathcal{T} \rightarrow \mathbb{R}$  defined by  $\|f\|(t) = \|f(t)\|$  is also strongly measurable, so we can define

$$\|f\|_{L^1(\mathcal{T}, Z)} := \int_{\mathcal{T}} \|f(t)\| d\mu(t) \in \mathbb{R} \cup \{+\infty\}.$$

We define then  $\mathcal{L}^1(\mathcal{T}, Z)$  the space of *Bochner integrable functions* as

$$\mathcal{L}^1(\mathcal{T}, Z) := \{f : \mathcal{T} \rightarrow Z \mid f \text{ is strongly measurable and } \|f\|_{L^1(\mathcal{T}, Z)} < +\infty\}.$$

We define  $L^1(\mathcal{T}, Z)$  as the quotient of  $\mathcal{L}^1(\mathcal{T}, Z)$  by this equivalence relation :

$$f \sim g \Leftrightarrow f(t) = g(t) \text{ for } \mu\text{-a.e. } t \in \mathcal{T}.$$

<sup>2</sup>In fact the original Pettis's result is stronger, see [137, Theorem II.1.2], but the one presented here is sufficient for us.

<sup>3</sup>This means that there exists a closed separated subspace  $Z_0$  of  $Z$  such that  $f(t) \in Z_0$  for  $\mu$ -a.e.  $t \in \mathcal{T}$ .

Similarly, we define for all  $p \in ]1, +\infty[$

$$\|f\|_{L^p(\mathcal{T}, Z)} := \left( \int_{\mathcal{T}} \|f(t)\|^p d\mu(t) \right)^{\frac{1}{p}} \in \mathbb{R} \cup \{+\infty\},$$

and

$$\|f\|_{L^\infty(\mathcal{T}, Z)} := \sup_{t \in \mathcal{T}} \|f(t)\|.$$

In an obvious way, we define the corresponding spaces of (classes of) strongly measurable functions  $L^p(\mathcal{T}, Z)$  and  $L^\infty(\mathcal{T}, Z)$ .

Next Proposition shows that the Bochner integral commutes with the bounded operators.

**Proposition B.2.4** ([5], Lemma 11.45). Let  $\phi : \mathcal{T} \rightarrow X$  be a Bochner integrable function, and  $A \in \mathcal{L}(X, Y)$  a continuous linear operator. Then  $A \circ \phi$  is also Bochner integrable, and

$$\int_{\mathcal{T}} A(\phi(t)) d\mu(t) = A \left( \int_{\mathcal{T}} \phi(t) d\mu(t) \right).$$

From now, we focus on the case  $\mathcal{T} = [0, T]$ ,  $T \in ]0, +\infty[$ , and give some properties of the spaces  $L^p([0, T], Z)$ .

**Theorem B.2.5.** For all  $p \in [1, +\infty]$ ,  $\|\cdot\|_{L^p([0, T], Z)}$  is a norm on  $L^p([0, T], Z)$ , which makes it a Banach space.

This result is exposed in [162, Theorem 2.100], [137, p.50, p.97] or [308, Theorem VI.1.1, Section VI.3]. One can also verify that  $L^2([0, T], Z)$  is Hilbert space if  $Z$  is a Hilbert space, if it is endowed with the inner product

$$\langle f, g \rangle_{L^2([0, T], Z)} := \int_0^T \langle f(t), g(t) \rangle_Z dt.$$

It is easy to see, as in the real case, that because of the finite measure of  $[0, T]$ , we have an ordered embedding of spaces :

$$L^\infty([0, T], Z) \subset L^q([0, T], Z) \subset L^p([0, T], Z) \subset L^1([0, T], Z) \quad \text{for all } 1 \leq p \leq q \leq +\infty.$$

We also have the useful:

**Proposition B.2.6.**  $C([0, T], Z) \subset L^\infty([0, T], Z)$  is an embedding.

*Proof.* Note that, by definition, the set of  $t \in [0, T]$  such that  $\|f\|_{L^\infty([0, T], Z)} < \|f(t)\|$  has null measure, and is open for a continuous function  $f$ . Both properties imply that this set is empty, and that  $\|f\|_{C([0, T], Z)} = \|f\|_{L^\infty([0, T], Z)}$ , so the result follows.  $\blacksquare$

An important point is to identify the topological dual of such  $L^p$  spaces. A first partial result can be found in [137, Theorem IV.1.1 (and the discussion above)], see also [308, Chapter VI] or [162, Theorem 2.112]:

**Theorem B.2.7.** Let  $p \in [1, +\infty[$ , and  $q \in ]1, +\infty]$  its conjugate exponent, i.e. such that  $\frac{1}{p} + \frac{1}{q} = 1$ . Then  $L^q([0, T], Z^*)$  can be isometrically viewed as a subset of  $L^p([0, T], Z)^*$ . Moreover, the equality  $L^q([0, T], Z^*) = L^p([0, T], Z)^*$  holds if  $Z$  is a reflexive<sup>4</sup> space. In both cases, the duality pairing is defined as follows:

$$\forall g \in L^q([0, T], Z^*), \forall f \in L^p([0, T], Z), \quad \langle g, f \rangle := \int_0^T \langle g(t), f(t) \rangle_{Z^* \times Z} dt.$$

<sup>4</sup>In fact, as it is stated in [137, Theorem IV.1.1], the equality  $L^q([0, T], Z^*) = L^p([0, T], Z)^*$  holds if and only if  $Z^*$  has the Radon-Nykodym property, which is also equivalent for  $Z$  to be an Asplund space, due to Stegall's Theorem [301]. This class includes the reflexive spaces, and the Banach spaces whose dual is separable.



**Remark B.2.8.** In Section 7.3, we need to work with the whole dual of  $L^1([0, T], Z)$ , for  $Z$  being not reflexive. In that case,  $L^1([0, T], Z)^*$  is strictly bigger than  $L^\infty([0, T], Z^*)$ , and its complete description pass through the consideration of weak\* measurable functions, see the next section.

We end this section with a classic result due to Rockafellar on the subdifferential of convex integrands [286, Theorem 4].

**Theorem B.2.9.** *Let  $X$  be an Euclidean space. Let  $f : [0, T] \times X \rightarrow \mathbb{R}$  be such that*

- *for all  $x \in X$ ,  $f_x := f(\cdot, x) : [0, T] \rightarrow \mathbb{R}$  is Borel measurable and has bounded values,*
- *for all  $t \in [0, T]$ ,  $f_t := f(t, \cdot) : X \rightarrow \mathbb{R}$  is convex.*

*Consider the associated convex integrand*

$$\begin{aligned} I_f : L^\infty([0, T], X) &\longrightarrow \mathbb{R} \\ \omega &\longmapsto \int_0^T f(t, \omega(t)) \, dt. \end{aligned}$$

*Then  $I_f$  is well-defined, convex and (strongly) continuous. Moreover its subdifferential enjoys the following characterization:  $\forall v \in L^1([0, T], X), \forall \omega \in L^\infty([0, T], X)$ ,*

$$v \in \partial I_f(\omega) \Leftrightarrow v(t) \in \partial f_t(\omega(t)) \text{ for a.e. } t \in [0, T].$$

**Remark B.2.10.** Note that in the above result, the subdifferential lives in the dual of  $L^\infty([0, T], X)$ , in which  $L^1([0, T], X)$  can be identified as a subspace, and where the subdifferential characterization holds.

Adapting the proof of the above theorem, we derive a useful corollary for multi-objective problems:

**Corollary B.2.11.** Let  $X$  be an Euclidean space, and let  $f_1, \dots, f_m : X \rightarrow \mathbb{R}$  be a finite family of convex continuous functions. Let  $\theta = (\theta_i) \in L^\infty([0, T], \mathbb{R}^m)$  be such that  $\theta(t) \in \Delta_m$  for a.e.  $t \in [0, T]$ . Consider the associated convex integrand

$$\begin{aligned} I : L^\infty([0, T], X) &\longrightarrow \mathbb{R} \\ \omega &\longmapsto \int_0^T \sum_{i=1}^m \theta_i(t) f_i(\omega(t)) \, dt. \end{aligned}$$

Then  $I$  is well-defined, convex and (strongly) continuous. Moreover, its subdifferential enjoys the following characterization:  $\forall v \in L^1([0, T], X), \forall \omega \in L^\infty([0, T], X)$ ,

$$v \in \partial I(\omega) \Leftrightarrow v(t) = \sum_{i=1}^m \theta_i(t) v_i(t) \text{ with } v_i(t) \in \partial f_i(\omega(t)) \text{ for a.e. } t \in [0, T].$$

*Proof.* By the Fenchel extremality relation,

$$(B.2) \quad v \in \partial I(\omega) \Leftrightarrow I(\omega) + I^*(v) - \langle \omega, v \rangle_{(L^\infty(0, T; \mathcal{H}), L^1(0, T; \mathcal{H}))} = 0.$$

Applying [286, Theorem 2] to the function  $f(t, x) := \sum_{i=1}^m \theta_i(t) f_i(x)$ , we obtain that the integrand  $I$  is well-defined, convex, continuous, and its Fenchel transform (see [51, 268]) is given by

$$I^*(v) = \int_0^T \left( \sum_{i=1}^m \theta_i(t) f_i \right)^* (v(t)) dt.$$

Let us analyze this last expression. Since the  $f_i$  are convex continuous functions, their conjugate are coercive functions, and

$$\left( \sum_{i=1}^m \theta_i(t) f_i \right)^* (v(t)) = \min \left\{ \sum_{i=1}^m (\theta_i(t) f_i)^* (z_i) \mid \sum_{i=1}^m z_i = v(t) \right\}.$$

The same measurable selection argument as the one used in Theorem 7.1.3 gives the existence of measurable functions  $z_i(\cdot)$  such that

$$(B.3) \quad \left( \sum_{i=1}^m \theta_i(t) f_i \right)^* (v(t)) = \sum_{i=1}^m (\theta_i(t) f_i)^* (z_i(t)), \quad \text{with } \sum_{i=1}^m z_i(t) = v(t).$$

Returning to (B.2), we obtain

$$v \in \partial I(\omega) \Leftrightarrow \int_0^T \sum_{i=1}^q (\theta_i(t) f_i)(\omega(t)) + (\theta_i(t) f_i)^* (z_i(t)) - \langle \omega(t), z_i(t) \rangle dt = 0.$$

Since each of the elements of this last sum expression is nonnegative, we deduce that, for each  $i \in \{1, \dots, m\}$  and for almost all  $t > 0$ ,

$$\theta_i(t) f_i(\omega(t)) + (\theta_i(t) f_i)^* (z_i(t)) - \langle \omega(t), z_i(t) \rangle = 0.$$

Equivalently,  $z_i(t) \in \partial (\theta_i(t) f_i)(\omega(t))$ . Since we know from (B.3) that  $v(t) = \sum_{i=1}^m z_i(t)$ , we need to verify that there exists  $v_i \in L^\infty([0, T], X)$  such that  $z_i(t) = \theta_i(t) v_i(t)$ .

Take some  $\tilde{z}_i \in L^\infty([0, T], X)$  such that  $\tilde{z}_i(t) \in \partial f_i(\omega(t))$  for almost all  $t > 0$ , (such an element exists, take for example  $\tilde{z}_i(t) = (\partial f_i)^0(\omega(t))$ ). We have

$$z_i(t) = \theta_i(t) v_i(t) \quad \text{for almost all } t > 0,$$

where

$$v_i(t) = \begin{cases} \frac{v(t)}{\theta_i(t)} & \text{if } \theta_i(t) > 0, \\ \tilde{z}_i(t) & \text{if } \theta_i(t) = 0. \end{cases}$$

This choice of  $v_i$  is measurable, and  $v_i(t) \in \partial f_i(\omega(t))$  for almost all  $t > 0$ . By continuity of  $f_i$ , we conclude that  $v_i \in L^\infty([0, T], X)$ . ■

### B.3 Weak\* measurability and $L_{w^*}^\infty$ space

Let  $f : [0, T] \rightarrow Z^*$ , where  $Z$  is a Banach space. We say that  $f$  is *weak\* measurable* if, for all  $z \in Z$ , the function

$$z \circ f : t \in [0, T] \mapsto \langle f(t), z \rangle \in \mathbb{R}$$

is strongly measurable. Note that we abusively note  $z \circ f$ , as if  $z$  was a linear functional on  $Z^*$ . In fact this notation makes sense if one sees  $z$  as an element of  $Z^{**}$  through the canonical embedding  $Z \hookrightarrow Z^{**}$ . For example, Borel measurable functions from  $[0, T]$  to  $Z^*$  are in particular weakly\* measurable

Define  $\mathcal{L}_{w^*}^\infty([0, T], Z^*)$  as the vector space of bounded weak\* measurable functions. Consider the following equivalence relation on  $\mathcal{L}_{w^*}^\infty([0, T], Z^*)$ :

$$f \sim g \Leftrightarrow \forall z \in Z, \text{ for a.e. } t \in [0, T], \langle f(t), z \rangle = \langle g(t), z \rangle.$$

Define then  $L_{w^*}^\infty([0, T], Z^*)$  as the quotient of  $\mathcal{L}_{w^*}^\infty([0, T], Z^*)$  by this equivalence relation. It is a Banach space, once equipped with the essential supremum norm (see [308, Corollary VI.4 p.78 and Remark VII.2 p.89]). As announced (see [308, Corollary VII.4 p.95]):

**Theorem B.3.1.**  $L_w^\infty([0, T], Z^*)$  is isometrically equal to the dual space  $L^1([0, T], Z)^*$ . The duality pairing is defined as follows:

$$\forall g \in L_w^\infty([0, T], Z^*), \forall f \in L^1([0, T], Z), \langle g, f \rangle := \int_0^T \langle g(t), f(t) \rangle_{Z^* \times Z} dt.$$

# Appendix C

## Others

**Theorem C.0.1** (Arzela-Ascoli's Theorem). *Let  $X$  be a compact topological vector space, and  $E$  a metric space. Let  $C(X, E)$  be the space of continuous functions from  $X$  to  $E$ , equipped with the uniform metric. Then a set  $A \subset C(X, E)$  is relatively compact if and only if the family  $A$  is equicontinuous, and  $A(x) := \{f(x) \mid f \in A\}$  is relatively compact in  $E$ .*

*Proof.* See [138, Theorem 7.5.7 p. 142], or [81, Theorem X.2.5.2 p. 290]. ■

**Proposition C.0.2.** Let  $Y$  be a separable Banach space, and  $\Theta$  a weakly\* compact subset of  $Y^*$ . Then,

- i)  $(\Theta, w^*)$  is a metrizable and separable topological vector space,
- ii)  $(C(\Theta), \|\cdot\|_{C(\Theta)})$  is a separable Banach space,
- iii)  $(\mathcal{P}^R(\Theta), w^*)$  is a metrizable separable topological space.

*Proof.* Since  $(Y, \|\cdot\|_Y)$  is separable, it follows from [153, Corollary 3.104] that  $(\mathbb{B}_{Y^*}, w^*)$  is separable, where  $\mathbb{B}_{Y^*}$  denotes the unit ball in  $Y^*$ . From [5, Theorem 6.30] it follows also that  $(\mathbb{B}_{Y^*}, w^*)$  is metrizable. Since  $\Theta$  is compact, we have  $\Theta \subset n\mathbb{B}_{Y^*}$  for some  $n \in \mathbb{N}$ . But  $(n\mathbb{B}_{Y^*}, w^*)$  is metrizable and separable, so any of its topological subspace is also metrizable and separable (see for instance [138, 3.10.9 p.43]). This proves that  $(\Theta, w^*)$  is metrizable and separable. Now use [153, Lemma 3.102] to obtain that  $(C(\Theta), \|\cdot\|_{C(\Theta)})$  is separable. Using now the same arguments for  $(\mathcal{P}^R(\Theta), w^*)$  than the ones for  $(\Theta, w^*)$ , we finally obtain that  $(\mathcal{P}^R(\Theta), w^*)$  is metrizable and separable (see also [5, Theorem 15.11]). ■

We give here the two integral forms of Gronwall's Lemma that we used in the proof of Theorem 8.1.5. They can be found in Brezis's book [83, Lemma A.4 & Lemma A.5, pp. 156–157].

**Lemma C.0.3** (Gronwall-Bellman). Let  $t_0 \in \mathbb{R}$  and  $T \in ]t_0, +\infty[$ . Let  $a \in [0, +\infty[$ , and  $g \in L^1([0, T], \mathbb{R})$  with  $g(t) \geq 0$  for a.e.  $t \in [0, T]$ . Let  $h \in C([0, T], \mathbb{R})$  such that

$$(C.1) \quad h(t) \leq a + \int_{t_0}^t g(s)h(s) \, ds \quad \text{for all } t \in [t_0, T].$$

Then  $h(t) \leq ae^{\int_{t_0}^t g(s) \, ds}$  for all  $t \in [t_0, T]$ .

**Lemma C.0.4.** Let  $t_0 \in \mathbb{R}$  and  $T \in ]t_0, +\infty[$ . Let  $a \in [0, +\infty[$ , and  $g \in L^1([0, T], \mathbb{R})$  with  $g(t) \geq 0$  for a.e.  $t \in [0, T]$ . Let  $h \in C([0, T], \mathbb{R})$  such that

$$(C.2) \quad \frac{1}{2}h^2(t) \leq \frac{a^2}{2} + \int_0^t g(s)h(s) \, ds \quad \text{for all } t \in [0, T],$$

then  $|h(t)| \leq a + \int_0^T g(s) \, ds$  for all  $t \in [0, T]$ .



# Bibliography

- [1] P.-A. Absil, R. Mahony and B. Andrews, *Convergence of the iterates of descent methods for analytic cost functions*, SIAM J. Optim., **16**, pp. 531–547, 2005.
- [2] S. Adly and H. Rammal, *A new method for solving Pareto eigenvalue complementarity problems*, Computational Optimization and Applications, **55**(3), pp. 703–731, 2013.
- [3] S. Adly and A. Seeger, *A nonsmooth algorithm for cone-constrained eigenvalue problems*, Computational Optimization and Applications, **49**(2), pp. 299–318, 2011.
- [4] N.E. Alaa and M. Pierre, *Convergence to equilibrium for discretized gradient-like systems with analytic features*, IMA Journal of Numerical Analysis, **33**(4), pp. 1291–1321, 2013.
- [5] C.D. Aliprantis and K.C. Border, *Infinite dimensional analysis, a Hitchhiker’s Guide*, Springer Science & Business Media, Third edition, 2006.
- [6] C.D. Aliprantis and O. Burkinshaw, *Principles of real analysis*, Academic Press, Third edition, 1998.
- [7] C.D. Aliprantis, M. Florenzano, V. F. Martins da Rocha and R. Tourky, *Equilibrium analysis in financial markets with countably many securities*, Journal of Mathematical Economics, **40**(6), pp. 683–699, 2004.
- [8] C.D. Aliprantis, M. Florenzano, and R. Tourky, *General equilibrium analysis in ordered topological vector spaces*, Journal of Mathematical Economics, **40**(3-4), pp. 247–269, 2004.
- [9] F. Alvarez, *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM J. Control Optim., **38**(4), pp. 1102–1119, 2000.
- [10] F. Alvarez, H. Attouch, J. Bolte and P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping: application to optimization and mechanics*, Journal de Mathématiques Pures et Appliquées, **81**(8), pp. 747–779, 2002.
- [11] F. Alvarez, J. Bolte and O. Brahic, *Hessian Riemannian gradient flows in convex programming*, SIAM J. Control Optim., **43**(2), pp. 477–501, 2004.
- [12] F. Alvarez, J. López and H. Ramírez, *Interior proximal algorithm with variable metric for second-order cone programming: applications to structural optimization and support vector machines*, Optim. Methods Softw., **25**(4-6), pp. 859–881, 2010.
- [13] F. Alvarez and J. Peypouquet, *Asymptotic almost-equivalence of Lipschitz evolution systems in Banach spaces*, Nonlinear Anal., **73** (9), pp. 3018–3033, 2010.
- [14] F. Alvarez and J. Peypouquet, *A unified approach to the asymptotic almost-equivalence of evolution systems without Lipschitz conditions*, Nonlinear Anal., **74**(11), pp. 3440–3444, 2011.

- [15] A.S. Antipin, *Second order proximal differential systems with feedback control*, Differential Equations, **29**(11), pp. 1597–1607, 1993.
- [16] L. Armijo, *Minimization of functions having Lipschitz continuous first partial derivatives*, Pacific Journal of Mathematics, **16**(1), pp. 1–3, 1966.
- [17] H. Attouch, *Variational convergence for functions and operators*, Pitman, Applicable Mathematical Series, 1984.
- [18] H. Attouch and J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Math. Program. Ser. B **116**, pp. 5–16, 2009.
- [19] H. Attouch, J. Bolte, P. Redont and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Łojasiewicz inequality*, Math. Oper. Res., **35**(2), pp. 438–457, 2010.
- [20] H. Attouch, J. Bolte and B.F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Math. Program., **137**(1-2), pp. 91–129, 2013.
- [21] H. Attouch, G. Buttazzo and G. Michaille, *Variational analysis in Sobolev and BV Spaces: Applications to PDEs and optimization*, MPS-SIAM Series on Optimization, 2006.
- [22] H. Attouch, Z. Chbani, J. Peypouquet and P. Redont, *Fast convergence of an inertial dynamical system with asymptotic vanishing damping*, to appear.
- [23] H. Attouch and R. Cominetti, *A dynamical approach to convex minimization coupling approximation with the steepest descent method*, Journal of Differential Equations, **128**(2), pp. 519–540, 1996.
- [24] H. Attouch and M.-O. Czarnecki, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*, Journal of Differential Equations, **248**(6), pp. 1315–1344, 2010.
- [25] H. Attouch, M.-O. Czarnecki and J. Peypouquet, *Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities*, SIAM Journal on Optimization, **21**(4), pp. 1251–1274, 2011.
- [26] H. Attouch and G. Garrigos, *Multiobjective optimization: an inertial dynamical approach to Pareto optima*, preprint available on arXiv:1506.02823, submitted on June 2015.
- [27] H. Attouch, G. Garrigos and X. Goudou, *A dynamic gradient approach to Pareto optimization with nonsmooth convex objective functions*, Journal of Mathematical Analysis and Applications, **422**(1), pp. 741–771, 2015.
- [28] H. Attouch and X. Goudou, *A continuous gradient-like dynamical approach to Pareto optimization in Hilbert spaces*, Set-Valued and Variational Analysis, **22**(1), pp. 189–219, 2014.
- [29] H. Attouch, X. Goudou and P. Redont, *The heavy ball with friction method I, the continuous dynamical system*, Communications in Contemporary Mathematics, **2**(1), pp. 1–34, 2000.
- [30] H. Attouch and P.-E. Maingé, *Asymptotic behavior of second-order dissipative evolution equations combining potential with non-potential effects*, ESAIM: Control, Optimisation and Calculus of Variations, **17**(3), pp. 836–857, 2011.
- [31] H. Attouch and B.F. Svaiter, *A continuous dynamical Newton-like approach to solving monotone inclusions*, SIAM J. Optim., **49**(2), pp. 574–598, 2011.

- [32] H. Attouch and R. Wets, *Quantitative stability of variational systems II, a framework for nonlinear conditioning*, SIAM J. Opt., **3**(2), pp. 359–381, 1993.
- [33] J.P. Aubin and I. Ekeland, *Applied nonlinear analysis*, Wiley, 1984.
- [34] A. Auslender, *Méthodes numériques pour la résolution des problèmes d'optimisation avec contraintes*, PhD thesis, Uni. Grenoble, 1969.
- [35] A. Auslender, *Optimisation: Méthodes numériques*, Masson, Paris, 1976.
- [36] D. Aussel, J.-N. Corvellec and M. Lassonde, *Mean value property and subdifferential criteria for lower semicontinuous functions*, Transactions of the American Mathematical Society, **347**(10), pp. 4147–4161, 1995.
- [37] D. Aussel and A. Daniilidis, *Normal characterization of the main classes of quasiconvex functions*, Set-valued analysis, **8**(3), pp. 219–236, 2000.
- [38] D. Aussel and A. Daniilidis, *Normal cones to sublevel sets: an axiomatic approach*, in Generalized Convexity and Generalized Monotonicity, Lecture Notes in Economics and Mathematical Systems Vol. **502**, Springer, 2001.
- [39] J.M. Ayerbe Toledano, T. Domínguez Benavides and G. López Acedo, *Measures of non-compactness in metric fixed point theory*, Springer Science & Business Media, Vol. 99, 1997.
- [40] J.-B. Baillon, P.L. Combettes and R. Cominetti, *There is no variational characterization of the cycles in the method of periodic projections*, Journal of Functional Analysis, **261**(1), pp. 400–408, 2012.
- [41] U. Bagchi *Simultaneous minimization of mean and variation of flow time and waiting time in single machine systems*, Oper. Res., **37**, pp. 118–125, 1989.
- [42] R. Baranuik, M. Davenport, R. DeVore and M. Wakin, *A simple proof of the restricted isometry property for random matrices*, Constructive Approximation, **28**(3), pp. 253–263, 2008.
- [43] E. Barone, *Il Ministero della Produzione hello Stato Collettivista*, Giornale degli economisti e rivista di statistica, **37**, pp. 267-293 and pp. 391-414, 1908. Translation as *The Ministry of production in the collectivist state*, F. A. Hayek, Collectivist Economic Planning, London, 1935.
- [44] T. Bárta, R. Chill and E. Fasangová, *Every ordinary differential equation with a strict Lyapunov function is a gradient system*, Monatshefte für Mathematik, **166**(1), pp. 57–72, 2012.
- [45] S. Banach, *Über die Baire'sche Kategorie gewisser Funktionenmengen*, Studia Mathematica, **3**(1), pp. 174–179, 1931.
- [46] T.Q. Bao and B.S. Mordukhovich, *Sufficient conditions for global weak Pareto solutions in multiobjective optimization*, Positivity, **16**(3), pp. 579–602, 2012.
- [47] L. Baudoin and J. Salomon, *Constructive solution of a bilinear optimal control problem for a Schrödinger equation*, Syst Contr Lett., **57**(6), pp. 453–464, 2008.
- [48] H.H. Bauschke and J.M. Borwein, *On the convergence of von Neumann's alternating projection algorithm for two sets*, Set-Valued Analysis, **1**(2), pp. 185–212, 1993.
- [49] H.H. Bauschke and J.M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Review, **38**(3), pp. 367–426, 1996.



- [50] H.H. Bauschke and P.L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces*, Mathematics of Operations Research, **26**, pp. 248–264, 2001.
- [51] H.H. Bauschke and P. Combettes, *Convex analysis and monotone operator theory*, CMS books in Mathematics, Springer, 2011.
- [52] R. Beausoleil, ‘*MOSS*’ *multiobjective scatter search applied to non-linear multiple criteria optimization*, European Journal of Operational Research, **169**(2), pp. 426–449, 2006.
- [53] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., **2**(1), pp. 183–202, 2009.
- [54] A. Beck and M. Teboulle, *Gradient-based algorithms with applications in signal recovery problems*, In Convex Optimization in Signal Processing and Communications, D. Palomar and Y. Eldar Eds., pp. 33–88. Cambridge University Press, 2010.
- [55] P. Bégout, J. Bolte and M.A. Jendoubi, *On damped second-order gradient systems*, Journal of Differential Equations, **259**(7), pp. 3115–3143, 2015.
- [56] R.E. Bellman, R.E. Kalaba and J. Lockett, *Numerical inversion of the Laplace transform*, Elsevier, New York, 1966.
- [57] P. Bénélian, M.G. Crandall and A. Pazy, “*Bonnes solutions*” *d’un problème d’évolution semi-linéaire*, C. R. Acad. Sci. Paris Sér. I Math., **306**(13), pp. 527–530, 1988.
- [58] H.P. Benson, *Optimization over the efficient set*, Journal of Mathematical Analysis and Applications, **98**(2), pp. 562–580, 1984.
- [59] G.C. Bento and J.X. Cruz Neto, *A Subgradient Method for Multiobjective Optimization on Riemannian Manifolds*, Journal of Optimization Theory and Applications, **159**(1), pp. 125–137, 2013.
- [60] G.C. Bento, J.X. Cruz Neto, P.R. Oliveira and A. Soubeyran, *The self regulation problem as an inexact steepest descent method for multicriteria optimization*, European Journal of Operational Research, **235**(3), pp. 494–502, 2014.
- [61] G.C. Bento, J.X. Cruz Neto, and A. Soubeyran, *A proximal point-type method for multicriteria optimization*, Set-Valued and Variational Analysis, **22**(3), pp. 557–573, 2014.
- [62] G.C. Bento, O.P. Ferreira and P.R. Oliveira, *Unconstrained steepest descent method for multicriteria optimization on riemannian manifolds*, Journal of Optimization Theory and Applications, **154**(1), pp. 88–107, 2012.
- [63] J.Y. Bello Cruz, P.R. Lucambio Pérez and J.G. Melo, *Convergence of the projected gradient method for quasiconvex multiobjective optimization*, Nonlinear Analysis: Theory, Methods and Applications, **74**(16), pp. 5268–5273, 2011.
- [64] D.P. Bertsekas, *Projected Newton methods for optimization problems with simple constraints*. SIAM J. on Control and Optim., **20**(2), pp. 221–246, 1982.
- [65] D. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont MA, 1999.
- [66] J. Bochnak, M. Coste and M.-F. Roy, *Géométrie algébrique réelle*, Ergebnisse der Mathematik und ihrer Grenzgebiete, 3 Folge, Band 12, Springer-Verlag, Berlin, Heidelberg, and New York, 1987.

- [67] S. Bolintineanu, *Optimality conditions for minimization over the (weakly or properly) efficient set*, Journal of Mathematical Analysis and Applications, **173**(2), pp. 523–541, 1993.
- [68] J. Bolte, A. Daniilidis and A.S. Lewis, *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., **17**, pp. 1205–1223, 2007.
- [69] J. Bolte, A. Daniilidis and A.S. Lewis, *Mathematical Programming*, **117**(1-2), pp. 5–19, 2009.
- [70] J. Bolte, A. Daniilidis, A.S. Lewis and M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM J. Optim., **18**(2), pp. 556–572, 2007.
- [71] J. Bolte, A. Daniilidis, O. Ley and L. Mazet, *Characterizations of Lojasiewicz inequalities: Subgradient flows, talweg, convexity*, Trans. Amer. Math. Soc., **362**, pp. 3319–3363, 2010.
- [72] J. Bolte, S. Sabach and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming, **146**(1-2), pp. 459–494, 2013.
- [73] H. Bonnel, A. Iusem and B.F. Svaiter, *Proximal methods in vector optimization*, SIAM Journal on Optimization, **15**(4), 953–970, 2005.
- [74] H. Bonnel and C.Y. Kaya, *Optimization over the efficient set of multi-objective convex optimal control problems*, Journal of Optimization Theory and Applications, **147**(1), pp. 93–112, 2010.
- [75] J. Borde and J.-P. Crouzeix, *Continuity properties of the normal cone to the level sets of a quasiconvex function*, Journal of Optimization Theory and Applications, **66**, pp. 415–429, 1990.
- [76] J.M. Borwein, *Continuity and differentiability properties of convex operators*, Proceedings of the London Mathematical Society, **s3-44**(3), pp. 420–444, 1982.
- [77] J.M. Borwein and A.S. Lewis, *Convex analysis and nonlinear optimization: Theory and examples*, CMS Books in Mathematics, Springer, 2006.
- [78] J.M. Borwein, W. Moors and X. Wang, *Generalized subdifferentials: a Baire categorical approach*, Transactions of the AMS, **353**(10), pp. 3875–3893, 2001.
- [79] J.M. Borwein and Q.J. Zhu, *Techniques of variational analysis*, Springer Science & Business Media, 2005.
- [80] R.I. Bot, S.-M. Grad and G. Wanka, *Duality in vector optimization*, Springer Science & Business Media, 2009.
- [81] N. Bourbaki, *General topology: Chapters 5-10*, Addison-Wesley, Springer, 1989.
- [82] H. Brézis, *Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations*, in Contributions to Nonlinear Functional Analysis (Madison, 1971), E. H. Zarantonello (ed.), Academic Press, New York, pp. 101–156, 1971.
- [83] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland/Elsevier, New-York, 1973.
- [84] H. Brézis, *Analyse fonctionnelle: Théorie et applications*, Masson, 1989.

- [85] H. Brézis, *Functional analysis, Sobolev spaces and partial differential equations*, Springer New York, 2010.
- [86] H. Brézis and P.-L. Lions, *Produits infinis de résolvantes*, Israel J. Math., **29**(4), pp. 329–345, 1978.
- [87] R.E. Bruck, *Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces*, Journal of Functional Analysis, **18**, pp. 15–26, 1975.
- [88] M. Brown and R.E. Smith, *Directed multi-objective optimisation*, Internat. J. Comput. Syst. Signals, **6**(1), pp. 3–17, 2005.
- [89] R. Burachik, L.M. Graña Drummond, A.N. Iusem and B.F. Svaiter, *Full convergence of the steepest descent method with inexact line searches*, Optimization, **32**(2), pp. 137–146, 1995.
- [90] A. Cabot, *The steepest descent dynamical system with control. Applications to constrained minimization*, ESAIM: Control, Optimisation and Calculus of Variations, **10**(2), pp. 243–258, 2004.
- [91] A. Cabot and L. Thibault, *Sequential formulae for the normal cone to sublevel sets*, Transactions of the American Mathematical Society, **366**(12), pp. 6591–6628, 2014.
- [92] A. Cabot, H. Engler and S. Gadat, *On the long time behavior of second order differential equations with asymptotically small dissipation*, Transactions of the American Mathematical Society, **361**, pp. 5983–6017, 2009.
- [93] E.J. Candès, *The restricted isometry property and its implications for compressed sensing*, Comptes Rendus Mathématique, **346**(9-10), pp. 589-592, 2008.
- [94] E. Casas, R. Herzog and G. Wachsmuth, *Approximation of sparse controls in semilinear equations by piecewise linear functions*, Numer. Math., **122**(4), pp. 645–669, 2012.
- [95] C. Castaing and M. Valadier, *Convex analysis and measurable multifunctions*, Lectures notes in mathematics, Springer, **580**, 1977.
- [96] A.-L. Cauchy, *Méthode générale pour la résolution des systèmes d'équations simultanées*, C. R. Acad. Sci. Paris, **25**, pp. 536–538, 1847.
- [97] T. Cazenave, *An introduction to semilinear elliptic equations*, Lecture notes, Editora do IM-UFRJ, Rio de Janeiro, 2006. Available on the author's webpage: [www.ljll.math.upmc.fr/cazenave](http://www.ljll.math.upmc.fr/cazenave)
- [98] L.-C. Ceng, B.S. Mordukhovich and J.-C. Yao, *Hybrid approximate proximal method with auxiliary variational inequality for vector optimization*, Journal of Optimization Theory and Applications, **146**(2), pp. 267–303, 2010.
- [99] L.-C. Ceng and J.-C. Yao, *Approximate proximal methods in vector optimization*, European Journal of Operational Research, **183**(1), pp. 1–19, 2007.
- [100] A. Chambolle and C. Dossal, *On the convergence of the iterates of Fista*, preprint available on HAL[hal-01060130], submitted on October 2014.
- [101] V. Chandrasekaran, S. Sanghavi, P.A. Parillo and A.S. Willsky, *Rank-sparsity incoherence for matrix decomposition*, SIAM J. Opt., **21**(2), pp. 572–596, 2011.

- [102] A. Charnes, and W.W. Cooper, *The strong Minkowski-Farkas-Weyl theorem for vector spaces over ordered fields*, Proceedings of the National Academy of Sciences of the United States of America, **44**(9), pp. 914–916, 1958.
- [103] R. Chill, *The Lojasiewicz-Simon gradient inequality in Hilbert spaces*, Proceedings of the 5th European-Maghrebian Workshop on Semigroup Theory, Evolution Equations, and Applications (M.-A. Jendoubi, ed.), pp. 25–36, 2006.
- [104] R. Chill and A. Fiorenza, *Convergence and decay rate to equilibrium of bounded solutions of quasilinear parabolic equations*, J. Differential Equations, **228**, pp. 611–632, 2006.
- [105] R. Chill and M.-A. Jendoubi, *Convergence to steady states in asymptotically autonomous semilinear evolution equations*, Nonlinear Anal., **53**, pp. 1017–1039, 2003.
- [106] E. Chouzenoux, J.-C. Pesquet and A. Repetti, *Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function*, J. Optim. Theory Appl., **162**(1), pp. 107–132, 2013.
- [107] E. Chouzenoux, J.-C. Pesquet and A. Repetti, *A block coordinate variable metric forward-backward algorithm*, preprint on HAL[hal-00945918], submitted on February 2014.
- [108] J. Christensen, *Topology and Borel structure*, Math. studies 10, Notas de Matemàtica, Elsevier, 1974.
- [109] J.X. Cruz Neto, G.J.P. Da Silva , O.P. Ferreira and J.O. Lopes, *A subgradient method for multiobjective optimization*, Computational Optimization and Applications, **54**(3), pp. 461–472, 2013.
- [110] F.H. Clarke, *A new approach to Lagrange multipliers*, Mathematics of Operations Research, **1**(2), pp.165-174, 1976.
- [111] F.H. Clarke, *Optimization and nonsmooth analysis*, Society for Industrial and Applied Mathematics, 1990.
- [112] P.L. Combettes, *A block-iterative surrogate constraint splitting method for quadratic signal recovery*, IEEE Transactions on Signal Processing, **51**, pp. 1771–1782, 2003.
- [113] P.L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-point algorithms for inverse problems in science and engineering, Springer New York, 2011.
- [114] R. Cominetti, J. Peypouquet and S. Sorin, *Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization*, Journal of Differential Equations, **245**(12), pp. 3753–3763, 2008.
- [115] B. Cornet, *Monotone planning procedures and accessibility of Pareto optima*, in: M. Aoki, A. Marzollo (Eds.), New Trends in Dynamic System Theory and Economics, Academic Press, pp. 337–349, 1979.
- [116] B. Cornet, *Contributions to the mathematical theory of dynamical mechanisms of resource allocation*, Thèse de doctorat d'état, Université Paris IX Dauphine, 1981.
- [117] B. Cornet, *Existence of slow solutions for a class of differential inclusions*, Journal of Mathematical Analysis and Applications, **96**, pp. 130–147, 1983.
- [118] R. Correa, P. Gajardo and L. Thibault, *Subdifferential representation formula and subdifferential criteria for the behavior of nonsmooth functions*, Nonlinear Analysis: Theory, Methods & Applications, **65**(4), pp. 864–891, 2006.

- [119] M.G. Crandall and T.M. Liggett, *Generation of semi-groups of nonlinear transformations on general Banach spaces*, American Journal of Mathematics, **93**(2), pp. 265–298, 1971.
- [120] G. Crombez, *Image recovery by convex combinations of projections*, Journal of Mathematical Analysis and Applications, **155**, pp. 413–419, 1991.
- [121] M. Crouzeix, *Une méthode multipas implicite-explicite pour l'approximation des équations d'évolution paraboliques*, Numerische Mathematik, **35**(3), pp. 257–276, 1980.
- [122] J.X. Cruz Neto, G.J.P Da Silva, O.P. Ferreira and J.O. Lopes, *A subgradient method for multiobjective optimization*, Computational Optimization and Applications, **54**(3), pp. 461–472, 2013.
- [123] H.B. Curry, *The method of steepest descent for nonlinear minimization problems*, The Quarterly Journal of Mechanics and Applied Mathematics, **2**, pp. 258–261, 1944.
- [124] M.-O. Czarnecki, *Asymptotic control and stabilization of nonlinear oscillators with non isolated equilibria, a note: from  $L^1$  to non  $L^1$* , Journal of Differential Equations, **217**(2), pp. 501–511, 2005.
- [125] M.-O. Czarnecki, N. Noun and J. Peypouquet, *Splitting forward-backward penalty scheme for constrained variational problems*, preprint on arXiv:1408.0974, published on Aug 2014.
- [126] I. Das and J.E. Dennis, *A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multi-criteria optimization problems*, **14**(1), pp. 63–69, 1997.
- [127] I. Daubechies, M. Defrise and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math, **57**(11), pp. 1413–1457, 2004.
- [128] J.P. Dauer and W. Stadler, *A survey of vector optimization in infinite-dimensional spaces: part II*, Journal of Optimization Theory and Applications, **51**(2), pp. 205–241, 1986.
- [129] K. Deb, *Multi-objective genetic algorithms: problem difficulties and construction of test problems*, Evol. Comput., **7**(3), pp. 205–230, 1999.
- [130] K. Deb, S. Agrawal, A. Pratap and T. Meyarivan, *A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II*, Lecture Notes in Computer Science, **1917**, pp. 849–858, 2000.
- [131] G. Debreu, *Theory of value*, John Wiley and Sons, New York, 1959.
- [132] G. Debreu, *Valuation equilibrium and Pareto optimum*, Proceedings of The National Academy of Sciences, **40**, pp. 588–592, 1954.
- [133] M. Degiovanni, R. Luccetti and N. Ribarska, *Critical point theory for vector valued functions*, Journal of Convex Analysis, **9**(2), pp. 415–428, 2002.
- [134] M. Degiovanni, A. Marino and M. Tosques, *Evolution equations with lack of convexity*, Nonlinear Analysis, **9**, pp. 1401–1443, 1985.
- [135] A.R. De Pierro and A.N. Iusem, *A parallel projection method for finding a common point of a family of convex sets*, Pesquisa Operacional, **5**, pp. 1–20, 1985.
- [136] F. Deutsch, *Best approximation in inner product spaces*, Springer New York, 2001.

- [137] J. Diestel and J. J. Uhl, *Vector measures*, American Mathematical Society, Mathematical Surveys No. 15, 1977.
- [138] J. Dieudonné, *Foundations of modern analysis*, New York : Academic press, Vol. 286, 1960.
- [139] L. Dinh The, *Theory of vector optimization*, Springer-Verlag, Berlin-New-York, 1988.
- [140] L. van den Dries, *A generalization of the Tarski-Seidenberg theorem, and some nondefinability results*, Bulletin of the American Mathematical Society, **15**(2), pp. 189–193, 1986.
- [141] L. van den Dries, *Tame Topology and O-minimal Structures*, Cambridge University Press, 1998.
- [142] L. van den Dries, *Tame topology and o-minimal structures*, Bull. of the AMS, **37**(3), pp. 351–357, 2000.
- [143] L. van den Dries, A. Macintyre and D. Marker, *The elementary theory of restricted analytic fields with exponentiation*, Annals of Mathematics, **140**(1), pp. 183–205, 1994.
- [144] L. van den Dries and C. Miller, *Geometric categories and o-minimal structures*, Duke Math. J., **84**, pp. 497–540, 1996.
- [145] L. van den Dries and P. Speissegger, *The real field with convergent generalized power series*, Transactions of the American Mathematical Society, **350**(11), pp. 4377–4421, 1998.
- [146] L. van den Dries and P. Speissegger, *The field of reals with multisummable series and the exponential function*, Proceedings of the London Mathematical Society, **81**(3), pp. 513–565, 2000.
- [147] D.L. Donoho, *High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension*, Discrete & Computational Geometry, **35**(4), pp. 617–652, 2005.
- [148] D.L. Donoho and V. Stodden, *Breakdown point of model selection when the number of variables exceeds the number of observations*, International Joint Conference on Neural Networks, pp. 1916–1921, 2006.
- [149] D.L. Donoho, *For most large underdetermined systems of linear equations the minimal  $\ell^1$ -norm solution is also the sparsest solution*, Comm. on Pure and Appl. Math., **59**(6), pp. 797–829, 2006.
- [150] D.L. Donoho and J. Tanner, *Counting the faces of randomly-projected hypercubes and orthants, with applications*, Discrete Comput. Geom., **43**, pp. 522–541, 2010.
- [151] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika, **1**(3), pp. 211–218, 1936.
- [152] F.Y. Edgeworth, *Mathematical psychics*, P. Keagan, England, 1881.
- [153] M. Fabian, P. Habala, P. Hájek, V. Montesinos and V. Zizler, *Banach space theory: The basis for linear and nonlinear analysis*, Springer Science & Business Media, 2011.
- [154] J. Farkas, *Theorie der einfachen Ungleichungen*, J. F. Reine u. Ang. Mat., **124**, 1–27, 1902.
- [155] E. Feireisl, F. Issard-Roch, and H. Petzeltová, *A non-smooth version of the Lojasiewicz-Simon theorem with applications to non-local phase-field systems*, Journal of Differential Equations, **199**(1), pp. 1–21, 2004.

- [156] E. Feireisl and F. Simondon, *Convergence for semilinear degenerate parabolic equations in several space dimensions*, J. Dynamics Differential Equations, **12**(3), pp. 647–673, 2000.
- [157] M. Ferris, *Finite termination of the proximal point algorithm*, Math. Program., **50**, pp. 359–366, 1991.
- [158] P.C. Fishburn, *Exceptional paper-lexicographic orders, utilities and decision rules: A survey*, Management science, **20**(11), pp. 1442–1471.
- [159] J. Fliege and B.F. Svaiter, *Steepest descent methods for multicriteria optimization*, Mathematical Methods of Operations Research, **51**(3), pp. 479–494, 2000.
- [160] J. Fliege, L.M. Graña Drummond and B.F. Svaiter, *Newton’s method for multiobjective optimization*, SIAM Journal on Optimization, **20**(2), pp. 602–626, 2009.
- [161] C.M. Fonseca and P.J. Flemming, *An overview of evolutionary algorithms in multiobjective optimization*, Evol. Comput., **3**(1), 1995.
- [162] I. Fonseca and G. Leoni, *Modern methods in the calculus of variations:  $L^p$  spaces*, Springer Science & Business Media, 2007.
- [163] P. Frankel, G. Garrigos and J. Peypouquet, *Splitting methods with variable metric for Kurdyka-Lojasiewicz functions and general convergence rates*, Journal of Optimization Theory and Applications, **165**(3), pp. 874–900, 2015.
- [164] E.H. Fukuda and L.M. Graña Drummond, *On the convergence of the projected gradient method for vector optimization*, Optimization, **60**(8-9), pp. 1009–1021, 2011.
- [165] E.H. Fukuda and L.M. Graña Drummond, *Inexact projected gradient method for vector optimization*, Computational Optimization and Applications, **54**(3), pp. 473–493, 2013.
- [166] A.M. Gabriélov, *Projections of semi-analytic sets*, Functional Analysis and Its Applications, **2**(4), pp. 282–291, 1968.
- [167] E.M. Gafni and D.P. Bertsekas, *Two-metric projection methods for constrained optimization*, SIAM J. on Control and Optim., **22**(6), pp. 936–964, 1984.
- [168] A. Gambier and E. Badreddin, *Multi-objective optimal control: An overview*, IEEE International Conference on Control Applications, pp. 170–175, 2007.
- [169] A. Ganesh, Z. Lin, J. Wright, L. Wu, M. Chen and Y. Ma, *Fast algorithms for recovering a corrupted low-rank matrix*, in Proceedings of the 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), pp. 213–216, 2009.
- [170] M.K. Gavurin, *Nonlinear functional equations and continuous analogues of iteration methods*, Izv. Vyssh. Uchevn. Zaved. Mat., **5**(6), pp. 18–31, 1958.
- [171] A.M. Geoffrion, *Proper efficiency and the theory of vector maximization*, Journal of Mathematical Analysis and Applications, **22**(3), pp. 618–630, 1968.
- [172] D. Gilbarg and N.S. Trudinger, *Elliptic partial differential equations of second order*, Springer Berlin Heidelberg, 1983.
- [173] L.M. Graña Drummond and A.N. Iusem, *A projected gradient method for vector optimization problems*, Computational Optimization and Applications, **28**(1), pp. 5–29, 2004.

- [174] M. Grasselli and M. Pierre, *Convergence to equilibrium of solutions of the backward Euler scheme for asymptotically autonomous second-order gradient-like systems*, Communications on Pure and Applied Analysis, **11**(6), pp. 2393–2416, 2012.
- [175] L.M. Graña Drummond, F.M.P. Raupp and B.F. Svaiter, *A quadratically convergent Newton method for vector optimization*, Optimization, **63**(5), pp. 661–677, 2014.
- [176] L.M. Graña Drummond and B.F. Svaiter, *A steepest descent method for vector optimization*, Journal of Computational and Applied Mathematics, **175**(2), pp. 395–414, 2005.
- [177] A. Göpfert, H. Riahi, C. Tammer and C. Zalinescu, *Variational Methods in Partially Ordered Spaces*, Springer Science & Business Media, 2006.
- [178] A.A. Goldstein, *Cauchy’s method of minimization*, Numerische Mathematik, **4**(1), pp. 146–150, 1962.
- [179] P.A. Gordan, *Über die Auflösung linearer Gleichungen mit reellen Coefficienten*, Mathematische Annalen, **6**, pp. 23–28, 1873.
- [180] X. Goudou, J. Munier, *The gradient and heavy ball with friction dynamical systems: the quasiconvex case*, Math Programming, **116**(1-2), pp. 173–191, 2009.
- [181] A. Grothendieck, *Esquisse d’un programme*, London Mathematical Society Lecture Note Series, pp. 5–48, 1997.
- [182] M. Haddou and P. Maheux, *Smoothing methods for nonlinear complementarity problems*, Journal of Optimization Theory and Applications, **160**(3), pp. 711–729, 2013.
- [183] Y.Y. Haimes, L.S. Lasdon and D.A. Wismer, *On a bicriterion formulation of the problems of integrated system identification and system optimization*, IEEE Transactions on Systems, Man and Cybernetics, **3**, 296–297, 1971.
- [184] E. Hansen and T. Stillfjord, *Convergence of the implicit-explicit Euler scheme applied to perturbed dissipative evolution equations*, Mathematics of Computation, **82**(284), pp. 1975–1985, 2013.
- [185] A. Haraux, *A hyperbolic variant of Simon’s convergence theorem. Evolution equations and their applications in physical and life sciences*, Lecture Notes in Pure and Appl. Math., **215**, pp. 255–264, 2001.
- [186] A. Haraux and M.A. Jendoubi, *Convergence of solutions of second-order gradient-like systems with analytic nonlinearities*, Journal of Differential Equations, **144**(2), pp. 313–320, 1998.
- [187] A. Haraux and M.A. Jendoubi, *Convergence of bounded weak solutions of the wave equation with dissipation and analytic nonlinearity*, Calculus of Variations and Partial Differential Equations, **9**(2), pp. 95–124, 1999.
- [188] A. Haraux and M.A. Jendoubi, *The Lojasiewicz gradient inequality in the infinite dimensional Hilbert space framework*, J. of Func. Anal., **260**(9), pp. 2826–2842, 2011.
- [189] M. Hasenjäger, B. Sendhoff, T. Sonoda and T. Arima, *Three dimensional evolutionary aerodynamic design optimization using single and multi-objective approaches*, in: R. Schilling, W. Haase, J. Periaux, H. Baier, G. Bugeđa (Eds.), Evolutionary and Deterministic Methods for Design, Optimization and Control with Applications to Industrial and Societal Problems EUROGEN 2005, FLM, Munich, 2005.



- [190] Y. Haugazeau, *Sur les inéquations variationnelles et la minimisation de fonctionnelles convexes*, PhD. , Univ. de Paris, 1968.
- [191] C. Henry, *An existence theorem for a class of differential equations with multivalued right-hand side*, J. Math. Anal. Appl, **41**, pp. 168–179, 1973.
- [192] J.-B. Hiriart-Urruty, *Refinements of necessary optimality conditions in nondifferentiable programming I*, Applied Mathematics and Optimization, **5**(1), pp.63–82, 1979.
- [193] J.-B. Hiriart-Urruty, *Generalized Differentiability / Duality and Optimization for Problems Dealing with Differences of Convex Functions*, in Lecture Notes in Economics and Mathematical Systems **256**, Springer Berlin Heidelberg, 1985.
- [194] J.-B. Hiriart-Urruty and H.Y. Le, *From Eckart and Young approximation to Moreau envelopes and vice versa*, RAIRO - Operations Research, **47**(3), pp. 299–310, 2013.
- [195] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms I: Fundamentals*, Springer Science & Business Media, 1993.
- [196] H. Hironaka, *Introduction to real-analytic sets and real-analytic maps*, Lecture notes, Istituto matematico "L. Tonelli" dell Università di Pisa, 1973.
- [197] K.-H. Hoffmann and P. Rybka, *Convergence of solutions to Cahn-Hilliard equation*, Comm. Partial Differential Equations, **24**, pp. 1055–1077, 1999.
- [198] M. Hong, Z.-Q. Luo and M. Razaviyayn, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, preprint available on arXiv:1410.1390, submitted on October 2014.
- [199] S.-Z. Huang, and P. Takáč, *Convergence in gradient-like systems which are asymptotically autonomous and analytic*, Nonlinear Analysis, Ser. A Theory Methods **46**, pp. 675–698, 2001.
- [200] W. Hundsdorfer and J. Verwer, *Numerical solution of time-dependent advection-diffusion-reaction equations*, Springer Berlin Heidelberg, 2003.
- [201] A.D. Ioffe, *Proximal analysis and approximate subdifferentials*, **s2-41**(1), pp. 175–192, 1990.
- [202] A.D. Ioffe, *Metric regularity and subdifferential calculus*, Russian Mathematical Surveys, **55**(3), pp. 501–558, 2000.
- [203] A.D. Ioffe, *A Sard theorem for tame set-valued mappings*, Journal of Mathematical Analysis and Applications, **335**(2), pp. 882–901, 2007.
- [204] A.D. Ioffe, *An Invitation to Tame Optimization*, SIAM Journal on Optimization, **19**(4), pp. 1894–1917, 2009.
- [205] A.D. Ioffe and A.S. Lewis, *Critical points of simple functions*, Optimization, **57**(1), pp. 3–16, 2008.
- [206] J. Jahn, *Vector optimization: Theory, applications, and extensions*, Springer Science & Business Media, 2004.
- [207] G. Jameson, *Ordered linear spaces*, Springer-Verlag, 1970.
- [208] R. Janssen, M. van Herwijnen, T.J. Stewart and J.C.J.H. Aerts, *Multiobjective decision support for land-use planning*, Environ. Plann. B, **35**, pp. 740–756, 2008.

- [209] O. Kavian, *Introduction à la théorie des points critiques - et applications aux problèmes elliptiques*, Springer-Verlag Berlin Heidelberg, 1993.
- [210] K.L. Kiran, D. Jayachandran and S. Lakshminarayanan, *Multi-objective optimization of cancer immuno-chemotherapy*, in: R. Magjarevic, C.T. Lim, J.C.H. Goh(Eds.), 13th International Conference on Biomedical Engineering, Springer, Berlin, Heidelberg, **23**, pp. 1337–1340, 2009.
- [211] G. Kreisselmeier and R. Steinhauser, *Application of vector performance optimization to a robust control loop design for a fighter aircraft*, International Journal of Control, **37**(2), pp. 251–284, 1983.
- [212] A.Y. Kruger, *On Fréchet subdifferentials*, Journal of Mathematical Sciences, **116**(3), pp. 3325–3358, 2003.
- [213] A.Y. Kruger, *Stationarity and regularity of set systems*, Pacific Journal of Optimization, **1**(1), pp. 101–126, 2005.
- [214] A.Y. Kruger, *About regularity of collections of sets*, Set-Valued Analysis, **14**(2), pp. 187–206, 2006.
- [215] G. Köthe, *Topological vector spaces I*, Springer Berlin Heidelberg, 1969.
- [216] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Ann. Inst. Fourier, **48**, pp. 769–783, 1998.
- [217] K. Kurdyka and A. Parusiński,  *$w_f$ -stratification of subanalytic functions and the Lojasiewicz inequality*, C. R. Acad. Paris, **318**, pp. 129–133, 1994.
- [218] C. Lageman, *Pointwise convergence of gradient-like systems*, Mathematische Nachrichten, **280**(13-14), pp. 1543–1558, 2007.
- [219] E.S. Levitin, B.T. Polyak, *Constrained minimization methods*, U.S.S.R. Comput. Math. and Math. Phys. **6**(5), pp. 1–50, 1966.
- [220] A.B. Levy, R. Poliquin and L. Thibault, *Partial extensions of Attouch’s theorem with applications to proto-derivatives of subgradient mappings*, Transactions of the American Mathematical Society, **347**(4), pp. 1269–1294, 1995.
- [221] A.S. Lewis, *Active sets, nonsmoothness, and sensitivity*, SIAM Journal on Optimization, **13**(3), pp. 702–725, 2002.
- [222] A.S. Lewis, D.R. Luke and J. Malick, *Local linear convergence for alternating and averaged nonconvex projections*, Foundations of Computational Mathematics, **9**(4), pp. 485–513, 2009.
- [223] A.S. Lewis and M. Overton, *Nonsmooth optimization via quasi-Newton methods*, Mathematical Programming, **141**(1-2), pp. 135–163, 2013.
- [224] D. Li, L.-P. Pang and S. Chen, *A proximal alternating linearization method for nonconvex optimization problems*, Optimization Methods & Software, **29**(4), 2014.
- [225] G. Li and T.K. Pong, *Global convergence of splitting methods for nonconvex composite optimization*, preprint available on arXiv:1407.0753, submitted on July 2014.
- [226] J. Liang, J. Fadili and G. Peyré, *Local linear convergence of Forward–Backward under partial smoothness*, in Advances in Neural Information Processing Systems, pp. 1970–1978, 2014.

- [227] A. Lipej, *Optimization method for the design of axial hydraulic turbines*, J. Power. Energ. A, **218**, pp. 43–50, 2004.
- [228] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, **16**, pp. 964–979, 1979.
- [229] G.P. Liu, J.-B. Yang and J.F. Whidborn, *Multiobjective optimization and control*, Research Studies Press, Exeter, 2003.
- [230] S. Lojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels*, in: Les Équations aux Dérivées Partielles, pp. 87–89, Éditions du centre National de la Recherche Scientifique, Paris, 1965.
- [231] S. Lojasiewicz, *Sur les trajectoires du gradient d'une fonction analytique*, in Seminari di Geometria 1982-1983, Istituto di Geometria, Dipartimento di Matematica, Università di Bologna, Bologna, Italy, pp. 115–117, 1984.
- [232] S. Lojasiewicz, *Sur la géométrie semi- et sous-analytique*, Ann. Inst. Fourier, **43**, pp. 1575–1595, 1993.
- [233] D.A. Lorenz and T. Pock, *An inertial forward-backward algorithm for monotone inclusions*, J. Math. Imaging Vision, **51**(2), pp. 311–325, 2015.
- [234] C. Malivert, *A descent method for Pareto optimization*, Journal of Mathematical Analysis and Applications, **88**(2), pp. 610–631, 1982.
- [235] S. Marcellin and L. Thibault, *Evolution problems associated with primal lower nice functions*, Journal of Convex Analysis, **13**(2), pp. 385–422, 2006.
- [236] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle, **4** Sér. R-3, pp. 154–158, 1970.
- [237] S. Mazurkiewicz, *Sur les fonctions non dérivables*, Studia Mathematica, **3**(1), pp. 92–94, 1931.
- [238] M. Maschler and B. Peleg, *Stable sets and stable points of set-valued dynamic systems with applications to game theory*, SIAM Journal on Control and Optimization, **14**(6), pp. 985–995, 1976.
- [239] K. Meng, V. Roshchina and X. Yang, *On local coincidence of a convex set and its tangent cone*, Journal of Optimization Theory and Applications, **164**(1), pp. 123–137, 2014.
- [240] B. Merlet, M. Pierre, *Convergence to equilibrium for the backward Euler scheme and applications*, Commun. Pure Appl. Anal, **9**, pp. 685–702, 2010.
- [241] A. Messac, G. Sundararaj, R. Tappeta and J.E. Renaud, *Ability of objective functions to generate points on nonconvex Pareto frontiers*, AIAA Journal, **38**(6), pp. 1084–1091, 2000.
- [242] R. Mifflin, *Semismooth and Semiconvex Functions in Constrained Optimization*, SIAM Journal on Control and Optimization, **15**(6), pp. 959–972, 1977.
- [243] C. Miller, *Expansions of the real field with power functions*, Annals of Pure and Applied Logic, **68**(1), pp. 79–94, 1994.
- [244] B. Mordukhovich, *Variational analysis and generalized differentiation I: Basic theory*, Springer Science & Business Media, 2006.

- [245] B. Mordukhovich, *Variational analysis and generalized differentiation II: Applications*, Springer Science & Business Media, 2006.
- [246] J. J. Moreau, *Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires*, C. R. Acad. Sci., **255**, pp. 238–240, 1962.
- [247] E. Miglierina, *Slow solutions of a differential inclusion and vector optimization*, Set-Valued Analysis, **12**(3), pp. 345–356, 2004
- [248] E. Miglierina, E. Molho and M. C. Recchionni, *Box-constrained multi-objective optimization, a gradient-like method without a priori scalarization*, European Journal of Operational Research, **188**(3), pp. 662–682, 2008.
- [249] B. Natarajan, *Sparse approximate solutions to linear systems*, SIAM Journal on Computing, **24**(2), pp. 227–234, 1995.
- [250] A.S. Nemirovsky and D.B. Yudin, *Problem complexity and method efficiency in optimization*, Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, New York, 1983.
- [251] Y. Nesterov, *A method of solving a convex programming problem with convergence rate  $O(1/k^2)$* , Soviet Mathematics Doklady, **27**, pp. 372–376, 1983.
- [252] J. von Neumann, *Functional operators: The geometry of orthogonal spaces*, Annals of Mathematics studies 22, Princeton University Press, 1950.
- [253] D. Noll, *Convergence of non-smooth descent methods using the Kurdyka-Lojasiewicz inequality*, J Optim Theory Appl., **160**, pp. 553–572, 2014.
- [254] P. Ochs, Y. Chen, T. Brox and T. Pock, *iPiano: inertial proximal algorithm for nonconvex optimization*, SIAM Journal on Imaging Sciences, **7**(2), pp. 1388–1419, 2014.
- [255] J.M. Ortega and W.C. Rheinboldt, *Iterative solutions of nonlinear equations in several variables*, Academic Press, New York, 1970.
- [256] A. Ostermann, *Stability of W-methods with applications to operator splitting and to geometric theory*, Applied Numerical Mathematics, **42**(1-3), pp. 353–366, 2002.
- [257] J. Palis Jr and W. de Melo, *Geometric theory of dynamical systems: An introduction*, Springer-Verlag New York, 1982.
- [258] V. Pareto, *Manuale di economia politica*, Società Editrice Libreria, Milano, 1906. Translated into French by Girard and Brière, as *Manuel d'économie politique*, Giard, Paris, France, First Edition, 1909 and Second Edition, 1927. Translated into English by A. S. Schwier, as *Manual of Political Economy*, The MacMillan Company, New York, 1971.
- [259] A. Pascoletti and P. Serafini, *An iterative procedure for vector optimization*, Journal of Mathematical Analysis and Applications, **89**(1), pp. 95–106, 1982.
- [260] G. Passty, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, J. Math. Anal. Appl., **72**(2), pp. 383–390, 1979.
- [261] B. Peleg and M.E. Yaari, *Efficiency prices in an infinite-dimensional space*, Journal of Economic Theory, **2**(1), pp. 41–85, 1970.
- [262] J.-P. Penot, *Favorable classes of mappings and multimappings in nonlinear analysis and optimization*, Journal of Convex Analysis, **3**(1), pp. 97–116, 1996.

- [263] J.-P. Penot, *Calculus without derivatives*, Springer Science & Business Media, 2012.
- [264] F. Petterson, N. Chakraborti and S.B. Singh, *Neural networks analysis of steel plate processing augmented by multi-objective genetic algorithms*, *Steel. Res. Int.*, **78**, pp. 890–898, 2007.
- [265] J. Peypouquet, *Analyse asymptotique de systèmes d'évolution et applications en optimisation*, PhD., Université Paris 6 & Universidad de Chile, 2007.
- [266] J. Peypouquet, *Asymptotic convergence to the optimal value of diagonal proximal iterations in convex minimization*, *J. Convex Anal.*, **16**(1), pp. 277–286, 2009.
- [267] J. Peypouquet, *Coupling the gradient method with a general exterior penalization scheme for convex minimization*, *Journal of Optimization Theory and Applications*, **153**(1), pp. 123–138, 2011.
- [268] J. Peypouquet, *Convex optimization in normed spaces. Theory, methods and examples.*, Springer Science & Business media, 2015.
- [269] J. Peypouquet and S. Sorin, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, *Journal of Convex Analysis*, **17**(3-4), pp. 1113–1163, 2010.
- [270] J. Philip, *Algorithms for the vector maximization problem*, *Mathematical Programming*, **2**(1), pp. 207–229, 1972.
- [271] A. Pillay and C. Steinhorn, *Definable sets in ordered structures. I*, *Transactions of the American Mathematical Society*, **295**(2), pp. 565–592, 1986.
- [272] A. Pinto da Costa, J.A.C. Martins, I.N. Figueiredo and J.J. Judice, *The directional instability problem in systems with frictional contacts*, *Computer Methods in Applied Mechanics and Engineering*, **193**(3-5), pp. 357–384, 2004.
- [273] R.A. Poliquin, *Integration of subdifferentials of nonconvex functions*, *Nonlinear Analysis: Theory, Methods & Applications*, **17**(4), pp. 385–398, 1991.
- [274] R.A. Poliquin and R.T. Rockafellar, *Prox-regular functions in variational analysis*, *Transactions of the American Mathematical Society*, **348**(5), pp. 1805–1838, 1996.
- [275] R. Poliquin, R.T. Rockafellar and L. Thibault, *Local differentiability of distance functions*, *Transactions of the American Mathematical Society*, **352**(11), pp. 5231–5249, 2000.
- [276] B.T. Polyack, *Some methods of speeding up the convergence of iterative methods*, *Z. Vychisl. Math. Fiz.*, **4**, pp. 1–17, 1964.
- [277] Z. Povalej, *Quasi-Newton's method for multiobjective optimization*, *Journal of Computational and Applied Mathematics*, **255**, pp. 767–777, 2014.
- [278] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, *Mathematical Programming*, **58**(3), pp. 353–367, 1993;
- [279] S. Qu, M. Goh and F.T.S. Chan, *Quasi-Newton methods for solving multiobjective optimization*, *Operations Research Letters*, **39**(5), pp. 397–399, 2011.
- [280] E.A.P. Quiroz and P.R. Oliveira, *Proximal point methods for functions involving Lojasiewicz, quasiconvex and convex properties on Hadamard manifolds*, submitted at [http://www.optimization-online.org/DB\\_HTML/2008/06/1996.html](http://www.optimization-online.org/DB_HTML/2008/06/1996.html) on March 2013.

- [281] R. Radner, *Efficiency prices for infinite horizon production programmes*, The Review of Economic Studies, **34**(1), pp. 51–66, 1967.
- [282] M. Recchioni, *A path following method for box-constrained multiobjective optimization with applications to goal programming problems*, **58**(1), pp. 69–85, 2003.
- [283] B. Recht, B. Fazel and P.A. Parillo, *Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization*, SIAM Rev., **52**, pp. 471–501, 2007.
- [284] B. Mohammadi and P. Redont, *Improving the identification of general Pareto fronts by global optimization*, C. R. Acad. Sci. Paris, Ser. I, **347**, pp. 327–331, 2009.
- [285] S. Reich, *Nonlinear evolution equations and nonlinear ergodic theorems*, Nonlinear Analysis, **1**, pp. 319–330, 1976.
- [286] R. T. Rockafellar, *Integrals which are convex functionals*, Pacific Journal of Mathematics, **23**(3) pp. 525–539, 1968.
- [287] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*. SIAM J. Control Optimization, **14**(5), pp. 877–898, 1976.
- [288] R.T. Rockafellar, *Clarke’s tangent cones and the boundaries of closed sets in  $\mathbb{R}^n$* , Nonlinear Analysis: Theory, Methods and Applications, **3**(1), pp. 145–154, 1979.
- [289] R.T. Rockafellar, *Favorable classes of Lipschitz continuous functions in subgradient optimization*, in: E. Nurminski (Ed.), Progress in Non-Differentiable Optimization, Pergamon Press, 1981.
- [290] R.T. Rockafellar and R.J.-B. Wets, *Variational analysis*, Springer, New York, 2009.
- [291] P.C. Rosenbloom, *The method of steepest descent*, Proc Symp Appl Math **6**, pp. 127–176, 1956.
- [292] S. Salzo and S. Villa, *Inexact and accelerated proximal point algorithms*, Journal of Convex Analysis, **19**(4), pp. 1167–1192, 2012.
- [293] P. A. Samuelson, *Foundations of economic analysis*, Athenum, New-York, 1970.
- [294] A. Seeger, *Eigenvalue analysis of equilibrium processes defined by linear complementarity conditions*, Linear Algebra and its Applications, **292**(1-3), pp. 1–14, 1999.
- [295] S. Schäffler, R. Schultz and K. Weinzierl, *Stochastic method for the solution of unconstrained vector optimization problems*, Journal of Optimization Theory and Applications, **114**(1), pp. 209–222, 2002.
- [296] L. Simon, *Asymptotics for a class of nonlinear evolution equations, with applications to geometric problems*, Ann. Math., **118**, pp. 525–571, 1983.
- [297] S. Smale, *Global analysis and economics I: Pareto optimum and a generalization of Morse theory*, Synthese, **31**(2), pp. 345–358, 1975.
- [298] P. Speissegger, *The Pfaffian closure of an  $o$ -minimal structure*, Journal für die reine und angewandte Mathematik (Crelles Journal), **508**, pp. 189–211, 1999.
- [299] S. Sra, S. Nowozin and S.J. Wright, *Optimization for machine learning*, MIT Press, 2011.
- [300] W. Stadler, *A survey of multicriteria optimization or the vector maximum problem, part I: 1776-1960*, Journal of Optimization Theory and Applications, **29**(1), pp. 1–52, 1979.

- [301] C. Stegall, *The duality between Asplund spaces and spaces with the Radon-Nikodym property*, Israël Journal of Mathematics, **29(4)** pp. 408–412, 1978.
- [302] R.E. Steuer, Y. Qi and M. Hirschberger, *Portfolio selection in the presence of multiple criteria*, in: C. Zopounidis, M. Doumpos, P.M. Pardalos (Eds.), Handbook of Financial Engineering, in: Springer Optimization and Its Applications, **18**, pp. 3–24, 2008.
- [303] W. Su, S. Boyd and E. J. Candès, *A differential equation for modeling Nesterov’s accelerated gradient method: Theory and Insights*, preprint available on arXiv:1503.01243, submitted on Mar 2015.
- [304] R. Temam, *Sur la stabilité et la convergence de la méthode des pas fractionnaires*, Annali di Matematica Pura ed Applicata, **79(1)**, pp. 191–379, 1968.
- [305] L. Thibault, *On generalized differentials and subdifferentials of Lipschitz vector-valued functions*, Nonlinear Analysis: Theory, Methods & Applications, **6(10)**, pp. 1037–1053, 1982.
- [306] L. Thibault and D. Zagrodny, *Integration of subdifferentials of lower semicontinuous functions on Banach spaces*, Journal of Mathematical Analysis and Applications, **189(1)**, pp. 33–58, 1995.
- [307] J.S. Treiman, *Clarke’s gradients and epsilon-subgradients in Banach spaces*, Transactions of the AMS, **294(1)**, pp. 75–78, 1986.
- [308] A. I. Tulcea and C. I. Tulcea, *Topics in the theory of lifting*, Springer Berlin Heidelberg, 1969.
- [309] H. Uzawa, *A Theorem on convex polyhedral cones*, in Studies in Linear and Non-Linear Programming: Stanford Mathematical Studies In The Social Sciences, No. 2, 1958.
- [310] M. Valadier, *Sous-différentiabilité de fonctions convexes à valeurs dans un espace vectoriel ordonné*, Mathematica Scandinavica, **30**, pp. 65–74, 1972.
- [311] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM Review, **38(1)**, pp. 49–95, 1996.
- [312] K. D. V. Villacorta and P. R. Oliveira, *An interior proximal method in vector optimization*, European Journal of Operational Research, **214(3)**, pp. 485–492, 2011.
- [313] K. D. V. Villacorta, P. R. Oliveira and A. Soubeyran, *A trust-region method for unconstrained multiobjective problems with applications in satisficing processes*, Journal of Optimization Theory and Applications, **160(3)**, 865–889, 2014.
- [314] T. Walras, *Éléments D’économie politique pure*, L. Corbaz and Companie, Lausanne, 1874–1877. Translation of the Definitive Edition into English by W. Jaffé as *Elements of Pure Economics*, George Allen and Unwin, London, 1954.
- [315] Y.-H. Wan, *On local Pareto optima*, Journal of Mathematical Economics, **2(1)**, pp. 35–42, 1975.
- [316] F. Wang, Z. Xu and H.-K. Xu, *Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems*, preprint available on arXiv:1410.8625, submitted on October 2014.
- [317] H. Weyl, *Elementare Theorie der konvexen Polyeder*, Comm. Helvet., **7**, pp. 290–306, 1935.

- [318] H. Whitney, *Elementary structure of real algebraic varieties*, Annals of Mathematics, **66**(3), pp. 545–556, 1957.
- [319] A. Wilkie, *Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function*, Journal of the American Mathematical Society, **9**(4), pp. 1051–1094, 1996.
- [320] D.J. White, *Epsilon-dominating solutions in mean-variance portfolio analysis*, Eur. J. Oper. Res., **105**, pp. 457–466, 1998.
- [321] Y. Xu and W. Yin, *A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion*, SIAM J. Imaging Sciences, **6**(3), pp. 1758–1789, 2013.
- [322] X. Yuan and J. Yang, *Sparse and low-rank matrix decomposition via alternating direction method*, Pacific Journal of Optimization, **9**(1), pp. 167–180, 2013.
- [323] L. Zadeh, *Optimality and non-scalar-valued performance criteria*, IEEE Transactions on Automatic Control, **8**(1), pp. 59–60, 1963.
- [324] W.I. Zangwill, *Convergence conditions for nonlinear programming algorithms*, Management Science, **16**(1), pp. 1–13, 1969.
- [325] C. Zalinescu, *Convex analysis in general vector spaces*, World Scientific, 2002.
- [326] E. Zeidler, *Nonlinear functional analysis and its applications, Part II: Monotone operators*, Springer-Verlag, New-York, 1990.
- [327] J. Zowe, *Subdifferentiability of convex functions with values in an ordered vector space*, Mathematica Scandinavica, **34**, pp. 69–83, 1974.





## Résumé.

Dans une première partie, nous nous intéressons aux systèmes dynamiques gradients gouvernés par des fonctions non lisses, mais aussi non convexes, satisfaisant l'inégalité de Kurdyka-Lojasiewicz. Après avoir obtenu quelques résultats préliminaires pour la dynamique de la plus grande pente continue, nous étudions un algorithme de descente général. Nous prouvons, sous une hypothèse de compacité, que tout suite générée par ce schéma général converge vers un point critique de la fonction à minimiser. Nous obtenons aussi de nouveaux résultats sur la vitesse de convergence, tant pour les valeurs que pour les itérés. Ce schéma général couvre en particulier des versions parallélisées de la méthode *forward-backward*, autorisant une métrique variable et des erreurs relatives. Cela nous permet par exemple de proposer une version non convexe non lisse de l'algorithme Levenberg-Marquardt. Enfin, nous proposons quelques applications de ces algorithmes aux problèmes de faisabilité, et aux problèmes inverses parcimonieux.

Dans une seconde partie, cette thèse développe une dynamique de descente associée à des problèmes d'optimisation vectoriels sous contrainte. Pour cela, nous adaptons la dynamique de la plus grande pente usuelle aux fonctions à valeurs dans un espace ordonné par un cône convexe fermé d'intérieur non vide. Cette dynamique peut être vue comme l'analogie continu de nombreux algorithmes développés ces dernières années. Nous avons un intérêt particulier pour les problèmes de décision multi-objectifs, pour lesquels cette dynamique de descente fait décroître toutes les fonctions objectif au cours du temps. Nous prouvons l'existence de trajectoires pour cette dynamique continue, ainsi que leur convergence vers des points faiblement efficaces. Finalement, nous explorons une nouvelle dynamique inertielle pour les problèmes multi-objectif, avec l'ambition de développer des méthodes rapides convergeant vers des équilibres de Pareto.

## Abstract.

In a first part, we focus on gradient dynamical systems governed by non-smooth but also non-convex functions, satisfying the so-called Kurdyka-Lojasiewicz inequality. After obtaining preliminary results for a continuous steepest descent dynamic, we study a general descent algorithm. We prove, under a compactness assumption, that any sequence generated by this general scheme converges to a critical point of the function to be minimized. We also obtain new convergence rates both for the values and the iterates. The analysis covers alternating versions of the forward-backward method, with variable metric and relative errors. As an example, a non-smooth and non-convex version of the Levenberg-Marquardt algorithm is detailed. Applications to non-convex feasibility problems, and to sparse inverse problems are discussed.

In a second part, the thesis explores descent dynamics associated to constrained vector optimization problems. For this, we adapt the classic steepest descent dynamic to functions with values in a vector space ordered by a closed convex cone with nonempty interior. It can be seen as the continuous analogue of various descent algorithms developed in the last years. We have a particular interest for multi-objective decision problems, for which the dynamic make decrease all the objective functions along time. We prove the existence of trajectories for this continuous dynamic, and show their convergence to weak efficient points. Then, we explore an inertial dynamic for multi-objective problems, with the aim to provide fast methods converging to Pareto points.