THESIS / THÈSE

DOCTOR OF SCIENCES

Towards interior proximal point methods for solving equilibrium problems

Nguyen, Thi Thu Van

Award date:
2008

Awarding institution:
University of Namur

Facultés Universitaires Notre-Dame de la Paix Namur
Faculté des Sciences
Département de Mathematique

Towards Interior Proximal Point Methods for Solving Equilibrium Problems

Dissertation présentée par
Nguyen Thi Thu Van
pour l'obtention du grade
de Docteur en Sciences

Composition du Jury:

Jean-Jacques STRODIOT (Promoteur)
Van Hien NGUYEN (Co-promoteur)
LE Dung Muu
Michel WILLEM
Joseph WINKIN

September 2008

Rempart de la Vierge, 13
B-5000 Namur (Belgique)

Imprimé en Belgique

ISBN-13 : 978-2-87037-614-0
Dépôt légal: D / 2008 / 1881 / 42

Acknowledgements

I am indebted to my PhD supervisor, Professor Jean-Jacques STRODIOT, for his guidance and assistance given during the preparation of this thesis. It is from Prof. STRODIOT that I have not only systematically learned functional analysis, convex analysis, optimization theory and numerical algorithms but also how to conduct research and to write up my findings coherently for publication. He has even demonstrated how to be a good teacher via teaching me how to write lesson plans and how to present scientific seminars. A debt I will not be able to repay but one I am most grateful for. The only thing I can do is to try my best to practice these skills and to pass on my new found knowledge to future students.

Secondly, I would like to express my deep gratitude to Professor Van Hien NGUYEN, my co-supervisor, for his guidance, continuing help and encouragement. I would probably not have had such a fortunate chance to study in Namur without his help. I really appreciate his useful advice on my thesis and especially thank him for the amount of time he spent reading my papers and providing valuable suggestions. It is also from Prof. Hien that I have learned to work in the spirit to willingly share time with others and to be helpful at heart.

I would like to thank my committee members, Professors LE Dung Muu, Michel WILLEM, and Joseph WINKIN for really practical and constructive comments.

I would also like to thank CIUF (Conseil Interuniversitaire de la Communauté Française) and CUD (Commission Universitaire pour le Développement) for financial support given during two training placements, 3 months in 2001 and 6 months in 2003, at the University of Namur. I further like to address my thanks to the University of Namur for the financial support received for my PhD research, from 2004 until 2008. I also want to thank the Department of Mathematics, especially the Unit of Optimization and Control for the generous help they have provided me. On this occasion, I want to thank my friends in the Department of Mathematics for their warm support and for their help during my stay in Namur, namely Jehan BOREUX, Delphine LAMBERT, Anne-Sophie LIBERT, Benoît NOYELLES, Simone RIGHI, Caroline SAINVITU, Geneviève SALMON, Stéphane VALK, Emilie WANUFELLE, Melissa WEBER MENDONÇA, and Sebastian XHONNEUX.

Last but not least, special thanks are also given to Professor NGUYEN Thanh Long of the University of Natural Sciences - Vietnam National University, Ho Chi Minh City for everything he has done for me. He has not only helped me to do research but also offered me many training courses which allowed me to earn my living. He always listens patiently to me and gives me valuable advice. His attitude in doing research motivates me to work harder.

Xin bày tỏ lòng biết ơn đến các Thầy Cô giáo tại khoa Toán - Tin học, trường Đại học Khoa học Tự Nhiên - Đại học Quốc Gia Thành phố Hồ Chí Minh và các Giáo sư tại Viện Toán học Hà Nội đã quan tâm và giúp đỡ tác giả trong thời gian qua.

Xin chân thành cảm ơn các chị, anh, em đang sinh sống, làm việc và học tập tại Bỉ và các bạn bè đồng nghiệp xa gần đã luôn bên cạnh động viên và giúp đỡ tác giả trong suốt quá trình học tập và nghiên cứu tại Bỉ.

Luận án này là món quà tinh thần tác giả xin kính tặng đến Gia đình của mình với tất cả lòng biết ơn, yêu thương và trân trọng.

Nguyễn Thị Thu Vân

Abstract: This work is devoted to study efficient numerical methods for solving nonsmooth convex equilibrium problems in the sense of Blum and Oettli. First we consider the auxiliary problem principle which is a generalization to equilibrium problems of the classical proximal point method for solving convex minimization problems. This method is based on a fixed point property. To make the algorithm implementable we introduce the concept of μ-approximation and we prove that the convergence of the algorithm is preserved when in the subproblems the nonsmooth convex functions are replaced by μ-approximations. Then we explain how to construct μ-approximations using the bundle concept and we report some numerical results to show the efficiency of the algorithm. In a second part, we suggest to use a barrier function method for solving the subproblems of the previous method. We obtain an interior proximal point algorithm that we apply first for solving nonsmooth convex minimization problems and then for solving equilibrium problems. In particular, two interior extragradient algorithms are studied and compared on some test problems.

Résumé: Ce travail est consacré à l'étude de méthodes numériques efficaces pour résoudre des problèmes d'équilibre convexes non différentiables au sens de Blum et Oettli. D'abord nous considérons le principe du problème auxiliaire qui est une généralisation aux problèmes d'équilibre de la méthode du point proximal pour résoudre des problèmes de minimisation convexes. Cette méthode est basée sur une propriété de points fixes. Pour rendre l'algorithme implémentable nous introduisons le concept de μ-approximation and nous montrons que la convergence de l'algorithme est préservée lorsque dans les sous problèmes la fonction convexe non différentiable est remplacée par une μ-approximation. Nous expliquons ensuite comment construire cette approximation en utilisant le concept de faisceaux et nous présentons des résultats numériques pour montrer l'efficacité de l'algorithme. Dans une seconde partie nous suggérons d'utiliser une méthode de type barrière pour résoudre les sous problèmes de la méthode précédente. Nous obtenons un algorithme de point proximal intérieur que nous appliquons à la résolution des problèmes de minimisation convexes non différentiables et ensuite à celle des problèmes d'équilibre. En particulier nous étudions deux algorithmes de type extragradient intérieurs que nous comparons sur des problèmes tests.

Contents

Chapter 1

Introduction

Equilibrium can be defined as a state of balance between opposing forces or influences. This concept is usually used in many scientific branches as physics, chemistry, economics and engineering. For example, in physics, the equilibrium state for a system, in terms of classical mechanics, means that the impact of all the forces on this system equals zero and that this state can be maintained for an indefinitely long period. In chemistry, it is a state where a forward chemical reaction and its reverse reaction proceed at equal rates.

In economics, the concept of an equilibrium is fundamental. A simple example is given by a market where consumers and producers buy and sell, respectively, a homogeneous commodity, their reaction depending on the current commodity price. More precisely, given a price p, the consumers determine their total demand $D(p)$ and the producers determine their total supply $S(p)$, so that the excess demand of the market is $E(p) = D(p) - S(p)$. If we consider a certain amount of transactions between consumers and producers then there exists the equality between the partial supply and demand at each price level, but the problem is to find the price which implies the equality between the total supply and demand, i.e., when $E(p^*) = 0$. This is called an equilibrium price model and corresponds to the classical static equilibrium concept, where the impact of all the forces equals zero, i.e., it is the same as in mechanics. Moreover, this price implies constant clearing of the market and may be maintained for an indefinitely long period. For a detailed study of Equilibrium Models, the reader is referred to the book by Konnov [49].

The equilibrium problem theory has been receiving growing interest by researchers, especially in economics. Many Nobel Prize winners, such as K.J. Arrow (1972), W.W. Leontief

(1973), L. Kantorovich and T. Koopmans (1975), G. Debreu (1983), H. Markovitz (1990), and J.F. Nash (1994), were awarded for their contributions in this field.

Recently the main concepts of optimization problems have also been extended to the field of equilibrium problems. This was motivated by the fact that optimization problems are not an adequate mathematical tool for modeling in situations of decision involving multiple agents as explained by A.S. Antipin in [4]: "*Optimization problems can be more or less adequate in situations where there is one person making decisions working with an alternative set, but in situations with many agents, each having their personal set and system of preferences on it and each working within the localized constraints of their specific situation, it becomes impossible to use the optimization model to produce an aggregate solution that will satisfy the global constraints that exist for the agents as a whole.*"

There exists a large number of different concepts of equilibrium models. These models are investigated and applied separately. They require to construct adequate tools both for the theory and for the solution methods. But, in the scope of a mathematical research, it is expected to present a general form which can unify some particular cases. Such an approach needs certain extensions of the usual concept of equilibrium and a presentation of unifying tools for investigating and solving these equilibrium models and meanwhile to drop some details in particular models. For that purpose, in this thesis we intend to consider the following class of equilibrium problem.

Let C be a nonempty closed convex subset of $\mathbb{R}^n$ and let $f : C \times C \to \mathbb{R}$ be an equilibrium bifunction, i.e., $f(x, x) = 0$ for all $x \in C$. The equilibrium problem (EP, for short) is to find a point $x^* \in C$ such that

$$f(x^*, y) \geq 0 \text{ for all } \ y \in C. \tag{EP}$$

This formulation was first considered by Nikaido and Isoda [70] as a generalization of the Nash equilibrium problem in non-cooperative many-person games. Subsequently, many authors have investigated this equilibrium model [4], [19], [20], [34], [40], [41], [42], [44], [46], [47], [48], [49], [62], [64], [66], [67], [72], [84], [85].

As mentioned by Blum and Oettli [20], this problem has numerous applications. Amongst them, it includes, as particular cases, the optimization problem, the variational inequality problem, the Nash equilibrium problem in noncooperative games, the fixed point problem, the nonlinear complementarity problem and the vector optimization problem. For the sake of clarity, let us introduce some more details on each of these problems. Note that in these examples we assume that $f(x, \cdot) : C \to \mathbb{R}$ is convex and lower semicontinuous for all $x \in C$ and that $f(\cdot, y) : C \to \mathbb{R}$ is upper semicontinuous for all $y \in C$.

Example 1.1. *(Convex minimization problem) Let $F : \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous convex function. Let C be a closed convex subset of $\mathbb{R}^n$. The convex minimization problem (CMP, for short) is to find $x^* \in C$ such that*

$$F(x^*) \leq F(y) \ \textit{for all}\ y \in C.$$

If we take $f(x, y) = F(y) - F(x)$ for all $x, y \in C$, then x^* is a solution to problem CMP if and only if x^* is a solution to problem EP.

Example 1.2. *(Nonlinear complementarity problem) Let $C \subset \mathbb{R}^n$ be a closed convex cone and let $C^+ = \{x \in \mathbb{R}^n \,|\, \langle x, y\rangle \geq 0 \ \textit{for all}\ y \in C\}$ be its polar cone. Let $T : C \to \mathbb{R}^n$ be a continuous mapping. The nonlinear complementarity problem (NCP, for short) is to find $x^* \in C$ such that*

$$T(x^*) \in C^+ \quad \textit{and} \quad \langle T(x^*), x^*\rangle = 0.$$

If we take $f(x, y) = \langle T(x), y - x\rangle$ for all $x, y \in C$, then x^* is a solution to problem NCP if and only if x^* is a solution to problem EP.

Example 1.3. *(Nash equilibrium problem in Noncooperative Games) Let*

- *I be a finite index set $\{1, \cdots, p\}$ (the set of p players),*
- *C_i be a nonempty closed convex set of $\mathbb{R}^n$ (the strategy set of the ith player) for each $i \in I$,*
- *$f_i : C_1 \times \cdots \times C_p \to \mathbb{R}$ be a continuous function (the loss function of the ith player, depending on the strategies of all players) for each $i \in I$.*

For $x = (x_1, \ldots, x_p)$, $y = (y_1, \ldots, y_p) \in C_1 \times \cdots \times C_p$, and $i \in I$, we define $x[y_i] \in C_1 \times \cdots \times C_p$ as

$$x[y_i] = \begin{cases} (x[y_i])_j = x_j \ \textit{for all components}\ j \neq i \\ (x[y_i])_i = y_i \ \textit{for the}\ i\textit{th component}. \end{cases}$$

If we take $C = C_1 \times \cdots \times C_p$, then C is a nonempty closed convex subset of $\mathbb{R}^n$. The Nash equilibrium problem (in Noncooperative Games) is to find $x^* \in C$ such that

$$f_i(x^*) \leq f_i(x^*[y_i]) \quad \textit{for all}\ \ i \in I \ \textit{and all}\ y \in C.$$

If we take $f : C \times C \to I\!R$ *defined as* $f(x, y) := \sum_{i=1}^{p} \{f_i(x[y_i]) - f_i(x)\}$ *for all* $x, y \in C$*, then* x^* *is a solution to the Nash equilibrium problem if and only if* x^* *is a solution to problem EP.*

Example 1.4. *(Vector minimization problem) Let* $K \subset I\!R^m$ *be a closed convex cone, such that both* K *and its polar cone* K^+ *have nonempty interior. Consider the partial order in* $I\!R^m$ *given by*

$$x \preceq y \text{ if and only if } y - x \in K$$

$$x \prec y \text{ if and only if } y - x \in int(K).$$

A function $F : C \subset I\!R^n \to I\!R^m$ *is said to be* $K-$*convex if* C *is convex and* $F(tx + (1-t)y) \preceq t\,F(x) + (1-t)\,F(y)$ *for all* $x, y \in C$ *and for all* $t \in (0, 1)$*. Let* $K \subset I\!R^m$ *be a closed convex cone with nonempty interior, and let* $F : C \to I\!R^m$ *be a* $K-$*convex mapping. The vector minimization problem (VMP, for short) is to find* $x^* \in C$ *such that* $F(y) \not\prec F(x^*)$ *for all* $y \in C$*.*

If we take $f(x, y) = \max_{\|z\|=1,\, z \in K^+} \langle z, F(y) - F(x) \rangle$*, then* x^* *is a solution to problem VMP if and only if* x^* *is a solution to problem EP.*

Example 1.5. *(Fixed point problem) Let* $T : I\!R^n \to 2^{I\!R^n}$ *be an upper semicontinuous point-to-set mapping such that* $T(x)$ *is a nonempty, convex compact subset of* C *for each* $x \in C$*. The fixed point problem (FPP, for short) is to find* $x^* \in C$ *such that* $x^* \in T(x^*)$*.*

If we take $f(x, y) = \max_{\xi \in T(x)} \langle x - \xi, y - x \rangle$ *for all* $x, y \in C$*, then* x^* *is a solution to problem FPP if and only if* x^* *is a solution to problem EP.*

Example 1.6. *(Variational inequality problem) Let* $T : C \to 2^{I\!R^n}$ *be an upper semicontinuous point-to-set mapping such that* $T(x)$ *is a nonempty compact set for all* $x \in C$*. The variational inequality problem (VIP, for short) is to find* $x^* \in C$ *and* $\xi \in T(x^*)$ *such that*

$$\langle \xi, y - x^* \rangle \geq 0 \text{ for all } y \in C.$$

If we take $f(x, y) = \max_{\xi \in T(x)} \langle \xi, y - x \rangle$ *for all* $x, y \in C$*, then* x^* *is a solution to problem VIP if and only if* x^* *is a solution to problem EP.*

Example 1.7. *Let* $C = I\!R^n_+$ *and* $f(x, y) = \langle Px + Qy + q, y - x \rangle$*, where* $q \in I\!R^n$ *and* P*,* Q *are two symmetric positive semidefinite matrices of dimension* n*. The corresponding equilibrium problem is a generalized form of an equilibrium problem defined by the Nash-Cournot oligopolistic market equilibrium model [67].*

Note that this problem is not a variational inequality problem.

As shown above by the examples, problem EP is a very general problem. Its interest is that it unifies all these particular problems in a convenient way. Therefore, many methods devoted to solving one of these problems can be extended, with suitable modifications, to solving the general equilibrium problem.

In this thesis two numerical methods will be mainly studied for solving equilibrium problems: the proximal point method and a method derived from the auxiliary problem principle. Both methods are based on a fixed point property associated with problem EP. Furthermore, the aim of the thesis is to go progressively from the classical proximal point method to an interior proximal point method for solving problem EP. So the title of the thesis: "Towards Interior Proximal Point Methods for Solving Equilibrium Problems". In a first part (Chapter 3), the proximal point method is studied in the case where f is convex and nonsmooth in the second argument. A special emphasis will be given on an implementable method, called the bundle method, for solving problem EP. In this method the constraint set is simply incorporated into each subproblem. In a second part (Chapters 4-5), the constraints are taken into account thanks to a barrier function associated with an entropy-like distance. The corresponding method is a generalization to problem EP of a method due to Auslender, Teboulle, and Ben-Tiba for solving convex minimization problems [9] and variational inequality problems [10]. We study the convergence of the new method with several variants (Chapter 4) and we consider a bundle-type implementation for the particular case of the constrained convex minimization (Chapter 5).

However before developing each of these methods, an entire chapter (Chapter 2) will be devoted to the basic notions and methods that are well known in the literature for solving equilibrium problems.

The main contribution of this thesis is contained in Chapters 3, 4 and 5. It has been the subject of three papers [83], [84] and [85] published in Journal of Convex Analysis, Mathematical Programming and Journal of Global Optimization, respectively.

For any undefined terms or usage concerning Convex Analysis, the readers are referred to the books [5], [74], and [86].

Chapter 2

Proximal Point Methods

In this thesis we are particularly interested in equilibrium problems where the function f is convex and nonsmooth in the second argument. One of the well-known methods for taking account of this situation is the proximal point method. This method due to Martinet [60] and developed by Rockafellar [73] has been first applied for solving a nonsmooth convex minimization problem. The basic idea is to replace the nonsmooth objective function by a smooth one in such a way that the minima of the two functions coincide. Practically nonsmooth strongly convex subproblems are considered whose solutions converge to a minimum of the nonsmooth objective function [28], [58]. This proximal point method has been generalized for solving variational inequality and equilibrium problems [66].

In order to make this method implementable, approximate solutions of each subproblem can be obtained using a bundle strategy [28], [58]. The subproblems become convex quadratic programming problems and can be solved very efficiently. This method first developed for solving minimization problems has been generalized for solving variational inequality problems [75].

The way the constraints are taken into account is also important. As usual two strategies can be used for dealing with constraints: the constraint is either directly included in the subproblem or treated thanks to a barrier function. This latter method has been intensively studied by Auslender, Teboulle, and Ben-Tiba [9], [10] for solving convex minimization problems and variational inequality problems over polyhedrons.

The aim of this chapter is to give a survey of all these methods. In a first section we consider the proximal point method for solving nonsmooth convex minimization problems. Then we examine its generalization to variational inequality problems and to equilibrium problems. Finally we present the main features of the barrier method also called the interior proximal point method.

2.1 Convex Minimization Problems

Consider the convex minimization problem:

$$\min_{y \in C} F(y), \tag{CMP}$$

where $F : I\!R^n \to I\!R \cup \{+\infty\}$ is a lower semicontinuous proper and convex function.

This problem, as mentioned above, is a particular case of problem EP. Besides, if $F \in C^1(C)$, then the solution set of problem CMP is equivalent to the one of the variational inequality problem $\langle \nabla F(x), y - x \rangle \geq 0$ for all $y \in C$. In this section, for the sake of simplicity, we consider $C = I\!R^n$.

When F is smooth, many numerical methods have been proposed to find a minimum of problem CMP like Newton's method, Conjugate direction methods, Quasi-Newton methods. More details about these methods can be found in [18], [81].

When F is nonsmooth, a strategy is to consider the proximal point method which is based on a fixed point property.

2.1.1 Classical Proximal Point Algorithm

The proximal point method, according to Rockafellar's terminology, is one of the most popular method for solving the nonsmooth convex optimization problem. It has been proposed by Martinet [60] for convex minimization problems and then developed by Rockafellar [73] for maximal monotone operator problems. More recently, a lot of works have been devoted to this method and nowadays it is still the object of intensive investigation (see, for example, [55], [77], [78], [77]). This method is based on a regularization function due to Moreau and Yosida (see, for example, [88]).

Definition 2.1. *Let $c > 0$. For each $x \in I\!R^n$, the function $J : I\!R^n \to I\!R$ defined by*

$$J(x) = \min_{y \in I\!R^n} \{ F(y) + \frac{1}{2\,c} \parallel y - x \parallel^2 \} \tag{2.1}$$

is called the Moreau - Yosida regularization of F.

The next proposition shows that the Moreau-Yosida regularization has many nice properties.

Proposition 2.1. ([37], Lemma 4.1.1 and Theorem 4.1.4, Volume II)

(a) The Moreau - Yosida regularization J is finite everywhere, convex and differentiable on $\mathbb{R}^n$,

(b) For each $x \in \mathbb{R}^n$, problem (2.1) *has a unique solution denoted $p_F(x)$,*

(c) The gradient of the Moreau - Yosida regularization is Lipschitz continuous on $\mathbb{R}^n$ with constant $1/c$, and

$$\nabla J(x) = \frac{1}{c}\left[x - p_F(x)\right] \;\in\; \partial F(p_F(x)) \;\textit{ for all }\; x \in \mathbb{R}^n,$$

(d) If F^* and J^* stand for the conjugate functions of F and J respectively, i.e., for each $y \in \mathbb{R}^n$, $F^*(y) = \sup_{x\in\mathbb{R}^n}\left\{\langle x, y\rangle - F(x)\right\}$ and $J^*(y) = \sup_{x\in\mathbb{R}^n}\left\{\langle x, y\rangle - J(x)\right\}$, then for each $s \in \mathbb{R}^n$, one has

$$J^*(s) = F^*(s) + \frac{c}{2}\|s\|^2.$$

Note that, because F, J are lower semicontinuous proper and convex, so are their conjugate functions.

It is useful to introduce here a simple example to illustrate the Moreau-Yosida regularization function J.

Example 2.1. *Let $F(x) = |x|$. The Moreau-Yosida regularization of F is*

$$J(x) = \begin{cases} \frac{1}{2\,c}x^2 & \textit{if } |x| \le c, \\ |x| - \frac{c}{2} & \textit{if } |x| > c. \end{cases}$$

Observe, from this example, that the minimum sets of F and J are the same. In fact, this result is true for any convex funtion F. Thanks to Proposition 2.1, we obtain the following properties of the Moreau-Yosida regularization.

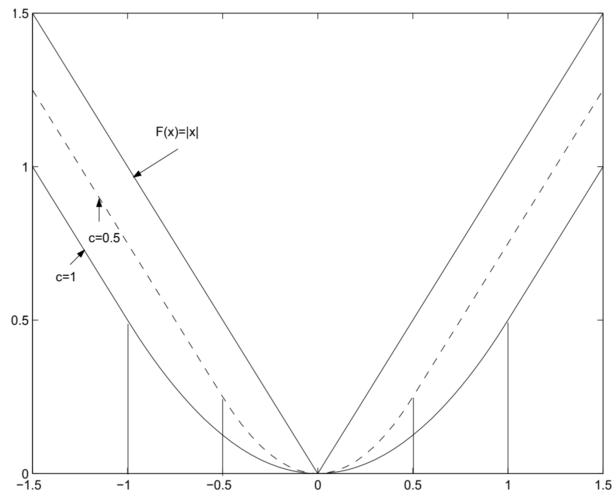

Figure 2.1: Moreau-Yosida regularization for different values of c

Theorem 2.1. ([37], Theorem 4.1.7, Volume II)

(a) $\inf_{y \in \mathbb{R}^n} J(y) = \inf_{y \in \mathbb{R}^n} F(y)$.

(b) The following statements are equivalent:

(i) x *minimizes* F,

(ii) $p_F(x) = x$,

(iii) x *minimizes* J,

(iv) $J(x) = F(x)$.

As such, Theorem 2.1 gives us some equivalent formulations to problem CMP. Amongst them, $(b.ii)$ is very interesting because it implies that solving problem CMP amounts to finding a fixed point of the *prox–operator* p_F. So we can easily derive the following algorithm from this fixed point property. This algorithm is called the classical proximal point algorithm.

Classical Proximal Point Algorithm

Data: Let $x^0 \in I\!R^n$ and let $\{c_k\}_{k\in I\!N}$ be a sequence of positive numbers.

Step 1. Set $k = 0$.

Step 2. Compute $x^{k+1} = p_F(x^k)$ by solving the problem

$$\min_{y\in I\!R^n}\{F(y) + \frac{1}{2\,c_k}\|y - x^k\|^2\} \tag{2.2}$$

Step 3. If $x^{k+1} = x^k$, then Stop: x^{k+1} is a minimum of F.

Step 4. Replace k by $k+1$ and go to Step 2.

Remark 2.1. **(a)** *If we take $c_k = c$ for all k, then $x^{k+1} = p_F(x^k)$ becomes $x^{k+1} = x^k - c\,\nabla J(x^k)$. So, in this case, the proximal point method is the gradient method applied to J with a constant step c.*

(b) *When x^{k+1} is the solution to subproblem* (2.2)*, we have, using the optimality condition, that*

$$\nabla(-\frac{1}{2\,c_k}\|\cdot - x^k\|^2)(x^{k+1}) \in \partial F(x^{k+1}).$$

In other words, the slope of the tangent of $-\frac{1}{2\,c_k}\|\cdot - x^k\|^2$ coincides with the slope of some subgradient of F at x^{k+1}. Consequently, x^{k+1} is the unique point at which the graph of the quadratic function $-\frac{1}{2\,c_k}\|\cdot - x^k\|^2$ raised up or down just touches the graph of $F(y)$.

The progress toward the minimum of F depends on the choice of the positive sequence $\{c_k\}_{k\in I\!N}$. When c_k is chosen large, the graph of the quadratic function is "blunt". In this case, solving subproblem (2.2) *is as difficult as solving CMP. However, the method makes slow the progress when c_k is small.*

The convergence result of the classical proximal point algorithm is described as follows.

Theorem 2.2. ([37], Theorem 4.2.4, Volume II) *Let $\{x^k\}_{k\in I\!N}$ be the sequence generated by the algorithm. If $\sum_{k=0}^{+\infty} c_k = +\infty$, then*

(a) $\lim_{k\to\infty} F(x^k) = F^* \equiv \inf_{y\in I\!R^n} F(y)$,

(b) The sequence $\{x^k\}$ converges to some minimum of F (if any).

In summary, as to problem CMP, we are not specific whether it has solution or not, and because of this, finding its solution seems to be silly. Oppositely, subproblem (2.2) always has a unique solution because of strong convexity. Nevertheless, it is only a conceptual algorithm because it is not identified how to carry out Step 2. To handle this problem, we introduce in the next subsection a strategy for approximating F. The resulting method is called the bundle method.

2.1.2 Bundle Proximal Point Algorithm

Our task is now to identify how to solve subproblem (2.2) when F is nonsmooth. Obviously in this case finding exactly x^{k+1} in (2.2) is very difficult. Therefore, it is interesting, from a numerical point of view, to solve approximately the subproblems. The strategy is to replace at each iteration the function F by a simpler convex function φ in such a way that the subproblems are easier to solve and that the convergence of the sequence of minima is preserved.

For example, if at iteration k, F is replaced by a piecewise linear convex function of the form $\varphi^k(x) = \max_{1\leq j\leq p}\{a_j^T x + b_j\}$, where $p \in I\!N_0$ and for all j, $a_j \in I\!R^n$ and $b_j \in I\!R$, then the subproblem $\min_{y\in I\!R^n}\{\varphi^k(y) + \frac{1}{2\,c_k}\,\|y - x^k\|^2\}$ is equivalent to the convex quadratic problem

$$\begin{cases} \min & v + \frac{1}{2\,c_k}\,\|y - x^k\|^2 \\ \text{s.t.} & a_j^T y + b_j \leq v, \quad j = 1 \dots p. \end{cases}$$

There is a large number of efficient methods for solving such a problem.

As usual, we assume that at x^k, only the value $F(x^k)$ and some subgradient $s(x^k) \in \partial F(x^k)$ are available thanks to an oracle [28], [58]. We also suppose that the function F is a finite–valued convex function.

To construct such a desired function φ^k, we have to impose some conditions on it. First let us introduce the following definition.

Definition 2.2. *Let $\mu \in (0,1)$ and $x^k \in I\!R^n$. A convex function φ^k is said to be a μ-approximation of F at x^k if $\varphi^k \leq F$ and*

$$F(x^k) - F(x^{k+1}) \geq \mu\,[\,F(x^k) - \varphi^k(x^{k+1})\,],$$

where x^{k+1} is the solution of the following problem

$$\min_{y \in I\!R^n} \{\varphi^k(y) + \frac{1}{2\,c_k}\,\|y - x^k\|^2\}. \tag{2.3}$$

When $\varphi^k(x^k) = F(x^k)$, this condition means that the actual reduction on F is at least a fraction of the reduction predicted by the model φ^k.

Bundle Proximal Point Algorithm

> Data: Let x^0, $\mu \in (0,1)$, and let $\{c_k\}_{k \in I\!N}$ be a sequence of positive numbers.
>
> Step 1. Set $k = 0$.
>
> Step 2. Find φ^k a μ-approximation of F at x^k and find x^{k+1} the unique solution of subproblem (2.3).
>
> Step 3. Replace k by $k+1$ and go to Step 2.

Theorem 2.3. ([28], Theorem 4.4) *Let $\{x^k\}$ be the sequence generated by the bundle proximal point algorithm.*

(a) If $\sum_{k=1}^{+\infty} c_k = +\infty$, then $F(x^k) \searrow F^* = \inf_{y \in I\!R^n} F(y)$.

(b) If, in addition, there exists $\bar{c} > 0$ such that $c_k \leq \bar{c}$ for all k, then $x^k \to x^*$ where x^* is a minimum of F (if any).

The next step is to explain how to build a μ-approximation. As we have seen, subproblem (2.3) is equivalent to a convex quadratic problem when φ^k is a piecewise linear convex function and, thus, there are many efficient numerical methods to solve such a problem. So, it is judicious to construct a piecewise linear convex function for the model function φ^k piece by piece by generating successive models

$$\varphi_i^k, i = 1, 2, \ldots$$

until (if possible) $\varphi_{i_k}^k$ is a μ-approximation of F at x^k for some $i_k \geq 1$. For $i = 1, 2, \ldots$, we denote by y_i^k the unique solution to the problem (P_i^k)

$$\min_{y \in I\!R^n} \{\varphi_i^k(y) + \frac{1}{2\,c_k}\|y - x^k\|^2\},$$

and we set $\varphi^k = \varphi_{i_k}^k$ and $x^{k+1} = y_{i_k}^k$.

In order to obtain a μ-approximation $\varphi^k_{i_k}$ of F at x^k, we have to impose some conditions on the successive models φ^k_i, $i = 1, 2, \ldots$. However, before presenting them, we need to define the affine functions $l^k_i,\ i = 1, 2, \ldots$ by

$$l^k_i(y) = \varphi^k_i(y^k_i) + \langle \gamma^k_i, y - y^k_i \rangle \ \text{ for all } \ y \in I\!R^n,$$

where $\gamma^k_i = \dfrac{1}{c_k}(x^k - y^k_i)$. By optimality of y^k_i, we have $\gamma^k_i \in \partial\varphi^k_i(y^k_i)$. Then it is easy to observe that, for $i = 1, 2, \ldots$

$$l^k_i(y^k_i) = \varphi^k_i(y^k_i) \quad \text{and} \quad l^k_i(y) \le \varphi^k_i(y) \ \text{ for all } \ y \in I\!R^n.$$

Now, we assume that the following conditions are satisfied by the convex models φ^k_i, for all $i = 1, 2, \ldots$

$(A1)$ $\varphi^k_i \le F$,

$(A2)$ $l^k_i \le \varphi^k_{i+1}$,

$(A3)$ $F(y^k_i) + \langle s(y^k_i), \cdot - y^k_i \rangle \le \varphi^k_{i+1}$,

where $s(y^k_i)$ denotes the subgradient of F available at y^k_i. These conditions have already been used in [28] for the standard proximal method.

Let us introduce several models fulfill these conditions. For example, for the first model φ^k_1, we can take the linear function

$$\varphi^k_1(y) = F(x^k) + \langle s(x^k), y - x^k \rangle \ \text{ for all } \ y \in I\!R^n.$$

Since $s(x^k) \in \partial F(x^k)$, $(A1)$ is satisfied for $i = 1$. For the next models $\varphi^k_i, i = 2, \ldots$, there exist several possibilities. A first example is to take for $i = 1, 2, \ldots$

$$\varphi^k_{i+1}(y) = \max\{l^k_i(y), F(y^k_i) + \langle s(y^k_i), y - y^k_i \rangle\} \ \text{ for all } \ y \in I\!R^n.$$

$(A2) - (A3)$ are obviously satisfied and $(A1)$ is also satisfied because each linear piece of these functions is below F.

Another example is to take for $i = 1, 2, \ldots$

$$\varphi^k_{i+1}(y) = \max_{0 \le j \le i} \{F(y^k_j) + \langle s(y^k_j), y - y^k_j \rangle\} \ \text{ for all } \ y \in I\!R^n, \tag{2.4}$$

where $y^k_0 = x^k$. Since $s(y^k_j) \in \partial F_k(y^k_j)$ for $j = 0, \ldots, i$ and since $\varphi^k_{i+1} \ge \varphi^k_i \ge l^k_i$, it is easy to see that $(A1) - (A3)$ are satisfied.

As usual in the bundle methods, we assume that at each $x \in I\!R^n$, one subgradient of F at x can be computed (this subgradient is denoted by $s(x)$ in the sequel). This assumption is realistic because computing the whole subdifferential is often very expensive or impossible while obtaining one subgradient is often easy. This situation occurs, for instance, if the function F is the dual function associated with a mathematical programming problem.

Now the algorithm allowing us to pass from x^k to x^{k+1}, i.e., to make what is called a serious step, can be expressed as follows.

Serious Step Algorithm

Data: Let $x^k \in I\!R^n$ and $\mu \in (0, 1)$.

Step 1. Set $i = 1$.

Step 2. Choose φ_i^k a convex function that satisfies $(A1) - (A3)$ and solve the subproblem (P_i^k) to get y_i^k.

Step 3. If $F(x^k) - F(y_i^k) \geq \mu\,[F(x^k) - \varphi_i^k(y_i^k)]$, then set $x^{k+1} = y_i^k$, $i_k = i$ and Stop: x^{k+1} is a serious step.

Step 4. Replace i by $i + 1$ and go to Step 2.

Our aim is now to prove that if x^k is not a minimum of F and if the models $\varphi_i^k, i = 1, \ldots$ satisfy $(A1) - (A2)$, then there exists $i_k \in I\!N_0$ such that $\varphi_{i_k}^k$ is a μ-approximation of F at x^k, i.e., that the Stop occurs at Step 3 after finitely many iterations.

In order to obtain this result we need the following proposition.

Proposition 2.2. ([28], Proposition 4.3) *Suppose that the models $\varphi_i^k, i = 1, 2, \ldots$ satisfy $(A1)-$ $(A3)$, and let, for each i, y_i^k be the unique solution of subproblem (P_i^k). Then*

(1) $F(y_i^k) - \varphi_i^k(y_i^k) \to 0$,

(2) $y_i^k \to p_F(x^k)$,

when $i \to +\infty$.

Theorem 2.4. ([28], Theorem 4.4) *If x^k is not a minimum of F, then the serious step algorithm stops after finitely many iterations i_k with $\varphi_{i_k}^k$ a μ-approximation of F at x^k and with $x^{k+1} = y_{i_k}^k$.*

Now we incorporate the serious step algorithm into Step 2 of the bundle proximal point algorithm. Then we obtain the following algorithm.

Bundle Proximal Point Algorithm I

Data: Let $x^0 \in C$, $\mu \in (0,1)$ and let $\{c_k\}_{k\in I\!N}$ be a sequence of positive numbers.

Step 1. Set $y_0^0 = x^0$ and $k = 0, i = 1$.

Step 2. Choose a piecewise linear convex function φ_i^k satisfying $(A1) - (A3)$ and solve

$$\min_{y\in I\!R^n} \{\varphi_i^k(y) + \frac{1}{2\,c_k}\|y - x^k\|^2\},$$

to obtain the unique optimal solution y_i^k.

Step 3. If

$$F(x^k) - F(y_i^k) \geq \mu\,[F(x^k) - \varphi_i^k(y_i^k)], \tag{2.5}$$

then set $x^{k+1} = y_i^k$, $y_0^{k+1} = x^{k+1}$, replace k by $k+1$ and set $i = 0$.

Step 4. Replace i by $i+1$ and go to Step 2.

From Theorems 2.3 and 2.4, we obtain the following convergence results.

Theorem 2.5. ([28], Theorem 4.4) *Suppose that $\sum_{k=0}^{+\infty} c_k = +\infty$ and that there exists $\bar{c} > 0$ such that $c_k \leq \bar{c}$ for all k. If the sequence $\{x^k\}$ generated by the bundle proximal point algorithm I is infinite, then $\{x^k\}$ converges to some minimum of F. If after some k has been reached, the criterion (2.5) is never satisfied, then x^k is a minimum of F.*

For practical implementation, it is necessary to define a stopping criterion. Let $\epsilon > 0$. Let us recall that $\bar{x}$ is an ε–stationary point of problem CMP if there exists $s \in \partial_\varepsilon F(\bar{x})$ with $\|s\| \leq \varepsilon$. Since, by optimality of y_i^k, $\gamma_i^k \in \partial\varphi_i^k(y_i^k)$, it is easy to prove that

$$\gamma_i^k \in \partial_{\varepsilon_i^k} F(y_i^k)$$

where $\varepsilon_i^k = F(y_i^k) - \varphi_i^k(y_i^k)$. Indeed, for all $y \in I\!R^n$, we have

$$\begin{aligned} F(y) &\geq \varphi_i^k(y) \geq \varphi_i^k(y_i^k) + \langle \gamma_i^k, y - y_i^k\rangle \\ &= F(y_i^k) + \langle \gamma_i^k, y - y_i^k\rangle - [F(y_i^k) - \varphi_i^k(y_i^k)]. \end{aligned}$$

Hence we introduce the stopping criterion: if $F(y_i^k) - \varphi_i^k(y_i^k) \leq \varepsilon$ and $\|\gamma_i^k\| \leq \varepsilon$, then y_i^k is an ε–stationary point.
In order to prove that the stopping criterion is satisfied after finitely many iterations, we need the following proposition.

Proposition 2.3. ([80], Proposition 7.5.2) *Suppose that there exist two positive parameters* $\underline{c}$ *and* $\bar{c}$ *such that* $0 < \underline{c} \leq c_k \leq \bar{c}$ *for all* k. *If the sequence* $\{x^k\}$ *generated by the bundle proximal point algorithm I is infinite, then* $F(y_i^k) - \varphi_i^k(y_i^k) \to 0$ *and* $\|\gamma_i^k\| \to 0$ *when* $k \to +\infty$. *If the sequence* $\{x^k\}$ *is finite with* k *the latest index, then* $F(y_i^k) - \varphi_i^k(y_i^k) \to 0$ *and* $\|\gamma_i^k\| \to 0$ *when* $i \to +\infty$.

We are now in a position to present the bundle proximal point algorithm with a stopping criterion.

Bundle Proximal Point Algorithm II

Data: Let $x^0 \in C$, $\mu \in (0,1)$, $\varepsilon > 0$, and let $\{c_k\}_{k \in \mathbb{N}}$ be a sequence of positive numbers.

Step 1. Set $y_0^0 = x^0$ and $k = 0, i = 1$.

Step 2. Choose a piecewise linear convex function φ_i^k satisfying $(A1) - (A3)$ and solve

$$\min_{y \in \mathbb{R}^n} \{\varphi_i^k(y) + \frac{1}{2\,c_k}\|y - x^k\|^2\}, \qquad (P_i^k)$$

to obtain the unique optimal solution y_i^k.
Compute $\gamma_i^k = (x^k - y_i^k)/c_k$.
If $\|\gamma_i^k\| \leq \varepsilon$ and $F(y_i^k) - \varphi_i^k(y_i^k) \leq \varepsilon$, then Stop: y_i^k is an ε–stationary point.

Step 3. If

$$F(x^k) - F(y_i^k) \geq \mu\,[F(x^k) - \varphi_i^k(y_i^k)], \qquad (2.6)$$

then set $x^{k+1} = y_i^k$, $y_0^{k+1} = x^{k+1}$, replace k by $k+1$ and set $i = 0$.

Step 4. Replace i by $i+1$ and go to Step 2.

Combining the results of Theorem 2.5 and Proposition 2.3, we obtain the following convergence result.

Theorem 2.6. ([80], Theorem 7.5.4) *Suppose that $0 < \underline{c} \leq c_k \leq \bar{c}$ for all k. The bundle proximal point algorithm II exits after finitely many iterations with an ε–stationary point. In other words, there exists k and i such that $\|\gamma_i^k\| \leq \varepsilon$ and $F(y_i^k) - \varphi_i^k(y_i^k) \leq \varepsilon$.*

2.2 Equilibrium Problems

This section is intended to review some methods for solving equilibrium problems and to shed light on the issues related to this thesis. Two important methods are presented here consisting in the proximal point method and a method based on the auxiliary problem principle. First we give convergence results concerning these methods and then we show how to make them implementable using what is called a gap function. Then to avoid strong assumptions on the equilibrium function f, we describe an extragradient method which combines the projection method with the auxiliary problem principle. Finally, we explain how to use an efficient barrier method to treat linear constraints. This method gives rise to the interior proximal point algorithms. From now on, we assume that problem EP has at least one solution.

2.2.1 Existence and Uniqueness of Solutions

This section presents a number of basic results about the existence and uniqueness of solutions of problem EP along with some related definitions. Because the existence and uniqueness of solutions is not the main issue studied in this thesis, we only mention concisely the most important results without any proof. The proofs can be found in the corresponding references.

To begin with, let us observe that proving the existence of solutions to problem EP amounts to show that $\cap_{y \in C} Q(y) \neq \emptyset$, where, for each $y \in C$, $Q(y) = \{x \in C \,|\, f(x,y) \geq 0\}$. For this reason, we can use the following fixed point theorem due to Ky Fan [31].

Theorem 2.7. ([31], Corollary 1) *Let C be a subset of $I\!R^n$. For each $y \in C$, let $Q(y)$ be a closed subset of $I\!R^n$ such that for every finite subset $\{y^1, \ldots y^n\}$ of C, one has*

$$\textit{conv}\,\{y^1, \ldots y^n\} \subset \bigcup_{i=1}^{n} Q(y^i). \tag{2.7}$$

If $Q(y)$ is compact for at least one $y \in C$, then $\bigcap_{y \in C} Q(y) \neq \emptyset$.

In order to employ this result, we need to introduce the following definitions.

Definition 2.3. *A function $F : C \to R$ is said to be convex if for each $x,\, y \in C$ and for all $\lambda \in [0,1]$*

$$F(\lambda x + (1-\lambda)y) \le \lambda F(x) + (1-\lambda)F(y),$$

strongly convex if there exists $\beta > 0$ such that for each $x,\, y \in C$ and for all $\lambda \in (0,1)$

$$F(\lambda x + (1-\lambda)y) \le \lambda F(x) + (1-\lambda)F(y) - \frac{1}{2}\beta(1-\beta)\|x-y\|^2$$

quasiconvex if for each $x,\, y \in C$ and for all $\lambda \in [0,1]$

$$F(\lambda x + (1-\lambda)y) \le \max\,\{\, F(x),\, F(y)\,\},$$

semistrictly quasiconvex if for each $x,\, y \in C$ such that $F(x) \neq F(y)$ and for all $\lambda \in (0,1)$

$$F(\lambda x + (1-\lambda)y) < \max\,\{\, F(x),\, F(y)\,\},$$

hemicontinuous if for each $x,\, y \in C$ and for all $\lambda \in [0,1]$

$$\lim_{\lambda \to 0^+} F(\lambda x + (1-\lambda)y) = F(y),$$

upper hemicontinuous if for each $x,\, y \in C$

$$\limsup_{\lambda \to 0^+} F(\lambda x + (1-\lambda)y) \le F(y),$$

lower semicontinuous at $x \in C$ if for any sequence $\{x_k\} \subset C$ converging to x,

$$\liminf_{k \to +\infty} F(x_k) \ge F(x),$$

upper semicontinuous at $x \in C$ if, for any sequence $\{x_k\} \subset C$ converging to x,

$$\limsup_{k \to +\infty} F(x_k) \le F(x).$$

Furthermore, F is said to be lower semicontinuous (upper semicontinuous) on C if F is lower semicontinuous (upper semicontinuous) at every $x \in C$.

This definition gives immediately that: (i) if F is convex, then it is also quasiconvex and semistrictly quasiconvex, (ii) if F is lower semicontinuous and upper semicontinuous, then F is continuous, and (iii) if F is hemicontinuous, then F is upper hemicontinuous.

Using Theorem 2.7, we can now present an existence result for problem EP, which is known as Ky Fan's inequality.

Theorem 2.8. ([30], Theorem 1) *Suppose that the following assumptions hold:*

a. C *is a compact,*

b. $f(x, \cdot) : C \to I\!R$ *is quasiconvex for each* $x \in C$*,*

c. $f(\cdot, y) : C \to I\!R$ *is upper semicontinuous for each* $y \in C$*.*

Then $\cap_{y \in C} Q(y) \neq \emptyset$*, i.e., problem EP is solvable.*

This theorem is a direct consequence of Theorem 2.7. Indeed, from assumptions $a.$ and $c.$, we deduce that $Q(y)$ is compact for all $y \in C$ and, from assumption $b.$, that condition (2.7) is satisfied.

However, Theorem 2.8 cannot be applied when C is not compact, which is very often the case in applications (for example when $C = I\!R^n_+$). To avoid this drawback, Brézis, Nirenberg, and Stampacchia [25] improved this result by replacing the compactness of C by the coercivity of f on C in the sense that *there exist a nonempty compact subset* $L \subset I\!R^n$ *and* $y_0 \in L \cap C$ *such that for every* $x \in C \setminus L$*,* $f(x, y_0) < 0$.

Theorem 2.9. ([25], Theorem 1) *Suppose that the following assumptions hold:*

a. f *is coercive on* C*,*

b. $f(x, \cdot) : C \to I\!R$ *is quasiconvex for each* $x \in C$*,*

c. $f(\cdot, y) : C \to I\!R$ *is upper semicontinuous for each* $y \in C$*.*

Then problem EP is solvable.

It is worthy noting that, for minimization problems, $F : C \to I\!R$ *is said to be coercive on* C *if there exists* $\alpha \in I\!R$ *such that the closure of the level set* $\{x \in C \mid F(x) \leq \alpha\}$ *is compact*. If $f(x, y) = F(y) - F(x)$, then the coercivity of f is equivalent to that of F.

Another popular approach of addressing the existence of solutions of problem EP is to consider the same question but for its dual formulation. The dual equilibrium problem (DEP, for short) is to find a point $x^* \in C$ such that

$$f(y, x^*) \leq 0 \quad \text{for all} \quad y \in C. \tag{DEP}$$

This problem can also be written as: find $x^* \in C$ such that $x^* \in \cap_{y \in C} L_f(y)$, where, for each $y \in C$, $L_f(y) = \{x \in C \,|\, f(y,x) \leq 0\}$. It is the convex feasibility problem studied by Iusem and Sosa [40].

Let us denote by S^* and S^d the solution sets of EP and DEP, respectively. Obviously, the strategy to solve EP by solving DEP is only interesting when $S^d \subset S^*$. For that purpose, we need to define the following monotonicity properties.

Definition 2.4. *The function f is said to be monotone, if for any $x, y \in C$*

$$f(x,y) + f(y,x) \leq 0,$$

strictly monotone, if for any $x, y \in C$ and $x \neq y$

$$f(x,y) + f(y,x) < 0,$$

strongly monotone with modulus $\gamma > 0$, if for all $x, y \in C$,

$$f(x,y) + f(y,x) \leq -\gamma \|x - y\|^2,$$

pseudomonotone, if for any $x, y \in C$

$$f(x,y) \geq 0 \Rightarrow f(y,x) \leq 0,$$

strictly pseudomonotone, if for any $x, y \in C$ and $x \neq y$

$$f(x,y) \geq 0 \Rightarrow f(y,x) < 0.$$

It is straightforward to see that if f is monotone, then f is pseudomonotone, and that if f is strictly pseudomonotone, then f is pseudomonotone. Moreover, if f is strongly monotone, then f is monotone. The relationships between S^* and S^d are given in the next lemma.

Lemma 2.1. ([19], Proposition 3.2)

a. If f is pseudomonotone, then $S^* \subset S^d$,

b. If $f(x, \cdot)$ is quasiconvex and semistrictly quasiconvex for each $x \in C$ and $f(\cdot, y)$ is hemicontinuous for each $y \in C$, then $S^d \subset S^*$.

Thanks to this lemma, Bianchi and Schaible [19], and Brézis, Nirenberg, and Stampacchia [25] proved the existence and uniqueness of solutions of problems EP and DEP.

Theorem 2.10. *Suppose that the following assumptions hold:*

a. Either C is compact or f is coercive on C,

b. $f(x, \cdot)$ is semistrictly quasiconvex and lower semicontinuous for each $x \in C$,

c. $f(\cdot, y)$ is hemicontinuous for each $y \in C$,

d. f is pseudomonotone.

Then, the solution sets of problems EP and DEP coincide and are nonempty, convex and compact. Moreover, if f is strictly pseudomonotone, then problems EP and DEP have at most one solution.

Remark 2.2. *Obviously the dual problem coincides with the equilibrium problem when it is the convex minimization problem (Example 1.1). In that case the duality is not interesting at all. Moreover, the dual problem is not related to the Fenchel-type dual problem introduced recently by Martinez-Legaz and Sosa [61].*

It should be noted that there exist a number of variant versions of the existence and uniqueness of the solution of problem EP, which are slight modifications of the results presented above. An excellent survey of these results can be found in [47].

2.2.2 Proximal Point Algorithms

Motivated by the efficiency of the classical proximal point algorithm, Moudafi [66] suggested the following proximal point algorithm for solving the equilibrium problems.

Proximal Point Algorithm

Data: Let $x^0 \in C$ and $c > 0$.

Step 1. Set $k = 0$.

Step 2. Find a solution $x^{k+1} \in C$ to the equilibrium subproblem

$$f(x^{k+1}, y) + \frac{1}{c}\left\langle x^{k+1} - x^k, y - x^{k+1}\right\rangle \geq 0 \text{ for all } y \in C. \tag{PEP}$$

Step 3. Replace k by $k+1$, and go to Step 2.

This algorithm can be seen as a general form of the classical proximal point algorithm. Indeed, if we take $C = I\!R^n$ and $f(x,y) = F(y) - F(x)$ where F is a lower semicontinuous proper and convex function on $I\!R^n$, then problem PEP reduces to

$$F(y) \geq F(x^{k+1}) + \frac{1}{c}\,\langle x^k - x^{k+1}, y - x^{k+1}\rangle \text{ for all } y \in I\!R^n,$$

i.e., $\frac{1}{c}\,(x^k - x^{k+1}) \in \partial F(x^{k+1})$. This is the optimality condition related to the convex problem

$$x^{k+1} = \arg\min_{y \in I\!R^n} \{\, F(y) + \frac{1}{2\,c}\|y - x^k\|^2 \,\}.$$

So, in that case, the proximal point algorithm coincides with the classical proximal point algorithm introduced by Martinet [60] for solving convex minimization problems. The convergence of the proximal point algorithm is given in the next theorem.

Theorem 2.11. ([66], Theorem 1) *Assume that f is monotone, that $f(\cdot, y)$ is upper hemicontinuous for all $y \in C$, and that $f(x, \cdot)$ is convex and lower semicontinuous on C for all $x \in C$. Then, for each k, problem PEP has a unique solution x^{k+1}, and the sequence $\{x^k\}$ generated by the proximal point algorithm converges to a solution to problem EP. If, in addition, f is strongly monotone, then the sequence $\{x^k\}$ generated by the algorithm converges to the unique solution to problem EP.*

When f is monotone, let us observe that for each k, the function $(x, y) \mapsto f(x, y) + \frac{1}{c}\,\langle x - x^k, y - x\rangle$ is strongly monotone. So for using the proximal point algorithm, we need an efficient algorithm for solving the strongly monotone equilibrium subproblems PEP. Such an algorithm will be described in Section 2.2.3.

Next it is also interesting, for numerical reasons, to show that that the convergence can be preserved when the subproblems are solved approximately. This was done by Konnov [46] where the following inexact version of the proximal point algorithm is proposed.

Inexact Proximal Point Algorithm

Data: Let $\bar{x}^0 \in C$, $c > 0$, and let $\{\epsilon_k\}$ be a sequence of positive numbers.

Step 1. Set $k = 0$.

Step 2. Find $\bar{x}^{k+1} \in C$ such that $\|\bar{x}^{k+1} - x^{k+1}\| \leq \epsilon_{k+1}$, where

$$x^{k+1} \in C_{k+1} = \{ x \in C \mid f(x, y) + \frac{1}{c} \langle x - \bar{x}^k, y - x \rangle \geq 0 \text{ for all } y \in C \}.$$

Step 3. Replace k by $k+1$, and go to Step 2.

Let us observe that each iterate $\bar{x}^{k+1}$ generated by this algorithm is an approximation of the exact solution x^{k+1} with accuracy ϵ_{k+1}.

Theorem 2.12. ([46], Theorem 2.1) *Let $\{\bar{x}^k\}$ be a sequence generated by the inexact proximal point algorithm. Suppose that $S^d \neq \emptyset$, $\sum_{k=0}^{\infty} \epsilon_k < \infty$, and that $C_k \neq \emptyset$ for $k = 1, 2, \ldots$. Then*

a. $\{x^k\}$ has limit points in C and all these limit points belong to S^*,

b. If $S^d = S^*$, then $\lim_{k\to\infty} x^k = x^* \in S^*$.

Let us note that, contrary to Theorem 2.11, it is not supposed that f is monotone to obtain the convergence, but only that $S^d = S^*$, which is true when f is pseudomonotone.

In order to make this algorithm implementable, it remains to explain how to stop the algorithm used for solving the subproblems to get the approximate solution $\bar{x}^{k+1}$ without computing the exact solution x^{k+1}. This will be carried out thanks to a gap function (see Section 2.2.4).

2.2.3 Auxiliary Problem Principle

Another way to solve problem EP is based on the following fixed point property: $x^* \in C$ is a solution to problem EP if and only if

$$x^* \in \arg\min_{y \in C} f(x^*, y). \tag{2.8}$$

Then the corresponding fixed point algorithm is the following one.

A General Algorithm

Data: Let $x^0 \in C$ and $\epsilon > 0$.
Step 1. Set $k = 0$.
Step 2. Find a solution $x^{k+1} \in C$ to the subproblem $$\min_{y \in C} \; f(x^k, y).$$
Step 3. If $x^{k+1} = x^k$, then Stop: x^k is a solution to problem EP. Replace k by $k+1$, and go to Step 2.

This algorithm is simple, but practically difficult to use because the subproblems in Step 2 may have several solutions or even no solution. To overcome this difficulty, Mastroeni [62] proposed to consider an auxiliary equilibrium problem (AuxEP, for short) instead of problem EP. This new problem is to find $x^* \in C$ such that

$$f(x^*, y) + \hbar(x^*, y) \geq 0 \;\text{ for all }\; y \in C, \qquad \text{(AuxEP)}$$

where $\hbar(\cdot, \cdot) : C \times C \to I\!R$ satisfies the following conditions:

B1. $\hbar$ is nonnegative and continuously differentiable on $C \times C$,

B2. $\hbar(x, x) = 0$ and $\nabla_y \hbar(x, x) = 0$ for all $x \in C$,

B3. $\hbar(x, \cdot)$ is strongly convex for all $x \in C$.

An example of such a function $\hbar$ is given by $\hbar(x, y) = \frac{1}{2}\|x - y\|^2$. This auxiliary principle problem generalizes the work of Cohen [26] for minimization problems [26] and for variational inequality problems [27]. Between the two problems EP and AuxEP, we have the following relationship.

Lemma 2.2. ([62], Corollary 2.1) x^* *is a solution to problem EP if and only if* x^* *is a solution to problem AuxEP.*

Thanks to this lemma, we can apply the general algorithm to the auxiliary equilibrium problem for finding a solution to problem EP. The corresponding algorithm is as follows.

Auxiliary Problem Principle Algorithm

> Data: Let $x^0 \in C$ and $c > 0$.
>
> Step 1. Set $k = 0$.
>
> Step 2. Find a solution $x^{k+1} \in C$ to the subproblem
>
> $$\min_{y \in C} \{c\, f(x^k, y) + \hbar(x^k, y)\}.$$
>
> Step 3. If $x^{k+1} = x^k$ then Stop: x^k is a solution to problem EP.
> Replace k by $k + 1$, and go to Step 2.

This algorithm is well-defined. Indeed, for each k, the function $c\, f(x^k, \cdot) + \hbar(x^k, \cdot)$ is strongly convex and thus each subproblem in Step 2 has a unique solution.

Theorem 2.13. ([62], Theorem 3.1) *Suppose that the following conditions are satisfied on the equilibrium function f:*

(a) $f(x, \cdot) : C \to I\!R$ is convex differentiable for all $x \in C$,

(b) $f(\cdot, y) : C \to I\!R$ is continuous for all $y \in C$,

(c) $f : C \times C \to I\!R$ is strongly monotone (with modulus $\gamma > 0$),

(d) There exist constants $d_1 > 0$ and $d_2 > 0$ such that, for all $x, y, z \in C$,

$$f(x, y) + f(y, z) \geq f(x, z) - d_1 \|y - x\|^2 - d_2 \|z - y\|^2. \tag{2.9}$$

Then the sequence $\{x^k\}$ generated by the auxiliary problem principle algorithm converges to the solution to problem EP provided that $c \leq d_1$ and $d_2 < \gamma$.

Remark 2.3. *Let us observe that the auxiliary problem principle algorithm is nothing else than the proximal point algorithm for convex minimization problems where, at each iteration k, we consider the objective function $f(x^k, \cdot)$. So when $f(x, y) = F(y) - F(x)$ and $\hbar(x, y) = \frac{1}{2}\|x - y\|^2$, the optimization problem in Step 2 is equivalent to*

$$\min_{y \in C} \{F(y) + \frac{1}{2\,c}\|y - x^k\|^2\},$$

i.e., the iteration $k + 1$ of the classical proximal point algorithm.

Also, the inequality (d) is a Lipschitz-type condition. Indeed, when $f(x,y) = \langle F(x), y-x\rangle$ with $F : I\!R^n \to I\!R^n$, problem EP amounts to the variational inequality problem: find $x^* \in C$ such that $\langle F(x^*), y - x^*\rangle \geq 0$ for all $y \in C$. In that case, $f(x,y) + f(y,z) - f(x,z) = \langle F(x) - F(y), y - z\rangle$ for all $x, y, z \in C$, and it is easy to see that if F is Lipschitz continuous on C (with constant $L > 0$), then for all $x, y, z \in C$,

$$|\langle F(x) - F(y), y - z\rangle| \leq L\|x - y\| \, \|y - z\| \leq \frac{L}{2}\,[\|x - y\|^2 + \|y - z\|^2],$$

and thus, f satisfies condition (2.9). Furthermore, when $z = x$, this condition becomes

$$f(x,y) + f(y,x) \geq -(d_1 + d_2)\,\|y - x\|^2 \text{ for all } x, y \in C.$$

This gives a lower bound on $f(x,y) + f(y,x)$ while the strong monotonicity gives an upper bound on $f(x,y) + f(y,x)$.

As seen above, the convergence result can only be reached, in general, when f is strongly monotone and Lipschitz continuous. So this algorithm can be used, for example, for solving subproblems PEP of the proximal point algorithm. However, these assumptions on f are too strong for many applications. To avoid them, Mastroeni modified the auxiliary problem principle algorithm introducing what is called a gap function.

2.2.4 Gap Function Approach

The gap function approach is based on the following lemma.

Lemma 2.3. ([63], Lemma 2.1) *Let $f : C \times C \to I\!R$ with $f(x,x) = 0$ for all $x \in C$. Then problem EP is equivalent to the problem of finding $x^* \in C$ such that*

$$\sup_{y\in C}\{\, -f(x^*, y)\,\} = \min_{x\in C}\{\,\sup_{y\in C}\{\, -f(x,y)\,\}\,\} = 0. \tag{2.10}$$

According to this lemma, the equilibrium problem can be transformed into a minimax problem whose optimal value is zero.
Setting $g(x) = \sup_{y\in C}\{\, -f(x,y)\,\}$, we immediately see that $g(x) \geq 0$ for all $x \in C$ and $g(x^*) = 0$ if and only if x^* is a solution to problem EP. This function is called a gap function. More generally, we introduce the following definition.

Definition 2.5. *A function $g : C \to I\!R$ is said to be a gap function for problem EP if*

a. $g(x) \geq 0$ *for all* $x \in C$,

b. $g(x^*) = 0$ *if and only if* x^* *is a solution to problem EP.*

Once a gap function is determined, a strategy for solving problem EP consists in minimizing this function until it is nearly equal to zero. The concept of gap function was first introduced by Auslender [6] for the variational inequality problem with the function $g(x) = \sup_{y \in C} \langle -F(x), y - x \rangle$. However, this gap function has two main disadvantages: it is in general not differentiable and it can be undefined when C is not compact.

The next proposition due to Mastroeni [64] gives sufficient conditions to ensure the differentiability of the gap function.

Proposition 2.4. ([64], Proposition 2.1) *Suppose that* $f(x, \cdot) : C \to I\!R$ *is a strongly convex function for every* $x \in C$*, that* f *is differentiable with respect to* x*, and that* $\nabla_x f(\cdot, \cdot)$ *is continuous on* $C \times C$*. Then the function*

$$g(x) = \sup_{y \in C} \{-f(x, y)\}$$

is a continuously differentiable gap function for problem EP whose gradient is given by

$$\nabla g(x) = -\nabla_x f(x, y(x)),$$

where $y(x) = \arg\min_{y \in C} f(x, y)$.

In this proposition, the strong convexity of $f(x, \cdot)$ is used to obtain a unique value for $y(x)$. However, this strong convexity on $f(x, \cdot)$ is not satisfied for important equilibrium problems as the variational inequality problems where $f(x, \cdot)$ is linear. To avoid this strong assumption, we consider problem AuxEP instead of problem EP and we apply Lemma 2.3 to this problem to obtain the following lemma.

Lemma 2.4. ([64], Proposition 2.2) x^* *is a solution to problem EP if and only if*

$$\sup_{y \in C} \{ -f(x^*, y) - \hbar(x^*, y) \} = \min_{x \in C} \{ \sup_{y \in C} \{ -f(x, y) - \hbar(x, y) \} \} = 0,$$

where $\hbar$ *satisfies conditions* $(B1) - (B3)$.

This lemma gives us the gap function $g(x) = \sup_{y \in C} \{-f(x, y) - \hbar(x, y)\}$. This time, the compound function $f(x, \cdot) + \hbar(x, \cdot)$ is strongly convex when $f(x, \cdot)$ is convex and the corresponding gap function is well-defined and differentiable as explained in the following theorem.

Theorem 2.14. ([64], Theorem 2.1) *Suppose that $f(x,\cdot) : C \to I\!R$ is a convex function for every $x \in C$, that f is differentiable with respect to x, and that $\nabla_x f(\cdot,\cdot)$ is continuous on $C \times C$. Suppose also that $\hbar$ satisfies conditions $(B1)-(B3)$. Then $g(x) = \sup_{y\in C}\{-f(x,y) - \hbar(x,y)\}$ is a continuously differentiable gap function for problem EP whose gradient is given by $\nabla g(x) = -\nabla_x f(x, y_x) - \nabla_x \hbar(x, y(x))$, where*

$$y(x) = \arg\min_{y\in C} \{\, f(x,y) + \hbar(x,y) \,\}.$$

Once a gap function g of class C^1 is determined, a simple method for solving problem EP consists in using a descent method for minimizing g. More precisely, let $x^k \in C$. First a descent direction d^k at x^k for g is computed and then a line search is performed along this direction to get the next iterate $x^{k+1} \in C$. Let us recall that d^k is a descent direction at x^k for g if $\nabla g(x^k)\, d^k < 0$. Such a direction is obtained using the next proposition.

Proposition 2.5. *Suppose that the hypotheses of Theorem 2.14 hold true, and in addition, that, for all $x, y \in C$,*

$$\langle \nabla_x f(x,y) + \nabla_x \hbar(x,y) + \nabla_y f(x,y) + \nabla_y \hbar(x,y), y - x\rangle \geq 0 \tag{2.11}$$

is verified. Then $d(x) = y(x) - x$ is a descent direction for the gap function at $x \in C$ provided that $y(x) \neq x$.

Proof. Apply Proposition 3.1 in [64] to the auxiliary equilibrium problem AuxEP (see more details in [69]). □

Remark 2.4. *Note that when $\hbar(x,y) = \frac{1}{2}\,\|x-y\|^2$, the assumption* (2.11) *is satisfied in the case of problem VIP, i.e., $f(x,y) = \langle F(x), y - x\rangle$ provided that $\nabla F(x)$ is a positive semidefinite matrix for all $x \in C$.*

Now we can formulate a line search method for solving problem EP.

A Line Search Algorithm

Data: Let $x^0 \in C$ and let $g(x) = \sup_{y\in C}\{-f(x,y) - \hbar(x,y)\}$.

Step 1. Set $k = 0$.

Step 2. Find $y(x^k)$ the solution of the optimization problem:

$$\min_{y\in C} \{\, f(x^k, y) + \hbar(x^k, y)\,\},$$

and set $d^k = y(x^k) - x^k$.

Step 3. Find t_k the solution of the line search problem:

$$\min_{t\in[0,1]} g(x^k + t\, d^k),$$

and set $x^{k+1} = x^k + t_k\, d^k$

Step 4. If $x^{k+1} = x^k$, then Stop: x^k is a solution to problem EP.
Replace k by $k+1$, and go to Step 2.

Let us note that the line search is carried out on the segments $[x^k, y(x^k)]$ which are included in C. So $x^k \in C$ for all k. The next theorem gives the convergence of this algorithm.

Theorem 2.15. ([62], Theorem 5.2) *Suppose that C is compact, that $f(x,\cdot)$ is convex for each $x \in C$, that f is differentiable with respect to x, and that $\nabla_x f$ is continuous on $C \times C$. If* (2.11) *is satisfied, then, for any starting point $x^0 \in C$, the sequence $\{x^k\}$ generated by the line search algorithm is contained in C, and each limit point of $\{x^k\}$ is a solution to problem EP.*

Finally, it should be pointed out that the gap function can also be used to check the accuracy of an approximate solution of an equilibrium problem. This strategy has been used by Konnov [46] to determine an approximate solution of the subproblems that must be solved in the Proximal Point algorithm. More precisely, given the iterate $\bar{x}^k \in C$ and an algorithm for solving subproblem PEP, namely, find $x \in C$ such that

$$f(x,y) + \frac{1}{c}\,\langle x - \bar{x}^k,\, y - x\rangle \geq 0 \text{ for all } y \in C,$$

the question is to decide when the algorithm must be stopped to get $\bar{x}^{k+1} \in C$ such that $\|\bar{x}^{k+1} - x^{k+1}\| \leq \epsilon_{k+1}$ where x^{k+1} is the exact solution of problem PEP.

Using the gap function $g(x) = \sup_{y \in C}\{-f(x,y) - \frac{1}{c}\langle x - \bar{x}^k, y - x\rangle - \frac{1}{2}\|x-y\|^2\}$ associated with subproblem PEP, Konnov proved [46] that if $g(\bar{x}^{k+1}) \le \sigma \epsilon_{k+1}^2$, then $\|\bar{x}^{k+1} - x^{k+1}\| \le \epsilon_{k+1}$. Here σ is the strong monotonicity modulus of the function $f(x,y) + \frac{1}{c}\langle x - \bar{x}^k,\, y - x\rangle$.

2.2.5 Extragradient Methods

Let us first consider the important case of the variational inequality problem (VIP, for short): find $x^* \in C$ such that $\langle F(x^*),\, x - x^*\rangle \ge 0$ for all $x \in C$ where $F : C \to I\!R^n$. In that case, with $\hbar(x,y) = \frac{1}{2}\|y - x\|^2$, the subproblems considered in the Auxiliary Problem Principle algorithm become:

$$x^{k+1} = \arg\min_{y \in C} \{\langle F(x^k),\, y - x^k\rangle + \frac{1}{2\,c}\|y - x^k\|^2\}.$$

Using the optimality condition associated with this problem, it is easy to see that x^{k+1} is the orthogonal projection over C of the vector $x^k - c\,F(x^k)$. So the Auxiliary Problem Principle algorithm can be rewritten as:

Basic Projection Algorithm

Data: Let $x^0 \in C$ and $c > 0$.

Step 1. Set $k = 0$.

Step 2. Compute $x^{k+1} = P_C(x^k - c\,F(x^k))$.

Step 3. If $x^{k+1} = x^k$ then Stop: x^k is a solution to problem VIP.

Step 4. Replace k by $k + 1$ and go to Step 2.

The convergence of this algorithm is a direct consequence of Theorem 2.13. Indeed, condition (2.9) is satisfied when F is Lipschitz continuous (see Remark 2.3) and f is strongly monotone with modulus $\gamma > 0$ if F is strongly monotone with modulus γ, i.e.,

$$\langle F(x) - F(y),\, x - y\rangle \ge \gamma\|x - y\|^2 \text{ for all } x,\, y \in C.$$

Proposition 2.6. ([33], Theorem 12.1.2) *Let $F : C \to I\!R^n$. Suppose that F is strongly monotone with modulus $\gamma > 0$ and that F is Lipschitz continuous with constant $L > 0$. If $L^2\,c < 2\,\gamma$, then any sequence $\{x^k\}$ generated by the Basic Projection algorithm converges to the unique solution of problem VIP.*

However, this algorithm has two disadvantages: (i) it is converging under very strong assumptions, in general, not satisfied, and (ii) the two parameters L and γ are unknown in practice.

To overcome the first obstacle, Zhu and Marcotte [90] proved the convergence of the basic projection algorithm under the assumption that F is co-coercive on C with modulus $\nu > 0$, i.e.,

$$\langle F(x) - F(y), x - y \rangle \geq \nu \, \|F(x) - F(y)\|^2 \text{ for all } x, y \in C.$$

It is easy to see that if F is strongly monotone ($\gamma > 0$) and Lipschitz continuous ($L > 0$), then F is co-coercive with modulus γ/L^2. Let us also note that the co-coercivity of F implies the monotonicity and the Lipschitz continuity of F. The corresponding algorithm with a variable step c_k can be stated as follows.

Projection Algorithm with Variable Steps

Data: Let $x^0 \in C$ and let $\{c_k\}$ be a sequence of positive numbers.

Step 1. Set $k = 0$.

Step 2. Compute $x^{k+1} = P_C(x^k - c_k \, F(x^k))$.

Step 3. If $x^{k+1} = x^k$, then Stop: x^k is a solution to problem VIP.

Step 4. Replace k by $k + 1$ and go to Step 2.

The convergence is given in the next proposition.

Proposition 2.7. ([33], Theorem 12.1.8) *Let $F : C \to I\!R^n$ be co-coercive with modulus $\nu > 0$. If $0 < \inf_k c_k \leq \sup_k c_k < 2\,\nu$, then the sequence $\{x^k\}$ generated by the projection algorithm with variable steps is converging to a solution of problem VIP.*

To avoid the co-coercivity assumption, Korpelevich [50] proposed a projection-type algorithm that executes two projections per iteration. It is based on the following result: x^* is a solution to problem VIP if and only if $x^* = P_C[x^* - c\,F(P_C[x^* - c\,F(x^*)])]$ for any $t > 0$.

The corresponding fixed point algorithm called the extragradient algorithm is given as follows.

Extragradient Algorithm for VIP

Data: Let $x^0 \in C$ and $c > 0$.
Step 1. Let $k = 0$.
Step 2. Compute $y^k = P_C(x^k - c\,F(x^k))$. If $y^k = x^k$, then Stop: x^k is a solution to problem VIP.
Step 3. Compute $x^{k+1} = P_C(x^k - c\,F(y^k))$.
Step 4. Replace k by $k+1$ and go to Step 2.

Proposition 2.8. ([33], Theorem 12.1.11) *Let $F : C \to I\!R^n$ be pseudomonotone on C and Lipschitz continuous on C with constant $L > 0$. If $0 < c < 1/L$, then the sequence $\{x^k\}$ generated by the extragradient algorithm converges to a solution of problem VIP.*

The extragradient algorithm requires two projections per iteration, but the benefit is significant because it is applicable to the class of pseudomonotone variational inequality problems. However, this algorithm still requires the Lipschitz condition which plays a role in controlling the step $c > 0$.

One way not to use the Lipschitz constant L is to proceed as follows: Given $x^k \in C$, we first compute the projection $y^k = P_C(x^k - c\,F(x^k))$ and next we use a simple Armijo-type line search to get a point z^k on the segment $[x^k, y^k]$ such that the hyperplane $H^k = \{x \in I\!R^n \mid \langle F(z^k),\, x - z^k\rangle = 0\}$ strictly separates x^k from the solution set of problem VIP. Then finally we project x^k onto H^k to obtain the point w^k, and the resulting point w^k onto C to obtain x^{k+1}. Doing that, the point x^{k+1} is closer to the solution set than x^k.

Hyperplane Projection Algorithm

Data: Let $x^0 \in C$, $\alpha \in (0,1)$, $\theta \in (0,1)$, and $c > 0$.

Step 1. Set $k = 0$.

Step 2. Compute $y^k = P_C(x^k - c\,F(x^k))$.
If $y^k = x^k$, then Stop: x^k is a solution to problem VIP.

Step 3. Find the smallest nonnegative integer m such that

$$\langle F(z^{k,m}), x^k - y^k \rangle \geq \frac{\alpha}{c}\, \|y^k - x^k\|^2,$$

where $z^{k,m} = (1-\theta^m)x^k + \theta^m y^k$. Set $z^k = z^{k,m}$ and go to Step 4.

Step 4. Compute $w^k = x^k - \dfrac{\langle F(z^k), x^k - z^k \rangle}{\|F(z^k)\|^2}\, F(z^k)$ and set $x^{k+1} = P_C(w^k)$.

Step 5. Replace k by $k+1$ and go to Step 2.

The next proposition gives the convergence of this algorithm.

Proposition 2.9. ([33], Theorem 12.1.16) *Let $F : C \to I\!R^n$ be continuous and pseudomonotone. Then the sequence generated by the hyperplane projection algorithm converges to a solution of problem VIP.*

The extragradient algorithm and the hyperplane projection algorithm have been recently adapted by Konnov [48], and Quoc, Muu, and Nguyen [72] for solving the equilibrium problem. More precisely, these algorithms become:

Extragradient Algorithm for Problem EP

Data: Let $x^0 \in C$ and $c > 0$.

Step 1. Set $k = 0$.

Step 2. Find y^k the solution of the problem:

$$\min_{y \in C} \{ f(x^k, y) + \frac{1}{2\,c} \|y - x^k\|^2 \}.$$

If $y^k = x^k$, then Stop: x^k is a solution to EP.

Step 3. Find x^{k+1} the solution of the problem:

$$\min_{y \in C} \{ f(y^k, y) + \frac{1}{2\,c} \|y - x^k\|^2 \}.$$

Step 4. Replace k by $k + 1$ and go to Step 2.

The next proposition gives the convergence of this algorithm.

Proposition 2.10. *([72], Theorem 3.2) Let $f : C \times C \to \mathbb{R}$. Assume that f is lower semicontinuous on $C \times C$, $f(x, \cdot)$ is convex and subdifferentiable on C for each $x \in C$, and $f(\cdot, y)$ is upper semicontinuous for each $y \in C$. Assume also that there exist two positive constants d_1 and d_2 such that* (2.9) *holds. Then the sequence $\{x^k\}$ generated by the extragradient algorithm is bounded, and any limit point of $\{x^k\}$ is a solution to problems EP and DEP. If, in addition, $S^d = S^*$ (in particular, if f is pseudomonotone on $C \times C$), then the whole sequence $\{x^k\}$ converges to a solution of problem EP.*

Next, the hyperplane projection algorithm becomes:

Hyperplane Projection Algorithm for Problem EP

Data: Let $x^0 \in C$, $\theta \in (0,1)$, $\alpha \in (0,1)$, $c > 0$ and let $\{\gamma_k\}$ be a sequence of positive numbers.

Step 1. Set $k = 0$.

Step 2. Find y^k the solution of the problem:

$$\min_{y \in C} \{ f(x^k, y) + \frac{1}{2c} \|y - x^k\|^2 \}. \tag{2.12}$$

If $y^k = x^k$, then Stop: x^k is a solution to EP.

Step 3. Find the smallest nonnegative integer m such that

$$f(z^{k,m}, x^k) - f(z^{k,m}, y^k) \geq \frac{\alpha}{c} \|y^k - x^k\|^2, \tag{2.13}$$

where $z^{k,m} = (1 - \theta^m) x^k + \theta^m y^k$. Set $z^k = z^{k,m}$ and go to Step 4.

Step 4. Take any $g^k \in \partial_2 f(z^k, x^k)$ and compute

$$\sigma_k = \frac{f(z^k, x^k)}{\|g^k\|^2} \text{ and } x^{k+1} = P_C(x^k - \gamma_k \, \sigma_k \, g^k).$$

Step 5. Replace k by $k + 1$ and go to Step 2.

The convergence result of the algorithm is given as follows.

Theorem 2.16. ([72], Theorem 4.7) *Assume that f is continuous on $C \times C$ and that $f(x, \cdot)$ is convex and differentiable on C for each $x \in C$. Then the sequence $\{x^k\}$ generated by the hyperplane projection algorithm for problem EP is bounded, and any limit point of $\{x^k\}$ is a solution to problems DEP and EP. If, in addition, $S^d = S^*$ (in particular, if f is pseudomonotone on $C \times C$), then the whole sequence $\{x^k\}$ converges to a solution of problem EP.*

Let us mention that Konnov [48] has also proposed an hyperplane projection algorithm for solving problem EP when $f(x, \cdot)$ is differentiable on C for each $x \in C$. Noting by $g_k(y) = \nabla f(x^k, \cdot)(y)$ the gradient of $f(x^k, \cdot)$ at y, Konnov approximated the function $f(x^k, \cdot)$ in the subproblems (2.12) by its linearization at x^k, i.e., for all $y \in C$, by

$$f(x^k, y) \simeq f(x^k, x^k) + \langle g(x^k), \, y - x^k \rangle = \langle g(x^k), \, y - x^k \rangle.$$

Then the subproblems (2.12) become

$$\min_{y \in C} \{ \langle g(x^k),\, y - x^k \rangle + \frac{1}{2\,c} \|y - x^k\|^2 \},$$

and can be easily solved when C is a polyhedron.

Konnov [45] also considers the case when f is a convex-concave function. In that case, the function $f(x^k, \cdot)$ is approximated by a piecewise linear convex function.

These ideas will be generalized in Chapter 4 where $f(x^k, \cdot)$ is only convex and nonsmooth. In that chapter, $f(x^k, \cdot)$ will also be approximated by a piecewise linear convex function giving rise to the so-called bundle method.

2.2.6 Interior Proximal Point Algorithm

All the methods presented in the previous sections assume that solving constrained subproblems can be done efficiently. But it is well known that the boundary of constraints can destroy some of the nice properties of unconstrained methods (see a discussion about this in [12]). Another way to take account of inequality constraints is to use a barrier method. This type of method has been very often considered for solving constrained minimization problems giving rise to the well-known interior point algorithms [68].

In this context and when $\operatorname{int} C \neq \emptyset$, Auslender, Teboulle, and Ben-Tiba proposed in [9] a new type of interior proximal method for solving convex programs by replacing in the subproblems the quadratic term $\frac{1}{2}\|x^k - y\|^2$ by some nonlinear function $D(y, x^k)$ composed of two parts: the first part is based on entropic proximal terms and will play a role of barrier function forcing the iterates $\{x^k\}$ to remain in the interior of C. The second part is a quadratic convex regularization based on the set C to preserve the nice properties of the auxiliary problem principle. So the classical difficulties associated with the boundary of the constraints are automatically eliminated. This way to transform a constrained problem into an unconstrained one has already been used by Antipin [34] but with a distance-like function $D(y, x^k)$ based on Bregman functions. Let us recall that a Bregman distance is a function of the form

$$D(x, y) = \psi(x) - \psi(y) - \langle \psi'(y),\, x - y \rangle,$$

where ψ is a differentiable strictly convex function.

Another distance-like function is based on the logarithmic-quadratic function:

$$\varphi(t) = \mu\, h(t) + \frac{\nu}{2}(t-1)^2 \text{ for all } t > 0,$$

where $\nu > \mu > 0$ and h is defined by

$$h(t) = t - \log t - 1 \text{ for all } t > 0.$$

This function is a differentiable strongly convex function on $I\!R_{++}$ with modulus $\nu > 0$. Furthermore, the conjugate of φ can be given explicitly and satisfies the property of being self-concordant with parameter 2, i.e.,

$$(\varphi^*)'''(s) \leq 2\,(\varphi^*)''^{3/2} \text{ for all } s \in I\!R.$$

This property is very important to get polynomial algorithms [68].

Associated with φ, we can consider the $\varphi-$divergence proximal distance

$$d_\varphi(x,y) = \sum_{j=1}^{n} y_j^2\, \varphi(\frac{x_j}{y_j}) \text{ for all } x, y \in I\!R^n_{++},$$

which can be written using the definition of φ, as

$$d_\varphi(x,y) = \frac{\nu}{2}\|x-y\|^2 + \mu\, d_h(x,y) \text{ for all } x, y \in I\!R^n_{++}.$$

We easily observe that if $x_j \to 0$ and $y_j > 0$ is fixed, then $h(\frac{x_j}{y_j}) \to +\infty$, and consequently $d_h(x,y) \to +\infty$ when x tends to the boundary of $I\!R^n_{++}$. This is the typical behavior of a barrier function.

Let us illustrate these ideas on the particular problem of minimizing a convex function $F : C \to I\!R$ over the nonnegative orthant $I\!R^n_+$. Using the barrier function $d_h(x,y)$, the strategy is to replace, in the classical proximal point algorithm, the constrained subproblem

$$x^{k+1} = \arg\min_{y \in I\!R^n_+} \{\, F(y) + \frac{1}{2}\,\|y - x^k\|^2 \,\}$$

by the unconstrained problem

$$x^{k+1} = \arg\min_{y \in I\!R^n_{++}} \{\, F(y) + \frac{\nu}{2}\,\|y - x^k\|^2 + \mu\, d_h(y, x^k) \,\}.$$

It is easy to see that x^{k+1} is well-defined and belongs to the open set $I\!R^n_{++}$. For this reason, the method that uses these unconstrained subproblems for generating a sequence $\{x^k\}$, is called an interior proximal point method. Furthermore, as a consequence of Theorem 2.2 in [9], we have that this sequence $\{x^k\}$ converges to a minimum of F over C when such a minimum exists.

As for the classical proximal point algorithm, solving the subproblems is not easy when F is nonsmooth. In Chapter 5, we propose to approximate F by a piecewise linear convex function in such a way that the subproblems become more tractable. We prove that the convergence is preserved and we report some numerical results to illustrate the behavior of the new algorithm.

This strategy can also be used for solving problem EP when $C = I\!R^n_+$. In that case, the subproblems associated with the Auxiliary Problem Principle algorithm become

$$\min_{y \in I\!R^n_{++}} \{\, f(x^k, y) + \frac{\nu}{2}\|y - x^k\|^2 + \mu\, d_h(y, x^k)\,\}.$$

In Chapter 5 we study in details the algorithms corresponding to the proximal extragradient algorithm and to the hyperplane projection algorithm. We prove the convergence of the two algorithms and we report some numerical results to illustrate the behavior of these algorithms on an equilibrium problem.

Chapter 3

Bundle Proximal Methods

In this chapter, we present a bundle method for solving nonsmooth convex equilibrium problems based on the auxiliary problem principle. First, we consider a general algorithm that we prove to be convergent. Then we explain how to make this algorithm implementable. The strategy is to approximate the nonsmooth convex functions by piecewise linear convex functions in such a way that the subproblems are easy to solve and the convergence is preserved. In particular, we introduce a stopping criterion which is satisfied after finitely many iterations and which gives rise to $\Delta-$stationary points. Finally, we apply our implementable algorithm for solving the particular case of singlevalued and multivalued variational inequalities and we find again the results obtained recently by Salmon et al. [75].

3.1 Preliminaries

As explained in Section 2.2.3, the auxiliary problem principle is based on the following fixed point property: $x^* \in C$ is a solution to problem EP if and only if x^* is a solution to the problem

$$\min_{y \in C} \{c\, f(x^*, y) + h(y) - h(x^*) - \langle \nabla h(x^*), y \rangle\}, \tag{3.1}$$

where $c > 0$ and $h : C \to I\!\!R$ is a strongly convex differentiable function. Here the function $\hbar$ introduced in the auxiliary problem principle (in Section 2.2.3) has been chosen as: $\hbar(x, y) = h(y) - h(x) - \langle \nabla h(x), y - x \rangle$ for all $x, y \in C$. Then the corresponding fixed point iteration is: *Given $x^k \in C$, find $x^{k+1} \in C$ the solution of*

$$(\mathrm{P}_k) \quad \min_{y \in C} \{c\, f(x^k, y) + h(y) - h(x^k) - \langle \nabla h(x^k), y \rangle\}.$$

A typical example of function h is $h(x) = \frac{1}{2}\|x\|^2$ for all $x \in C$. With this function, problem (P_k) is equivalent to $\min_{y\in C}\{c\, f(x^k, y) + \frac{1}{2}\|x - y\|^2\}$. In Chapter 2, it was this problem which has been considered for the sake of simplicity.

Observe that problem (P_k) has a unique solution since h is strongly convex. This algorithm has been introduced by Mastroeni who proved its convergence in [62], Theorem 3.1 under the assumptions that f is strongly monotone and satisfies (2.9).

When

$$f(x, y) = \langle F(x), y - x\rangle + \varphi(y) - \varphi(x) \text{ for all } x, y \in C \tag{3.2}$$

with $F : C \to I\!R^n$ a continuous mapping and $\varphi : C \to I\!R$ a continuous convex function, problem EP is reduced to the generalized variational inequality problem (GVIP, for short):

$$\text{Find } x^* \in C \text{ such that, for all } y \in C, \quad \langle F(x^*), y - x^*\rangle + \varphi(y) - \varphi(x^*) \geq 0.$$

In that case, the auxiliary equilibrium problem principle algorithm becomes: *Given $x^k \in C$, find $x^{k+1} \in C$ the solution to the problem*

$$\min_{y\in C}\{\, c\,[\varphi(y) + \langle F(x^k), y - x^k\rangle] + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\,\}. \tag{3.3}$$

It is easy to see that f is strongly monotone and condition (2.9) is satisfied when F is strongly monotone and Lipschitz continuous, respectively.

However these assumptions are very strong. In the case of problem GVIP, Zhu and Marcotte ([90], Theorem 3.2) proved that the sequence $\{x^k\}$ generated by the auxiliary problem principle converges to a solution when F is co-coercive on C in the sense that

$$\exists\, \gamma > 0 \quad \forall x, y \in C \quad \langle F(y) - F(x), y - x\rangle \geq \gamma\, \|F(y) - F(x)\|^2. \tag{3.4}$$

It is obvious that F co-coercive on C does not imply, in general, that the corresponding function f defined by (3.2) is strongly monotone (for instance, take $F = 0$ and observe that $f(x, y) + f(y, x) = 0$). So one of the aims of this chapter is to obtain the convergence of Mastroeni's algorithm under assumptions weaker than the strong monotonicity of f and (2.9) in such a way that Zhu and Marcotte's result can be derived as a particular case.

Concerning the implementation of the previous algorithm, the subproblems (P_k) can be difficult to solve when the convex function $f(x^k, \cdot)$ is nonsmooth. It is the case when f is given by (3.2) with φ a nonsmooth convex function. In that case, our strategy is to approximate the function $f(x^k, \cdot)$ by another convex function so that the subproblems (P_k) become easy to solve and the convergence is preserved under the same assumptions as in the exact case. The approximation will be done by using an extension of the bundle method developed in [75] for problem GVIP.

Let us mention that this strategy has been used by Konnov [44] at the lower level of a combined relaxation method for finding equilibrium points. More precisely, given $x^k \in C$, Konnov considers successive linearizations of the function $f(x^k, \cdot)$ in order to construct a convex piecewise linear approximation $\bar{f}_k$ of $f(x^k, \cdot)$ such that the solution y^k of subproblem (P_k) with $f(x^k, \cdot)$ replaced by $\bar{f}_k$ satisfies the property:

$$f(x^k, y^k) \leq \mu \bar{f}_k(y^k) \quad (0 < \mu < 1). \tag{3.5}$$

Then this solution y^k is used to compute a direction g^k in the subdifferential of the function $-f(\cdot, y^k)$ at x^k, and a steplength $\sigma_k = f(x^k, y^k)/\|g^k\|^2$ (if $g^k \neq 0$). Finally the next iterate x^{k+1} is defined as the projection over C of the vector $x^k - \gamma_k \sigma_k g^k$ where $0 < \gamma_k < 2$. Observe that this step is well defined when $f(\cdot, y)$ is concave on C for all $y \in C$. In this chapter we do not assume this property, so we do not consider Konnov's projection step and instead of this step, we set $x^{k+1} = y^k$. In other terms, our method is simply an implementable version of Mastroeni's auxiliary problem principle.

To summarize our approach, first we study the convergence of the algorithm when $f(x^k, \cdot)$ is approximated from below by any function $\bar{f}_k$ which satisfies the inequality (3.5) and then we present an implementable method to construct a broad class of convex piecewise linear functions $\bar{f}_k$ approximating $f(x^k, \cdot)$. An advantage of our approach is that it allows to limit the size of the bundle used to obtain $\bar{f}_k$.

Another way for solving problem EP is to transform it into a variational inequality problem (see [43], Thm 2.1.2) and to use a bundle type method for solving this equivalent problem. This method is interesting when C is compact because in that case there exist efficient variants of the bundle method allowing to obtain a complexity analysis. In these methods the level sets of the piecewise linear models are used to construct the successive iterates (see [37] and [54] for more details). This approach has been used by Gol'shtein [36] for solving problem EP when C is compact and $f(x, \cdot)$ satisfies a Lipschitz condition with a constant L independent on x. These conditions can be taken into account by our convergence theory.

Finally to show the interest of our general algorithm, first we apply it to problem GVIP with the purpose to find again the convergence theorem obtained in [75]. Then we consider the following multivalued variational inequality problem (MVIP, for short):

$$\text{Find } x^* \in C \text{ and } r^* \in F(x^*) \text{ such that, for all } y \in C, \quad \langle r^*, y - x^* \rangle \geq 0,$$

where C is a nonempty closed convex subset of $\mathbb{R}^n$ and $F : C \to 2^{\mathbb{R}^n}$ is a co-coercive continuous multivalued mapping with compact values. This problem is a particular instance of problem EP when the function f is defined, for all $x, y \in C$, by

$$f(x, y) = \sup_{\xi \in F(x)} \langle \xi, y - x \rangle.$$

For this problem, we use a very simple approximating function and we derive a convergence result from our general theory.

The chapter is organized as follows. In section 2, we consider a general algorithm for solving problem EP where the convex function $f(x^k, \cdot)$ is approximated, and we prove that it is convergent to a solution to problem EP. In section 3, we present an implementable version of this general algorithm by using a bundle strategy. In particular, we introduce a stopping criterion and we study the convergence properties of the resulting algorithm. Finally, in section 4, first we find again the convergence results obtained in [75] for problem GVIP and then we present a realization of the general algorithm for solving problem MVIP.

3.2 Proximal Algorithm

From now on, we impose that the gradient ∇h is Lipschitz continuous on C with constant $\Lambda > 0$. We also denote by $\beta > 0$ the modulus of the strongly convex function h. In this section, we consider the general equilibrium problem EP and the algorithm introduced by Mastroeni for solving it where the parameter $c = c_k > 0$ is allowed to vary at each iteration. This algorithm can be expressed as follows: *Given $x^k \in C$, find $x^{k+1} \in C$ the solution to problem (P_k).*

As explained before in Section 1, the function $f(x^k, \cdot)$, denoted f_k in the sequel, is replaced in problem (P_k) by another convex function $\bar{f}_k$ in such a way that the new problem

$$(\bar{P}_k) \quad \min_{y \in C} \{ c_k \bar{f}_k(y) + h(y) - h(x^k) - \langle \nabla h(x^k), y \rangle \}$$

is easier to solve and that the corresponding algorithm:

$$\textit{Given } x^k \in C, \textit{ find } x^{k+1} \in C \textit{ the solution to problem } (\bar{P}_k)$$

generates a sequence $\{x^k\}$ converging to some solution to problem EP.

To obtain the convergence of this algorithm, we introduce some conditions on the approximating function $\bar{f}_k$.

Definition 3.1. *Let $\mu \in (0, 1]$ and $x^k \in C$. A convex function $\bar{f}_k : C \to I\!R$ is a $\mu-$approximation of f_k at x^k if $\bar{f}_k \leq f_k$ on C and if*

$$f_k(y^k) \leq \mu \bar{f}_k(y^k), \tag{3.6}$$

where y^k is the unique solution to problem $(\bar{P}_k)$.

Since $f_k(x^k) = 0$, and $\bar{f}_k(x^k) \leq f_k(x^k)$, inequality (3.6) implies that $f_k(x^k) - f_k(y^k) \geq \mu[\bar{f}_k(x^k) - \bar{f}_k(y^k)]$, i.e., that the reduction on f_k is greater than a fraction of the reduction obtained by using the approximating function $\bar{f}_k$. This is motivated by the fact that, at iteration k, the objective is to minimize the function f_k (see (2.8)). Moreover, we observe that $\bar{f}_k = f_k$ is a 1-approximation of f_k at x^k.

Using this definition, the approximate auxiliary equilibrium principle algorithm can be expressed as follows:

Proximal Algorithm

> Data: Let $x^0 \in C$ and $\mu \in (0, 1]$.
>
> Step 1. Set $k = 0$.
>
> Step 2. Find $\bar{f}_k$ a $\mu-$approximation of f_k at x^k and denote by x^{k+1} the unique solution to problem $(\bar{P}_k)$.
>
> Step 3. Replace k by $k + 1$ and go to Step 2.

The convergence of this general algorithm is established in two steps. First we examine the convergence of the algorithm when the sequence $\{x^k\}$ is bounded and $\|x^{k+1} - x^k\| \to 0$. Then in a second theorem, we give conditions to obtain that these two properties are satisfied.

Theorem 3.1. *Suppose that $c_k \geq \underline{c} > 0$ for all $k \in I\!N$. If the sequence $\{x^k\}$ generated by the proximal algorithm is bounded and is such that $\|x^{k+1} - x^k\| \to 0$, $k \in I\!N$, then every limit point of $\{x^k\}_{k \in I\!N}$ is a solution to problem EP.*

Proof. Let x^* be a limit point of $\{x^k\}_{k\in I\!N}$ and let $\{x^k\}_{k\in K\subset I\!N}$ be some subsequence converging to x^*. Since $\|x^{k+1}-x^k\| \to 0$, we also have $\{x^{k+1}\}_{k\in K} \to x^*$. Hence, as $\bar{f}_k \le f_k$ and $f_k(x^{k+1}) \le \mu \bar{f}_k(x^{k+1})$, we obtain

$$\frac{1}{\mu} f_k(x^{k+1}) \le \bar{f}_k(x^{k+1}) \le f_k(x^{k+1}).$$

Now $f_k(x^{k+1}) = f(x^k, x^{k+1}) \to f(x^*, x^*) = 0$ for $k \to +\infty$ because $x^k \to x^*$, $x^{k+1} \to x^*$ for $k \to +\infty$, $k \in K$, and f is continuous. Hence $\bar{f}_k(x^{k+1}) \to 0$ for $k \to +\infty$. On the other hand, since x^{k+1} solves the convex optimization problem $(\bar{P}_k)$, we have

$$0 \in \partial\{\, c_k(\bar{f}_k + \psi_C)\}(x^{k+1}) - \nabla h(x^k) + \nabla h(x^{k+1}),$$

i.e.,

$$\nabla h(x^k) - \nabla h(x^{k+1}) \in \partial\{\, c_k(\bar{f}_k + \psi_C)\}(x^{k+1}),$$

where ψ_C denotes the indicator function associated with C ($\psi_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise). Using the definition of the subdifferential, we obtain

$$\forall y \in C \quad \bar{f}_k(y) - \bar{f}_k(x^{k+1}) \ge \frac{1}{c_k}\langle \nabla h(x^k) - \nabla h(x^{k+1}), y - x^{k+1}\rangle. \tag{3.7}$$

Applying the Cauchy-Schwarz inequality and the properties $\bar{f}_k \le f_k$ and ∇h is Lipschitz continuous on C with constant $\Lambda > 0$, we obtain successively for all $y \in C$,

$$\begin{aligned} f_k(y) - \bar{f}_k(x^{k+1}) &\ge -\frac{1}{c_k}\|\nabla h(x^k) - \nabla h(x^{k+1})\|\,\|y - x^{k+1}\| \\ &\ge -\frac{\Lambda}{c_k}\|x^k - x^{k+1}\|\,\|y - x^{k+1}\|. \end{aligned}$$

Taking the limit on $k \in K$, we deduce

$$\forall y \in C \quad f(x^*, y) \ge 0,$$

because f is continuous, $\bar{f}_k(x^{k+1}) \to 0$, $\|x^k - x^{k+1}\| \to 0$, $\|y - x^{k+1}\| \to \|y - x^*\|$ and $c_k \ge \underline{c} > 0$. But this means that x^* is a solution to problem EP. □

In the next theorem, we give conditions to obtain that the sequence $\{x^k\}$ is bounded and that $\|x^{k+1} - x^k\| \to 0$.

Theorem 3.2. *Suppose that there exist* $\gamma,\ d_1,\ d_2 > 0$ *and a nonnegative function* $g : C \times C \to I\!R$ *such that for all* $x, y, z \in C$,

$$(i) \quad f(x,y) \geq 0 \Rightarrow f(y,x) \leq -\gamma\, g(x,y);$$

$$(ii) \quad f(x,z) - f(y,z) - f(x,y) \leq d_1\, g(x,y) + d_2\, \|z-y\|^2.$$

If the sequence $\{c_k\}_{k \in I\!N}$ *is nonincreasing and* $c_k < \dfrac{\beta\mu}{2\, d_2}$ *for all* k *and if* $\dfrac{d_1}{\gamma} \leq \mu \leq 1$, *then the sequence* $\{x^k\}_{k\in I\!N}$ *generated by the proximal algorithm is bounded and* $\lim_{k\to+\infty} \|x^{k+1} - x^k\| = 0$.

Proof. Let x^* be a solution to problem EP and consider for each $k \in I\!N$ the Lyapounov function $\Gamma^k : C \times C \to I\!R$ defined for all $y, z \in C$, by

$$\Gamma^k(y,z) = h(z) - h(y) - \langle \nabla h(y), z-y\rangle + \frac{c_k}{\mu} f(z,y). \tag{3.8}$$

Since h is strongly convex with modulus $\beta > 0$, we have immediately that, for all $x \in C$,

$$\Gamma^k(x^k, x^*) \geq \frac{\beta}{2}\|x^k - x^*\|^2. \tag{3.9}$$

Noticing that $c_{k+1} \leq c_k$ for all $k \in I\!N$, the difference $\Gamma^{k+1}(x^k, x^*) - \Gamma^k(x^k, x^*)$ can then be evaluated as follows:

$$\begin{aligned}
\Gamma^{k+1}(x^{k+1}, x^*) - \Gamma^k(x^k, x^*) &\leq h(x^k) - h(x^{k+1}) + \langle \nabla h(x^k), x^{k+1} - x^k\rangle \\
&\quad + \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^* - x^{k+1}\rangle \\
&\quad + \frac{c_k}{\mu}\{f(x^*, x^{k+1}) - f(x^*, x^k)\} \\
&= s_1 + s_2 + s_3,
\end{aligned}$$

with

$$\begin{aligned}
s_1 &= h(x^k) - h(x^{k+1}) + \langle \nabla h(x^k), x^{k+1} - x^k\rangle, \\
s_2 &= \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^* - x^{k+1}\rangle, \\
s_3 &= \frac{c_k}{\mu}\{f(x^*, x^{k+1}) - f(x^*, x^k)\}.
\end{aligned}$$

For s_1, we easily derive from the strong convexity of h that

$$s_1 \leq -\frac{\beta}{2}\|x^{k+1} - x^k\|^2. \tag{3.10}$$

For s_2, we obtain, taking $y = x^*$ in (3.7)

$$s_2 = \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^* - x^{k+1}\rangle \leq c_k\{\bar{f}_k(x^*) - \bar{f}_k(x^{k+1})\} \leq c_k\{f(x^k, x^*) - \tfrac{1}{\mu} f(x^k, x^{k+1})\},$$

because $\bar{f}_k \leq f(x^k, \cdot)$ and (3.6) hold. Then, using assumption (ii), we deduce that

$$\begin{aligned} s_2 + s_3 &\leq c_k\{f(x^k, x^*) - \frac{1}{\mu} f(x^k, x^{k+1})\} + \frac{c_k}{\mu}\{f(x^*, x^{k+1}) - f(x^*, x^k)\} \\ &= \frac{c_k}{\mu}\{f(x^*, x^{k+1}) - f(x^*, x^k) - f(x^k, x^{k+1})\} + c_k f(x^k, x^*) \\ &\leq \frac{c_k}{\mu}\{d_1\, g(x^*, x^k) + d_2\, \|x^{k+1} - x^k\|^2\} + c_k f(x^k, x^*). \end{aligned}$$

Consequently, we have

$$\begin{aligned} \Gamma^{k+1}(x^{k+1}, x^*) - \Gamma^k(x^k, x^*) \leq\ & \epsilon_k f(x^k, x^*) - \frac{1}{2}(\beta - 2\frac{c_k d_2}{\mu})\|x^{k+1} - x^k\|^2 \\ & + \frac{c_k d_1}{\mu} g(x^*, x^k). \end{aligned}$$

Applying assumption (i) with $x = x^*$ and $y = x^k$, since $f(x^*, x^k) \geq 0$, we obtain

$$f(x^k, x^*) \leq -\gamma\, g(x^*, x^k).$$

Finally, we have that

$$\Gamma^{k+1}(x^{k+1}, x^*) - \Gamma^k(x^k, x^*) \leq -\frac{1}{2}(\beta - 2\frac{c_k\, d_2}{\mu})\|x^{k+1} - x^k\|^2 - c_k(\gamma - \frac{d_1}{\mu}) g(x^*, x^k). \quad (3.11)$$

Since $c_k < \dfrac{\beta\, \mu}{2\, d_2}$ for all k and $\mu \geq \dfrac{d_1}{\gamma}$, from (3.9) and (3.11), it follows that $\{\Gamma^k(x^k, x^*)\}_{k\in I\!N}$ is a nonincreasing sequence bounded below by 0. Hence, it is convergent in $I\!R$. Using again (3.9), we deduce that the sequence $\{x^k\}_{k\in I\!N}$ is bounded and, passing to the limit in (3.11), that the sequence $\{\|x^{k+1} - x^k\|\}_{k\in I\!N}$ converges to zero. □

Combining Theorems 3.1 and 3.2, we deduce the following theorem.

Theorem 3.3. *Suppose that $c_k \geq \underline{c} > 0$ for all $k \in I\!N$ and that all assumptions of Theorem 3.2 are fulfilled, then the sequence $\{x^k\}_{k\in I\!N}$ generated by the proximal algorithm converges to a solution to problem EP.*

Remark 3.1. *The same result as Theorem 3.2 can also be obtained when condition (ii) is replaced by the following condition:*

(iii) $$f(x,z) - f(y,z) - f(x,y) \le d_1\, g(x,y) + d_2\, \|z - y\|,$$

and when the series $\sum_{k=0}^{+\infty} (c_k)^2$ *is convergent. If, in addition,* $g(x,y) = 0$ *and* $\sum_{k=0}^{+\infty} c_k = +\infty$*, then the convergence of the sequence* $\{x^k\}$ *to a solution to problem EP can be proved as in [75] by using the gap function* $l(x) = -f(x, x^*)$ *where* x^* *is a solution to EP.*

So in order to obtain the convergence of the proximal algorithm, we need conditions (i) and (ii) or conditions (i) and (iii). Condition (i) is a monotonicity condition. Indeed, when $g = 0$, this condition means that f is pseudomonotone and when $g(x,y) = \|x - y\|^2$ that f is strongly pseudomonotone with modulus γ. Conditions (ii) and (iii) are Lipschitz-type conditions. The link between conditions (i) and (ii) or (iii) is made by the function g whose choice depends on the structure of the problem. So, for example, when $f(x,y) = \varphi(x) - \varphi(y)$ with $\varphi : C \to I\!R$ a continuous convex function, i.e., when problem (EP) is a constrained convex optimization problem, it suffices to choose $g(x,y) = 0$ for all $x, y \in C$ to obtain that (i), (ii) and (iii) are satisfied.

Other sufficient conditions to get conditions $(i), (ii)$ and (iii) are given in the next two propositions.

Proposition 3.1. *If* f *is pseudomonotone and* $f(x,\cdot)$ *is Lipschitz continuous on* C *uniformly in* x*, then conditions* (i) *and* (iii) *are satisfied with* $g(x,y) = 0$*.*

Proof. Let $x, y, z \in C$. Since $f(y,y) = 0$, we have

$$\begin{aligned} f(x,z) - f(y,z) - f(x,y) &= f(x,z) - f(x,y) + f(y,y) - f(y,z) \\ &\le 2L\|z - y\|, \end{aligned}$$

where L denotes the Lipschitz constant of $f(x,\cdot)$. □

Proposition 3.2. *If* f *is strongly monotone and if* (2.9) *holds, then conditions* (i) *and* (ii) *are satisfied with* $g(x,y) = \|x - y\|^2$*.*

Proof. If $f(x,y) \ge 0$, then by the strong monotonicity of f, we have

$$f(y,x) \le -f(x,y) - \gamma\|x - y\|^2 \le -\gamma\|x - y\|^2 = -\gamma\, g(x,y).$$

Condition (ii) is immediate from (2.9). □

As a consequence of this proposition, Theorem 3.3 is also valid under assumptions the strong monotonicity of f and (2.9). In particular, when $\mu = 1$, the conditions imposed on the parameters are $c_k < \frac{\beta}{2\,d_2}$ for all k and $\frac{d_1}{\gamma} \leq 1$, and Theorem 3.1 of Mastroeni [62] is recovered. So, when $\mu = 1$, Theorem 3.3 can be considered as a generalization of this theorem.

Finally, we consider the case where f is given by (3.2) and we introduce the following definition: F is $\varphi-$co-coercive on C if there exists $\gamma > 0$ such that for all $x, y \in C$, if $\langle F(x), y - x\rangle + \varphi(y) - \varphi(x) \geq 0$ holds, then

$$\langle F(y), y - x\rangle + \varphi(y) - \varphi(x) \geq \gamma \, \|F(y) - F(x)\|^2. \tag{3.12}$$

It is easy to prove that if F is co-coercive on C, then F is φ-co-coercive on C. Indeed, if F is co-coercive on C, then there exists $\gamma > 0$ such that

$$\forall x, y \in C \quad \langle F(x) - F(y), x - y\rangle \geq \gamma \, \|F(x) - F(y)\|^2.$$

But then, if $\langle F(x), y - x\rangle + \varphi(y) - \varphi(x) \geq 0$, we have

$$\begin{aligned}
\langle F(y), y - x\rangle + \varphi(y) - \varphi(x) &= \langle F(y) - F(x), y - x\rangle + \langle F(x), y - x\rangle \\
&+ \varphi(y) - \varphi(x) \\
&\geq \gamma \, \|F(y) - F(x)\|^2,
\end{aligned}$$

i.e., inequality (3.12).

Now in order to find again Zhu and Marcotte's convergence result ([90] Theorem 3.2) from our Theorem 3.3, we need the following proposition where another choice of g is necessary to obtain (i) and (ii).

Proposition 3.3. *Let $f(x, y) = \langle F(x), y - x\rangle + \varphi(y) - \varphi(x)$ where $F : C \to I\!R^n$ is continuous and $\varphi : C \to I\!R$ is convex. If F is $\varphi-$co-coercive on C, then there exist a nonnegative function $g : C \times C \to I\!R$ and $\gamma > 0$ such that for all $x, y, z \in C$ and for all $\nu > 0$,*

$$f(x, y) \geq 0 \Rightarrow f(y, x) \leq -\gamma \, g(x, y),$$

$$f(x, z) - f(y, z) - f(x, y) \leq \frac{1}{2\nu} \, g(x, y) + \frac{\nu}{2} \, \|z - y\|^2.$$

Proof. Using the definition of f and the $\varphi-$co-coercivity of F on C, there exists $\gamma > 0$ such that for all $x \in C$

$$f(x, y) \geq 0 \Rightarrow f(y, x) \leq -\gamma \|F(y) - F(x)\|^2.$$

On the other hand, we have for any $\nu > 0$,

$$f(x,z) - f(y,z) - f(x,y) = \langle F(x) - F(y), z - y\rangle \leq \frac{1}{2\nu}\|F(x) - F(y)\|^2 + \frac{\nu}{2}\|z - y\|^2.$$

So, with $g(x,y) = \|F(y) - F(x)\|^2$, we obtain the two inequalities. □

Using this proposition, Theorem 3.2 of [90] can be derived from our Theorem 3.3 with $\mu = 1$. Indeed, by choosing $\nu = \frac{1}{2\gamma}$, we obtain $d_1 = \frac{1}{2\nu} = \gamma$ and $d_2 = \frac{\nu}{2} = \frac{1}{4\gamma}$. Then conditions $\frac{d_1}{\gamma} \leq 1$ and $c_k < \frac{\beta}{2d_2}$ of Theorem 3.3 reduce to $c_k < 2\beta\gamma$, which is exactly the condition imposed by Zhu and Marcotte in their convergence theorem.

3.3 Bundle Proximal Algorithm

In order to obtain an implementable algorithm, we have now to say how to construct a $\mu-$approximation $\bar{f}_k$ of f_k at x^k such that problem $(\bar{P}_k)$ is easier to solve than problem (P_k). Here we assume that $\mu \in (0,1)$. In that purpose, we observe that if $\bar{f}_k$ is a piecewise linear convex function of the form

$$\bar{f}_k(y) = \max_{1\leq j\leq p}\{a_j^T y + b_j\},$$

where $a_j \in I\!R^n,\ b_j \in I\!R$ for $j = 1,\ldots,p$, the problem $(\bar{P}_k)$ is equivalent to the problem

$$(\text{QP}_k) \quad \begin{cases} \min & \{c_k\, v + h(y) - h(x^k) - \langle\nabla h(x^k), y - x^k\rangle\} \\ \text{s.t.} & v \geq a_j^T y + b_j,\ j = 1,\ldots,p \\ & y \in C. \end{cases}$$

When h is the squared norm and C is a closed convex polyhedron, this problem becomes quadratic.

There exist many efficient numerical methods for solving such a problem. When $\bar{f}_k$ is a piecewise linear convex function, it is judicious to construct $\bar{f}_k$, piece by piece, by generating successive models

$$\bar{f}_k^i,\ i = 1,2,\ldots$$

until (if possible) $\bar{f}_k^{i_k}$ is a $\mu-$approximation of f_k at x^k for some $i_k \geq 1$. For $i = 1,2,\ldots,$ we denote by y_k^i the unique solution to the problem

$$(P_k^i) \quad \min_{y\in C}\{c_k\bar{f}_k^i(y) + h(y) - h(x^k) - \langle\nabla h(x^k), y\rangle\},$$

and we set $\bar{f}_k = \bar{f}_k^{i_k}$ and $x^{k+1} = y_k^{i_k}$.

In order to obtain a $\mu-$approximation $\bar{f}_k^{i_k}$ of f_k at x^k, we have to impose some conditions on the successive models $\bar{f}_k^i$, $i = 1, 2, \dots$. However, before presenting them, we need to define the affine functions l_k^i, $i = 1, 2, \dots$ by

$$l_k^i(y) = \bar{f}_k^i(y_k^i) + \langle \gamma_k^i, y - y_k^i \rangle \text{ for all } y \in C,$$

where $\gamma_k^i = \dfrac{1}{c_k}[\nabla h(x^k) - \nabla h(y_k^i)]$. By optimality of y_k^i, we have

$$\gamma_k^i \in \partial(\bar{f}_k^i + \psi_C)(y_k^i). \tag{3.13}$$

It is then easy to observe that

$$l_k^i(y_k^i) = \bar{f}_k^i(y_k^i) \quad \text{and} \quad l_k^i(y) \leq \bar{f}_k^i(y) \text{ for all } y \in C. \tag{3.14}$$

Now we assume that the following conditions inspired for [28] are satisfied by the convex models $\bar{f}_k^i$,

$(C1)$ $\bar{f}_k^i \leq f_k$ on C for $i = 1, 2, \dots$

$(C2)$ $\bar{f}_k^{i+1} \geq f_k(y_k^i) + \langle s(y_k^i), \cdot - y_k^i \rangle$ on C for $i = 1, 2, \dots$

$(C3)$ $\bar{f}_k^{i+1} \geq l_k^i$ on C for $i = 1, 2, \dots,$

where $s(y_k^i)$ denotes the subgradient of f_k available at y_k^i.

Several models fulfill these conditions. For example, for the first model $\bar{f}_k^1$, we can take the linear function

$$\bar{f}_k^1(y) = f_k(x^k) + \langle s(x^k), y - x^k \rangle \text{ for all } y \in C.$$

Since $s(x^k) \in \partial f_k(x^k)$, condition $(C1)$ is satisfied for $i = 1$. For the next models $\bar{f}_k^i$, $i = 2, \dots,$ there exist several possibilities. A first example is to take for $i = 1, 2, \dots$

$$\bar{f}_k^{i+1}(y) = \max\{l_k^i(y), f_k(y_k^i) + \langle s(y_k^i), y - y_k^i \rangle\}. \tag{3.15}$$

Conditions $(C2), (C3)$ are obviously satisfied and condition $(C1)$ is also satisfied for $i = 2, 3, \dots,$ because each linear piece of these functions are below f_k. Another example is to take for $i = 1, 2, \dots$

$$\bar{f}_k^{i+1}(y) = \max_{0 \leq j \leq i} \{f_k(y_k^j) + \langle s(y_k^j), y - y_k^j \rangle\} \tag{3.16}$$

where $y_k^0 = x^k$. Since $s(y_k^j) \in \partial f_k(y_k^j)$ for $j = 0, \dots, i$ and since $\bar{f}_k^{i+1} \geq \bar{f}_k^i \geq l_k^i$, it is easy to see that conditions $(C1) - (C3)$ are satisfied.

Comparing (3.15) and (3.16), we can say that l_k^i plays the same role as the i linear functions $f_k(y_k^j) + \langle s(y_k^j), y - y_k^j \rangle,\ j = 0, \ldots, i-1$. It is the reason why this function l_k^i is called the aggregate affine function (see, e.g., [28]). The first example (3.15) is interesting from the numerical point of view, because its use allows to limit the number of linear constraints in subproblems (QP_k).

Now the algorithm allowing to pass from x^k to x^{k+1}, i.e., to make what is called a serious step, can be expressed as follows.

Serious Step Algorithm

Data: Let $x^k \in C$ and $\mu \in (0, 1)$.

Step 1. Set $i = 1$.

Step 2. Choose $\bar{f}_k^i$ a convex function that satisfies $(C1) - (C3)$ and solve problem (P_k^i) to get y_k^i.

Step 3. If $f_k(y_k^i) \leq \mu \bar{f}_k^i(y_k^i)$, then set $x^{k+1} = y_k^i$, $i_k = i$ and Stop: x^{k+1} is a serious step.

Step 4. Replace i by $i + 1$ and go to Step 2.

Our aim is now to prove that if x^k is not a solution to problem EP and if the models $\bar{f}_k^i, i = 1, \ldots$ satisfy $(C1) - (C3)$, then there exists $i_k \geq 1$ such that $\bar{f}_k^{i_k}$ is a $\mu-$approximation of f_k at x^k, i.e., that the Stop occurs at Step 2 after finitely many iterations.

To prove that, we need a lemma whose proof uses the following functions:

$$\tilde{l}_k^i(y) = l_k^i(y) + \frac{1}{c_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle\},$$
$$\tilde{f}_k^i(y) = \bar{f}_k^i(y) + \frac{1}{c_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle\}.$$

Using (3.13) and (3.14), we obtain:

$$
\begin{aligned}
\tilde{l}_k^i(y) - \tilde{l}_k^i(y_k^i) &= l_k^i(y) + \frac{1}{c_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle\} \\
&\quad - l_k^i(y_k^i) - \frac{1}{c_k}\{h(y_k^i) - h(x^k) - \langle \nabla h(x^k), y_k^i - x^k \rangle\} \\
&= \bar{f}_k^i(y_k^i) + \langle \gamma_k^i, y - y_k^i \rangle \\
&\quad + \frac{1}{c_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle\} \\
&\quad - \bar{f}_k^i(y_k^i) - \frac{1}{c_k}\{h(y_k^i) - h(x^k) - \langle \nabla h(x^k), y_k^i - x^k \rangle\} \\
&= \frac{1}{c_k}\{h(y) - h(y_k^i) - \langle \nabla h(y_k^i), y - y_k^i \rangle\}.
\end{aligned}
$$

Consequently, we obtain

$$
\tilde{l}_k^i(y) = \tilde{l}_k^i(y_k^i) + \frac{1}{c_k}\{h(y) - h(y_k^i) - \langle \nabla h(y_k^i), y - y_k^i \rangle\}. \tag{3.17}
$$

Moreover from (3.14) and $(C3)$, we have

$$
\tilde{f}_k^i(x^k) = \bar{f}_k^i(x^k) \tag{3.18}
$$

$$
\tilde{l}_k^i(y_k^i) = \tilde{f}_k^i(y_k^i) \tag{3.19}
$$

$$
\tilde{l}_k^i \le \tilde{f}_k^{i+1} \quad \text{on } C. \tag{3.20}
$$

Lemma 3.1. *Suppose that the models $\bar{f}_k^i$, $i \in I\!N_0$ satisfy conditions $(C1) - (C3)$ and let, for each i, y_k^i be the unique solution to problem (P_k^i). Then*

$$
\begin{aligned}
&(i) \quad f_k(y_k^i) - \bar{f}_k^i(y_k^i) \to 0, \\
&(ii) \quad y_k^i \to \bar{y}_k \equiv \arg\min_{y \in C}\{c_k f_k(y) + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle\},
\end{aligned}
$$

where $i \to +\infty$.

Proof. (i) To obtain the first statement, we use the following three steps.

(1) The sequence $\{\tilde{l}_k^i(y_k^i)\}_{i \in I\!N_0}$ is convergent and $y_k^{i+1} - y_k^i \to 0$ when $i \to +\infty$.
For all i, we have

$$
\begin{array}{rcll}
0 = f_k(x^k) & \geq & \bar{f}_k^{i+1}(x^k) & \text{(by } C1) \\
& = & \tilde{f}_k^{i+1}(x^k) & \text{(by (3.18))} \\
& \geq & \tilde{f}_k^{i+1}(y_k^{i+1}) & \text{(by definition of } y_k^{i+1}) \\
& = & \tilde{l}_k^{i+1}(y_k^{i+1}) & \text{(by (3.19))} \\
& \geq & \tilde{l}_k^{i}(y_k^{i+1}) & \text{(by (3.20))} \\
& = & \tilde{l}_k^{i}(y_k^{i}) + \dfrac{1}{c_k} D_h(y_k^{i+1}, y_k^{i}) & \text{(by (3.17))} \\
& \geq & \tilde{l}_k^{i}(y_k^{i}) + \dfrac{\beta}{2\, c_k}\|y_k^{i+1} - y_k^{i}\|^2 & \text{(by strong convexity of } h \text{ on } C) \\
& \geq & \tilde{l}_k^{i}(y_k^{i})
\end{array}
$$

where $D_h(y,z) = h(y) - h(z) - \langle \nabla h(z), y - z \rangle$. From these relations, we have for all i, that

$$\tilde{l}_k^{i+1}(y_k^{i+1}) \geq \tilde{l}_k^{i}(y_k^{i}).$$

So, the sequence $\{\tilde{l}_k^{i}(y_k^{i})\}_{i \in I\!N_0}$ is nonincreasing and bounded above by 0. Consequently $\{\tilde{l}_k^{i}(y_k^{i})\}_{i \in I\!N_0}$ is convergent and $y_k^{i+1} - y_k^{i} \to 0$ when $i \to +\infty$.

(2) The sequence $\{y_k^{i}\}_{i \in I\!N_0}$ is bounded.
We have (for y fixed)

$$
\begin{aligned}
& f_k(y) + \frac{1}{c_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle\} \\
& \geq \bar{f}_k^{i+1}(y) + \frac{1}{c_k}\{h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle\} \quad \text{(by } C1) \\
& = \tilde{f}_k^{i+1}(y) \\
& \geq \tilde{l}_k^{i}(y) \quad \text{(by (3.20))} \\
& = \tilde{l}_k^{i}(y_k^{i}) + \frac{1}{c_k}\{h(y) - h(y_k^{i}) - \langle \nabla h(y_k^{i}), y - y_k^{i} \rangle\} \quad \text{(by (3.17))} \\
& \geq \tilde{l}_k^{i}(y_k^{i}) + \frac{\beta}{2\, c_k}\|y - y_k^{i}\|^2 \quad \text{(by } h \text{ is strongly convex on } C).
\end{aligned}
$$

Since the sequence $\{\tilde{l}_k^{i}(y_k^{i})\}_{i \in I\!N_0}$ is convergent, the sequence $\{y - y_k^{i}\}_{i \in I\!N_0}$ is bounded and thus the sequence $\{y_k^{i}\}_{i \in I\!N_0}$ is also bounded.

(3) $f_k(y_k^{i+1}) - \bar{f}_k^{i+1}(y_k^{i+1}) \to 0$.
We have successively

$$
\begin{array}{rcll}
\langle s(y_k^{i}), y_k^{i+1} - y_k^{i} \rangle & \leq & \bar{f}_k^{i+1}(y_k^{i+1}) - f_k(y_k^{i}) & \text{(by } C2) \\
& \leq & f_k(y_k^{i+1}) - f_k(y_k^{i}) & \text{(by } C1) \\
& \leq & \langle s(y_k^{i+1}), y_k^{i+1} - y_k^{i} \rangle & \text{(by definition of } s(y_k^{i+1})).
\end{array}
$$

Since $\{y_k^i\}_{i\in\mathbb{N}_0}$ is bounded, then, by Theorem 24.7 in [74], the set $\cup_i \partial f_k(y_k^i)$ is bounded and thus the sequence $\{s(y_k^i)\}_{i\in\mathbb{N}_0}$ is bounded. So, we obtain

$$\bar{f}_k^{i+1}(y_k^{i+1}) - f_k(y_k^i) \to 0 \text{ and } f_k(y_k^{i+1}) - f_k(y_k^i) \ \to\ 0,$$

and consequently,

$$f_k(y_k^{i+1}) - \bar{f}_k^{i+1}(y_k^{i+1}) = f_k(y_k^{i+1}) - f_k(y_k^i) + f_k(y_k^i) - \bar{f}_k^{i+1}(y_k^{i+1}) \ \to\ 0.$$

(ii) $y_k^i \to \bar{y}_k \equiv \arg\min_{y\in C}\{c_k f_k(y) + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k\rangle\}$.
Since the sequence $\{y_k^i\}_{i\in\mathbb{N}_0}$ is bounded, it remains to prove that every limit point y_k^* of this sequence is equal to $\bar{y}_k$, i.e., that

$$\frac{1}{c_k}\{\nabla h(x^k) - \nabla h(y_k^*)\} \in \partial(f_k + \psi_C)(y_k^*)$$

or, by definition of the subdifferential, we obtain from (3.13) and $(C1)$ that

$$\forall y \in C \quad f_k(y) \geq \bar{f}_k^i(y) \geq \bar{f}_k^i(y_k^i) + \frac{1}{c_k}\langle \nabla h(x^k) - \nabla h(y_k^i), y - y_k^i\rangle,$$

i.e.,

$$\begin{aligned}\forall y \in C \quad f_k(y) \ &\geq\ [\bar{f}_k^i(y_k^i) - f_k(y_k^i)] + [f_k(y_k^i) - f_k(y_k^*)] \\ &\quad + f_k(y_k^*) + \frac{1}{c_k}\langle \nabla h(x^k) - \nabla h(y_k^i), y - y_k^i\rangle.\end{aligned} \tag{3.21}$$

Since y_k^* is a limit point of $\{y_k^i\}_{i\in\mathbb{N}_0}$, there exists $K \subseteq \mathbb{N}_0$ such that

$$y_k^i \to y_k^* \text{ for } i \in K, i \to +\infty.$$

Taking the limit (for $i \in K$) of both sides of (3.21), we obtain, for all $y \in C$, that

$$\begin{aligned} f_k(y) \geq \lim_i [\bar{f}_k^i(y_k^i) - f_k(y_k^i)] \ &+\ \lim_i [f_k(y_k^i) - f_k(y_k^*)] + f_k(y_k^*) \\ &+\ \frac{1}{c_k}\lim_i \langle \nabla h(x^k) - \nabla h(y_k^i), y - y_k^i\rangle.\end{aligned}$$

Since $\lim_i [\bar{f}_k^i(y_k^i) - f_k(y_k^i)] = 0$ by (i), $\lim_i [f_k(y_k^i) - f_k(y_k^*)] = 0$ because f_k is continuous, and ∇h is continuous at y_k^*, we deduce that

$$f_k(y) \geq f_k(y_k^*) + \frac{1}{c_k}\langle \nabla h(x^k) - \nabla h(y_k^*), y - y_k^*\rangle \text{ for all } y \in C.$$

This completes the proof. □

Theorem 3.4. *Suppose x^k is not a solution to problem EP. Then the serious step algorithm stops after finitely many iterations i_k with $\bar{f}_k^{i_k}$ a $\mu-$approximation of f_k at x_k and with $x^{k+1} = y_k^{i_k}$.*

Proof. Suppose, to get a contradiction, that the Stop never occurs. Then

$$f_k(y_k^i) > \mu \bar{f}_k^i(y_k^i) \geq \mu \bar{f}_k^i(y_k^i) \quad \text{for all } i \in I\!N_0. \tag{3.22}$$

Moreover, by Lemma 3.1, $y_k^i \to \bar{y}_k$. Then taking the limit of both members of (3.22), we obtain

$$f_k(\bar{y}_k) \geq \mu f_k(\bar{y}_k)$$

because f_k is continuous over C and $f_k(y_k^i) - \bar{f}_k^i(y_k^i) \to 0$. Hence, since $\mu < 1$, we deduce that $f_k(\bar{y}_k) \geq 0$.

On the other hand, by definition of $\bar{y}_k$ (see Lemma 3.1), we have, for all $y \in C$, that

$$\begin{aligned} c_k f_k(\bar{y}_k) + h(\bar{y}_k) - h(x^k) - \langle \nabla h(x^k), \bar{y}_k - x^k \rangle \quad &\leq \quad c_k\, f_k(y) + h(y) \\ &- \quad h(x^k) - \langle \nabla h(x^k), y - x^k \rangle. \end{aligned}$$

If we choose $y = x^k$ and observe that $f_k(x^k) = 0$, then this inequality becomes

$$c_k f_k(\bar{y}_k) \leq -h(\bar{y}_k) + h(x^k) + \langle \nabla h(x^k), \bar{y}_k - x^k \rangle.$$

Finally, using the strong convexity of h yields

$$0 \leq \; c_k f_k(\bar{y}_k) \leq -\frac{\beta}{2} \|\bar{y}_k - x^k\|.$$

Consequently $\|\bar{y}_k - x^k\| = 0$ and thus $x^k = \bar{y}_k$. But this means that x^k is a solution to problem EP, which contradicts the assumption of the theorem. So the serious step algorithm stops after finitely many iterations. □

Incorporating the serious step algorithm in Step 2 of the proximal algorithm, we obtain the following algorithm.

Bundle Algorithm for problem EP

Data: Let $x^0 \in C$ and $\mu \in (0,1)$, and let $\{c_k\}_{k\in I\!N}$ be a sequence of positive numbers.

Step 1. Set $y_0^0 = x^0$ and $k = 0, i = 1$.

Step 2. Choose a piecewise linear convex function $\bar{f}_k^i$ satisfying $(C1) - (C3)$ and solve

$$(P_k^i) \qquad \min_{y\in C} \{ c_k \bar{f}_k^i(y) + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle \}$$

to obtain the unique optimal solution $y_k^i \in C$.

Step 3. If

$$f_k(y_k^i) \le \mu \bar{f}_k^i(y_k^i), \tag{3.23}$$

then set $x^{k+1} = y_k^i$, $y_{k+1}^0 = x^{k+1}$, increase k by 1 and set $i = 0$.

Step 4. Replace i by $i+1$ and go to Step 2.

From Theorems 3.4 and 3.3, we obtain the following convergence results.

Theorem 3.5. *If after some k has been reached, the criterion (3.23) is never satisfied, then x^k is a solution to problem EP.*

Theorem 3.6. *Suppose that $c_k \ge \underline{c} > 0$ for all $k \in I\!N$ and that all assumptions of Theorem 3.2 are fulfilled, and that the sequence $\{x^k\}$ generated by the bundle proximal algorithm is infinite. Then $\{x^k\}$ converges to some solution to problem EP.*

For practical implementation, it is necessary to give a stopping criterion. In order to present it, we introduce the definition of a stationary point.

Definition 3.2. *Let $\Delta \ge 0$. A point $x^* \in I\!R^n$ is called a $\Delta-$stationary point of problem EP if $x^* \in C$ and if*

$$\exists\, \gamma \in \partial_\Delta (f_{x^*} + \psi_C)(x^*) \text{ such that } \|\gamma\| \le \Delta.$$

Using the definition of the $\Delta-$subdifferential of the convex function $f_{x^*} + \psi_C$, we obtain that if x^* is a $\Delta-$stationary point of problem EP, then

$$\forall y \in C \quad f_{x^*}(y) \ge f_{x^*}(x^*) + \langle \gamma, y - x^* \rangle - \epsilon \ge -\Delta \|y - x^*\| - \Delta,$$

where we have used $f_{x^*}(x^*) = 0$, the Cauchy-Schwarz inequality and $\|\gamma\| \le \epsilon$.
Observe that if $\Delta = 0$, then a $\Delta-$stationary point of problem EP is a solution to problem EP.

Now to prove that the iterate x^k generated by the bundle algorithm is a Δ-stationary point of problem EP for k large enough, we need the following results.

Proposition 3.4. *Let y_k^i be the solution to problem (P_k^i) and let*

$$\gamma_k^i = \frac{1}{c_k}[\nabla h(x^k) - \nabla h(y_k^i)] \text{ and } \delta_k^i := \langle \gamma_k^i, y_k^i - x^k \rangle - \bar{f}_k^i(y_k^i). \tag{3.24}$$

Then

$$\delta_k^i \geq 0 \text{ and } \gamma_k^i \in \partial_{\delta_k^i}(f_k + \psi_C)(x^k).$$

Proof. By optimality of y_k^i, we obtain that

$$0 \in c_k \partial(\bar{f}_k^i + \psi_C)(y_k^i) + \nabla h(y_k^i) - \nabla h(x^k),$$

i.e.,

$$\gamma_k^i \in \partial(\bar{f}_k^i + \psi_C)(y_k^i).$$

Hence by definition of the subdifferential and since $\bar{f}_k^i \leq f_k$, we have, for all $x \in C$

$$f_k(x) \geq \bar{f}_k^i(x) \geq \bar{f}_k^i(y_k^i) + \langle \gamma_k^i, x - y_k^i \rangle. \tag{3.25}$$

In particular for $x = x^k$, and noting that $f_k(x^k) = 0$, we deduce that

$$0 \geq \bar{f}_k^i(y_k^i) + \langle \gamma_k^i, x^k - y_k^i \rangle,$$

i.e., that $\delta_k^i \geq 0$.

On the other hand, from (3.25) and the definition of δ_k^i, we can write for all $x \in C$,

$$f_k(x) \geq \bar{f}_k^i(y_k^i) + \langle \gamma_k^i, x - y_k^i \rangle = f_k(x^k) + \langle \gamma_k^i, x - x^k \rangle - \delta_k^i,$$

i.e., that $\gamma_k^i \in \partial_{\delta_k^i}(f_k + \psi_C)(x^k)$. □

Theorem 3.7. *Suppose that $c_k \geq \underline{c} > 0$ for all $k \in I\!N$ and that all assumptions of Theorem 3.2 hold. Let $\{x^k\}$ be the sequence generated by the bundle proximal algorithm.*
(i) If $\{x^k\}$ is infinite, then the sequences $\{\gamma_k^{i_k}\}_k$ and $\{\delta_k^{i_k}\}_k$ converge to zero.
(ii) If $\{x^k\}$ is finite with k the latest index, then the sequences $\{\gamma_k^i\}_i$ and $\{\delta_k^i\}_i$ converge to zero.

Proof. (i) Since $\{x^k\}_k$ is infinite, it follows from Theorem 3.6 that $\{x^k\}$ converges to some solution x^* to problem EP.

On the other hand, we have, for all k

$$0 \leq \|\gamma_k^{i_k}\| = \|\frac{\nabla h(x^k) - \nabla h(y_k^{i_k})}{c_k}\| \leq \frac{\Lambda}{\underline{c}}\|x^k - y_k^{i_k}\| = \frac{\Lambda}{\underline{c}}\|x^k - x^{k+1}\|,$$

because ∇h is Lipschitz-continuous with constant $\Lambda > 0, c_k \geq \underline{c} > 0$ and $y_k^{i_k} = x^{k+1}$. Since $\|x^{k+1} - x^k\| \to 0$, we obtain that the sequence $\{\gamma_k^{i_k}\}_k$ converges to zero. Moreover, since

$$|\langle \gamma_k^{i_k}, y_k^{i_k} - x^k \rangle| \leq \|\gamma_k^{i_k}\| \, \|y_k^{i_k} - x^k\| = \|\gamma_k^{i_k}\| \, \|x^{k+1} - x^k\|,$$

we also obtain that $\langle \gamma_k^{i_k}, y_k^{i_k} - x^k \rangle \to 0$ when $k \to +\infty$. Finally, by definition of x^{k+1} and (3.23), we have

$$\frac{1}{\mu} f_k(x^{k+1}) \leq \bar{f}_k^{i_k}(x^{k+1}) \leq f_k(x^{k+1}). \tag{3.26}$$

But $f_k(x^{k+1}) = f(x^k, x^{k+1}) \to f(x^*, x^*) = 0$ by continuity of f_k, so that (3.26) implies that $\bar{f}_k^{i_k}(x^{k+1}) \to 0$. Consequently, we obtain that $\delta_k^i \to 0$ when $k \to +\infty$.

(ii) Let k be the latest index of the sequence $\{x^k\}$. Then x^k is a solution to problem EP by Theorem 3.5 and $\{y_k^i\}_i$ converges to $\bar{y}_k$ when $i \to +\infty$ by Lemma 3.1. Hence $x^k = \bar{y}_k$ and $\|x^k - y_k^i\| \to 0$ when $i \to +\infty$. But this means that $\{\gamma_k^i\}_i$ converges to zero. Moreover, by Lemma 3.1, for $i \to +\infty$, we have $f_k(y_k^i) - \bar{f}_k^i(y_k^i) \to 0$ and thus $\bar{f}_k^i(y_k^i) = \bar{f}_k^i(y_k^i) - f_k(y_k^i) + f_k(y_k^i) \to 0$ because f_k is continuous and $f_k(y_k^i) = f(x^k, y_k^i) \to f(x^k, x^k) = 0$. Consequently $\delta_k^i \to 0$ when $i \to +\infty$. □

Thanks to Proposition 3.4 and Theorem 3.7, we can easily introduce a stopping criterion in the bundle proximal algorithm just after Step 1 as follows.
Compute γ_k^i and δ_k^i by using (3.24). If $\|\gamma_k^i\| \leq \Delta$ and $\delta_k^i \leq \Delta$, then Stop: x^k is a $\Delta-$stationary point of problem EP. Otherwise, go to Step 2 of the bundle proximal algorithm.

Let us mention that this criterion is a generalization of the classical stopping test for bundle methods in optimization (see, e.g., [58]).

3.4 Application to Variational Inequality Problems

First we apply the bundle proximal algorithm for solving problem GVIP under the assumption that $F : I\!R^n \to I\!R^n$ is a continuous mapping and $\varphi : I\!R^n \to I\!R$ a convex function. As we know it, this problem is a particular case of problem EP corresponding to the function f defined, for all $x, y \in I\!R^n$, by $f(x, y) = \langle F(x), y - x \rangle + \varphi(y) - \varphi(x)$. Since the function φ may be nondifferentiable, we choose as model $\bar{f}_k^i$, the function

$$\bar{f}_k^i(y) = \theta_k^i(y) - \varphi(x^k) + \langle F(x^k), y - x^k \rangle,$$

where θ_k^i is a piecewise linear convex function which approximates φ at x^k. Moreover, we assume that this function θ_k^i satisfies the three following conditions:

$(C'1)$ $\theta_k^i \leq \varphi$ on C for $i = 1, 2, \ldots$

$(C'2)$ $\theta_k^{i+1} \geq \varphi(y_k^i) + \langle s'(y_k^i), \cdot - y_k^i \rangle$ on C for $i = 1, 2, \ldots$

$(C'3)$ $\theta_k^{i+1} \geq {l'}_k^i$ on C for $i = 1, 2, \ldots$

where $\gamma_k^i = \frac{1}{c_k}\left[\nabla h(x^k) - \nabla h(y_k^i)\right]$, ${l'}_k^i(y) = \theta_k^i(y_k^i) + \langle \gamma_k^i - F(x^k), y - y_k^i \rangle$, and $s(y_k^i)$ denotes the subgradient of φ available at y_k^i.

With these choices, problem (P_k^i) is equivalent to the problem

$$\min_{y \in C} \{\, c_k \theta_k^i(y) + c_k \langle F(x^k), y - x^k \rangle + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle \},$$

and (3.23) becomes

$$\varphi(x^k) - \varphi(y_k^i) \geq \mu\left[\varphi(x^k) - \theta_k^i(y_k^i)\right] + (1-\mu)\langle F(x^k), y_k^i - x^k \rangle.$$

Finally, the bundle proximal algorithm can be particularized as follows:

Bundle Algorithm for problem GVIP

Data: Let $x^0 \in C$, $\mu \in (0, 1)$, and let $\{c_k\}_{k \in I\!N}$ be a sequence of positive numbers.

Step 1. Set $y_0^0 = x^0$ and $k = 0,\ i = 1$.

Step 2. Choose a piecewise linear convex function θ_k^i satisfying $(C'1) - (C'3)$ and solve

$$\min_{y \in C} \{\, c_k \theta_k^i(y) + c_k \langle F(x^k), y - x^k \rangle + h(y) - h(x^k) - \langle \nabla h(x^k), y - x^k \rangle \} \qquad (3.27)$$

to obtain the unique optimal solution $y_k^i \in C$.

Step 3. If

$$\varphi(x^k) - \varphi(y_k^i) \geq \mu\left[\varphi(x^k) - \theta_k^i(y_k^i)\right] + (1-\mu)\,\langle F(x^k), y_k^i - x^k \rangle, \qquad (3.28)$$

then set $x^{k+1} = y_k^i, y_{k+1}^0 = x^{k+1}$, replace k by $k+1$ and set $i = 0$.

Step 4. Replace i by $i+1$ and go to Step 2.

This algorithm was presented by Salmon et al. in [75] and proven to be convergent under the assumption that F is $\varphi-$co-coercive on C. Thanks to Proposition 3.3, we can deduce from

Theorem 3.6 the convergence theorems obtained in [75] for the bundle method applied for solving problem GVIP (see Theorems 4.2 and 4.3 in [75]).

Theorem 3.8. *Suppose that the sequence $\{c_k\}$ is nonincreasing and satisfies $0 < \underline{c} \leq c_k$ for all k.*
If F is $\varphi-$co-coercive on C with $\gamma > \dfrac{c_0}{2\,\beta\,\mu^2}$, and if the sequence $\{x^k\}$ generated by the bundle proximal algorithm for solving GVIP is infinite, then the sequence $\{x^k\}$ converges to some solution to problem GVIP.

Proof. From Theorem 3.6 and Proposition 3.3, we only have to prove that if $\gamma > \dfrac{c_0}{2\beta\mu^2}$ then there exists $\tau > 0$ such that $c_0 < \dfrac{\beta\mu}{\tau}$ and $\mu \geq \dfrac{1}{2\tau\gamma}$. Since $c_0 < 2\beta\mu^2\gamma$, it is sufficient to set $\tau = \dfrac{1}{2\mu\,\gamma} > 0$ to obtain the two inequalities. □

As a second application, we apply the general algorithm to problem MVIP. This problem corresponds to problem EP with the function f defined, for all $x, y \in C$, by $f(x,y) = \sup_{\xi \in F(x)} \langle \xi, y - x \rangle$ where $F : C \to 2^{I\!R^n}$ is a continuous multivalued mapping with compact values. Thanks to Proposition 23 in [5], it is easy to see that f is continuous on $C \times C$. At iteration k, we consider the approximating function $\bar{f}_k(y) = \langle \xi^k, y - x^k \rangle$ with $\xi^k \in F(x^k)$. Here, we assume that at least one element of $F(x)$ is available for each $x \in C$. When h is the squared norm, the subproblem $(\bar{P}_k)$ becomes

$$\min_{y \in C} \{c_k \langle \xi^k, y - x^k \rangle + \frac{1}{2}\|y - x^k\|^2\}. \tag{3.29}$$

We observe that the optimality conditions associated with (3.29) are

$$\langle c_k \xi^k + y^k - x^k, y - y^k \rangle \geq 0 \text{ for all } y \in C, \tag{3.30}$$

where y^k is a solution to problem $(\bar{P}_k)$. In other words, y^k is the orthogonal projection of the vector $x^k - c_k\,\xi^k$ over C. This problem is a particular convex quadratic programming problem whose solution can be found explicitly when C has a special structure as a box, a ball, $\ldots$. Without loss of generality, we can assume that $y^k \neq x^k$. Indeed, if $y^k = x^k$, then it is easy to see that x^k is a solution to problem MVIP.

Our aim is first to find conditions to ensure that the function $\bar{f}_k$ defined above is a μ-approximation of f_k at x^k and then to apply Theorem 3.3 to get the convergence of the sequence $\{x^k\}$. In that purpose, we introduce the following definitions.

Definition 3.3. *Let* $F : C \to 2^{I\!R^n}$.

(i) F *is strongly monotone on* C *if* $\exists\, \alpha > 0$ *such that* $\forall x, y \in C$ $\forall \xi_1 \in F(x)$ *for all* $\xi_2 \in F(y)$, *one has*

$$\langle \xi_1 - \xi_2, x - y\rangle \geq \alpha \|x - y\|^2.$$

(ii) F *is Lipschitz continuous on* C *if* $\exists\, L > 0$ *such that* $\forall x, y \in C$, *one has*

$$g(x, y) \leq L\|x - y\|,$$

where

$$g(x, y) := \sup_{\xi_1 \in F(x)} \inf_{\xi_2 \in F(y)} \|\xi_1 - \xi_2\|^2. \tag{3.31}$$

(iii) F *is co-coercive on* C *if* $\exists\, \gamma > 0$ *such that* $\forall x, y \in C\, \forall \xi_1 \in F(x)\, \forall \xi_2 \in F(y)$, *one has*

$$\langle \xi_1 - \xi_2, x - y\rangle \geq \gamma g(x, y)\,.$$

In the next proposition, we present the main property of the function $\bar{f}_k$.

Proposition 3.5. *Suppose* F *is co-coercive on* C *with constant* $\gamma > 0$. *Let* $\mu \in (0, 1)$ *and* $x^k \in C$. *If* $c_k \leq 4\gamma(1 - \mu)$, *then the function* $\bar{f}_k(y) = \langle \xi_k, y - x^k\rangle$ *with* $\xi_k \in F(x^k)$ *is a* μ*-approximation of* f_k *at* x^k, *i.e.,* $\bar{f}_k \leq f_k$ *and* $f_k(y^k) \leq \mu \bar{f}_k(y^k)$ *where* y^k *is a solution of problem* $(\bar{P}_k)$.

Proof. Let $\mu \in (0, 1)$ and $\xi_k,\ \xi\ \in F(x^k)$. From (3.30) with $y = x^k$, we deduce that

$$c_k\langle \xi_k, y^k - x^k\rangle \leq -\|x^k - y^k\|^2 < 0. \tag{3.32}$$

Using successively the co-coercivity of F, the definition of g in (3.31) and the Cauchy Schwarz inequality. We have, for every $\eta \in F(y^k)$ and for any $\nu > 0$, that

$$\begin{aligned}
\langle \xi - \xi_k, y^k - x^k\rangle &= \langle \xi - \eta, y^k - x^k\rangle + \langle \eta - \xi_k, y^k - x^k\rangle \\
&\leq -\gamma g(x^k, y^k) + \|\eta - \xi_k\|\, \|y^k - x^k\| \\
&\leq -\gamma g(x^k, y^k) + (1/2\nu)\|\eta - \xi_k\|^2 + \nu/2\|y^k - x^k\|^2.
\end{aligned}$$

Taking the infimum on $\eta \in F(y^k)$ and using (3.32), we obtain

$$\begin{aligned}
\langle \xi - \xi_k, y^k - x^k\rangle &\leq -\gamma g(x^k, y^k) + (1/2\nu) \inf_{\eta \in F(y^k)} \|\eta - \xi_k\|^2 \\
&+ \nu/2\|y^k - x^k\|^2 \\
&\leq (\tfrac{1}{2\nu} - \gamma)\, g(x^k, y^k) - \frac{\nu c_k}{2}\, \langle \xi_k, y^k - x^k\rangle
\end{aligned}$$

for all $\nu > 0$. Choosing $\nu = 1/(2\gamma)$, we can write

$$\langle \xi, y^k - x^k \rangle \leq (1 - \frac{c_k}{4\gamma}) \langle \xi_k, y^k - x^k \rangle.$$

Finally, taking the supremum on $\xi \in F(x^k)$, and using the condition $c_k < 4\gamma(1-\mu)$, we deduce the thesis. □

Since $\bar{f}_k$ is a μ-approximation of f_k at x^k for a suitable value of c_k, using this approximating function, the general algorithm becomes:
Given $x^k \in C$ and $c_k > 0$, choose $\xi_k \in F(x^k)$ and solve the problem

$$\min_{y \in C} \{c_k \langle \xi_k, y - x^k \rangle + \frac{1}{2}\|y - x^k\|^2\}$$

to get x^{k+1}.

In particular case, when F is co-coercive on C, the assumptions (i) and (ii) of Theorem 3.2 are satisfied.

Proposition 3.6. *Let $f(x, y) = \sup_{\xi \in F(x)} \langle \xi, y - x \rangle$ and g defined by (3.31). Then*
(i) for every $x, y, z \in C$ and for any $\nu > 0$,

$$f(x, z) - f(y, z) - f(x, y) \leq \frac{1}{2\nu} g(x, y) + \frac{\nu}{2} \|z - y\|^2,$$

(ii) if F is co-coercive on C with constant γ, then for every $x, y \in C$,

$$f(x, y) \geq 0 \Rightarrow f(y, x) \leq -\gamma g(x, y).$$

Finally, for the sequence $\{x^k\}$ generated by this algorithm, we obtain the following convergence theorem.

Theorem 3.9. *Suppose F is co-coercive on C with constant $\gamma > 0$. Let $\{c_k\}$ be a nonincreasing sequence bounded away from 0. If $c_k < 4(2-\sqrt{3})\,\gamma$ for all k, then the sequence $\{x^k\}$ converges to some solution x^* to problem MVIP.*

Proof. Since $\beta = 1$, from Propositions 3.5 and 3.6, and from Theorem 3.3, we only have to prove that there exist $\mu \in (0, 1)$, and $\nu > 0$ such that

$$c_k \leq 4\gamma(1 - \mu), \quad c_k < \frac{\mu}{\nu}, \quad \frac{1}{2\nu\gamma} \leq \mu.$$

Choosing the smallest possible ν, we obtain $\nu = 1/(2\mu\gamma)$. Then the previous conditions become:

$$c_k \leq 4\gamma(1-\mu) \quad \text{and} \quad c_k < 2\mu^2\gamma. \tag{3.33}$$

It is easy to see that the maximum of the function $r(\mu) = \min\{4\gamma(1-\mu), 2\mu^2\gamma\}$ occurs at $\mu = \sqrt{3}-1$ and has $4(2-\sqrt{3})\gamma$ for optimal value. So the conditions (3.33) are satisfied with this value of μ if $c_k < 4(2-\sqrt{3})\,\gamma$. □

When F is singlevalued, the approximating function $\bar{f}_k$ coincides with f_k. In that case, $\mu = 1$ and Proposition 3.5 must not be considered. This means that only the second inequality in (3.33) must be retained, i.e., $c_k < 2\gamma$.

An interesting particular case is when F is strongly monotone (with constant $\alpha > 0$) and Lipschitz continuous (with constant $L > 0$) on C. Then, for all $x, y \in C$,

$$g(x,y) \leq L^2\,\|x-y\|^2.$$

Hence, F being strongly monotone, we have, for all $x, y \in C$ and $\xi \in F(x), \eta \in F(y)$, that

$$\langle \xi - \eta, x - y\rangle \geq \alpha\,\|x-y\|^2 \geq \frac{\alpha}{L^2}\,g(x,y).$$

But this means that F is co-coercive on C with constant $\gamma = \alpha/L^2$. Then Theorem 3.9 becomes:

Theorem 3.10. *Suppose F is strongly monotone (with constant $\alpha > 0$) and Lipschitz continuous (with constant $L > 0$) on C. Let $\{c_k\}$ be a nonincreasing sequence bounded away from 0. If $c_k < 4(2-\sqrt{3})\,\frac{\alpha}{L^2}$ for all k, then the sequence $\{x^k\}$ converges to the unique solution x^* to problem MVIP. When F is singlevalued, the same property holds but with $c_k < \frac{2\alpha}{L^2}$ for all k.*

Let us mention that when F is singlevalued, we retrieve a classical result for variational inequalities (see, for instance, [1], [2], [38], [39]). In the multivalued case, our algorithm has been studied by El Farouq in [29] but under the assumption that the series $\sum c_k^2$ is convergent, and thus that the sequence $\{c_k\}$ converges to 0.

Chapter 4

Interior Proximal Extragradient Methods

In this chapter we present a new and efficient method for solving equilibrium problems on polyhedra. The method is based on an interior-quadratic proximal term which replaces the usual quadratic proximal term. This leads to an interior proximal type algorithm. Each iteration consists in a prediction step followed by a correction step as in the extragradient method. In a first algorithm each of these steps is obtained by solving an unconstrained minimization problem, while in a second algorithm the correction step is replaced by an Armijo-backtracking line-search followed by an hyperplane projection step. We prove that our algorithms are convergent under mild assumptions: pseudomonotonicity for the two algorithms and a Lipschitz property for the first one. Finally we present some numerical experiments to illustrate the behavior of the proposed algorithms.

4.1 Preliminaries

In this chapter, we consider problem EP where C is a polyhedral set with a nonempty interior given by $C = \{x \,|\, A\,x \leq b\}$ with A an $m \times n$ $(m \geq n)$ matrix of full rank with row a_i, and b a vector in $\mathbb{R}^m$ with rows b_i.

An important example of such a C is the nonnegative orthant of $\mathbb{R}^n$. We also assume that f is continuous on $C \times C$ and that $f(x, \cdot)$ is convex and subdifferentiable on C for all $x \in C$. In order to take account of the constraints, we consider a distance-like function, denoted $D_\varphi(x, y)$. This function is constructed from a class of functions $\varphi : \mathbb{R} \to (-\infty, +\infty]$ of the form

$$\varphi(t) = \mu h(t) + \frac{\nu}{2}(t-1)^2, \tag{4.1}$$

where $\nu > \mu > 0$ and h is a closed and proper convex function satisfying the following additional properties:

(a) h is twice continuously differentiable on $(0, +\infty)$, the interior of its domain,

(b) h is strictly convex on its domain,

(c) $\lim_{t \to 0^+} h^{'}(t) = -\infty$,

(d) $h(1) = h^{'}(1) = 0$ and $h^{''}(1) = 1$, and

(e) For $t > 0$, $1 - t^{-1} \leq h^{'}(t) \leq t - 1$.

Amongst all the functions h satisfying properties $(a) - (e)$, let us mention the following one:

$$h(t) = \begin{cases} t - \log t - 1 & \text{if} \quad t > 0, \\ +\infty & \text{otherwise.} \end{cases}$$

The corresponding function φ is called the logarithmic-quadratic function. It enjoys attractive properties for developing efficient algorithms (see [12] and [16] for the properties of this function).

Another function h which is also often used in the literature (see, for example, [21] and [71]) is

$$h(t) = \begin{cases} t \log t - t + 1 & \text{if} \quad t > 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Associated with φ, we consider the $\varphi-$divergence proximal distance

$$d_\varphi(x, y) = \sum_{j=1}^{n} y_j^2 \varphi\left(\frac{x_j}{y_j}\right) \text{ for all } x, y \in I\!R^n_{++},$$

and for any $x, y \in \operatorname{int} C$, we define the distance-like function D_φ by

$$D_\varphi(x, y) = d_\varphi(l(x), l(y)) \text{ for all } x, y \in \operatorname{int} C,$$

where $l(x) = (l_1(x), \ldots, l_n(x))$ and $l_j(x) = b_j - \langle a_j, x\rangle$, $j = 1, \ldots, n$.
It is easy to see that

$$D_\varphi(x, y) = \mu D_h(x, y) + \frac{\nu}{2}\|A(x - y)\|^2 \text{ for all } x, y \in \operatorname{int} C,$$

showing the barrier and regularization terms. Note that A being of full rank, the function $(u, v) \to \langle A^T Au, v\rangle$ defines on $I\!R^n$ an inner product denoted $\langle u, v\rangle_A$ with $\|u\|_A := \|Au\| = \langle Au, Au\rangle^{\frac{1}{2}}$, so that we can write

$$D_\varphi(x, y) = \mu D_h(x, y) + \frac{\nu}{2}\|x - y\|_A^2 \quad \text{for all } x, y \in \operatorname{int} C. \tag{4.2}$$

With this distance, the basic iteration of our method can be written as follows: *Given* $x^k \in \mathrm{int}\, C$*, find* $x^{k+1} \in \mathrm{int}\, C$*, the solution of the unconstrained problem*

$$(P_k) \qquad \min_y\{c_k f(x^k, y) + D_\varphi(y, x^k)\}.$$

This method has been intensively studied by Auslender et al. for solving particular equilibrium problems as the convex optimization problems (see, for example, [9], [12], [14], [15]) and the variational inequality problems (see, for example, [10], [11], [13], [15]). See also [21], [22], [24], [83], [87].

Our aim in this chapter is to study extragradient methods based on problem (P_k) for solving problem EP where $C = \{x \mid Ax \leq b\}$. In the next two sections, we assume that φ is of the form (4.1) with h a function satisfying properties $(a) - (e)$.

4.2 Interior Proximal Extragradient Algorithm

Let us recall some preliminary results which will be used later in our analysis. First, for all $x, y, z \in I\!R^n$, it is easy to see that

$$\|x - y\|_A^2 + \|x - z\|_A^2 = \|y - z\|_A^2 + 2\langle x - z, x - y\rangle_A. \tag{4.3}$$

Next, let us introduce a lemma that plays a key role in the convergence analysis.

Lemma 4.1. *For all* $x, y \in$ *int* C *and* $z \in C$*, it holds that*
(i) $D_\varphi(\cdot, y)$ *is differentiable and strongly convex on int* C *with modulus* ν*, i.e.,*

$$\langle \nabla_1 D_\varphi(x, p) - \nabla_1 D_\varphi(y, p),\, x - y\,\rangle \geq \nu\, \|x - y\|_A^2 \text{ for all } p \in \text{ int}\, C,$$

where $\nabla_1 D_\varphi(x, p)$ *denotes the gradient of* $D_\varphi(\cdot, p)$ *at* x.
(ii) $D_\varphi(x, y) = 0$ *if and only if* $x = y$*,*
(iii) $\nabla_1 D_\varphi(x, y) = 0$ *if and only if* $x = y$*,*
(iv) $\langle \nabla_1 D_\varphi(x, y), x - z\rangle \geq \left(\dfrac{\nu + \mu}{2}\right)(\|x - z\|_A^2 - \|y - z\|_A^2) + \left(\dfrac{\nu - \mu}{2}\right)\|x - y\|_A^2.$

Proof. See Proposition 2.1. in [12] and Proposition 4.1 in [24]. □

The next result is crucial to establish the existence and the characterization of a solution to subproblem (P_k).

Theorem 4.1. *Let $F : C \to I\!R \cup \{+\infty\}$ be a closed proper convex function such that $domF \cap int\, C \neq \emptyset$. Given $x \in int\, C$ and $c_k > 0$. Then there exists a unique $y \in int\, C$ such that*

$$y = \arg\min_z \{c_k F(z) + D_\varphi(z, x)\}$$

and

$$0 \in c_k \partial F(y) + \nabla_1 D_\varphi(y, x),$$

where $\partial F(y)$ denotes the subdifferential of F at y.

Proof. See Lemma 3.3 in [9]. □

Now we present a first interior proximal extragradient algorithm for solving problem EP.

Interior Proximal Extragradient Algorithm (IPE)

Data: Let $x^0 \in C$, choose $c_0 > 0$ and a couple of positive parameters (ν, μ) such that $\nu > \mu$. The corresponding distance function is denoted D_φ.

Step 1. Set $k = 0$.

Step 2. Solve the interior proximal convex program

$$\min_y \{c_k f(x^k, y) + D_\varphi(y, x^k)\} \tag{4.4}$$

to obtain its unique solution y^k. If $y^k = x^k$, then Stop: x^k is a solution to problem EP. Otherwise, go to Step 3.

Step 3. Solve the interior proximal convex program

$$\min_y \{c_k f(y^k, y) + D_\varphi(y, x^k)\} \tag{4.5}$$

to obtain its unique solution x^{k+1}.

Step 4. Replace k by $k + 1$, choose $c_k > 0$ and return to Step 2.

First observe that the algorithm is well defined. Indeed, thanks to Theorem 4.1 with function F defined by $f(x^k, \cdot)$ and $f(y^k, \cdot)$, respectively, the subproblems (4.4) and (4.5) have a unique solution and

$$0 \in c_k \partial_2 f(x^k, y^k) + \nabla_1 D_\varphi(y^k, x^k) \text{ and } 0 \in c_k \partial_2 f(y^k, x^{k+1}) + \nabla_1 D_\varphi(x^{k+1}, x^k),$$

where $\partial_2 f(x,y)$ denotes the subdifferential of $f(x,\cdot)$ at y.

Consequently, using the definition of the subdifferential, we can write

$$c_k f(x^k,y) \geq c_k f(x^k,y^k) + \langle \nabla_1 D_\varphi(y^k,x^k), y^k - y\rangle \text{ for all } y \in C, \tag{4.6}$$

and

$$c_k f(y^k,y) \geq c_k f(y^k,x^{k+1}) + \langle \nabla_1 D_\varphi(x^{k+1},x^k), x^{k+1} - y\rangle \text{ for all } y \in C. \tag{4.7}$$

In the next proposition, we justify the stopping criterion.

Proposition 4.1. *If $y^k = x^k$, then x^k is a solution to problem EP.*

Proof. When $y^k = x^k$, the inequality (4.6) becomes

$$c_k f(x^k,y) \geq c_k f(x^k,x^k) + \langle \nabla_1 D_\varphi(x^k,x^k), x^k - y\rangle \text{ for all } y \in C.$$

Since $f(x^k,x^k) = 0$ and $\nabla_1 D_\varphi(x^k,x^k) = 0$ (by Lemma 4.1(iii)), it follows that

$$c_k f(x^k,y) \geq 0 \text{ for all } y \in C,$$

i.e., that x^k is a solution to problem EP. □

Now we are in a position to prove the convergence of the IPE algorithm.

Theorem 4.2. *Suppose that $\nu > 5\mu$ and that there exist two positive parameters d_1 and d_2 such that*

$$\forall x,y,z \in C \quad f(x,y) + f(y,z) \geq f(x,z) - d_1\|y-x\|_A^2 - d_2\|z-y\|_A^2. \tag{4.8}$$

Then the following statements hold:
(i) If $x^* \in S_d^*$, then

$$\begin{aligned} \Delta(x^k) - \Delta(x^{k+1}) \;\geq\; & \left(\tfrac{1}{2} - \tfrac{2\mu + c_k\, d_1}{\nu - \mu}\right)\|y^k - x^k\|_A^2 \\ +\; & \left(\tfrac{1}{2} - \tfrac{\mu + c_k\, d_2}{\nu-\mu}\right)\|x^{k+1} - y^k\|_A^2, \end{aligned} \tag{4.9}$$

where $\Delta(x) = \left(\frac{1}{2} + \dfrac{\mu}{\nu-\mu}\right)\|x - x^*\|_A^2$;
(ii) If $0 < c \leq c_k < \min\left\{\dfrac{\nu - 5\mu}{2\,d_1}, \dfrac{\nu-3\mu}{2\,d_2}\right\}$, then the sequence $\{x^k\}$ is bounded and every limit point of $\{x^k\}$ is a solution to problem EP. In addition, if $S_d^* = S^*$, then the whole sequence $\{x^k\}$ tends to a solution of problem EP.

Proof. (i) Take any $x^* \in S_d^*$ and consider the inequality (4.6) with $y = x^{k+1}$. Then

$$c_k f(x^k, x^{k+1}) - c_k f(x^k, y^k) \geq \langle \nabla_1 D_\varphi(y^k, x^k), y^k - x^{k+1} \rangle.$$

Using first Lemma 4.1 (iv) to the right hand side of this inequality and then the equality (4.3) with $x = y^k, y = x^{k+1}$ and $z = x^k$, we obtain successively

$$\begin{aligned} c_k f(x^k, x^{k+1}) - c_k f(x^k, y^k) &\geq \theta \left(\|y^k - x^{k+1}\|_A^2 - \|x^k - x^{k+1}\|_A^2 \right) \\ &+ (\nu - \theta)\|y^k - x^k\|_A^2 \\ &= \theta \left(-\|y^k - x^k\|_A^2 + 2\langle y^k - x^k, y^k - x^{k+1} \rangle_A \right) \\ &+ (\nu - \theta)\|y^k - x^k\|_A^2 \qquad (4.10) \\ &= -\mu \|y^k - x^k\|_A^2 \\ &+ (\nu + \mu)\langle y^k - x^k, y^k - x^{k+1} \rangle_A, \end{aligned}$$

where $\theta = \frac{\nu+\mu}{2}$.

On the other hand, considering the inequality (4.7) with $y = x^*$, we have

$$c_k f(y^k, x^*) - c_k f(y^k, x^{k+1}) \geq \langle \nabla_1 D_\varphi(x^{k+1}, x^k), x^{k+1} - x^* \rangle.$$

Using again Lemma 4.1 (iv) and the equality (4.3) with $x = x^{k+1} \in \operatorname{int} C, y = x^k \in \operatorname{int} C$, $z = x^* \in C$, we obtain that

$$\begin{aligned} c_k f(y^k, x^*) - c_k f(y^k, x^{k+1}) &\geq \theta(\|x^{k+1} - x^*\|_A^2 - \|x^k - x^*\|_A^2) \\ &+ (\nu - \theta)\|x^{k+1} - x^k\|_A^2 \\ &= \theta(\|x^{k+1} - x^*\|_A^2 - \|x^k - x^*\|_A^2) \\ &+ (\nu - \theta)\|x^k - x^*\|_A^2 \\ &- (\nu - \theta)\|x^{k+1} - x^*\|_A^2 \\ &+ 2(\nu - \theta)\langle x^{k+1} - x^k, x^{k+1} - x^* \rangle_A \\ &= \mu\|x^{k+1} - x^*\|_A^2 - \mu\|x^k - x^*\|_A^2 \\ &+ (\nu - \mu)\langle x^{k+1} - x^k, x^{k+1} - x^* \rangle_A, \end{aligned}$$

Noting that $\nu - \mu > 0$ and $f(y^k, x^*) \leq 0$ because $x^* \in S_d^*$, we deduce from the above inequality that

$$\begin{aligned}
\langle x^{k+1} - x^k, x^* - x^{k+1}\rangle_A &\geq \frac{c_k}{\nu - \mu} f(y^k, x^{k+1}) \\
&+ \frac{\mu}{\nu - \mu}\|x^{k+1} - x^*\|_A^2 - \frac{\mu}{\nu - \mu}\|x^k - x^*\|_A^2 \\
&\geq \frac{c_k}{\nu - \mu}[f(x^k, x^{k+1}) - f(x^k, y^k)] \qquad (4.11) \\
&- \frac{c_k d_1}{\nu - \mu}\|y^k - x^k\|_A^2 - \frac{c_k d_2}{\nu - \mu}\|x^{k+1} - y^k\|_A^2 \\
&+ \frac{\mu}{\nu - \mu}\|x^{k+1} - x^*\|_A^2 - \frac{\mu}{\nu - \mu}\|x^k - x^*\|_A^2
\end{aligned}$$

where the second inequality is obtained after using assumption (4.8) with $x = x^k, y = y^k$ and $z = x^{k+1}$.

On the other hand, from equality (4.3) with $x = x^{k+1}$, $y = x^*$, $z = x^k$ and then with $x = y^k$, $y = x^{k+1}$ and $z = x^k$, we deduce

$$\|x^k - x^*\|_A^2 - \|x^{k+1} - x^*\|_A^2 = \|x^{k+1} - x^k\|_A^2 + 2\langle x^{k+1} - x^k, x^* - x^{k+1}\rangle_A, \qquad (4.12)$$

and

$$\|x^{k+1} - x^k\|_A^2 = -2\langle y^k - x^k, y^k - x^{k+1}\rangle_A + \|x^{k+1} - y^k\|_A^2 + \|y^k - x^k\|_A^2. \qquad (4.13)$$

Finally, using successively (4.12), (4.11), (4.10), (4.13), and the inequality

$$\begin{aligned}
\langle y^k - x^k, y^k - x^{k+1}\rangle_A &\geq -\|y^k - x^k\|_A \, \|y^k - x^{k+1}\|_A \\
&\geq -\tfrac{1}{2}\,\|y^k - x^k\|_A^2 - \tfrac{1}{2}\,\|y^k - x^{k+1}\|_A^2,
\end{aligned}$$

we obtain the following equalities and inequalities

$$\begin{array}{rcl}\Delta(x^k) - \Delta(x^{k+1}) & = & \frac{1}{2}\|x^{k+1} - x^k\|_A^2 + \langle x^{k+1} - x^k, x^* - x^{k+1}\rangle_A \\ & + & \frac{\mu}{\nu-\mu}\|x^k - x^*\|_A^2 - \frac{\mu}{\nu-\mu}\|x^{k+1} - x^*\|_A^2 \\ & \geq & \frac{1}{2}\|x^{k+1} - x^k\|_A^2 + \frac{c_k}{\nu-\mu}[f(x^k, x^{k+1}) - f(x^k, y^k)] \\ & - & \frac{c_k d_1}{\nu-\mu}\|y^k - x^k\|_A^2 - \frac{c_k d_2}{\nu-\mu}\|x^{k+1} - y^k\|_A^2 \\ & \geq & \frac{1}{2}\|x^{k+1} - x^k\|_A^2 + \frac{\nu+\mu}{\nu-\mu}\langle y^k - x^k, y^k - x^{k+1}\rangle_A \\ & - & \frac{\mu+c_k d_1}{\nu-\mu}\|y^k - x^k\|_A^2 - \frac{c_k d_2}{\nu-\mu}\|x^{k+1} - y^k\|_A^2 \\ & = & (\frac{1}{2} - \frac{\mu+c_k d_1}{\nu-\mu})\|y^k - x^k\|_A^2 + (\frac{1}{2} - \frac{c_k d_2}{\nu-\mu})\|x^{k+1} - y^k\|_A^2 \\ & + & \frac{2\mu}{\nu-\mu}\langle y^k - x^k, y^k - x^{k+1}\rangle_A \\ & \geq & (\frac{1}{2} - \frac{2\mu+c_k d_1}{\nu-\mu})\|y^k - x^k\|_A^2 \\ & + & (\frac{1}{2} - \frac{\mu+c_k d_2}{\nu-\mu})\|x^{k+1} - y^k\|_A^2\end{array}$$

(ii) Since $\nu > 5\mu$ and $0 < c_k < \min\{\dfrac{\nu - 5\mu}{2\,d_1}, \dfrac{\nu - 3\mu}{2\,d_2}\}$, we have $\dfrac{1}{2} - \dfrac{2\mu + c_k\, d_1}{\nu - \mu} > 0$ and $\dfrac{1}{2} - \dfrac{\mu + c_k d_2}{\nu - \mu} > 0$. Consequently, from part (i), we obtain that

$$\Delta(x^k) - \Delta(x^{k+1}) \geq \left(\frac{1}{2} - \frac{2\mu + c_k\, d_1}{\nu - \mu}\right)\|y^k - x^k\|_A^2 \geq 0 \text{ for all } k.$$

This implies that the positive sequence $\{\Delta(x^k)\}$ is nonincreasing. Hence this sequence converges in $I\!R$ and consequently, is bounded and such that

$$\lim_{k\to+\infty}\|y^k - x^k\|_A^2 = 0. \tag{4.14}$$

Let $\bar{x}$ be a limit point of $\{x^k\}$. Then $\bar{x} = \lim_{j\to+\infty} x^{k_j}$, and, by (4.14), $\bar{x} = \lim_{j\to+\infty} y^{k_j}$. Using (4.6) and Lemma 4.1(iv), we have for all $y \in C$ and all j that

$$\begin{array}{rcl}c_{k_j} f(x^{k_j}, y) - c_{k_j} f(x^{k_j}, y^{k_j}) & \geq & \langle \nabla_1 D_\varphi(y^{k_j}, x^{k_j}), y^{k_j} - y\rangle \\ & \geq & \dfrac{\nu+\mu}{2}(\|y^{k_j} - y\|_A^2 - \|x^{k_j} - y\|_A^2).\end{array} \tag{4.15}$$

Taking $j \to +\infty$ in (4.15) and noting that $f(\bar{x}, \bar{x}) = 0$ and $0 < c < c_k \leq \min\{\frac{\nu-5\mu}{2d_1}, \frac{\nu-3\mu}{2d_2}\}$, we obtain

$$f(\bar{x}, y) \geq 0 \text{ for all } y \in C,$$

which means that $\bar{x}$ is a solution to problem EP.
Suppose now that $S_d^* = S^*$. Then the whole sequence $\{x^k\}$ converges to $\bar{x}$. Indeed, defining $\Delta(x^k)$ with $x^* = \bar{x} \in S_d^*$, we have $\Delta(x^{k_j}) \to 0$ because $x^{k_j} \to \bar{x}$. So the sequence $\Delta(x^k)$ being nonincreasing, the whole sequence $\{\Delta(x^k)\}$ also converges to 0 and thus $\|x^k - \bar{x}\|_A \to 0$, i.e., $x^k \to \bar{x}$. □

When the function $f(x, \cdot)$ is nonsmooth, it can be difficult to solve subproblems (4.4) and (4.5). In that case, we can use a bundle strategy as in nonsmooth optimization [12] (see also [83], [84]). For subproblem (4.4), the idea is to approximate the function $f(x^k, \cdot)$ from below by a piecewise linear convex function ψ^k and to take for the next iterate the solution y^k of the following subproblem

$$\min_y \{c_k \psi^k(y) + D_\varphi(y, x^k)\}. \tag{4.16}$$

More precisely, ψ^k is constructed, thanks to a sequence ψ_i^k, $i = 1, 2, \ldots$ as follows:

The starting data are $y_0^k = x^k$, $g_0^k \in \partial_2 f(x^k, y_0^k)$ and $\psi_1^k(y) = f(x^k, y_0^k) + \langle g_0^k, y - y_0^k \rangle$ for all $y \in I\!R^n$.

Suppose at iteration $i \geq 1$ that ψ_i^k is known. Then ψ_{i+1}^k is obtained by the following steps:

Step 1: Solve subproblem (4.16) with ψ^k replaced by ψ_i^k to get y_i^k; set $d_i^k = -\nabla_1 D_\varphi(y_i^k, x^k)$ and $l_i^k(y) = \psi_i^k(y_i^k) + \langle d_i^k, y - y_i^k \rangle$.
Step 2: Choose $\psi_{i+1}^k : I\!R^n \to I\!R$ as a piecewise linear convex function satisfying the conditions:
(C1) $l_i^k \leq \psi_{i+1}^k \leq f(x^k, \cdot)$
(C2) $f(x^k, y_i^k) + \langle g_i^k, y - y_i^k \rangle \leq \psi_{i+1}^k(y)$ for all $y \in I\!R^n$ with $g_i^k \in \partial_2 f(x^k, y_i^k)$.

It can be proven (see Theorem 3.2 in [12]) that after finitely many steps i, this algorithm gives a point y_i^k and a model ψ_i^k such that

$$f(x^k, y_i^k) \leq \eta\, \psi_i^k(y_i^k) \quad (0 < \eta < 1).$$

In that case we consider that the approximate function ψ_i^k is appropriate and we set $\psi^k = \psi_i^k$ and $y^k = y_i^k$.
Next, in order to obtain an efficient algorithm, the functions ψ_i^k must be chosen in such a way that the subproblems (4.16) (with ψ^k replaced by ψ_i^k) can be easily solved. In [12] it is shown that for $C = I\!R_+^n$ and

$$\psi_{i+1}^k(y) = \max\{l_i^k(y), f(x^k, y_i^k) + \langle g_i^k, y - y_i^k \rangle\} \quad \text{for all } y \in I\!R^n,$$

the conditions (C1) and (C2) are satisfied and the subproblems (4.16) can be simplified and reduced to minimizing a function of a single variable (see also [84], for other examples and properties of the models ψ_i^k). Since a similar strategy can be developed for solving subproblem (4.5), we finally obtain an implementable algorithm whose convergence results can be proven exactly as in [84].

4.3 Interior Proximal Linesearch Extragradient Method

Convergence of the IPE algorithm requires that the function f satisfies condition (4.8). This condition depends on two positive parameters d_1 and d_2 and in some cases, they are unknown or difficult to approximate. So in this section, we modify the second step of the algorithm using a linesearch and an hyperplane projection step in order to obtain the convergence without assuming that condition (4.8) is satisfied. When a quadratic regularization term is used, this strategy has been initiated by Konnov [42, 43, 44] in the particular case where f is differentiable. The non differentiable convex case has been recently considered by Quoc et al. [72]. In this section, we replace the usual quadratic proximal distance by the φ-divergence proximal distance D_φ defined in (4.2), and as in [72], we suppose that

(D1) C is contained in an open convex set $\Lambda \subset I\!R^n$,

(D2) $f : \Lambda \times \Lambda \to I\!R$ is a continuous function satisfying $f(x, x) = 0$ for each $x \in \Lambda$ and $f(x, \cdot)$ is convex for each $x \in \Lambda$.

Before giving the algorithm and in order to obtain more flexibility in the choice of the steplength, we introduce a sequence $\{\gamma_k\}$ which satisfies the properties

$$\gamma_k \in (0, 2) \quad \forall k = 0, 1, \ldots \text{ and } \liminf_{k \to +\infty} \gamma_k (2 - \gamma_k) > 0. \tag{4.17}$$

Obviously, $\gamma_k = 1$ for all k is an example of such a sequence.

The Interior Proximal Linesearch Extragradient Algorithm (IPLE)

Data: Let $x^0 \in \operatorname{int} C$, choose $\theta \in (0,1)$, $\tau \in (0,1)$, $\alpha \in (0,1)$, $c > 0$, $c_0 \geq c > 0$ and choose positive parameters ν, μ such that $\nu > \mu$.

Step 1. Set $k = 0$.

Step 2. Solve the convex program

$$\min_y \{c_k f(x^k, y) + D_\varphi(y, x^k)\} \tag{4.18}$$

to obtain its unique solution y^k. If $y^k = x^k$, then Stop: x^k is a solution to problem EP. Otherwise, go to Step 3.

Step 3. Find the smallest nonnegative integer m such that

$$f(z^{k,m}, x^k) - f(z^{k,m}, y^k) \geq \frac{\alpha}{c_k} D_\varphi(y^k, x^k), \tag{4.19}$$

where $z^{k,m} = (1-\theta^m)x^k + \theta^m y^k$. Set $z^k = z^{k,m}$ and go to Step 4.

Step 4. Take $g^k \in \partial_2 f(z^k, x^k)$.
Compute $\sigma_k = \dfrac{f(z^k, x^k)}{\|g^k\|^2}$ and $x^{k+1} = (1-\tau)x^k + \tau P_C(x^k - \gamma_k \sigma_k g^k)$,
where $P_C(z)$ denotes the orthogonal projection of z over C.

Step 5. Replace k by $k+1$, choose $c_k \geq c > 0$ and go to Step 2.

Remark 4.1. Algorithm (IPLE) is an extension of the combined relaxation method proposed by Konnov [44] for solving a differentiable monotone equilibrium problem. The Armijo-backtracking linesearch (Step 2) is slightly different from Konnov's one to take account of the φ-divergence proximal distance and of the fact that f is non differentiable. The hyperplane projection step (Step 3) is similar, a subgradient g_k of $f(z^k, \cdot)$ replacing the gradient of $f(z^k, \cdot)$.

In order to see that Algorithm (IPLE) is well defined, first observe that, by Theorem 4.1, the solution y^k of problem (4.18) exists and is unique. Furthermore, if $x^k \in \operatorname{int} C$, then x^{k+1} also belongs to $\operatorname{int} C$ because $\tau \in (0,1)$. Finally to state that the linesearch is also well defined, we introduce the following lemma:

Lemma 4.2. *Suppose that $y^k \neq x^k$ for some k. Then the next three properties hold:*
(i) *There exists a nonnegative integer m satisfying (4.19);*
(ii) $f(z^k, x^k) > 0$*;*
(iii) $0 \notin \partial_2 f(z^k, x^k)$.

Proof. (i) By contradiction, we suppose that statement (i) is not true, i.e., that for all nonnegative integer m, we have the inequality

$$f(z^{k,m}, x^k) - f(z^{k,m}, y^k) < \frac{\alpha}{c_k} D_\varphi(y^k, x^k).$$

Let $m \to +\infty$. Then $z^{k,m} \to x^k$ and because f is continuous on $C \times C$ and $f(x, x) = 0$ for all $x \in C$, we obtain

$$c_k f(x^k, y^k) + \alpha D_\varphi(y^k, x^k) \geq 0. \tag{4.20}$$

On the other hand, because y^k is a solution of (4.18), we have

$$c_k f(x^k, y^k) + D_\varphi(y^k, x^k) \leq c_k f(x^k, y) + D_\varphi(y, x^k) \text{ for all } y \in \operatorname{int} C.$$

Taking $y = x^k$ in this inequality and noting that $f(x^k, x^k) = 0$ and $D_\varphi(x^k, x^k) = 0$, we deduce

$$c_k f(x^k, y^k) + D_\varphi(y^k, x^k) \leq 0.$$

Combining this inequality and (4.20) and noting that $D_\varphi(y^k, x^k) > 0$ because $y^k \neq x^k$, we obtain $\alpha \geq 1$. But this contradicts the assumption and thus there exists a nonnegative integer m satisfying (4.19).
(ii) Because f is convex with respect to the second argument, it follows from the definition of z^k that

$$0 = f(z^k, z^k) \leq (1 - \theta^m) f(z^k, x^k) + \theta^m f(z^k, y^k). \tag{4.21}$$

Hence, using (4.19), we obtain

$$f(z^k, x^k) \geq \theta^m (f(z^k, x^k) - f(z^k, y^k)) \geq \frac{\alpha \theta^m}{c_k} D_\varphi(y^k, x^k) > 0.$$

(iii) By contradiction, let us suppose that $0 \in \partial_2 f(z^k, x^k)$, i.e., that

$$f(z^k, y) \geq f(z^k, x^k) \text{ for all } y \in C.$$

Taking $y = z^k$, we obtain that $f(z^k, x^k) \leq 0$. This contradicts (ii), and so (iii) holds. □

The following lemmas are the key results in our analysis of the convergence of the algorithm (IPLE).

Lemma 4.3. (i) *The sequence* $\{x^k\}$ *is bounded and for every solution* $x^* \in S_d^*$*, the following inequality holds*

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \tau\gamma_k(2-\gamma_k)(\sigma_k\|g^k\|)^2.$$

(ii) $\displaystyle\sum_{k=0}^{+\infty} \gamma_k(2-\gamma_k)(\sigma_k\|g^k\|)^2 < +\infty.$

Proof. (i) Take $x^* \in S_d^*$. Using successively the definition of x^{k+1}, the convexity of $\|\cdot\|^2$ and the nonexpansiveness of the projection, we have

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 &= \|(1-\tau)x^k + \tau P_C(x^k - \gamma_k\sigma_k g^k) - x^*\|^2 \qquad (4.22)\\
&= \|(1-\tau)(x^k - x^*) + \tau[P_C(x^k - \gamma_k\sigma_k g^k) - x^*]\|^2\\
&\leq (1-\tau)\|x^k - x^*\|^2 + \tau\|P_C(x^k - \gamma_k\sigma_k g^k) - x^*\|^2\\
&\leq (1-\tau)\|x^k - x^*\|^2 + \tau\|x^k - \gamma_k\sigma_k g^k - x^*\|^2\\
&= \|x^k - x^*\|^2 + \tau\|\gamma_k\sigma_k g^k\|^2 - 2\tau\langle\gamma_k\sigma_k g^k, x^k - x^*\rangle.
\end{aligned}$$

On the other hand, because $g^k \in \partial_2 f(z^k, x^k)$, it follows that

$$f(z^k, x^*) \geq f(z^k, x^k) + \langle g^k, x^* - x^k\rangle.$$

Furthermore, since $f(z^k, x^*) \leq 0$ and $\sigma_k = \dfrac{f(z^k, x^k)}{\|g^k\|^2}$, we obtain from the previous inequality that

$$\langle g^k, x^k - x^*\rangle \geq \sigma_k\|g^k\|^2.$$

Using this inequality in (4.22), we deduce that

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 + \tau\|\gamma_k\sigma_k g^k\|^2 - 2\tau\gamma_k\|\sigma_k g^k\|^2\\
&= \|x^k - x^*\|^2 - \tau\gamma_k(2-\gamma_k)(\sigma_k\|g^k\|)^2.
\end{aligned}$$

In particular, this implies that the sequence $\{x^k\}$ is bounded.
(ii) We easily deduce from part (i) that for all $m \in I\!N$, we have

$$0 \leq \sum_{k=0}^{m} \tau\gamma_k(2-\gamma_k)(\sigma_k\|g^k\|)^2 \leq \|x^0 - x^*\|^2 - \|x^{m+1} - x^*\|^2 \leq \|x^0 - x^*\|^2.$$

So, taking $m \to +\infty$, we obtain

$$\sum_{k=0}^{+\infty} \gamma_k(2-\gamma_k)(\sigma_k\|g^k\|)^2 < +\infty.$$

□

Lemma 4.4. *Let $\bar{x}$ be a limit point of $\{x^k\}$ and let $x^{k_j} \to \bar{x}$. Then the sequences $\{y^{k_j}\}$, $\{z^{k_j}\}$ and $\{g^{k_j}\}$ are bounded providing that $c_{k_j} \leq \bar{c}$ for all j.*

Proof. Since the sequence $\{x^k\}$ is bounded, it suffices to prove that there exists M such that $\|x^{k_j} - y^{k_j}\| \leq M$ for j large enough to obtain that the sequence $\{y^{k_j}\}$ is bounded. Without loss of generality, we suppose, that $y^{k_j} \neq x^{k_j}$ for all j, and we set $S(y) = c_{k_j} f(x^{k_j}, y) + \bar{D}_\varphi(y, x^{k_j})$.

Since $f(x^{k_j}, \cdot)$ is convex and since, by Lemma 4.1(i), the function $D_\varphi(\cdot, x^{k_j})$ is strongly convex on $\operatorname{int} C$ with modulus $\nu > 0$, we have, for all $y_1, y_2 \in \operatorname{int} C$, $g_1 \in \partial S(y_1)$ and $g_2 \in \partial S(y_2)$ that

$$\langle g_1 - g_2, y_1 - y_2 \rangle \geq \nu \|y_1 - y_2\|_A^2 \geq \nu\, \lambda_{min}(A^T A)\|y_1 - y_2\|^2,$$

where $\lambda_{min}(A^T A)$ denotes the smallest eigenvalue of the matrix $A^T A$.

Taking $y_1 = x^{k_j}$ and $y_2 = y^{k_j}$ and noting that $0 \in \partial S(y^{k_j})$ by definition of y^{k_j}, we deduce from the previous inequality that

$$\forall g_j \in \partial S(x^{k_j}) \quad \nu\, \lambda_{min}(A^T A)\|x^{k_j} - y^{k_j}\|^2 \leq \langle g_j, x^{k_j} - y^{k_j} \rangle \leq \|g_j\|\, \|x^{k_j} - y^{k_j}\|.$$

Since $y^{k_j} \neq x^{k_j}$, and since, by Lemma 4.1(iii), $\nabla_1 D_\varphi(x^{k_j}, x^{k_j}) = 0$, we can write

$$\forall g_j \in \partial_2 f(x^{k_j}, x^{k_j}) \quad \nu\, \lambda_{min}(A^T A)\|x^{k_j} - y^{k_j}\| \leq \|g_j\|. \tag{4.23}$$

On the other hand, let the sequence $\{f_j\}_{j \in I\!N}$ be defined for all $j \in I\!N$ by $f_j = f(x^{k_j}, \cdot)$. By continuity of f, this sequence of convex functions converges pointwise to the convex function $f(\bar{x}, \cdot)$. Since $x^{k_j} \to \bar{x} \in \Lambda$ and since $f(\bar{x}, \cdot)$ is finite on Λ, it follows from Theorem 24.5 in [74] that there exists an index j_0 such that

$$\forall j \geq j_0 \quad \partial f(x^{k_j}, x^{k_j}) \subset \partial f(\bar{x}, \bar{x}) + B,$$

where B is the closed Euclidean unit ball of $I\!R^n$. Since $g_j \in \partial_2 f(x^{k_j}, x^{k_j})$ for all j and $\partial_2 f(\bar{x}, \bar{x})$ is bounded, this inclusion implies that the right-hand side of (4.23) is bounded. So there exists

$M > 0$ such that $\|x^{k_j} - y^{k_j}\| \leq M$ for all $j \geq j_0$, and the sequence $\{y^{k_j}\}$ is bounded. The sequence $\{z^{k_j}\}$ being a convex combination of x^{k_j} and y^{k_j}, it is very easy to see that the sequence $\{z^{k_j}\}$ is also bounded and that there exists a subsequence of $\{z^{k_j}\}$, again denoted $\{z^{k_j}\}$, that converges to $\bar{z} \in C$.

Finally, to prove that the sequence $\{g^{k_j}\}$ is bounded, we proceed exactly as for the sequence $\{g_j\}$ but this time with the sequence $\{f_j\}_{j\in I\!N}$ defined for all $j \in I\!N$ by $f_j = f(z^{k_j}, \cdot)$. □

Thanks to Lemmas 4.3 and 4.4, we can deduce the following convergence result.

Theorem 4.3. *Suppose that the properties* $(D1)$ *and* $(D2)$ *are satisfied and that* $0 < c \leq c_k \leq \bar{c}$ *for all* k*. Then the following statements hold:*
(i) *Every limit point of* $\{x^k\}$ *is a solution to problem EP.*
(ii) *If* $S^* = S^*_d$ *then the whole sequence* $\{x^k\}$ *converges to a solution of problem EP.*

Proof. (i) Let $\bar{x}$ be a limit point of $\{x^k\}$ and $x^{k_j} \to \bar{x}$. Applying Lemma 4.3 (ii) and (4.17), we deduce that

$$\sigma_{k_j}\|g^{k_j}\| \to 0,$$

i.e., by using the definition of σ_{k_j}, that

$$\frac{f(z^{k_j}, x^{k_j})}{\|g^{k_j}\|} \to 0.$$

Since, by Lemma 4.4, the sequence $\{g^{k_j}\}$ is bounded, we obtain that $f(z^{k_j}, x^{k_j}) \to 0$ as $j \to +\infty$. Furthermore, it follows from (4.21) that for all j,

$$f(z^{k_j}, x^{k_j}) - f(z^{k_j}, y^{k_j}) \leq \frac{1}{\theta^m} f(z^{k_j}, x^{k_j}).$$

Combining this inequality with (4.19) and noting, from (4.2), that $D_\varphi(y^{k_j}, x^{k_j}) \geq \frac{\nu}{2}\|y^{k_j} - x^{k_j}\|_A^2$, we have

$$\frac{\alpha\,\nu}{2c_{k_j}}\|y^{k_j} - x^{k_j}\|_A^2 \leq \frac{1}{\theta^m} f(z^{k_j}, x^{k_j}).$$

Consequently, since $c_{k_j} \leq \bar{c}$ for all j and $f(z^{k_j}, x^{k_j}) \to 0$ as $j \to +\infty$, we have

$$\lim_{j\to+\infty} \|y^{k_j} - x^{k_j}\|_A^2 = 0,$$

and $y^{k_j} \to \bar{x}$ because $x^{k_j} \to \bar{x}$. Finally, using Theorem 4.1 and Lemma 4.1, we obtain again inequality (4.15). Taking the limit $j \to +\infty$ in (4.15), using the continuity of f and observing that $f(\bar{x}, \bar{x}) = 0$ and $0 < c \leq c_{k_j} \leq \bar{c}$ for all j, we deduce immediately that $f(\bar{x}, y) \geq 0$ for all

$y \in C$, i.e., $\bar{x}$ is a solution to problem EP.
(ii) Let $\bar{x} \in S^*$ be a limit point of the sequence $\{x^k\}$. Because $S^* = S^*_d$, it follows that $\bar{x} \in S^*_d$. Applying Lemma 4.3 (i), we have that the sequence $\{\|x^k - \bar{x}\|\}_k$ is nonincreasing and since it has a subsequence converging to 0, it converges to zero. Hence, the whole sequence $\{x^k\}$ converges to $\bar{x} \in S^*$. □

Remark. The (IPE) and (IPLE) algorithms can be interpreted as prediction-correction methods. Indeed, Step 1 gives a prediction step while Step 2 for (IPE) and Step 3 for (IPLE) bring a correction step. Recently, such strategies have been intensively used for solving nonlinear complementarity problems (NLC), i.e., problems where the constraint set and the equilibrium function are given by

$$C = I\!R^n_+ \quad \text{and} \quad f(x,y) = \langle F(x), y - x\rangle \text{ for all } x, y \in C, \tag{4.24}$$

with $F : I\!R^n \to I\!R^n$ a (pseudo)monotone and continuous mapping (see, for example, [21], [71], [87] and [89]).

In these papers, the proximal-point iteration is used in the prediction step and consists, given x^k, in finding a solution $\tilde{x}^k$ of the system in x:

$$c_k F(x) + x - (1-\mu)x^k - \mu X_k^2 x^{-1} = \xi^k$$

when $\varphi(t) = \frac{\nu}{2}(t-1)^2 + \mu(t - \log t - 1)$ and of the system

$$c_k F(x) + x - x^k + \mu X_k \log \frac{x}{x^k} = \xi^k$$

when $\varphi(t) = \frac{1}{2}(t-1)^2 + \mu(t \log t - t + 1)$.
Here X_k =diag(x^k) and x^{-1} denotes the vector $(x_1^{-1}, \ldots, x_n^{-1})$. Furthermore, the error ξ^k must satisfy the condition: $\|\xi^k\| \leq \eta\|x^k - \tilde{x}^k\|$, $0 < \eta < 1$.

For the NLC problem, a practical choice for ξ^k is to take $\xi^k = c_k F(x) - c_k F(x^k)$ so that the two previous systems become

$$c_k F(x^k) + x - (1-\mu)x^k - \mu X_k^2 x^{-1} = 0 \text{ and } c_k F(x^k) + x - x^k + \mu X_k \log \frac{x}{x^k} = 0,$$

following the choice of φ. These systems are in fact the optimality conditions associated with Step 1 in our two algorithms when C and f are as in (4.24). Let us also observe that the correction step in [71] is similar to Step 2 in (IPE) and the ones in [21, 87, 89] to Step 3 in

(IPLE) when no Armijo-backtracking linesearch is done. In this sense, we can say that our algorithms can be considered as generalizations of the algorithms mentioned above for solving the NLC problem. To end this section, let us also notice that a comparison with Solodov and Svaiter's method [79] is developed in [21].

4.4 Numerical Results

The aim of this section is to illustrate the proposed algorithms on a class of equilibrium problems where $C = \mathbb{R}^n_+$ and the equilibrium function
$f : C \times C \to \mathbb{R}$ is of the form

$$f(x,y) = \langle Px + Qy + q, y - x \rangle,$$

with P and Q two matrices of dimension n. The corresponding equilibrium problem is a generalized form of an equilibrium problem defined by the Nash-Cournot equilibrium model considered in [67]. Let us also notice that this problem, in general, is not a variational inequality problem.

In order to fulfill the assumptions imposed in the previous sections, we suppose that the matrices P and Q are chosen such that Q is symmetric positive definite and $Q - P$ is negative semidefinite. Under these assumptions, it can be proven (see [72], p.23) that f is continuous and monotone, that $f(x, \cdot)$ is differentiable and convex for all $x \in C$ and that condition (4.8) is satisfied with $d_1 = d_2 = \frac{1}{2}\|P - Q\|$.

With this choice of function f, solving subproblem (4.4) amounts to solving the subproblem

$$\min_{y \in \mathbb{R}^n} \{ \, g(y) + D_\varphi(y, x^k) \, \} \tag{4.25}$$

where $g(y) = c_k y^T Q y + c_k b^T y$ and $b = (P - Q)x + q$. The domain of the objective function of this problem is $\mathbb{R}^n_+$. So it is advisable to first solve its Fenchel dual

$$\min_{u \in \mathbb{R}^n} \{ \, g^*(u) + D_\varphi(\cdot, x^k)^*(-u) \, \}$$

for the reason that its objective function is finite everywhere. Indeed the domain of φ^* is equal to $\mathbb{R}$ and for all $u \in \mathbb{R}^n$, the functions

$$g^*(u) = \frac{1}{4c_k} \langle u - c_k\, b, Q^{-1}(u - c_k\, b) \rangle \text{ and } D_\varphi(\cdot, x^k)^*(u) = \sum_{j=1}^{n} (x^k_j)^2 \, \varphi^*(\frac{u_j}{x^k_j})$$

are finite.

Furthermore, since $\varphi^*(t)$ and $(\varphi^*)'(t)$ can be explicitly computed [12], it is possible to solve the Fenchel dual by using an efficient unconstrained optimization method. Let u^* denote the solution of this problem. Then the solution y^k of the subproblem (5.12) can be recovered by using the formula

$$(y^k)_j = x_j^k(\varphi^*)'(-\frac{u_j^*}{x_j^k}) \text{ for all } j = 1, \dots, n.$$

To illustrate our two algorithms, we introduce three academic numerical tests of small size. Our purpose is to compare the behavior of the two algorithms. The data are the following ones: for the first two examples, the matrix Q and the vector q are

$$Q = \begin{bmatrix} 1.6 & 1 & 0 & 0 & 0 \\ 1 & 1.6 & 0 & 0 & 0 \\ 0 & 0 & 1.5 & 1 & 0 \\ 0 & 0 & 1 & 1.5 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix} \quad \text{and} \quad q = \begin{bmatrix} -1 \\ -2 \\ -1 \\ 2 \\ -1 \end{bmatrix},$$

while for the third example, they are

$$Q = \begin{bmatrix} 2.3550 & 1.6364 & 1.8430 & 2.1540 & 0.7586 \\ 1.6364 & 1.6620 & 1.5323 & 1.4876 & 0.2901 \\ 1.8430 & 1.5323 & 2.4317 & 2.2961 & 1.0964 \\ 2.1540 & 1.4876 & 2.2961 & 2.8473 & 1.2273 \\ 0.7586 & 0.2901 & 1.0964 & 1.2273 & 0.8085 \end{bmatrix} \quad \text{and} \quad q = \begin{bmatrix} -1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The matrix P is chosen successively equal to

$$\begin{bmatrix} 3.1 & 2.0 & 0 & 0 & 0 \\ 2.0 & 3.6 & 0 & 0 & 0 \\ 0 & 0 & 3.5 & 2 & 0 \\ 0 & 0 & 2 & 3.3 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}, \quad \begin{bmatrix} 3.1 & 2.0 & 0 & 0 & 0 \\ 2.0 & 3.6 & 0 & 0 & 0 \\ 0 & 0 & 3.5 & 2 & 0 \\ 0 & 0 & 2 & 3.3 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix} \quad \text{and} \quad 10I,$$

where I denotes the identity matrix. The parameters are fixed to $\nu = 7$, $\mu = 1$, $c_k = 1/d_1$ for Algorithm IPE, and $\nu = 2$, $\mu = 1$, $\theta = 0.99$, $\alpha = 0.49$, $\tau = 0.999$ for Algorithm IPLE. For this algorithm, c_k is equal to 0.7 for the first two examples and to 0.1 for the third one. Finally the starting point is $x_0 = (1, 3, 1, 1, 2)$ for all the tests. The results are reported in the table below:

	Example 1		Example 2		Example 3	
Algorithm	IPE	IPLE	IPE	IPLE	IPE	IPLE
it	19	1305	20	1342	40	228
cpu (sec.)	1.078	26.89	1.296	27.64	10.875	13.25
optimality	-0.00000	-0.00257	-0.00000	-0.00237	-0.00006	-0.00152

where 'it' and 'cpu' stand for the number of iterations and the cpu time, respectively.

The two algorithms give the same solution for each example: for instance, in Example 1 the solution obtained from the algorithm IPE is $(0.00000, 0.38463, 0.20002, 0.00000, 0.19995)$ and the solution obtained from the algorithm ILPE is $(0.00000, 0.40733, 0.19610, 0.00000, 0.19777)$. As have been seen, two constraints are active at the solution for Examples 1 and 2. Three constraints are active for the third example as well. Furthermore for checking the quality of the solution x obtained by each algorithm, we solve the minimization problem $\min_{y \geq 0} f(x, y)$ whose optimal value must be equal to zero when x is the exact solution to problem EP. This optimal value is denoted 'optimality' in the table.

From the preliminary numerical results reported in the table, the first algorithm seems to be the most efficient. For each example, the total number of iterations is much smaller for this algorithm than for the second one as well as the cpu time. Furthermore it is also the most robust in the sense that the quality of the solution is the best. But this could be due to the fact that for the second algorithm, an unconstrained minimization problem is replaced at each iteration by a gradient step which usually slows down the convergence.

Chapter 5

Bundle Interior Proximal Algorithm for Convex Minimization Problems

In this chapter, we extend the bundle proximal point method for finding the minimum of a convex not necessarily differentiable function on the nonnegative orthant. The strategy consists in approximating the objective function by a piecewise linear convex function and using distance–like functions based on second order homogeneous kernels. First we prove the convergence of this new bundle interior proximal method under the same assumptions as for the standard bundle method and then we report some preliminary numerical experiences for a particular distance function.

5.1 Preliminaries

Let $F : I\!R^n \to I\!R \cup \{+\infty\}$ be a lower semicontinuous proper convex function such that $I\!R^n_+ \equiv \{\, x \in I\!R^n \,:\, x \geq 0 \,\} \subseteq \text{int dom}\, F$. The problem is to find the minimum of F on $I\!R^n_+$. For solving this problem, we use the proximal–like scheme defined by

$$x^{k+1} = \arg \min_{y \in I\!R^n_{++}} \{F(x) + \frac{1}{c_k} d_\varphi(y, x^k)\}, \tag{5.1}$$

where $\{c_k\}_{k \in I\!N}$ is a sequence of positive real numbers and the distance-function $d_\varphi(x, y) = \sum_{j=1}^n y_j^2 \varphi(\frac{x_j}{y_j})$ for all $x, y \in I\!R^n_{++} \equiv \{\, x \in I\!R^n \,:\, x > 0 \,\}$ based on a function $\varphi \in \Phi$. Here the class Φ contains all the lower semicontinuous, proper and convex functions $\varphi : I\!R \to I\!R \cup \{+\infty\}$ that satisfy the following properties:

1. dom $\varphi \subseteq [0, +\infty)$;

2. φ is twice continuously differentiable on int(domφ) = $(0, +\infty)$;

3. φ is strictly convex on its domain;

4. $\lim_{t \to 0^+} \varphi'(t) = -\infty$;

5. $\varphi(1) = \varphi'(1) = 0$ and $\varphi''(1) > 0$.

It is easy to see that the function d_φ has the following basic properties:

1. d_φ is an homogeneous function of order 2, i.e.,

$$d_\varphi(\alpha x, \alpha y) = \alpha^2 \, d_\varphi(x, y) \text{ for all } \alpha > 0 \text{ for all } x, y \in I\!R^n_{++},$$

2. $d_\varphi(x, y) \geq 0$ for all $x, y \in I\!R^n_{++}$,

3. $d_\varphi(x, y) = 0$ if and only if $x = y$.

The function φ being differentiable and convex on $(0, +\infty)$, the function $d_\varphi(\cdot, y)$ is differentiable and convex on $I\!R^n_{++}$ for any $y \in I\!R^n_{++}$. Hence x^k is a minimum of the function $F + d_\varphi(\cdot, x^{k-1})$ if and only if

$$0 \in \partial F(x^k) + \frac{1}{c_k} \Psi(x^k, x^{k-1}),$$

where ∂F denotes the subdifferential of F and

$$\Psi(a, b) = \left(b_1 \, \varphi' \left(\frac{a_1}{b_1} \right), \; b_2 \, \varphi' \left(\frac{a_2}{b_2} \right), ..., b_n \, \varphi' \left(\frac{a_n}{b_n} \right) \right) \text{ for all } a, b \in I\!R^n_{++}. \tag{5.2}$$

With these definitions, the basic iteration scheme (BIS) introduced by Auslender et al. [9] for finding the minimum of F on $I\!R^n_+$ can be expressed as

Given $\varphi \in \Phi$, $x^0 \in I\!R^n_{++}$, $\varepsilon_k \geq 0, c_k > 0$, generate the sequences $\{x^k\} \subseteq I\!R^n_{++}$ and $\{g^k\}$ satisfying

$$g^k \in \partial_{\varepsilon_k} F(x^k) \quad \textit{and} \quad c_k g^k + \Psi(x^k, x^{k-1}) = 0, \tag{5.3}$$

where $\partial_{\varepsilon_k} F(x^k)$ denotes the ε_k–subdifferential of F at x^k.

From (5.3), we have that

$$0 \in \partial_{\epsilon_k} F(x^k) + \frac{1}{c_k} \Psi(x^k, x^{k-1}).$$

This means that x^k is an ϵ_k–minimum of the function $F + d_\varphi(\cdot, x^{k-1})$.

Our aim is to present a bundle version of this algorithm and to study its convergence. In that purpose, we need to introduce a subclass of Φ defined by

$$\Phi_0 = \{h \in \Phi : h''(1)(1 - \frac{1}{t}) \leq h'(t) \leq h''(1)(t-1) \text{ for all } t > 0\}, \tag{5.4}$$

and to consider a specific choice for the functions φ we will use, namely,

$$\varphi(t) := \mu h(t) + \frac{\nu}{2}(t-1)^2, \tag{5.5}$$

where $\mu > 0$, $\nu > 0$ and $h \in \Phi_0$.

The kernel h is used to enforce the iterates to stay in the interior of the nonnegative orthant while the quadratic term $(t-1)^2$ gives rise to the usual term used in "regularization". It is easy to see that the following functions belong to Φ_0:

$$h_1(t) = t \log t - t + 1, \quad \text{dom}\, h_1 = [\,0, +\infty);$$

$$h_2(t) = -\log t + t - 1, \quad \text{dom}\, h_2 = (0, +\infty);$$

$$h_3(t) = 2(\sqrt{t} - 1)^2, \quad \text{dom}\, h_3 = [0, +\infty).$$

When the functions φ are defined by (5.5) with $h \in \Phi_0$ and $\nu \geq \mu h''(1) > 0$, Auslender and al. ([9], Theorem 3.2) proved that the sequence $\{x^k\}$ generated by the BIS algorithm converges to a minimum of F on $\mathbb{R}^n_+$ provided that $\sum c_k = +\infty$ and $\sum c_k \varepsilon_k < +\infty$.

5.2 Bundle Interior Proximal Algorithm

Let $F : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Since F can be nondifferentiable, we observe that finding x^{k+1} by using (5.1) is often as difficult as finding the minimum of F over $\mathbb{R}^n_+$. So the strategy to get an implementable algorithm, is to approximate F at iteration k by a simpler convex function in such a way that the resulting problem is easy to solve. Using the same strategy has been used for the classical proximal method, we propose a new method that incorporates the bundle proximal point algorithm and the BIS scheme.

Bundle Interior Proximal Algorithm

Data: Let $x^0 \in C$, $\sigma \in (0,1)$ and let $\{c_k\}_{k\in I\!N}$ be a sequence of positive numbers.

Step 1. Set $y_0^0 = x^0$ and $k = 0, i = 1$.

Step 2. Choose a piecewise linear convex function φ_i^k satisfying $(A1)-(A3)$ and solve

$$\min_{y\in I\!R^n_{++}} \{\varphi_i^k(y) + \frac{1}{c_k} d_\varphi(y, x^k)\},$$

to obtain the unique optimal solution y_i^k.

Step 3. If

$$F(x^k) - F(y_i^k) \geq \sigma[F(x^k) - \varphi_i^k(y_i^k)], \tag{5.6}$$

then set $x^{k+1} = y_i^k$, $y_0^{k+1} = x^{k+1}$, replace k by $k+1$ and set $i = 0$.

Step 4. Replace i by $i+1$ and go to Step 2.

Remark 5.1. *Observe that the reduction predicted by the model φ_i^k, namely $F(x^k) - \varphi_i^k(y_i^k)$ is nonnegative. Indeed, since $F \geq \varphi_i^k$ and $\gamma_i^k \in \partial\varphi_i^k(y_i^k)$, we have*

$$\begin{aligned} F(x^k) - \varphi_i^k(y_i^k) &\geq \varphi_i^k(x^k) - \varphi_i^k(y_i^k) \geq \langle \gamma_i^k, x^k - y_i^k \rangle \\ &= -\tfrac{1}{c_k}\langle \Psi(y_i^k, x^k), x^k - y_i^k \rangle \geq 0. \end{aligned}$$

To prove the convergence of the bundle interior proximal method, we also need to introduce the following notations

$$\tilde{l}^i(y) = l_i^k(y) + \frac{1}{c_k} d_\varphi(y, x^k),$$
$$\tilde{F}^i(y) = \varphi_i^k(y) + \frac{1}{c_k} d_\varphi(y, x^k).$$

Then we have

$$\tilde{F}^i(x^k) = \varphi_i^k(x^k) \text{ and } \tilde{l}^i(y_i^k) = \tilde{F}^i(y_i^k). \tag{5.7}$$

Indeed, $d_\varphi(x^k, x^k) = 0$ and

$$\tilde{l}^i(y_i^k) = l_i^k(y_i^k) + \frac{1}{c_k} d_\varphi(y_i^k, x^k) = \varphi_i^k(y_i^k) + \frac{1}{c_k} d_\varphi(y_i^k, x^k) = \tilde{F}^i(y_i^k).$$

Lemma 5.1. *There exists $\beta > 0$ such that, for all i,*

$$\tilde{l}^i(y) \geq \tilde{l}^i(y_i^k) + \frac{\beta}{2\,c_k}\|y - y_i^k\|^2.$$

Proof. By definition of $\tilde{l}^i$, we have

$$\begin{aligned}\tilde{l}^i(y) - \tilde{l}^i(y_i^k) &= l_i^k(y) + \tfrac{1}{c_k} d_\varphi(y, x^k) - l_i^k(y_i^k) - \tfrac{1}{c_k} d_\varphi(y_i^k, x^k) \\ &= \langle \gamma_i^k, y - y_i^k \rangle + \tfrac{1}{c_k} [\, d_\varphi(y, x^k) - d_\varphi(y_i^k, x^k) \,].\end{aligned} \tag{5.8}$$

Since $d_\varphi(x, x^k) = \displaystyle\sum_{i=1}^n (x_i^k)^2 \varphi(\frac{x_i}{x_i^k})$ and φ is strongly convex on $\{t \in I\!R \,|\, t > 0\}$, the function d_φ is itself strongly convex on $I\!R^n_{++}$, i.e., there exists $\beta > 0$ such that, for all $y \in I\!R^n_{++}$,

$$d_\varphi(y, x^k) - d_\varphi(y_i^k, x^k) \geq \langle \Psi(y_i^k, x^k), y - y_i^k \rangle + \frac{\beta}{2} \|y - y_i^k\|^2.$$

Using this inequality in (5.8) and noting that $\Psi(y_i^k, x^k) = -c_k \gamma_i^k$, we obtain

$$\tilde{l}^i(y) - \tilde{l}^i(y_i^k) \geq \frac{\beta}{2\, c_k} \|y - y_i^k\|^2.$$

□

Proposition 5.1. *Suppose that after x^k has been obtained in the bundle interior proximal algorithm, the test of sufficient reduction is suppressed : only null–steps are made. If the sequence $\{\varphi_i^k\}$ satisfies conditions (A1) and (A2), then*

$$(1) \quad F(y_i^k) - \varphi_i^k(y_i^k) \to 0,$$

$$(2) \quad y_i^k \to y^* = \arg\min_{x>0} \{F(x) + \tfrac{1}{c_k} d_\varphi(x, x^k)\}.$$

Proof

(1) We use three steps to prove this part.

(i) $\tilde{l}^i(y_i^k)$ is convergent and $y_{i+1}^k - y_i^k \to 0$.
For $i = 1, \ldots,$ we have

$$\begin{aligned} F(x^k) &\geq \varphi_{i+1}^k(x^k) && \text{by (A1)} \\ &= \tilde{F}^{i+1}(x^k) && \text{by (5.7)} \\ &\geq \tilde{F}^{i+1}(y_{i+1}^k) && \text{by definition of } y_{i+1}^k \\ &= \tilde{l}^{i+1}(y_{i+1}^k) && \text{by (5.7)} \\ &\geq \tilde{l}^i(y_{i+1}^k) && \text{by (A2)} \\ &\geq \tilde{l}^i(y_i^k) + \tfrac{\beta}{2c_k} \|y_{i+1}^k - y_i^k\|^2 && \text{by Lemma 5.1 with } y = y_{i+1}^k. \end{aligned}$$

From these relations, we have for all i,

$$\tilde{l}^i(y_i^k) \leq \tilde{l}^{i+1}(y_{i+1}^k) \text{ and } \tilde{l}^i(y_i^k) \leq F(x^k).$$

Hence the sequence $\{\tilde{l}^i(y_i^k)\}$ is convergent in $I\!R$. Moreover, by Lemma 5.1, we have

$$\tilde{l}^{i+1}(y_{i+1}^k) - \tilde{l}^i(y_i^k) \geq \frac{\beta}{2c_k}\|y_{i+1}^k - y_i^k\|^2 \geq 0.$$

Hence $y_{i+1}^k - y_i^k \to 0$.

(ii) The sequence $\{y_i^k\}$ is bounded.
We have (for y fixed)

$$\begin{array}{rcll} F(y) + \frac{1}{c_k} d_\varphi(y, x^k) & \geq & \varphi_{i+1}^k(y) + \frac{1}{c_k} d_\varphi(y, x^k) & \text{by (A1)} \\ & = & \tilde{F}^{i+1}(y) & \text{by definition of } \tilde{F}^{i+1} \\ & \geq & \tilde{l}^i(y) & \text{by (A2)} \\ & \geq & \tilde{l}^i(y_i^k) + \frac{\beta}{2c_k}\|y - y_i^k\|^2 & \text{by Lemma 5.1.} \end{array}$$

Since the sequence $\{\tilde{l}^i(y_i^k)\}$ is convergent, it is bounded and thus also the sequences $\{\|y - y_i^k\|^2\}$ and $\{y^i\}$.

(iii) $F(y_{i+1}^k) - \varphi_{i+1}^k(y_{i+1}^k) \to 0$.
By definition of $s(y_{i+1}^k)$, we have

$$\langle s(y_i^k), y_{i+1}^k - y_i^k \rangle \leq \varphi_{i+1}^k(y_{i+1}^k) - F(y_i^k) \leq F(y_{i+1}^k) - F(y_i^k) \leq \langle s(y_{i+1}^k), y_{i+1}^k - y_i^k \rangle.$$

Since the subdifferential ∂F is bounded on bounded subsets of $I\!R_{++}^n$ and the sequence $\{y_i^k\}$ is bounded, then the sequence $\{s(y_i^k)\}$ is also bounded. Taking the limit of the opposite sides of the previous inequalities, we obtain

$$\langle s(y_i^k), y_{i+1}^k - y_i^k \rangle \to 0 \quad \text{and} \quad \langle s(y_{i+1}^k), y_{i+1}^k - y_i^k \rangle \to 0,$$

and hence

$$\varphi_{i+1}^k(y_{i+1}^k) - F(y_i^k) \to 0 \quad \text{and} \quad F(y_{i+1}^k) - F(y_i^k) \to 0.$$

So

$$F(y^{i+1}) - \varphi_{i+1}^k(y_{i+1}^k) = F(y_{i+1}^k) - F(y_i^k) + F(y_i^k) - \varphi_{i+1}^k(y_{i+1}^k) \;\to\; 0.$$

(2) We also use three steps to prove this part.

(i) Any limit point $\bar{y}$ of $\{y_i^k\}$ is such that $\bar{y}_j > 0$ for all $j = 1, \ldots, n$.
Let $\{y_i^k\}_{i \in K}$ be a subsequence of $\{y_i^k\}$ converging to $\bar{y}$ and suppose, to get a contradiction that $J := \{\, j \mid \bar{y}_j = 0 \,\}$ is nonempty. By definition of γ_i^k, we can write

$$\gamma_i^k = -\frac{1}{c_k}\Psi(y_i^k, x^k) \in \partial\varphi_i^k(y_i^k).$$

Then, since $F \geq \varphi_i^k$, we have

$$\forall y \in I\!R^n \quad F(y) \geq \varphi_i^k(y) \geq \varphi_i^k(y_i^k) + \langle \gamma_i^k, y - y_i^k \rangle,$$

i.e.,

$$\forall y \in I\!R^n \quad F(y) \geq \varphi_i^k(y_i^k) - \frac{1}{c_k}\langle \Psi(y_i^k, x^k), y - y_i^k \rangle. \tag{5.9}$$

On the other hand

$$\Psi(y_i^k, x^k) = \left(x_1^k \varphi' \left(\frac{y_{i,1}^k}{x_1^k} \right), ..., x_j^k \varphi' \left(\frac{y_{i,j}^k}{x_j^i} \right), \ldots, x_n^k \varphi' \left(\frac{y_{i,n}^k}{x_n^k} \right) \right).$$

Then it is easy to see that for all $j \in J$, we have

$$\frac{y_{i,j}^k}{x_j^k} \rightarrow 0^+ \quad \text{and} \quad \varphi' \left(\frac{y_{i,j}^k}{x_j^k} \right) \rightarrow -\infty \text{ (by property (iv) of } \varphi) \tag{5.10}$$

while, for $j \notin J$, we obtain

$$\varphi' \left(\frac{y_{i,j}^k}{x_j^k} \right) \rightarrow \varphi' \left(\frac{\bar{y}_j^i}{x_j^k} \right) \in I\!R. \tag{5.11}$$

Choose $y = (1, \ldots, 1)^T$. Then, for all i, we deduce from (5.9) that

$$\begin{aligned} F(y) \geq \varphi_i^k(y_i^k) - F(y_i^k) + F(y_i^k) \quad & - \quad \tfrac{1}{c_k} \sum_{j \in J} x_j^k \varphi' \left(\frac{y_{i,j}^k}{x_j^k} \right) (1 - y_{i,j}^k) \\ & - \quad \tfrac{1}{c_k} \sum_{j \notin J} x_j^k \varphi' \left(\frac{y_{i,j}^k}{x_j^k} \right) (1 - y_{i,j}^k). \end{aligned}$$

Taking the limit as $i \rightarrow +\infty$, using part (1), the continuity of F, (5.10) and (5.11), we obtain

$$F(y) \geq +\infty.$$

This is impossible, so J is empty.

(ii) Any limit point $\bar{y}$ of $\{y_i^k\}$ is a solution of

$$\begin{cases} \text{minimize } \; F(x) + \frac{1}{c_k} d_\varphi(x, x^k), \\ \text{subject to } x \; > 0. \end{cases}$$

Let $y_i^k \rightarrow \bar{y}$, $i \in K \subseteq I\!N$. By part 2(i), $\bar{y}_j > 0$ for all $j = 1, \ldots, n$. To obtain that $\bar{y}$ is a minimum of $F + \frac{1}{c_k} d_\varphi(\cdot, x^k)$, we have to prove that $0 \in \partial F(\bar{y}) + \frac{1}{c_k}\Psi(\bar{y}, x^k)$, i.e.,

$$\forall y \in I\!R^n \quad F(y) \geq F(\bar{y}) - \frac{1}{c_k}\langle \Psi(\bar{y}, x^k), y - \bar{y} \rangle. \tag{5.12}$$

Let then $y \in I\!\!R^n$. By definition of $\gamma_i^k \in \partial\varphi_i^k(y_i^k)$ and since $F \geq \varphi_i^k$, we have

$$F(y) \geq \varphi_i^k(y) \geq \varphi_i^k(y_i^k) - \frac{1}{c_k}\langle \Psi(y_i^k, x^k), y - y_i^k \rangle,$$

i.e.,

$$F(y) \geq \varphi_i^k(y_i^k) - F(y_i^k) + F(y_i^k) - \frac{1}{c_k}\langle \Psi(y_i^k, x^k), y - y_i^k \rangle.$$

Taking the limit as $i \to +\infty$, using part (1)(iii), the continuity of F and $\Psi(\cdot, x^k)$, we obtain the required inequality (5.12).

(iii) $y_i^k \to y^* = \text{arg min}\{F(x) + \frac{1}{c_k} d_\varphi(x, x^k)\}$ when $i \to +\infty$.

By part 2(ii), any limit point of $\{y_i^k\}$ is a solution of the problem

$$\begin{cases} \text{minimize} & F(x) + \frac{1}{c_k} d_\varphi(x, x^k) \\ \text{subject to} & x > 0. \end{cases}$$

However, this problem has exactly one solution because $d_\varphi(\cdot, x^k)$ is strongly convex. So, all the limit points of $\{y_i^k\}$ coincide and thus the whole sequence $\{y_i^k\}$ converges to y^*. □

Now we can apply these results to prove the convergence of the Bundle Interior Proximal method. But first we need a lemma.

Lemma 5.2. *If* $x^k = \arg\min\{F(x) + \frac{1}{c_k} d_\varphi(x, x^k)/x > 0\}$ *then* x^k *is a minimum of* F *on* $I\!\!R_+^n$.

Proof. By optimality of x^k, we have

$$0 \in \partial F(x^k) + \frac{1}{c_k}\Psi(x^k, x^k).$$

Since $\Psi(x^k, x^k) = 0$ by definition of φ, we obtain $0 \in \partial F(x^k)$ and, since $x^k > 0$, x^k is a minimum of F over $I\!\!R_+^n$. □

Theorem 5.1. *Let* $\varphi(t) = \mu h(t) + (\nu/2)(t-1)^2$, *with* $h \in \Phi_0$, $\mu > 0$ *and* $\nu \geq \mu h''(1) > 0$. *Then in the bundle interior proximal algorithm, there are two possibilities*

(1) The index k *remains fixed, i.e., only null steps are made from* x^k. *In this case,* x^k *is a minimum of* F *on* $I\!\!R_+^n$.

(2) The index $k \to +\infty$. *Then*

- $\displaystyle\sum_{k=1}^{+\infty} c_k = +\infty \Longrightarrow F(x^k) \to \bar{F} := \inf_{x \in I\!\!R_+^n} F(x).$
- *If, in addition,* $\{c_k\}$ *is bounded, then* $x^k \to x^*$, *minimum of* F *(if there exists some minimum).*

Proof.

(1) Let i_k be the iteration index that has produced x^k. Since only null–steps are made from x^k, we have

$$\forall i > i_k \quad F(x^k) - F(y_i^k) < \sigma[F(x^k) - \varphi_i^k(y_i^k)]. \tag{5.13}$$

By Proposition 5.1, we have $y_i^k \to y^* \equiv \arg\min\{F(x)+\frac{1}{c_k}d_\varphi(x,x^k)\}$ and $F(y_i^k)-\varphi_i^k(y_i^k) \to 0$. Taking the limit in (5.13), we obtain

$$F(x^k) - F(y^*) \leq \sigma[F(x^k) - F(y^*)],$$

because $\varphi_i^k(y_i^k) = \varphi_i^k(y_i^k) - F(y_i^k) + F(y_i^k) \to F(y^*)$ and F is continuous. Hence

$$(1-\sigma)[F(x^k) - F(y^*)] \leq 0.$$

Since $1-\sigma > 0$, we have $F(x^k) \leq F(y^*)$, or again by definition of $d_\varphi(.,x^k)$,

$$F(x^k) + \frac{1}{c_k}d_\varphi(x^k,x^k) = F(x^k) \leq F(y^*) \leq F(y^*) + \frac{1}{c_k}d_\varphi(y^*,x^k).$$

Since the solution y^* is unique, we deduce that $x^k = y^*$ and by Lemma 5.2, x^k is a minimum of F on $I\!R_+^n$.

(2) Denote by $i(k)$ the iteration index where x^k is updated. Then we have $y_{i(k)}^k = x^{k+1}$. Let us also define $\gamma^k \equiv \gamma_{i(k)}^k \in \partial\varphi_{i(k)}^k(x^{k+1})$. We know that

$$\gamma^k = -\frac{1}{c_k}\Psi(x^{k+1},x^k).$$

With these notations, we prove the following assertions.

a. $\{F(x^k)\}$ is nonincreasing.

Since

$$F(x^k) - F(x^{k+1}) \geq \sigma[F(x^k) - \varphi_{i(k)}^k(x^{k+1})] \tag{5.14}$$

and since, by Remark 5.1, the reduction $F(x^k) - \varphi_{i(k)}^k(x^{k+1})$ predicted by the model is nonnegative, it follows that $\{F(x^k)\}$ is nonincreasing. In the sequel, we suppose that $\{F(x^k)\}$ is bounded from below (otherwise $F(x^k) \to -\infty$ and the proof is finished).

b. $\gamma^k \in \partial_{\varepsilon_k}F(x^k)$ with

$$\varepsilon_k = F(x^k) - \varphi_{i(k)}^k(x^{k+1}) + \frac{1}{c_k}\langle\Psi(x^{k+1},x^k), x^k - x^{k+1}\rangle.$$

By definition of γ^k, we observe immediately that

$$\varepsilon_k = F(x^k) - \varphi_{i(k)}^k(x^{k+1}) - \langle\gamma^k, x^k - x^{k+1}\rangle.$$

Moreover, since $F \geq \varphi^k_{i(k)}$ and $\gamma^k \in \partial\varphi^k_{i(k)}(x^{k+1})$, we have for all y, that

$$F(y) \geq \varphi^k_{i(k)}(y) \geq \varphi^k_{i(k)}(x^{k+1}) + \langle \gamma^k, y - x^{k+1} \rangle. \tag{5.15}$$

In particular, for $y = x^k$, we obtain that $\varepsilon_k \geq 0$. Now, from (5.15), we also have for all y, that

$$F(y) \geq F(x^k) + \varphi^k_{i(k)}(x^{k+1}) - F(x^k) + \langle \gamma^k, y - x^k \rangle + \langle \gamma^k, x^k - x^{k+1} \rangle,$$

i.e., $\gamma^k \in \partial_{\varepsilon_k} F(x^k)$.

c. $\sum_{k=1}^{+\infty} \{\varepsilon_k - \frac{1}{c_k} \langle \Psi(x^{k+1}, x^k), x^k - x^{k+1} \rangle\} < +\infty$.

By (5.14) we have

$$\begin{aligned} \varepsilon_k &= F(x^k) - \varphi^k_{i(k)}(x^{k+1}) + \tfrac{1}{c_k}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1} \rangle \\ &\leq \sigma^{-1}[F(x^k) - F(x^{k+1})] + \tfrac{1}{c_k}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1} \rangle. \end{aligned}$$

Thus

$$\begin{aligned} \sum_{k=1}^{p} \{\varepsilon_k - \frac{1}{c_k} < \Psi(x^{k+1}, x^k), x^k - x^{k+1} >\} &\leq \sigma^{-1} \sum_{k=1}^{p} [F(x^k) - F(x^{k+1})] \\ &= \sigma^{-1}[F(x^1) - F(x^{p+1})]. \end{aligned}$$

Since F is bounded from below, then

$$\sum_{k=1}^{+\infty} \{\varepsilon_k - \frac{1}{c_k} \langle \Psi(x^{k+1}, x^k), x^k - x^{k+1} \rangle\} < +\infty.$$

d. $F(x^k) \ \rightarrow \ \bar{F} = \inf\{F(x) \,|\, x \geq 0\}$.

Since the sequence $\{F(x^k)\}$ is nonincreasing, it converges to some $\bar{F}$. Suppose now, to get a contradiction, that $\bar{F} > F^* := \inf_{x \geq 0} F(x)$. Then there exist $y \in I\!R^n_+$ and $\delta > 0$ such that, for all k, $F(y) + \delta < F(x^k)$. Since

$$\varepsilon_k - \frac{1}{c_k} \langle \Psi(x^{k+1}, x^k), x^k - x^{k+1} \rangle \rightarrow \ 0,$$

there exists k_0 such that, for $k \geq k_0$,

$$\varepsilon_k - \frac{1}{c_k} \langle \Psi(x^{k+1}, x^k), x^k - x^{k+1} \rangle < \frac{\delta}{2}. \tag{5.16}$$

Using Lemma 3.4 of [9] with $a = x^k, b = x^{k+1}$ and $c = y$, we obtain

$$\|y - x^{k+1}\|^2 - \|y - x^k\|^2 \leq -\theta \langle y - x^{k+1}, \Psi(x^{k+1}, x^k) \rangle$$

$$= -\theta\langle y - x^k, \Psi(x^{k+1}, x^k)\rangle - \theta\langle x^k - x^{k+1}, \Psi(x^{k+1}, x^k)\rangle, \tag{5.17}$$

where $\theta = [(\nu + \mu h''(1))/2]^{-1}$.

By (5.16), we have immediately

$$-\theta\langle x^k - x^{k+1}, \Psi(x^{k+1}, x^k)\rangle < c_k\theta\left(\frac{\delta}{2} - \varepsilon_k\right). \tag{5.18}$$

On the other hand, by definition of $\gamma^k = -\frac{1}{c_k}\Psi(x^{k+1}, x^k) \in \partial_{\varepsilon_k} F(x^k)$, and by part b., we have successively

$$-\theta\langle y - x^k, \Psi(x^{k+1}, x^k)\rangle = c_k\theta\langle \gamma^k, y - x^k\rangle \tag{5.19}$$

and

$$F(x^k) - \delta > F(y) \geq F(x^k) + \langle \gamma^k, y - x^k\rangle - \varepsilon_k. \tag{5.20}$$

Combining (5.19) and (5.20) yields

$$-\theta\langle y - x^k, \Psi(x^{k+1}, x^k)\rangle < c_k\theta[-\delta + \epsilon_k]. \tag{5.21}$$

Finally we obtain, from (5.17), (5.18) and (5.21), that

$$\|y - x^{k+1}\|^2 \leq \|y - x^k\|^2 + c_k\theta[\frac{\delta}{2} - \varepsilon_k - \delta + \varepsilon_k] = \|y - x^k\|^2 - c_k\theta\frac{\delta}{2}.$$

Summing up, we obtain, for all $k > k_0$,

$$0 \leq \|x^k - y\|^2 \leq \|x^{k_0} - y\|^2 - \frac{\delta\theta}{2}\sum_{k=k_0}^{k-1} c_k.$$

Taking the limit as $k \to +\infty$, we have $\sum_{k=k_0}^{+\infty} c_k \leq 2\theta^{-1}\delta^{-1}\|x^{k_0} - y\|^2 < +\infty$ which contradicts the assumption $\sum_{k=1}^{+\infty} c_k = +\infty$.

Now suppose that F has a minimum $\bar{x}$ on $I\!R^n_+$ and that the sequence $\{c_k\}$ is bounded.

e. $\{x^k\}$ is bounded.

Using inequality (5.17) with $y = \bar{x}$, we obtain

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \theta\langle \bar{x} - x^k, \Psi(x^{k+1}, x^k)\rangle - \theta\langle x^k - x^{k+1}, \Psi(x^{k+1}, x^k)\rangle.$$

By definition of $\gamma^k = -\frac{1}{c_k}\Psi(x^{k+1}, x^k) \in \partial_{\varepsilon_k} F(x^k)$, we have

$$\begin{aligned} -\theta\langle \bar{x} - x^k, \Psi(x^{k+1}, x^k)\rangle &= c_k\theta\langle \gamma^k, \bar{x} - x^k\rangle \\ &\leq c_k\theta[F(\bar{x}) - F(x^k) + \varepsilon_k] \leq c_k\theta\varepsilon_k. \end{aligned}$$

So

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 + c_k\theta[\epsilon_k - \frac{1}{c_k}\langle\Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle]. \tag{5.22}$$

Since $\{c_k\}$ is bounded (by assumption) and since, by part c., we have

$$\sum_{k=1}^{+\infty} c_k\theta[\varepsilon_k - \frac{1}{c_k}\langle\Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle] < +\infty.$$

Using Lemma 3.1 of [9], we deduce that the sequence $\{\|x^k - x\|\}_k$ is convergent. Thus $\{x^k\}$ is bounded.

f. Any limit point x^* of $\{x^k\}$ is a minimum of F on $I\!R^n_+$ and $x^k \to x^*$.

Let $x^{n_k} \to x^*$. By continuity of F, we have $F(x^{n_k}) \to F(x^*)$. By part d., $F(x^k) \to \bar{F} = \inf_{x\geq 0} F(x)$. So $F(x^{n_k}) \to \bar{F}$ and, by the uniqueness of the limit, $F(x^*) = \bar{F}$. Since $x^* \in I\!R^n_+$, then x^* is a minimum of F on $I\!R^n_+$. Using (5.22) with x^* instead of $\bar{x}$, we obtain

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + \delta_k,$$

where

$$\delta_k = c_k\theta[\varepsilon_k - \frac{1}{c_k}\langle\Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle].$$

Since $\sum_{k=1}^{+\infty} \delta_k < +\infty$, it follows from Proposition 1.3 of [28] that the whole sequence $\{x^k\}$ converges to x^*. □

5.3 Numerical Results

To obtain that the bundle interior proximal algorithm is implementable, it remains to explain how to solve the subproblem

$$\begin{cases} \text{minimize } \ \varphi_i^k(y) + \frac{1}{c_k} d_\varphi(y, x^k) \\ \text{subject to } y \ \in \ I\!R^n_{++}. \end{cases}$$

Since $\varphi_i^k(y) = \max\{F(y^j) + \langle s(y^j), y - y^j\rangle \mid j = 0, \ldots, i-1\}$, this problem is equivalent to

$$(SP)_{k,i} \begin{cases} \text{minimize} & v + \frac{1}{c_k} d_\varphi(y, x^k) \\ \text{subject to} & v \geq F(y^j) + \langle s(y^j), y - y^j\rangle \quad j = 0, \ldots, i-1 \\ & y \in I\!R^n_{++}. \end{cases}$$

Observe that if (y_i^k, v^i) is a solution of this problem, then

$$v^i = \max_{0 \le j \le i-1} \{F(y^j) + \langle s(y^j), y - y^j \rangle\}$$

so that the stopping criterion for the inner iterations is

$$F(x^k) - F(y_i^k) \ge \sigma[F(x^k) - v^i].$$

Since $\varphi(t) = \mu h(t) + \frac{\nu}{2}(t-1)^2$, the objective function of $(SP)_{k,i}$ is highly nonlinear and finding the solution of $(SP)_{k,i}$ can be very hard. However, if we observe that

$$d_\varphi(y, x^k) = \sum_{m=1}^{n} (x_m^k)^2 \, \varphi\left(\frac{y_m}{x_m^k}\right),$$

then the objective function is separable and one way of solving such a problem is to solve its dual. Setting $z_m = \frac{y_m}{x_m^k}$ for all $m = 1, \ldots, n$ and $z = (z_m)$, problem $(SP)_{k,i}$ can be expressed as

$$(MSP)_{k,i} \begin{cases} \text{minimize} \quad v + \displaystyle\sum_{m=1}^{n} \alpha_m \varphi(z_m) \\ \text{subject to} \quad \langle s^j, z \rangle - v \le b_j \quad j = 0, \ldots, i-1, \end{cases}$$

where $\alpha_m = \frac{1}{c_k}(x_m^k)^2, s_m^j = s(y^j)_m x_m^k$, $m = 1, \ldots, n$ and $b_j = \langle s(y^j), y^j \rangle - F(y^j), j = 0, \ldots, i-1$. Then the Lagrangian function associated with $(MSP)_{k,i}$ is

$$L(v, z, \lambda) = v + \sum_{m=1}^{n} \alpha_m \varphi(z_m) + \sum_{j=0}^{i-1} \lambda_j \left[\langle s^j, z \rangle - v - b_j \right]$$

and the dual function

$$\begin{aligned} d(\lambda) &= \inf L(v, z, \lambda) \\ &= \begin{cases} \inf v + \displaystyle\sum_{m=1}^{n} \alpha_m \varphi(z_m) + \sum_{j=0}^{i-1} \lambda_j [\langle s^j, z \rangle - v - b_j] & \text{if } \sum \lambda_j = 1, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

So the dual problem is

$$(D) \begin{cases} \text{maximize} \quad d(\lambda) \\ \text{subject to} \quad \sum \lambda_j = 1 \quad \lambda_j \ge 0, \; j = 0, \ldots, i-1, \end{cases}$$

where

$$d(\lambda) = \sum_{m=1}^{n} d_m(\lambda) - \sum_{j=0}^{i-1} \lambda_j b_j \quad \text{with} \quad d_m(\lambda) = \inf_{z_m} \{\alpha_m \varphi(z_m) + \sum_{j=0}^{i-1} \lambda_j s_m^j z_m\}.$$

Moreover, each function d_m is differentiable and

$$\nabla d(\lambda) = \left(\sum_{m=1}^{n} s_m^j \tilde{z}_m\right)_{0\le j\le i-1} - b,$$

where, for each m, $\tilde{z}_m = \arg\min_{z_m}\{\alpha_m \varphi(z_m) + \sum_{j=0}^{i-1} \lambda_j s_m^j z_m\}$.

Since (D) is a smooth problem whose objective function is easily evaluated, we can use any classical method for solving it. Let λ^* be the solution of (D). Then the vector $z^* = (z_m^*)$ where, for each m,

$$z_m^* = \text{arg min } \{\alpha_m^* \varphi(z_m) + \sum_{j=0}^{i-1} \lambda_j^* s_m^j z_m\}$$

and the scalar

$$v^* = \langle \sum_{j=0}^{i-1} \lambda_j^* s^j, z^* \rangle - \sum_{j=0}^{i-1} \lambda_j^* b_j$$

are solutions of problem $(SP)_{k,i}$. Indeed by the complementarity conditions

$$\sum_{j=0}^{i-1} \lambda_j^* [\langle s^j, z^* \rangle - v^* - b_j] = 0,$$

i.e., since $\sum_{j=0}^{i-1} \lambda_j^* = 1$,

$$v^* = \sum_{j=0}^{i-1} \lambda_j^* v^* = \langle \sum_{j=0}^{i-1} \lambda_j^* s^j, z^* \rangle - \sum_{j=0}^{i-1} \lambda_j^* b_j.$$

The computational results presented here are obtained by using the MATLAB environment. The function F used in the tests, is defined on $I\!R^{10}$ and is the maximum of five quadratic functions:

$$q_j(x) = x^T C^j x - d^{jT} x, \; j = 1, \ldots, 5,$$

where C^j is a 10×10 symmetric matrix defined by

$$C_{ik}^j = \exp(\frac{i}{k}) \cos(ik) \sin j, \; i < k \quad \text{and} \quad C_{ii}^j = \frac{i}{n} \mid \sin j \mid + \sum_{i \ne k} \mid C_{ik}^j \mid,$$

and d^j is a vector in $I\!R^{10}$ whose components are $d_i^j = \exp(i/k)\sin(ij)$. This function F is well–known in nonsmooth optimization ([56], Test problem 1: Maxquad, n.151).

The parameters of the method and the function h are chosen as follows: $\nu = 2$, $\mu = 1$, $c_k = 0.1$ for all k and $h(t) = -\log t + t - 1$ for all $t > 0$ so that the function φ becomes

$$\varphi(t) = h(t) + (t-1)^2 = t^2 - t - \log t \text{ for all } t > 0.$$

The stopping criterion for the outer iterations is $\|x^{k+1} - x^k\| \leq \varepsilon$ where $\varepsilon = 10^{-3}$. Two values for the parameter σ are used in the numerical experiences, $\sigma = 0.1$ and $\sigma = 0.05$. The results are reported in Table 1 where for each outer iteration (denoted by k), the number of subproblems to be solved is mentioned. As in nonsmooth convex optimization, let us mention that it is possible to limit the size of the bundle, i.e., the number of constraints in the subproblems, by using aggregation [51]. Although our convergence theorems allow us to use this technique (see (2.4)), we have not applied it to illustrate the behavior of our method given the small size of the test problems.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
σ_1	4	1	3	1	2	3	4	5	5	6	6	10	8	12	9	10	15	12	49
σ_2	3	1	2	1	1	2	2	3	2	4	5	6	7	8	13	8	10	16	26

Table 5.1: The bundle interior proximal method. Number of inner iterations (for $\sigma_1 = 1$ and $\sigma_2 = 0.5$) for each outer iteration denoted by k.

From this table, we can observe that the number of subproblems per outer iteration is relatively small. Furthermore, for fixed k, each subproblem is the previous one with an additional linear inequality constraint. So, these problems can be solved very efficiently if the solution of a subproblem is the starting point of the next one. We also observe that the number of subproblems become smaller when the value of the parameter σ is reduced. The smaller is the value of σ, the faster is the stopping criterion satisfied for inner iterations. Contrary to the standard proximal methods, the subproblems are no more quadratic and the way to solve them is crucial for the rate of convergence of the algorithm. The preliminary results are encouraging but more efforts should be devoted to design appropriate numerical methods for solving them. This could be the subject of a future research.

Chapter 6

Conclusions and Further Work

The aim of the thesis was to present and to study efficient numerical methods for solving equilibrium problems in the sense of Blum and Oettli. The thesis was divided into two parts. In the first part (Chapters 1 and 2), we have recalled what is the proximal point method for solving an equilibrium problem while in the second part (Chapters 3-5), we have given our main contribution.

In Chapter 1 we have explained what is an equilibrium problem and we have shown, by giving many examples, that many important problems are in fact particular equilibrium problems: so are the convex minimization problem, the variational inequality problem and the Nash equilibrium problem. In the second chapter, we have recalled some fundamental results related to the proximal point method. Starting from convex minimization problems, we have shown how to generalize the proximal point method to equilibrium problems. In particular we have derived the auxiliary problem principle which can give rise to implementable and efficient algorithms. Up to now, most of the methods used in the literature were based upon the assumption that solving the subproblems to obtain the sequence of iterates $\{x^k\}$ is not too hard. However, that is far to be the case when the function f is nonsmooth in the second variable. So our aim in this thesis was to consider approximate subproblems, that are easy to solve and that guarantee the convergence of the sequence $\{x^k\}$ to a solution of the equilibrium problem.

Chapters 3, 4 and 5 are the main contributions of the thesis. In Chapter 3, the proximal point method is studied in the framework of an equilibrium problem where the equilibrium function f is convex and nonsmooth in the second argument. A special emphasis is put on an implementable method, called the bundle method, for solving that problem. In this method the constraint set is simply incorporated into each subproblem. At the end of the chapter, applications to multivalued variational inequalities and variational inequalities are given. In Chapters

4 and 5, the constraints are taken into account in another way thanks to a barrier function associated with an entropy-like distance. The corresponding method is a generalization to problem EP of a method due to Auslender, Teboulle, and Ben-Tiba for solving convex minimization problems [9] and variational inequality problems [10]. Some numerical results are given to demonstrate the efficiency of our algorithm. We study the convergence of the new method with several variants. In Chapter 5, we have considered a bundle-type implementation for the particular case of the constrained convex minimization. We have also introduced numerical results for proving the efficiency of our algorithm.

Several questions have not been considered in the thesis through lack of time. However we think that it is worth looking at them. Let us briefly mention them here.

One of these questions is the management of the parameter c_k in relationship with the behavior of the algorithm. It is easy to see that this parameter is related to the radius r_k of some trust region around the current point, and that the largest c_k is, the largest the radius r_k is. So, the usual trust region rule could be applied: if a serious step has been done (i.e., if the reduction of the function $f(x^k, \cdot)$ is large enough), then the radius r_k and thus c_k is increased. If several null steps are done, then it is interesting to decrease the radius and thus the parameter c_k in order to reduce the trust region. Some work has already been done in this direction but in the case of minimization problems (see, for example [76]). We think that a similar study should be done for equilibrium problems.

Another subject of future research is the following one: in this thesis we have only considered the general equilibrium problem. But recently many papers have been devoted to the very important particular case of Nash equilibrium problems and more specifically to generalized Nash equilibrium problems (see [32] for a survey). We think that it is worth studying bundle proximal point methods for solving those problems when the equilibrium function is nonsmooth.

Finally let us mention another interesting subject of research: the nonconvex case. Many proximal point methods have been proposed for solving nonconvex minimization problems and for solving nonconvex equilibrium problems, but of class C^1. The study of the nonconvex nonsmooth case, although very difficult at first glance, could be a research subject in the continuation of the work done in this thesis.

Bibliography

[1] Anh, P.N. and Muu, L.D.: Coupling the Banach Contraction Mapping Principle and the Proximal Point Algorithm for Solving Monotone Variational Inequalities, Acta Math. Vietnam **29**, 119–133 (2004)

[2] Anh, P.N., Muu, L.D., Nguyen, V.H., and Strodiot, J.J.: On the Contraction and Nonexpansiveness Properties of the Marginal Mappings in Generalized Variational Inequalities Involving Co-Coercive Operators. In: Andrew E., Nicolas H., and Luc, D.T. (eds.), Generalized Convexity, Generalized Monotonicity and Applications, Springer, Berlin, 89–111 (2005)

[3] Antipin, A.S.: The Convergence of Proximal Methods to Fixed Points of Extremal Mappings and Estimates of their Rates of Convergence, Comput. Maths Math. Phys. **35**, 539–551 (1995)

[4] Antipin, A.S.: From Optima to Equilibria, Dynamics of Non-Homogeneous System, Proceedings of ISA RAS **3**, 35–64 (2000)

[5] Aubin, J.P. and Ekeland, I.: *Applied Nonlinear Analysis*, John Wiley and Sons Inc., New York (1984)

[6] Auslender, A.: *Optimisation: Méthodes Numériques*, Masson, Paris (1976)

[7] Auslender, A., Cominetti, R., and Haddou, M.: Asymptotic Analysis for Penalty and Barrier Methods in Convex and Linear Programming, Math. Oper. Res. **22**, 43–62 (1997)

[8] Auslender, A. and Haddou, M.: An Interior Proximal Method for Convex Linearly Constrained Problems and Its Extension to Variational Inequalities, Math. Program. **71**, 77–100 (1995)

[9] Auslender, A., Teboulle, M., and Ben-Tiba, S.: Interior Proximal and Multiplier Methods Based On Second Order Homogeneous Kernels, Math. Oper. Res. **24**, 654–668 (1999)

[10] Auslender, A., Teboulle, M., and Ben-Tiba, S.: A Logarithmic-Quadratic Proximal Method for Variational Inequalities, Comput. Optim. Appl. **12**, 31–40 (1999)

[11] Auslender, A. and Teboulle, M.: Lagrangian Duality and Related Multiplier Methods for Variational Inequality Problems, SIAM J. Optim. **10**, 1097–1115 (2000)

[12] Auslender, A. and Teboulle, M.: Interior Gradient and Epsilon-Subgradient Descent Methods for Constrained Convex Minimization, Math. Oper. Res. **29**, 1–26 (2004)

[13] Auslender, A. and Teboulle, M.: Interior Projection-like Methods for Monotone Variational Inequalities, Math. Program. **104**, 39–68 (2005)

[14] Auslender, A. and Teboulle, M.: Interior Gradient and Proximal Methods for Convex and Conic Optimization, SIAM J. Optim. **16**, 697–725 (2006)

[15] Auslender, A. and Teboulle, M.: Entropic Proximal Decomposition Methods for Convex Programs and Variational Inequalities, Math. Program. **91**, 33–47 (2001)

[16] Auslender, A. and Teboulle, M.: The Log-quadratic Proximal Methodology in Convex Optimization Algorithms and Variational Inequalities. In: Daniele, P., Gianessi, F., and Maugeri, A. (eds.), Equilibrium Problems and Variational Models., Kluwer Academic Publishers, Dordrecht, The Netherlands, 19–52 (2003)

[17] Ben-Tal, A. and Zibulevsky, M.: Penalty/Barrier Multiplier Methods for Convex Programming Problems, SIAM J. Optim. **7**, 347–366 (1997)

[18] Bertsekas, D.P.: *Nonlinear Programming*, Athena Scientific, Belmont, Massachusetts (1995)

[19] Bianchi, M. and Schaible, S.: Generalized Monotone Bifunctions and Equilibrium Problems, J. Optim. Theory Appl. **90**, 31–43 (1996)

[20] Blum, E. and Oettli, W.: From Optimization and Variational Inequalities to Equilibrium Problems, The Math. Stud. **63**, 123–145 (1994)

[21] Bnouhachem, A.: An LQP Method for Pseudomonotone Variational Inequalities, J. Global Optim. **36**, 351–363 (2006)

[22] Bnouhachem, A.: A New Inexactness Criterion for Approximate Logarithmic-Quadratic Proximal Methods, Numer. Math. J. Chinese Univ. (English Ser.) **15**, 74–81 (2006)

[23] Burachik R.S. and Iusem, A. N.: A Generalized Proximal Point Algorithm for the Variational Inequality Problem in a Hilbert Space, SIAM J. Optim. **8**, 197–216 (1998)

[24] Burachik, R. S. and Svaiter B.F.: A Relative Error Tolerance for a Family of Generalized Proximal Point Methods, Math. Method Oper. Res. **26**, 816–831 (2001)

[25] Brézis, H., Nirenberg, L., and Stampacchia, G.: A Remark on Ky Fan's Minimax Principle, Boll. Un. Mat. Ital. **6**, 293–300 (1972)

[26] Cohen, G.: Auxiliary Problem Principle and Decomposition of Optimization Problems, J. Optim. Theory Appl. **32**, 277–305 (1980)

[27] Cohen, G.: Auxiliary Problem Principle Extended to Variational Inequalities, J. Optim. Theory Appl. **59**, 325–333 (1988)

[28] Correa, R. and Lemaréchal, C.: Convergence of Some Algorithms for Convex Minimization, Math. Program. **62**, 261–275 (1993)

[29] El Farouq, N.: Pseudomonotone Variational Inequalities: Convergence of the Auxiliary Problem Method, J. Optim. Theory Appl. **111**, 305–326 (2001)

[30] Fan, K.: A Minimax Inequality and Applications. In: O. Shisha (ed.), Inequality III, Academic Press, New York, 103–113 (1972)

[31] Fan, K.: Some Properties of Convex Sets Related to Fixed Point Theorems, Math. Ann. **266**, 519–537 (1984)

[32] Facchinei, F. and Kanzow, C.: Generalized Nash Equilibrium Problems, 4OR **5**, 173–210 (2007)

[33] Facchinei, F. and Pang, J.S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York (2003)

[34] Flåm, S.D. and Antipin, A.S.: Equilibrium Programming Using Proximal-Like Algorithms, Math. Program. **78**, 29–41 (1997)

[35] Fukushima, M.: Equivalent Differentiable Optimization Problems and Descent Methods for Asymmetric Variational Inequality Problems, Math. Program. **53**, 99–110 (1992)

[36] Gol'shtein, E.G.: A Method for Solving Variational Inequalities Defined by Monotone Mappings, Comput. Math. Math. Phys. **42**, 921–930 (2002)

[37] Hiriart–Urruty, J.B. and Lemaréchal, C. : *Convex Analysis and Minimization Algorithms*, Volumes I and II, Springer, Berlin (1993)

[38] Hue, T.T., Strodiot, J.J., and Nguyen, V.H.: Convergence of the Approximate Auxiliary Problem Method for Solving Generalized Variational Inequalities, J. Optim. Theory Appl. **121**, 119–145 (2004)

[39] Hue, T.T. and Strodiot, J.J.: Convergence Analysis of a Relaxed Extragradient-Proximal Point Algorithm, Application to Variational Inequalities, Optimization **54**, 191–213 (2005)

[40] Iusem, A. and Sosa, W.: New Existence Results for Equilibrium Problem, Nonlinear Anal., Theory Methods Appl. **52**, 621–635 (2003)

[41] Iusem, A. and Sosa, W.: Iterative Algorithms for Equilibrium Problems, Optimization **52**, 301–316 (2003)

[42] Konnov, I.: Combined Relaxation Methods for Finding Equilibrium Points and Solving Related Problems, Russ. Math. **37**, 44–51 (1993)

[43] Konnov, I.V.: *Combined Relaxation Methods for Variational Inequalities*, Springer, Berlin (2001)

[44] Konnov, I.V.: The Application of a Linearization Method to Solving Nonsmooth Equilibrium Problems, Russ. Math. **40**, 54–62 (1996)

[45] Konnov, I.V.: Combined Relaxation Method for Generalized Variational Inequalities, Russ. Math. **45**, 43–51 (2001)

[46] Konnov, I. V.: Application of the Proximal Point Method to Nonmonotone Equilibrium Problems, J. Optim. Theory Appl. **119**, 317–333 (2003)

[47] Konnov, I. V.: Generalized Monotone Equilibrium Problems and Variational inequalities. In: Handbook of Generalized Convexity and Generalized Monotonicity, Springer, Netherlands, 559–618 (2005)

[48] Konnov, I.V. : Combined Relaxation Method for Monotone Equilibrium Problems, J. Optim. Theory Appl. **111**, 327–340 (2001)

[49] Konnov, I.V.: *Equilibrium Models and Variational Inequalities*, Elsevier, Amsterdam (2007)

[50] Korpelevich, G.M.: The Extragradient Method for Finding Saddle Points and Other Problems, Ekonomika i Matematcheskie Metody **12**, 747–756 (1976)

[51] Kiwiel, K.C.: Approximations in Proximal Bundle Methods and Decomposition of Convex Programs, J. Optim. Theory Appl. **84**, 529–548 (1995)

[52] Kiwiel, K.C.: Proximal Minimization Methods with Generalized Bregman Functions, SIAM J. Control Optim. **35**, 1142–1168 (1997)

[53] Kiwiel, K.C.: A Bundle Bregman Proximal Method for Convex Nondifferentiable Minimization, Math. Program. **85**, 241–258 (1999)

[54] Kiwiel, K.C.: Proximal Level Bundle Methods for Convex Nondifferentiable Optimization, Saddle Point Problems and Variational Inequalities, Math. Program. **69**, 89–109 (1995)

[55] Lemaire, B.: The Proximal Algorithm. In: Penot, J.P. (ed.), International Series of Numerical Mathematics, Birkhäuser, Basel, 73–87 (1989)

[56] Lemaréchal, C. and Mifflin, R.: *Nonsmooth Optimization*, IIASA Proceedings Series, Volume 3, Pergamon Press, Oxford (1978)

[57] Lemaréchal, C., Nemirovskii, A., and Nesterov, Y.: New Variants of Bundle Methods, Math. Program. **69**, 111–147 (1995)

[58] Lemaréchal, C., Strodiot, J.J., and Bihain, A.: On a Bundle Method for Nonsmooth Optimization. In: Mangasarian, O.L., Meyer, R.R., and Robinson, S.M. (eds.), Nonlinear Programming 4, Academic Press, 245–282 (1981)

[59] Martinet, B.: Régularisation d'Inéquations Variationnelles par Approximations Successives, R.A.I.R.O. , 154–159, (1970)

[60] Martinet, B.: *Algorithmes pour la Résolution de Problèmes d'Optimisation et de Minimax*, Thèse d'Etat, Univ. Grenoble, France (1972)

[61] Martinez-Legaz, J. E. and Sosa, W.: Duality for Equilibrium Problems, J. Global Optim. **35**, 311–319 (2006)

[62] Mastroeni, G.: On Auxiliary Principle for Equilibrium Problems. In: Daniele, P., Giannessi, F., and Maugeri, A. (eds.), Equilibrium Problems and Variational Models. Kluwer Academic Publishers, Dordrecht, 289–298 (2003)

[63] Mastroeni, G.: Minimax and Extremum Problems Associated to a Variational Inequality, Supplemento ai Rendiconti del Circolo Matematico di Parlemo, Suppl. **58**, 185–196 (1999)

[64] Mastroeni, G.: Gap Functions for Equilibrium Problems, J. Global Optim. **27**, 411–426 (2003)

[65] Moreau, J.J.: Proximité et Dualité dans un Espace Hilbertien, Bull. Soc. Math. France **93**, 273–299 (1965)

[66] Moudafi, A.: Proximal Point Algorithm Extended to Equilibrium Problems, J. Nat. Geom. **15**, 91–100 (1999)

[67] Muu, L.D., Nguyen, V.H., and Quy, N.V.: On Nash-Cournot Oligopolistic Market Equilibrium Models with Concave Cost Functions, J. Global Optim. **41**, 351–364 (2007)

[68] Nesterov, Y. and Nemirovskii, A.: *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM Studies in Applied Mathematics (1994)

[69] Nguyen, V.H: *Lecture Notes on Equilibrium Problems*, Nha Trang University of Pedagogy, Nha Trang City, Vietnam (2002)

[70] Nikaido, H. and Isoda, K.: Note on Noncooperative Convex Games, Pacific J. Math. **5**, 807–815 (1955)

[71] Noor, M.A. and Bnouhachem, A.: Modified Proximal-Point Method for Nonlinear Complementarity Problems, J. Comput. Appl. Math. **197**, 395–405 (2006)

[72] Quoc, T.D., Muu, L.D., and Nguyen, V.H.: Extragradient Algorithms Extended to Equilibrium Problems, Optimization (DOI: 10.1080/02331930601122876)

[73] Rockafellar, R. T.: Monotone Operators and the Proximal Point Algorithm, SIAM J. Control Optim. **14**, 877–898 (1976)

[74] Rockafellar, R. T.: *Convex Analysis*, Princeton University Press, Princeton, New Jersey (1976)

[75] Salmon, G., Strodiot, J.J., and Nguyen, V.H.: A Bundle Method for Solving Variational Inequalities, SIAM J. Optim. **14**, 869–893 (2004)

[76] Schramm, H. and Zowe, J.: A Version of the Bundle Idea for Minimizing a Nonsmooth Function: Conceptual Idea, Convergence Analysis, Numerical Results, SIAM J. Optim. **2**, 121–152 (1992)

[77] Solodov, M. and Svaiter, B.: A New Projection Method for Variational Inequality Problems, SIAM J. Control Optim. **37**, 765–776 (1999)

[78] Solodov, M. and Svaiter, B.: A Unified Framework for Some Inexact Proximal Point Algorithms, Numer. Funct. Anal. Optim. **22**, 1013–1035 (2001)

[79] Solodov, M. and Svaiter, B.: Error Bounds for Proximal Point Subproblems and Associated Inexact Proximal Point Algorithms, Math. Program. **88**, 371–389 (2000)

[80] Strodiot, J.J.: *An Introduction to Nonsmooth Optimization*, University of Natural Sciences, Ho Chi Minh City, Vietnam (2004)

[81] Strodiot, J.J.: *Numerical Methods in Optimization*, University of Natural Sciences, Ho Chi Minh City, Vietnam (2008)

[82] Tseng, P. and Bertsekas, D. P. : On the Convergence of the Exponential Multiplier Method for Convex Programming, Math. Program. **60**, 1–19 (1993)

[83] Van, N.T.T., Nguyen, V.H., and Strodiot, J.J.: A Bundle Interior Proximal Method for Solving Convex Minimization Problems, J. Convex Anal. **12**, 95–111 (2005)

[84] Van, N.T.T., Strodiot, J.J., and Nguyen, V.H.: A Bundle Method for Solving Equilibrium Problems, Math. Program. **116**, 529–552 (2009)

[85] Van, N.T.T, Strodiot, J.J., and Nguyen, V.H.: The Logarithmic Quadratic Extragradient Method for Solving Equilibrium Problems, J. Global Optim. (DOI 10.1007/s10898-008-9311-0)

[86] Willem, M.: *Analyse Convexe et Optimisation*, Ciaco, Bruxelles (1989)

[87] Xu, Y., He, B., and Yuan, X.: A Hybrid Inexact Logarithmic-Quadratic Proximal Method for Nonlinear Complementarity Problems, J. Math. Anal. Appl. **322**, 276–287 (2006)

[88] Yosida, K.: *Functional Analysis*, Springer, Berlin (1964)

[89] Yuan, X.: The Prediction-Correction Approach to Nonlinear Complementarity Problems, European J. Oper. Res. **176**, 1357–1370 (2007)

[90] Zhu, D. and Marcotte, P.: Co-coercivity and Its Role in the Convergence of Iterative Schemes for Solving Variational Inequalities, SIAM J. Optim. **6**, 714–726 (1996)

[91] Zhu, D. and Marcotte, P.: An Extended Descent Framework for Variational Inequalities, J. Optim. Theory Appl. **80**, 349–366 (1994)