# Sapienza Università di Roma 

PhD Thesis<br>Penalty methods for the solution of generalized Nash equilibrium problems and hemivariational inequalities with VI constraints<br>Department of Computer, Control and Management Engineering

PhD in System engineering

Candidate<br>Lorenzo Lampariello

Advisor
Professor Francisco Facchinei

God forbid that Truth should be confined to Mathematical Demonstration!

William Blake, Notes on Reynold's Discourses

## Contents

Introduction ..... 1
1 Mathematical background ..... 5
1.1 Preliminary definitions ..... 5
1.2 Some aspects of penalty methods for optimization problems ..... 10
1.3 Variational Inequalities ..... 12
1.3.1 Generalities ..... 13
1.3.2 Existence results ..... 14
1.3.3 Solution algorithms ..... 15
2 Partial penalization for the solution of GNEPs ..... 21
2.1 Preliminaries ..... 21
2.2 Brief historical background ..... 29
2.3 Penalty methods for GNEPs: the state of the art ..... 31
2.4 The partial penalization approach ..... 33
2.5 The original problem and the penalized problem ..... 34
2.6 How to update the penalty parameters ..... 41
2.7 Algorithm $\mathcal{A}$ ..... 44
2.8 Convergence of the smoothing procedure ..... 46
$2.9 \operatorname{VI}\left(\mathbf{K}, F_{\varepsilon_{i}}(\mathbf{x})\right)$ ..... 47
3 Solution algorithms for VI-constrained hemivariational inequalities ..... 55
3.1 Preliminaries ..... 55
3.2 Background Results ..... 58
3.3 Penalization of HVIs with Side Constraints ..... 62
3.4 The Main Algorithm: Centralized Version ..... 68
3.5 Numerical experiments ..... 76
3.5.1 Examined problems ..... 76
3.5.2 The implemented algorithm ..... 77
3.5.3 Numerical results and comments ..... 82
3.6 Distributed Solution of the Hemivariational Inequality ..... 86
3.7 Numerical experiments: centralized vs distributed algorithm ..... 93
3.7.1 Examined problems ..... 93
3.7.2 The implemented algorithm ..... 93
3.7.3 Numerical results and comments ..... 95
3.8 VI-C Minimization Problems ..... 98
3.9 An Application to a Rate Maximization Game ..... 99
Conclusions ..... 109
A Some considerations on constraint qualifications ..... 111
B Test problems ..... 115
B. 1 Lower feasible set K of $\mathrm{VI}(\mathrm{K}, \mathrm{F})$ ..... 115
B. 2 Lower function F of $\mathrm{VI}(\mathrm{K}, \mathrm{F})$ ..... 119
B. 3 Upper function $\Phi$ ..... 123
C Algorithm performances ..... 127
C. 1 Centralized exact version ..... 127
C. 2 Centralized inexact version ..... 127
C. 3 Centralized version vs distributed version ..... 128
Bibliography ..... 146
References ..... 147

## List of Figures

2.1 In bold the solution set $\mathcal{S}(\mathbf{x})$ of Example 1. . . . . . . . . . . . . . . . . 24
2.2 The three classes of GNEP. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
2.3 The solution set $\mathcal{S}(\mathbf{x})$ of Example 2. . . . . . . . . . . . . . . . . . . . . . 26
2.4 The only variational equilibrium ................................ . . 27
2.5 Example $2, \mathbf{x}=(1 / 2,1 / 2)$ and $\mathcal{F}(\overline{\mathbf{x}})=$
$[-\sqrt{3} / 2, \sqrt{3} / 2] \times[-\sqrt{3} / 2, \sqrt{3} / 2]$ (dashed set) neither
includes nor is included in $\mathbf{F}$ (gray set) . . . . . . . . . . . . . . . . . 27
2.6 In bold the set $\mathcal{S}(\mathbf{x})$ of Example 3; in red the variational equilibrium.28
2.7 The algorithmic scheme . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
3.1 HVI problem: we remark that, in case of continuously differentiable objective functions, the convex optimization set is actually a subset of the VI problem set. . . . . . . . . . . . . . . . . . . 56
3.2 VI-C HVI problem. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
3.3 Algorithm 3 scheme ................................................... . . 71
3.4 The centralized algorithmic scheme tested . . . . . . . . . . . . . . . . . 78
3.5 The centralized algorithmic scheme implemented ............. . 82
3.6 Different updating rules for $e_{k}: e_{k}^{1} \propto 1 / k^{6}$ (magenta -o- line), $e_{k}^{2} \propto 1 / \exp (k)\left(\right.$ red $-\triangleright$ - line),$e_{k}^{3} \propto 1 / 10 k^{3}($ ciano $-\square$ - line $)$, $e_{k}^{4} \propto 1 / 100 k^{2}$ (blue -*- line) and $e_{k}^{5} \propto 1 / 10 k^{2}$ (green - $\triangleleft$ - line) . 85
3.7 The distributed version of the algorithm; in particular, the inner Jacobi scheme is given in detail (red boxes). . . . . . . . . . . . 92
3.8 The distributed algorithmic scheme implemented . . . . . . . . . . . 97
3.9 Comparison of distributed algorithms solving the game $\mathcal{G}$ : Sum-rate of the users versus the iteration index, achieved by the simultaneous IWFA (dashed-dot line curve), Algorithm 3 based on the outer function $\phi(p)$ in (3.50) (solid line curve), and Algorithm 3 based on the outer function $-\phi(p)$ (dashed line curve), in the low/medium interference regime (i.e., $M$ is positive definite) and high interference regime (i.e., $M$ is positive semidefinite).

3.10 Average sum-rate versus the SNR [subplot (a)] and
outage sum-rate [subplot(b)] at the Nash equilibria of $\mathcal{G}$ :
simultaneous IWFA (dashed-dot line curve), Algorithm 3
based on the outer function $\phi(p)$ in (3.50) (solid line curve),
and Algorithm 3 based on the outer function $-\phi(p)$ (dashed
line curve). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 108
A. 1 Some constraint qualifications' relations. ..... 113
C. 1 Centralized exact version, problem K1F13Фr, $n=50$. Number of outer iterations. ..... 129
C. 2 Centralized exact version, problem K1F13Фr, $n=50$. Number of inner iterations ..... 130
C. 3 Centralized exact version, problem K1F13Фr, $n=50$. Solutions error. ..... 131
C. 4 Centralized exact version, problem K2F21 $\Phi 3, n=50$. Number of outer iterations. ..... 132
C. 5 Centralized exact version, problem K2F21 $\Phi 3, n=50$. Number of inner iterations. ..... 133
C. 6 Centralized exact version, problem K2F21 $\$ 3, n=50$. Solutions error. ..... 134
C. 7 Centralized exact version, problem K5F52 $22, n=50$. Number of outer iterations ..... 135
C. 8 Centralized exact version, problem K5F52 $22, n=50$. Number of inner iterations. ..... 136
C. 9 Centralized exact version, problem K5F52 2 2, $n=50$.
Solutions error. ..... 137
C. 10 Centralized inexact version, problem K2F7 $\Phi \mathrm{r}$, $n=50$ ..... 138
C. 11 Centralized inexact version, problem K5F52 $22, n=50$. ..... 138
C. 12 Centralized inexact version, problem K5F52 14 , $n=50$. ..... 139
C. 13 Centralized inexact version, problem K5F52 $55, n=50$. ..... 139
C. 14 Centralized inexact version, problem K6F61 22 , $n=50$ ..... 140
C. 15 Problem K2F13 $66, n=50$ : for each $Q$, the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$. In this case, $\alpha_{1}=0.3, \alpha_{2}=6.0183, \alpha_{3}=9.7825, \alpha_{4}=13.2523$, $\alpha_{5}=16.7221, \alpha_{6}=29.2889, \alpha_{R}=5 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$
C. 16 Problem K2F21 $\Phi 2, n=50$ : for each $Q$, the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$. In this case, $\alpha_{1}=0.3, \alpha_{2}=0.8, \alpha_{3}=1.6, \alpha_{4}=2.4, \alpha_{5}=3.2$, $\alpha_{6}=7.2, \alpha_{R}=0.3$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 142
C. 17 Problem K2F61 $\Phi$ r, $n=50$ : for each $Q$, the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$. In this case, $\alpha_{1}=0.3, \alpha_{2}=4.7438, \alpha_{3}=9.4877, \alpha_{4}=14.2315$, $\alpha_{5}=18.9753, \alpha_{6}=42.6945, \alpha_{R}=4 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$.
C. 18 Problem $\operatorname{K2Fr} \Phi 2, n=50$ : for each $Q$, the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$. In this case, $\alpha_{1}=0.3, \alpha_{2}=2.093, \alpha_{3}=3.3734, \alpha_{4}=4.4545$, $\alpha_{5}=5.3132, \alpha_{6}=9.7138, \alpha_{R}=1 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$
C. 19 Problem $\operatorname{K2Fr} \Phi 7, n=50$ : for each $Q$, the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$. In this case, $\alpha_{1}=0.3, \alpha_{2}=2.3113, \alpha_{3}=3.8099, \alpha_{4}=5.1092$, $\alpha_{5}=6.1863, \alpha_{6}=11.6781, \alpha_{R}=0.3 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . .$.

## List of Tables

1.1 Monotonicity properties ..... 6
1.2 Relations between classes of monotone functions ..... 7
3.1 Centralized exact version with fixed values of $\varepsilon_{0}=6.4$ and $\alpha=0.3$ : numerical results in terms of number of inner and outer iterations and solution accuracy ..... 84
3.2 Centralized vs decomposed versions: numerical results in terms of number of inner and outer iterations and execution time for $Q=5$ ..... 96

## Introduction

In this thesis we propose solution methods for Generalized Nash Equilibrium Problems (GNEPs) and Hemivariational Inequalities (HVIs) where the feasible set is given by the intersection of a closed convex set with the solution set of a lower-level monotone Variational Inequality (VI).

The Generalized Nash Equilibrium Problem was first formally introduced in 1952 by Debreu in [21], where the term social equilibrium was coined. The GNEP has soon become a central model in game theory that has been used actively in many fields in the past fifty years. The GNEP models conflicts among players, each behaving selfishly to optimize one's own well-being subject to resource limitations and other constraints that may depend on the rivals' actions ("coupling" constraints). Therefore, the GNEP is a natural extension of the well known standard Nash Equilibrium Problem (NEP), if the players share some common resources (a communication link, an electrical transmission line, a transportation link etc.) or limitations (for example a common limit on the total pollution in a certain area), i.e. if they interact also at the level of the feasible sets.

It is worth mentioning that, in general, GNEPs usually have a manifold of solutions (that sometimes could be disconnected) which is one of the many peculiarities that make these problems so difficult to analyze and to solve. Indeed, until very recently, there were no provably convergent algorithms for the solution of GNEPs: here we mention KKT methods (see e.g. [23]), the Nikaido-Isoda reformulation of the GNEP (see e.g. [25]) and other reformulation approaches (see e.g. [31, 33]). In this work we consider one of the most promising approaches for which we are able to prove convergence results: penalty methods.

Penalty methods have originally been developed and studied for the solution of constrained (standard) optimization problems. The simple idea, which is the core of these approaches, is that one can recover a solution of the original constrained problem by solving (exactly or sequentially with a varying penalty parameter) a suitable equivalent "simpler" problem. In particular,
the use of penalty algorithms seems to be a promising method to cope with the complicated "coupling" constraints of a GNEP. In Chapter 2 we consider a "partial" exact penalty method which allows us to reduce the original GNEP to a nonsmooth standard NEP. Suitable constraints qualifications are investigated (to obtain the convergence of our scheme) and some relevant general theoretical results are established: among others a complete equivalence between the original GNEP and the penalized NEP. Furthermore, in order to tackle the nonsmooth penalized problem for fixed values of the penalty parameters, we opt for smoothing techniques. Finally, based on the previous results, classes of GNEPs for which this penalty approach is guaranteed to converge to a solution of the original game are identified: NEPs, jointly convex GNEPs and general GNEPs with separable coupling constraints and for which a "technical" condition holds on the the structure of the players' objective functions. Thus, for the first time, classes of GNEPs for which the presented scheme is provably convergent could be established.

HemiVariational Inequalities (HVI)s are a powerful modeling tool that encompasses both (convex) optimization and variational inequalities as particular instances. In their full generality, HVIs have been mainly considered in infinite-dimensional settings, (see e.g. [69, 72, 77]); nevertheless, finitedimensional HVIs have recently attracted attention in the mathematical programming literature, (see e.g. [1, 57, 68]). In Chapter 3, in particular, we consider the Variational Inequality-Constrained HemiVariational Inequality (VI-C HVI) with side constraints, thus an HVI for which the feasible set is implicitly defined as the intersection of a closed convex set with the solution set of a lower-level monotone VI. We remark that VI-C HVIs include also, as special case, the problem of selecting a particular equilibrium solution to optimize an auxiliary ("upper") objective function. Under suitable assumptions, this problem may arise, for example, when one tries to select a solution of a standard NEP that optimizes certain criteria.

In Chapter 3 we propose centralized and distributed schemes for the numerical solution of such HVIs. Firstly, in order to deal with the presence of further side constraints other than the solution set of the lower-level VI, we establish an exact penalization result which allows us to state, for suitable values of the penalty parameter, the equivalence between the original HVI (with side constraints) and a HVI without additional side constraints. Secondly we present a centralized solution method for solving the VI-C HVI and establish its convergence. Thirdly we present a distributed algorithm for solving a "partitioned" VI-C HVI. The algorithms consist of a main loop wherein a sequence of one-level and strongly monotone HVIs is solved that involves the penalization of the non-VI constraint and a combination of proximal and Tikhonov regularization to handle the lower-level VI constraints. Subsequently, the methods developed are used to successfully solve a new power control problem in ad-hoc networks. Finally the analysis is equipped with a series of numerical tests and results which show the outstanding behavior of the proposed approach.

To the best of our knowledge, these contributions are new and considerably expand existing results. The proposed distributed algorithm, in which we are interested as motivated by applications in non-cooperative game problems, see [34, Chapter 12], is novel even for a hierarchical optimization problem. Furthermore, the power control problem analyzed in the sequel is new and our results expand the applicability and flexibility of game-theoretic models in ad-hoc networks and also bring considerable gains over existing techniques.

The material presented in this work has led to the following publications:
G. Scutari, F. Facchinei, J-S Pang, and L. Lampariello, Equilibrium Selection in Power Control Games on the Interference Channel, to appear in the Proceedings of the IEEE 2012 International Conference on Computer Communications, Orlando, Florida USA, march 2012
F. Facchinei, J-S Pang, G. Scutari, and L. Lampariello, VI-constrained Hemivariational Inequalities: Distributed Algorithms and Power Control in AdHoc Networks, accepted (minor revision) in Mathematical Programming, 2011
F. Facchinei, L. Lampariello, and S. Sagratella, Recent advancements in the numerical solution of generalized Nash equilibrium problems, submitted to Quaderni di Matematica - Volume in ricordo di Marco DApuzzo, 2011
F. Facchinei and L. Lampariello, Partial penalization for the solution of generalized Nash equilibrium problems, Journal of Global Optimization 50, 1, 39-57, 2011

## Chapter 1 <br> Mathematical background

In this chapter we recall some relevant results that form the theoretical basis of the following developments.

In Section 1.1 we recall the definitions of concepts and mathematical tools that play a role in this work. Section 1.2 and Section 1.3 contain, respectively, some basic results about the penalty approaches for optimization problems and a brief survey on Variational Inequalities.

Most of the material of this chapter is taken from the fundamental books $[8,16,32]$.

### 1.1 Preliminary definitions

From now on, let be $K$ a subset of $\mathbb{R}^{n}$.
Definition 1. Function $F: K \rightarrow \mathbb{R}^{n}$ is globally Lipschitz continuous if, for some non negative scalar $L$, one has

$$
\begin{equation*}
\|F(x)-F(y)\| \leq L\|x-y\| \tag{1.1}
\end{equation*}
$$

for all $x$ and $y$ in $K$.
Moreover, if $F$ satisfy the Lipschitz condition (1.1) within a neighborhood of $x$, then we shall say that $F$ is locally Lipschitz at $x$.

We recall also that a function $F: K \rightarrow \mathbb{R}^{n}$ is non expansive if

$$
\|F(x)-F(y)\| \leq\|x-y\|, \quad \forall x, y \in K
$$

while $F$ is a contraction if there exists a constant $\eta \in(0,1)$ such that

$$
\|F(x)-F(y)\| \leq \eta\|x-y\|, \quad \forall x, y \in K
$$

## Monotonicity properties

Now we introduce several monotonicity properties of vector functions.
Definition 2. A mapping $F: K \subseteq \mathbb{R}^{n} \rightarrow \mathbb{R}^{n}$ is said to be
(a) pseudo monotone on $K$ if for all vectors $x$ and $y$ in $K$,

$$
(x-y)^{T} F(y) \geq 0 \Rightarrow(x-y)^{T} F(x) \geq 0
$$

(b) monotone on $K$ if

$$
(F(x)-F(y))^{T}(x-y) \geq 0, \forall x, y \in K
$$

(c) strictly monotone on $K$ if

$$
(F(x)-F(y))^{T}(x-y)>0, \forall x, y \in K \text { and } x \neq y
$$

(d) strongly monotone on $K$ if there exists a constant $c>0$ such that

$$
(F(x)-F(y))^{T}(x-y) \geq c\|x-y\|^{2}, \forall x, y \in K
$$

In the following table we report some elementary relations.


Table 1.1: Monotonicity properties

If $F$ is a continuously differentiable function defined on an open convex set, there is a connection between the above monotonicity properties and the positive semidefiniteness of the Jacobian of F.

Proposition 1. Let $F: \mathcal{D} \subseteq \mathbb{R}^{n} \rightarrow \mathbb{R}^{n}$ be continuously differentiable on the open convex set $\mathcal{D}$. The following statements are valid.
(a) $F$ is monotone on $\mathcal{D}$ if and only if $J F(x)$ is positive semidefinite for all $x \in \mathcal{D}$;
(b) $F$ is strictly monotone on $\mathcal{D}$ if $J F(x)$ is positive definite for all $x \in \mathcal{D}$;
(c) $F$ is strongly monotone on $\mathcal{D}$ if and only if $J F(x)$ is uniformly positive definite for all $x \in \mathcal{D}$; i.e., there exists a constant $c^{\prime}>0$ such that

$$
y^{T} J F(x) y \geq c^{\prime}\|y\|^{2}, \forall y \in \mathbb{R}^{n}
$$

for all $x \in \mathcal{D}$.

The monotone functions' characterization given by Proposition 1 is motivated by the corresponding properties of the gradients of convex functions. Indeed, the following result holds.

Proposition 2. Let $\theta: \mathcal{D} \subseteq \mathbb{R}^{n} \rightarrow \mathbb{R}$ continuously differentiable on the open convex set $\mathcal{D}$. The following statements are valid.
(a) $\nabla \theta$ is monotone on $\mathcal{D}$ if and only if $\theta$ is convex on $\mathcal{D}$.
(b) $\nabla \theta$ is strictly monotone on $\mathcal{D}$ if and only if $\theta$ is strictly convex on $\mathcal{D}$.
(c) $\nabla \theta$ is strongly monotone on $\mathcal{D}$ if and only if $\theta$ is strongly convex on $\mathcal{D}$

We can say that the class of monotone vector functions plays a similarly important role in the variational inequalities problem as the class of convex functions in optimization.

Another important definition is the following.
Definition 3. A mapping $F: K \subseteq \mathbb{R}^{n} \rightarrow \mathbb{R}^{n}$ is said to be
(a) pseudo monotone plus on $K$ if it is pseudo monotone on $K$ and for all vectors $x$ and $y$ in $K$,

$$
\left[(x-y)^{T} F(y) \geq 0 \text { and }(x-y)^{T} F(x) \geq 0\right] \Rightarrow F(x)=F(y)
$$

(b) monotone plus (or paramonotone) on $K$ if it is monotone on $K$ and for all vectors $x$ and $y$ in $K$

$$
(F(x)-F(y))^{T}(x-y)=0 \Rightarrow F(x)=F(y)
$$

(c) co-coercive on $K$ if there exists a constant $c>0$ such that

$$
(F(x)-F(y))^{T}(x-y) \geq c\|F(x)-F(y)\|^{2}, \forall x, y \in K
$$

The following diagram summarizes the relations between those classes of functions. In particular, we recall that a differentiable function $F$ with symmetric Jacobian $J F$ is symmetric. In addition it is worth mentioning that every Lipschitz continuous, strongly monotone function is co-coercive.


Table 1.2: Relations between classes of monotone functions

## Point-to-set maps

A point-to-set map, also called a multifunction or a set-valued map, is a map $\Phi$ from $\mathbb{R}^{n}$ into the power set of $\mathbb{R}^{n}$, i.e., for every $x \in \mathbb{R}^{n}, \Phi(x)$ is a (possibly empty) subset of $\Phi$. The domain of $\Phi$, denoted dom $\Phi$, the range of $\Phi$, denoted $\operatorname{ran} \Phi$, and the graph of $\Phi$, denoted $\operatorname{gph} \Phi$, are, respectively, the sets:

$$
\begin{aligned}
& \operatorname{dom} \Phi:=\left\{x \in \mathbb{R}^{n}: \Phi(x) \neq \emptyset\right\} \\
& \operatorname{ran} \Phi:=\bigcup_{x \in \operatorname{dom} \Phi} \Phi(x) \\
& \operatorname{gph} \Phi:=\left\{(x, y) \in \mathbb{R}^{2 n}: y \in \Phi(x)\right\}
\end{aligned}
$$

The following definition contains several classical concepts relevant to a setvalued map.
Definition 4. A set-valued map $\Phi: \mathbb{R}^{n} \rightarrow \mathbb{R}^{n}$ is said to be
(a) closed at $\bar{x}$, if

$$
\left.\begin{array}{cc}
\left\{x_{k}\right\} & \rightarrow \bar{x} \\
y_{k} \in \Phi\left(x_{k}\right) & \forall k \\
\left\{y_{k}\right\} & \rightarrow \bar{y}
\end{array}\right\} \Rightarrow \bar{y} \in \Phi(\bar{x}) ;
$$

(b) locally bounded at $\bar{x}$ if there exists an open neighborhood $\mathcal{N}$ of $\bar{x}$ such that the set:

$$
\bigcup_{x \in \mathcal{N} \cap \operatorname{dom} \Phi} \Phi(x)
$$

is bounded;
(c) lower semicontinuous at $\bar{x}$ if for every open set $\mathcal{U}$ such that $\Phi(\bar{x}) \cap \mathcal{U} \neq \emptyset$, there exists an open neighborhood $\mathcal{N}$ of $\bar{x}$ such that, for each $x \in \mathcal{N}$, $\Phi(x) \cap \mathcal{U} \neq \emptyset ;$
(d) upper semicontinuous at $\bar{x}$ if for every open set $\mathcal{V}$ containing $\Phi(\bar{x})$, there exists an open neighborhood $\mathcal{N}$ of $\bar{x}$ such that, for each $x \in \mathcal{N}, \mathcal{V}$ contains $\Phi(x) ;$
(e) continuous at $\bar{x}$ if $\Phi$ is both lower and upper semicontinuous at $\bar{x}$;
(f) closed on a set $S$ if $\Phi$ is closed at every point in $S$;
(g) (lower, upper semi)continuous on a set $S$ if $\Phi$ is (respectively, lower, upper semi)continuous at every point in $S$.

Some useful properties are consequences of these definitions. We mention here the fact that if $\Phi$ is closed at $\bar{x}$, then $\Phi(\bar{x})$ is a closed set; moreover, if $\Phi(\bar{x})$ is a closed set and $\Phi$ is upper semicontinuous at $\bar{x}$, then $\Phi$ is closed at $\bar{x}$. Other interesting results can be found in [32].

## Geometric concepts: tangent and normal cones

The concept of tangent cone is central in the forthcoming sections. The tangent cone of $K$ at $x \in K$, denoted $T_{K}(x)$, consists of all vectors $u \in \mathbb{R}^{n}$, called tangent vectors to $K$ at $x$, for which there exist a sequence of vectors $\left\{y_{\nu}\right\} \subset K$ and a sequence of positive scalars $\left\{\tau_{\nu}\right\}$ such that

$$
\lim _{\nu \rightarrow \infty} y_{\nu}=x, \quad \lim _{\nu \rightarrow \infty} \tau_{\nu}=0 \quad \text { and } \quad \lim _{\nu \rightarrow \infty} \frac{y_{\nu}-x}{\tau_{\nu}}=u
$$

We can define the normal cone to $K$ at $x$ by polarity with $T_{K}(x)$ :

$$
N_{K}(x):=\left\{v: v^{T} u \leq 0, \forall u \in T_{K}(x)\right\},
$$

We recall that the dual cone of a cone $N$ is defined as:

$$
N^{*}:=\left\{d \in \mathbb{R}^{n}: v^{T} d \geq 0 \forall v \in N\right\} .
$$

and thus $T_{K}(x)^{*}=-N_{K}(x)$. Furthermore, it is not hard to see that, if $K$ is convex, then $N_{K}(x)=\left\{v: v^{T}(y-x) \leq 0, \forall y \in K\right\}$.

## Some basic aspects of non smooth analysis

Finally we would like to recall some elementary and basic aspects of non smooth analysis. Let us consider, for sake of simplicity, a function $f: \mathbb{R}^{n} \rightarrow \mathbb{R}$. Let $f$ be locally Lipschitz near $x$. We can define the generalized directional derivative of $f$ which, when evaluated at $x$ in the direction $v$, is given by

$$
f^{\circ}(x, v):=\limsup _{y \rightarrow x, \lambda \rightarrow 0} \frac{f(y+\lambda v)-f(y)}{\lambda}
$$

The generalized gradient of $f$ at $x$ is the following non empty set:

$$
\begin{equation*}
\partial f(x):=\left\{\zeta \in \mathbb{R}^{n}: f^{\circ}(x, v) \geq v^{T} \zeta, \forall v \in \mathbb{R}^{n}\right\} \tag{1.2}
\end{equation*}
$$

It is possible to show that $\partial f(x)$ is a nonempty convex compact subset of $\mathbb{R}^{n}$.

We remark that if $f$ is smooth, then $\partial f(x)$ reduces to the singleton $\{\nabla f(x)\}$. When $f$ is convex, then $\partial f(x)$ coincides with the subdifferential of convex analysis, i.e. the set of vectors $\zeta \in \mathbb{R}^{n}$ such that:

$$
f(x+y)-f(x) \geq \zeta^{T} y, \quad \forall y \in \mathbb{R}^{n} .
$$

Another characterization of the generalized gradient can be obtained thanks to Rademacher's Theorem which asserts that a locally Lipschitz function is differentiable almost everywhere. Thus let $\Omega_{f}$ be the set of points in a
neighborhood of $x$ at which $f$ fails to be differentiable and let $S$ be any other set of measure zero. Then

$$
\partial f(x)=\operatorname{co}\left\{\lim _{k \rightarrow \infty} \nabla f\left(x_{k}\right): x_{k} \rightarrow x, x_{k} \notin S, x_{k} \notin \Omega_{f}\right\} .
$$

Hence, $\partial f(x)$ is the convex hull of all point of the form $\lim \nabla f\left(x_{k}\right)$, where $\left\{x_{k}\right\}$ is any sequence which converges to $x$ while avoiding $S \cup \Omega_{f}$.

For a comprehensive analysis of the properties of the generalized gradient, we refer to [16]. Here we would like to recall the following result which is a corollary of the subsequent Proposition 5.

Proposition 3. If $f$ is locally Lipschitz near $x$ and attains a minimum over $X$ at $x$, then $0 \in \partial f(x)+N_{X}(x)$.

If we consider a vector-valued function $F: \mathbb{R}^{n} \rightarrow \mathbb{R}^{m}$, the analysis becomes only slightly different from the one reported here.

### 1.2 Some aspects of penalty methods for optimization problems

Penalty methods are one of the most important and well studied classes of algorithms for the solution of constrained optimization problems (see $[19,37,82,83,106])$. The simple idea, which is the core of these methods, is that one can solve the original constrained problem by considering a (sequence of) suitable equivalent unconstrained problem(s). In order to define the equivalent penalized problem(s), in some sense, we "promote" the constraints to the objective function level by introducing a term which can measure and then penalize the feasibility violation with respect to those constraints.

Let us consider the optimization problem:

$$
\begin{equation*}
\operatorname{minimize}_{x} f(x) \quad \text { subject to } \quad x \in X \tag{1.3}
\end{equation*}
$$

where the feasible set $X \in \mathbb{R}^{n}$ is a closed subset of $\mathbb{R}^{n}$ and $f: \mathbb{R}^{n} \rightarrow \mathbb{R}$ is at least continuous.
If we can define a continuous penalty term $\psi: \mathbb{R}^{n} \rightarrow \mathbb{R}_{+}$such that:

$$
\psi(x) \begin{cases}=0 & \text { if } x \in X  \tag{1.4}\\ >0 & \text { if } x \notin X\end{cases}
$$

we introduce the following penalty function:

$$
\begin{equation*}
P(x, c):=f(x)+c \psi(x) \tag{1.5}
\end{equation*}
$$

where $c>0$ is the penalty parameter. We note that function $\psi$ could be nondifferentiable and that term $\psi(x)$ is a measure of the violation of the constraints defined by the set $X$.
With the penalization methods we aim at solving the original problem (1.3) by solving the following "simpler" unconstrained problem:

$$
\begin{equation*}
\operatorname{minimize}_{x} f(x)+c \psi(x) \tag{1.6}
\end{equation*}
$$

thus trying to "force" feasibility (at the objective function level) thanks to the presence of term $c \psi(x)$. We refer to problem (1.6) as the Penalized Optimization Problem with parameter $c, \mathrm{POP}_{c}$.

The "weight" of feasibility violation $c \in \mathbb{R}^{n}$ plays a central role in this kind of approaches. Indeed, based on it, two main classes of penalization methods could be distinguished.

- Sequential penalty approach: in this case, in order to obtain a solution of (1.3), an infinite sequence of "simpler" problems $\mathrm{POP}_{c_{k}}$, each obtained by considering different values $c_{k}$ of the penalty parameter $c$ and letting $c_{k} \rightarrow \infty$, is solved.
- Exact penalty approach: in this case problem (1.3) is recast and solved as a unique "simpler" problem $\mathrm{POP}_{c}$ with $c$ such that $c \geq \bar{c}$ for a suitable $\bar{c}$.

In a Sequential approach, we have to cope with the iterative solution of problem (1.6) with $c=c_{k}$ and $c_{k} \rightarrow \infty$. Function $P\left(x, c_{k}\right)$ is an external penalty function: indeed, for fixed values of the penalty parameter $c=c_{k}$, the solution $x_{k}=\operatorname{argmin} P\left(x, c_{k}\right)$ of problem (1.6) in general is not feasible (i.e. it does not lie in $X$ ).

It can be shown that, under mild assumptions, the sequence (after a suitable renumeration, if the case) of solutions $x_{k}$ of problem (1.6), with $c=c_{k}$ for $c_{k} \rightarrow \infty$ monotone sequence, converges to a solution of (1.3).

Proposition 4. Suppose that $f$ and $\psi$ are continuous functions; assume that problem (1.3) has an optimal solution and, for every fixed $c_{k}>0$, there exists $x_{k} \in \mathbb{R}^{n}$ such that $P\left(x_{k}, c_{k}\right)=\min _{x} P\left(x, c_{k}\right)$. Let $\left\{x_{k}\right\}$ be the sequence of such points and let $\left\{c_{k}\right\}$ be such that $c_{k+1}>c_{k}$ and $\lim _{k} c_{k}=+\infty$. If, for every $k, x_{k} \in D$ compact subset of $\mathbb{R}^{n}$, then any limit point of the sequence $\left\{x_{k}\right\}$ is a solution of the original problem (1.3).

It turns out that the main issues of this method are the choice of a suitable function $\psi$ (differentiable or nondifferentiable) and the solution of the sequence of unconstrained problems $\mathrm{POP}_{c_{k}}$.

In an Exact approach, suitable differentiable (see [38] or [82]) or non differentiable (see for example [106]) penalty terms can be employed. The aim of these methods is to solve the original constrained problem by requiring only one unconstrained minimization. It can be proved that, under suitable
assumptions, there exists $\bar{c} \in \mathbb{R}_{+}$such that, if $c>\bar{c}$, local (global) minima of penalized problem (1.6) coincide with those of problem (1.3). Many different approaches, based on the choice for a suitable penalty term, can be considered. We refer, for a complete study of these methods, to [8].

We remark here that one of the critical points of the exact approaches is the knowledge of the threshold value $\bar{c}$. A possible approach is to choose an initial value $c_{0}$ and increase it as necessary if the algorithm indicates that the current value is inadequate (see [8]).

## The distance function

Consider, now, the non differentiable but globally Lipschitz distance function $\operatorname{dist}_{X}(x): X \rightarrow \mathbb{R}$ defined by

$$
\operatorname{dist}_{X}(x):=\inf _{y \in X}\|x-y\|
$$

If $X$ is a closed set then, $x \in X \Leftrightarrow \operatorname{dist}_{X}(x)=0$. Furthermore, if $X$ is convex, then $\operatorname{dist}_{X}(\bullet)$ is convex. We recall also that the following property holds:

$$
N_{X}(x)=\operatorname{cl}\left\{\bigcup_{\lambda \geq 0} \lambda \partial \operatorname{dist}_{X}(x)\right\}
$$

where cl denotes weak closure, and

$$
v \in T_{X}(x) \Leftrightarrow v^{T} \zeta \leq 0, \text { for every } \zeta \in \partial \operatorname{dist}_{X}(x)
$$

Distance function is "naturally" connected to the exact penalization. Indeed the following result can be proved (see [16]).

Proposition 5. Let $f$ be Lipschitz with constant $L$ on a set $S$. Let $x$ belong to a set $X \subset S$ and suppose that $f$ attains a minimum over $X$ at $x$. Then for any $c \geq L$, the function $P(y, c):=f(y)+c \operatorname{dist}_{X}(y)$ attains a minimum over $S$ at $x$. If $c>L$ and $X$ is closed, then any other point minimizing $P$ over $S$ must also lie in $X$.

### 1.3 Variational Inequalities

The finite-dimensional variational inequality (VI) provides a broad unifying setting for the study of optimization and equilibrium problems and serves as the main computational framework for the practical solution of a host of continuum problems in the mathematical sciences. Needless to say, VI is a powerful modeling tool for diverse equilibrium phenomena (problems from engineering, economics and finance).

### 1.3.1 Generalities

Given a subset $K$ of $\mathbb{R}^{n}$ and a mapping $F: K \rightarrow \mathbb{R}^{n}$, the variational inequality $\mathrm{VI}(K, F)$ is to find a vector $x \in K$ such that

$$
\begin{equation*}
(y-x)^{T} F(x) \geq 0, \quad \forall y \in K \tag{1.7}
\end{equation*}
$$

The set of solutions to this problem is denoted $\operatorname{SOL}(K, F)$.
In the following sections set K is assumed to be closed and function $F$ continuous. As consequence of these assumptions, $\operatorname{SOL}(K, F)$ is a closed set.

From a geometric point of view, $\bar{x} \in K$ is a solution of $\mathrm{VI}(K, F)$ if and only if $F(\bar{x})$ forms a non-obtuse angle with every vector of the form $y-x$ for every $y \in K$. Hence, $\bar{x} \in K$ solves the $\operatorname{VI}(K, F)$ if and only if

$$
-F(\bar{x}) \in N_{K}(\bar{x}) \Leftrightarrow 0 \in F(\bar{x})+N_{K}(\bar{x})
$$

By the defining inequality (1.7) it is easy to see that $\bar{x}$ is a solution to $\mathrm{VI}(K, F)$ if and only if $\bar{x}$ is a solution of the optimization problem:

$$
\begin{equation*}
\operatorname{minimize}_{x} y^{T} F(\bar{x}) \quad \text { subject to } \quad y \in X \tag{1.8}
\end{equation*}
$$

We remark that VIs include many special relevant cases. The problem of solving systems of nonlinear equations corresponds to the case where $K$ is equal to $\mathbb{R}^{n}$. More generally, if $\bar{x}$ is a solution of the $\mathrm{VI}(K, F)$ and $\bar{x} \in \operatorname{int} K$, then $F(x)=0$.

If K is a cone, i.e. $x \in K \Rightarrow \tau x \in K, \forall \tau \geq 0$, the VI is equivalent to a complementarity problem (see [32]).

Moreover, consider the problem:

$$
\begin{equation*}
\text { minimize } f(x) \text { subject to } \quad x \in X \tag{1.9}
\end{equation*}
$$

where $f$ is continuously differentiable. By the minimum principle, if the set $K$ is convex, any solution $\bar{x}$ of (1.9) is such that:

$$
\begin{equation*}
(y-\bar{x})^{T} \nabla f(\bar{x}) \geq 0, \quad \forall y \in K \tag{1.10}
\end{equation*}
$$

which is nothing else but the $\mathrm{VI}(K, \nabla f)$. In order to understand that VI setting is richer than that of a constrained (with a convex feasible set) "simple" optimization problem, we recall that a continuously differentiable function $F$ is a gradient map and, thus, $F=\nabla f$ for some suitable function $f$, if and only if its Jacobian $J F(x)$ is symmetric.

We have observed that VIs includes as a special case the problem of solving systems of equations. It turns out that VIs can be equivalently reformulated in terms of system equations and optimization problems. For the purpose of this work we report here only the equivalent nonsmooth equation formulation of the VI with the natural map $\mathbf{F}_{K}^{n a t}$.

Proposition 6. If $K$ is closed convex and $F: K \rightarrow \mathbb{R}^{n}$ is arbitrary then

$$
\begin{equation*}
[\bar{x} \in \operatorname{SOL}(K, F)] \Leftrightarrow\left[\mathbf{F}_{K}^{n a t}(\bar{x})=0\right] \tag{1.11}
\end{equation*}
$$

where $\mathbf{F}_{K}^{n a t}(v):=v-\operatorname{Proj}_{K}(v-F(v))$ and $\operatorname{Proj}_{K}$ is the Euclidean projector onto the closed convex set $K$.

We recall that, if $K$ is a closed convex set, there exists a unique vector $\tilde{x} \in K$ that is closest to $x$ in the Euclidean norm. This vector is the projection of $s$ onto $K, \operatorname{Proj}_{K}(x)$ which is the unique solution of the following problem:

$$
\begin{equation*}
\text { minimize } \quad \frac{1}{2}\|y-x\|^{2} \quad \text { subject to } \quad y \in K \tag{1.12}
\end{equation*}
$$

When $K$ is a polyhedron, the above optimization problem is a strictly convex quadratic programming. When $K$ is not polyhedral, computing the projection onto $K$ could be in general a non trivial task (a fact that will play a role when numerical experiments will be considered).

For sake of clarity, we recall some properties of the Euclidean projector onto $K$, $\operatorname{Proj}_{k}$.
Theorem 1. Let $K \subseteq \mathbb{R}^{n}$ be nonempty, closed and convex. Then
(a) For each $x \in \mathbb{R}^{n}, \operatorname{Proj}_{K}(x)$ exists and is unique.
(b) For each $x \in \mathbb{R}^{n}$, $\operatorname{Proj}_{K}(x)$ is the unique vector $\tilde{x} \in K$ satisfying the following inequality:

$$
(y-\tilde{x})^{T}(\tilde{x}-x) \geq 0, \quad \forall y \in K
$$

(c) For any two vectors $u$ and $v$ in $\mathbb{R}^{n}$,

$$
\left(\operatorname{Proj}_{K}(u)-\operatorname{Proj}_{K}(v)\right)^{T}(u-v) \geq\left\|\operatorname{Proj}_{K}(u)-\operatorname{Proj}_{K}(v)\right\|^{2}
$$

thus Proj $_{K}$ is a co-coercive function.
(d) Function $\operatorname{Proj}_{K}(x)$ is nonexpansive; that is for any two vectors $u$ and $v$ in $\mathbb{R}^{n}$,

$$
\left\|\operatorname{Proj}_{K}(u)-\operatorname{Proj}_{K}(v)\right\| \leq\|u-v\|
$$

thus Proj $_{K}$ is a globally Lipschitz continuous function.
Instead of the Euclidean norm, we may use a vector norm induced by a symmetric positive definite matrix to define the projection operator. This results in a skewed projector, which leads to a generalization of the natural map of a VI (for the details see [32]).

### 1.3.2 Existence results

One of the main existence results for VIs is based on the following proposition.

Proposition 7. Let $K \subseteq \mathbb{R}^{n}$ be closed convex and $F: K \rightarrow \mathbb{R}^{n}$ be continuous. Consider the following statements:
(a) There exixts $x^{r e f} \in K$ such that the set:

$$
L_{<}:=\left\{x \in K: F(x)^{T}\left(x-x^{r e f}\right)<0\right\}
$$

is bounded.
(b) There exists a bounded open set $\Omega$ and $x^{r e f} \in K \cap \Omega$ such that

$$
F(x)^{T}\left(x-x^{r e f}\right) \geq 0, \forall x \in K \cap b d \Omega
$$

(c) The $\operatorname{VI}(K, F)$ has a solution.

It holds that $(a) \Rightarrow(b) \Rightarrow(c)$. Moreover, if the set

$$
L_{\leq}:=\left\{x \in K: F(x)^{T}\left(x-x^{r e f}\right) \leq 0\right\}
$$

is bounded, then $\operatorname{SOL}(K, F)$ is nonempty and compact.
We remark that, in the proof of this proposition, the reformulation (1.11) plays a central role. A consequence of the previous proposition is that, if $K$ is compact and convex and $F$ is continuous, then $\operatorname{SOL}(K, F)$ is nonempty and compact. Moreover, if we drop the compactness assumption on set $K$, we can demonstrate the following theorem.

Theorem 2. Let $K \subseteq \mathbb{R}^{n}$ be closed convex and $F: K \rightarrow \mathbb{R}^{n}$ be continuous.
(a) If $F$ is strictly monotone on $K$, the $V I(K, F)$ has at most one solution.
(b) If $F$ is strongly monotone on $K$, the $\operatorname{VI}(K, F)$ has a unique solution.

Furthermore, it can be proved (see [32]) that, if $F$ is pseudo monotone on $K$ closed and convex then $\operatorname{SOL}(K, F)$ is convex, while for pseudo monotone plus $\mathrm{VI}(K, F), F(\operatorname{SOL}(K, F))$ is a singleton.

### 1.3.3 Solution algorithms

There are many approaches for the solution of VIs. In this brief survey we mention the KKT-based algorithms which "simply" try to find a KKT triple of the $\mathrm{VI}(K, F)$ and algorithms based on the definition of suitable merit functions for the VI. The former methods are very attractive because they are simple and convergence results can be established by exploiting the particular structure of the KKT system. The main drawback of these schemes is probably the fact that KKT algorithms could not fully exploit the properties that a problem may have (for example monotonicity). The latter methods are attractive theoretically but we recall that the evaluation of the merit function is, in general, a non trivial task by itself. For a most complete review of such methods we refer to [32].

In the following part of this section we recall some methods that take advantage of the presence of some kind of monotonicity in the VI problem, namely projection methods, Tikhonov and proximal point methods.

Projection methods require the ability to efficiently calculate the projection onto the closed convex set $K$; this feature certainly limits the applicability of these methods. Another peculiarity of these algorithms is that they do not require the use of the derivatives of $F$ and do not involve any complex computation besides the projection on $K$ : needless to say the use of no derivative information prevents these methods from being fast. Thus, for sets $K$ such that the projection can be easily carried out, projection methods can be applied to the solution of very large problems because of their simplicity.

Projection methods are based on the Banach fixed-point theorem (see Theorem 2.1.21 in [32]). We know, indeed, by (1.11) that, if $K$ is a closed convex subset of $\mathbb{R}^{n}, \bar{x}$ is a solution of $\mathrm{VI}(K, F)$ if and only if

$$
\begin{equation*}
\bar{x}=\operatorname{Proj}_{K}(\bar{x}-\tau F(\bar{x})), \tag{1.13}
\end{equation*}
$$

for positive scalar $\tau$. Hence, fixed points of the mapping

$$
x \rightarrow \operatorname{Proj}_{K}(x-\tau F(x))
$$

are solutions of the $\mathrm{VI}(K, F)$ and vice versa. If the latter map is a contraction, a fixed-point contraction algorithm can be used to calculate a solution of the VI. Below we rephrase the fixed-point contraction algorithm in this context and we get the Basic Projection Algorithm (BPA).

```
Algorithm 1: BPA
(S.0) : \(x_{0} \in K\), Set \(k:=0\).
(S.1) : If \(x_{k}=\operatorname{Proj}_{K}\left(x_{k}-\tau F\left(x_{k}\right)\right):\) STOP
(S.2) : Set \(x_{k+1}:=\operatorname{Proj}_{K}\left(x_{k}-\tau F\left(x_{k}\right)\right), k \leftarrow k+1\); go to (S.1).
```

The following result holds.
Theorem 3. Let $F: K \rightarrow \mathbb{R}^{n}$, where $K$ be closed convex subset of $\mathbb{R}^{n}$. Suppose that $L$ and $\mu$ are such that for any $x$ and $y$ in $K$,

$$
\begin{equation*}
(F(x)-F(y))^{T}(x-y) \geq \mu\|x-y\|^{2} \tag{1.14}
\end{equation*}
$$

and

$$
\begin{equation*}
\|F(x)-F(y)\| \leq L\|x-y\| \tag{1.15}
\end{equation*}
$$

If $\frac{1}{\tau}>\frac{L^{2}}{2 \mu}, \operatorname{Proj}_{K}(x-\tau F(x))$ is a contraction from $K$ to $K$; therefore every sequence $\left\{x_{k}\right\}$ produced by BPA converges to the unique solution of the $V I(K, F)$.

The results here reported can be generalized if we consider another norm $\|\bullet\|_{D}$ (see [32]). Hence, if $F$ is strongly monotone and Lipschitz continuous and the
constants $L$ and $\mu$ are available and $\tau$ chosen as described, the sequence $\left\{x_{k}\right\}$ such that:

$$
x_{k+1}:=\operatorname{Proj}_{K}\left(x_{k}-\tau F\left(x_{k}\right)\right)
$$

converges to the unique solution of the $\operatorname{VI}(K, F)$. Two are the main disadvantages of this basic approach. Firstly function $F$ must be strongly monotone, secondly the knowledge of $L$ and $\mu$ is required while these constants are in general unknown a priori. For these reasons three other projection-type algorithms were developed: the Projection Algorithm with Variable Steps (PAVS), the Extragradient method (EgA) and the Hyperplane Projection method (HPA).

In a variable-step projection scheme (PAVS), the step size $\tau$ is allowed to vary from one iteration to the next; however we remark that the resulting algorithm does not perform a line-search routine in order to calculate the variable-step size $\tau_{k}$.

It can be proved that if $F$ is co-coercive with constant $c$, then $\left\{\tau_{k}\right\}$ can be chosen so that $x_{k}$ converges to a solution of the VI. We remind that every strongly monotone, Lipschitz continuous function is co-coercive but not vice versa; then the convergence properties of this projection algorithm involve a broader class of VIs.

Theorem 4. Let $K$ be a closed convex subset of $\mathbb{R}^{n}$ and $F: K \rightarrow \mathbb{R}^{n}$ be co-coercive on $K$ with constant $c$. Suppose that $\operatorname{SOL}(K, F) \neq \emptyset$. If

$$
0<\inf _{k} \tau_{k} \leq \sup _{k} \tau_{k}<2 c
$$

PAVS generates a sequence $\left\{x_{k}\right\}$ converging to a solution of the $\operatorname{VI}(K, F)$.
Here it suffices to say that the demonstration of the previous theorem relies on the co-coercivity of the natural map $\mathbf{F}_{K, \tau}^{\mathrm{nat}}$ of the $\mathrm{VI}(K, F)$ : indeed it can be proved that $\mathbf{F}_{K, \tau}^{\text {nat }}$ is co-coercive on $K$ with constant $1-\tau / 4 c$ if $\tau \in(0,4 c)$.

The Extragradient method (EgA) consists in a projection algorithm that executes two projections per iteration. Although this undoubtedly requires twice the amount of computations, the benefit is significant because the resulting algorithm is applicable to the class of pseudo monotone VIs. However, the function $F$ is still required to be Lipschitz continuos and an estimate of its Lipschitz constant is needed. The EgA takes its name from the extra evaluation of F (and the extra projection) that is called for in each iteration. The name originates from the case of a symmetric VI. In this case, the VI represents the optimality condition of a differentiable optimization problem (see (1.10)), the extra evaluation of F corresponds to an extra evaluation of the gradient of the objective function, thus extragradient. At each iteration, this method calculates $y^{k}:=\operatorname{Proj}_{K}\left(x_{k}-\tau F\left(x_{k}\right)\right)$ and a second projection $x^{k+1}:=\operatorname{Proj}_{K}\left(x_{k}-\tau F\left(y_{k}\right)\right)$. It turns out that, if $K$ is a closed convex set and $F$ a mapping from $K$ into $\mathbb{R}^{n}$ that is pseudo monotone on $K$ with respect to $\operatorname{SOL}(K, F)$ and Lipschitz continuous on $K$ with constant $L$ and if $\tau<1 / L$,
then the sequence $x_{k}$ generated by the EgA converges to a solution of the $\mathrm{VI}(K, F)$.

The extragradient method still requires the knowledge of the Lipschitz constant $L$ of $F$, which is usually not known. Then, we introduce an enhanced extragradient-like method, the HPA, that neither requires $F$ to be Lipschitz continuous nor calls for the knowledge of potentially unknown constants. Let $K$ be a closed convex set and F be a continuous mapping from $K$ into $\mathbb{R}^{n}$ that is pseudo monotone on $K$ with respect to $\operatorname{SOL}(K, F)$. Let $\tau>0$ be a fixed scalar. The algorithm can be described geometrically as follows. Let $x_{k} \in K$ be given. First compute the point $\operatorname{Proj}_{K}\left(x_{k}-\tau F\left(x_{k}\right)\right)$. Next search the line segment joining $x_{k}$ and $\operatorname{Proj}_{K}\left(x_{k}-\tau F\left(x_{k}\right)\right)$, by a simple Armijo-type line search routine, for a point $z_{k}$ such that the hyperplane $H_{k}=\left\{x \in \mathbb{R}^{n}: F\left(z_{k}\right)^{T}\left(x-z_{k}\right)=0\right\}$ strictly separates $x_{k}$ from $\operatorname{SOL}(K, F)$. We then project $x_{k}$ onto $H_{k}$ and the resulting point onto $K$, obtaining $x_{k+1}$. It can be shown that $x_{k+1}$ is closer to $\operatorname{SOL}(K, F)$ than $x_{k}$. The resulting method requires three projections per iteration, two onto $K$ and one onto $H_{k}$. The latter projection is easily performed and is given by an explicit formula and thus, also from a computational point of view, is less heavy. It can be shown that, if $F$ is a continuous mapping that is pseudo monotone on $K$ with respect to $\operatorname{SOL}(K, F)$, the sequence produced by HPA converges to a solution of the $\mathrm{VI}(K, F)$.

For a more detailed analysis of these approaches and other refinements, we refer to [32].

With the Tikhonov (see [97, 98]) regularization we aim at solving the VI by considering a sequence of problems that are better behaved. The regularization of the $\mathrm{VI}(K, F)$ in terms of the family of perturbed strongly monotone VIs $\left(K, F_{\varepsilon}\right)$, where $F_{\varepsilon}:=F+\varepsilon I$ and $\varepsilon$ is a positive parameter, was defined for monotone VIs because a monotone problem generally lacks the kind of strong stability properties that are present in a strongly monotone problem.

For each $\varepsilon>0$, let $x(\varepsilon)$ be the unique solution of the strongly monotone $\mathrm{VI}\left(K, F_{\varepsilon}\right)$; the family of solutions

$$
\begin{equation*}
\{x(\varepsilon): \varepsilon>0\} \tag{1.16}
\end{equation*}
$$

is called the Tikhonov trajectory of the $\mathrm{VI}(K, F)$. We remark that unlike the $\mathrm{VI}\left(K, F_{\varepsilon}\right)$, the original $\mathrm{VI}(K, F)$ may have multiple or no solutions. Thus, if we consider the limit $\lim _{\varepsilon \rightarrow 0} x(\varepsilon)$, it does not always exist. The following result paves the way for the practical use of the Tikhonov approach.

Theorem 5. Let $K \in \mathbb{R}^{n}$ be closed convex and $F$ be continuous and monotone on $K$. Let (1.16) be the Tikhonov trajectory. The following three statements are equivalent:
(a) $\lim _{\varepsilon \rightarrow 0} x(\varepsilon)$ exists;
(b) $\lim \sup _{\varepsilon \rightarrow 0} x(\varepsilon)<\infty$;
(c) $\operatorname{SOL}(K, F) \neq \emptyset$.

It is worth mentioning that Theorem 5 can be generalized by considering the family:

$$
\{\mathrm{VI}(K, F+\varepsilon G): \varepsilon>0\}
$$

where $G: \mathbb{R}^{n} \rightarrow \mathbb{R}^{n}$ is a strongly monotone (on $K$ ) mapping. Furthermore, the previous result can be generalized by weakening the monotonicity of $F$ (see [32]).

The Tikhonov regularization leads to an iterative algorithm for approximating a solution of a VI, provided that such a solution exists: the method involves the solution of a sequence of sub-VIs. The rationale is quite simple: once we have fixed a decreasing sequence $\left\{\varepsilon_{k}\right\}$ of positive numbers converging to zero, for each $k$ we in turn use an appropriate algorithm to compute $x\left(\varepsilon_{k+1}\right)$ solution (possibly inexact) of the strongly monotone sub-VI $\left(K, F+\varepsilon_{k} I\right)$.

The Tikhonov regularization algorithm may have a computational drawback: when $\varepsilon_{k}$ goes to zero, the perturbed problems approach the original problem and thus it may become more and more difficult to solve them. In order to overcome this difficulty, in Proximal point methods, a sequence of subproblems is still (approximately) solved, but at step $k$ the perturbing function is $c_{k}\left(x-x_{k-1}\right)$, for some $c_{k}>0$, instead of $\varepsilon_{k} x$. Intuitively, if the sequence $\left\{x_{k}\right\}$ converges, the term $c_{k}\left(x_{k}-x_{k-1}\right)$ approaches zero provided that $c_{k}$ remains bounded; thus $c_{k}$ does not need to go to zero. As result, the core step of such methods is the (approximate) solution of the strongly monotone $\mathrm{VI}\left(K, F^{k}\right)$, where $F^{k}(x):=c_{k}\left(x-x^{k}\right)+F(x)$. For a complete and elegant theoretical analysis of Proximal point algorithms we refer to the fundamental [32].

Needless to say, there exist many other methods for the solution of VIs. Here we briefly reviewed those that form the basis of the developments in the next chapters.

# Chapter 2 <br> Partial penalization for the solution of GNEPs 

### 2.1 Preliminaries

Non-cooperative game theory is a branch of game theory for the resolution of conflicts among players (or economic agents), each behaving selfishly to optimize one's own well-being subject to resource limitations and other constraints that may depend on the rivals' actions. The Generalized Nash Equilibrium Problem (GNEP) is a central model in game theory that has been used actively in many fields in the past fifty years. But it is only since the mid-nineties that research on this topic gained momentum, especially in the operations research (OR) community. The GNEP lies at the intersection of many different disciplines (e.g. economics, engineering, mathematics, computer science, OR), and sometimes researchers in different fields worked independently and unaware of existing results. This explains why this problem has a number of different names in the literature including pseudo-game, social equilibrium problem, equilibrium programming, coupled constraint equilibrium problem, and abstract economy. We will use the term generalized Nash equilibrium problem that seems the one favorite by OR researchers in recent years.

Formally, the GNEP consists of $N$ players, each player $\nu$ controlling the variables $x^{\nu} \in \mathbb{R}^{n_{\nu}}$. We denote by $\mathbf{x}$ the vector formed by all these decision variables

$$
\mathbf{x}:=\left(\begin{array}{c}
x^{1} \\
\vdots \\
x^{N}
\end{array}\right),
$$

which has dimension $n:=\sum_{\nu=1}^{N} n_{\nu}$, and by $\mathbf{x}^{-\nu}$ the vector formed by all the players' decision variables except those of player $\nu$. For sake of clarity, we sometimes write ( $x^{\nu}, \mathbf{x}^{-\nu}$ ) instead of $\mathbf{x}$.

Each player's strategy belongs to a set $\mathcal{F}_{\nu}\left(\mathbf{x}^{-\nu}\right) \subseteq \mathbb{R}^{n_{\nu}}$ that depends on the rival players' strategies. The aim of player $\nu$, given the other players' strategies $\mathbf{x}^{-\nu}$, is to choose a strategy $x^{\nu}$ that solves the minimization prob-
lem

$$
\begin{equation*}
\operatorname{minimize}_{x^{\nu}} \theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right) \quad \text { subject to } \quad x^{\nu} \in \mathcal{F}_{\nu}\left(\mathbf{x}^{-\nu}\right) \tag{2.1}
\end{equation*}
$$

The GNEP is the problem of finding a vector $\overline{\mathbf{x}}$ such that each player's strategy $\bar{x}^{\nu}$ solves the player's problem (given $\overline{\mathbf{x}}^{-\nu}$ ):

$$
\theta_{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right) \leq \theta_{\nu}\left(y^{\nu}, \overline{\mathbf{x}}^{-\nu}\right), \quad \forall y^{\nu} \in \mathcal{F}_{\nu}\left(\overline{\mathbf{x}}^{-\nu}\right)
$$

Such a point $\overline{\mathbf{x}}$ is called a (generalized) Nash equilibrium or, more simply, a solution of the GNEP. For any $\mathbf{x}^{-\nu}$, the solution set of problem (2.1) is denoted by $\mathcal{S}_{\nu}\left(\mathbf{x}^{-\nu}\right)$. Then, we see that a point $\overline{\mathbf{x}}$ is a solution if and only if

$$
\bar{x}^{\nu} \in \mathcal{S}_{\nu}\left(\overline{\mathbf{x}}^{-\nu}\right) \quad \text { for all } \nu=1, \ldots, N
$$

If we denote by $\mathcal{S}(\mathbf{x})$ the set $\mathcal{S}(\mathbf{x}):=\Pi_{\nu=1}^{N} \mathcal{S}_{\nu}\left(\mathbf{x}^{-\nu}\right)$, we see that we can say that $\overline{\mathbf{x}}$ is a solution if and only if $\overline{\mathbf{x}} \in \mathcal{S}(\overline{\mathbf{x}})$, i.e. if and only if $\overline{\mathbf{x}}$ is a fixed point of the point-to-set mapping $\mathcal{S}$.

If the feasible sets are actually fixed, i.e. if $\mathcal{F}_{\nu}\left(\mathbf{x}^{-\nu}\right)=\mathcal{F}_{\nu}$ for some given sets $\mathcal{F}_{\nu}$, the problem reduces to a standard Nash Equilibrium Problem (NEP) which is, to a certain extent, a much simpler problem. An intermediate case between NEPs and GNEPs is the so called jointly convex GNEP, where for all $\nu$

$$
\mathcal{F}_{\nu}\left(\mathbf{x}^{-\nu}\right):=\left\{x^{\nu} \in \mathbb{R}^{n_{\nu}}:\left(x^{\nu}, \mathbf{x}^{-\nu}\right) \in \mathbf{F}\right\}
$$

with $\mathbf{F} \subseteq \mathbb{R}^{n}$ a nonempty closed convex set.
We assume that the feasible sets of the players are defined by
$\mathcal{F}_{\nu}\left(\mathbf{x}^{-\nu}\right)=K_{\nu} \cap X_{\nu}\left(\mathbf{x}^{-\nu}\right), \quad$ where $X_{\nu}\left(\mathbf{x}^{-\nu}\right):=\left\{x^{\nu}: g^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right) \leq 0\right\}$,
thus distinguishing, as usual, between the "private" constraints represented by $K_{\nu}$ and those constraints $g^{\nu}$ that depend on both the player's variables and its rivals' ones. In general, private constraints can be such that $K_{\nu}=\mathbb{R}^{n_{\nu}}$ : this means, with a different point of view, that sometimes we include the private constraints among the $g^{\nu}$ S (the coupling ones). In the setting described so far, the players' problems become

$$
\begin{align*}
\min _{x^{\nu}} & \theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right) \\
& g^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right) \leq 0  \tag{2.2}\\
& x^{\nu} \in K_{\nu} .
\end{align*}
$$

The Debreu paper [21] gives the first existence theorem for the solutions of GNEPs. This existence result is based on fixed-point arguments, by exploiting the fact that a point $\overline{\mathbf{x}}$ is an equilibrium if $\overline{\mathbf{x}} \in \mathcal{S}(\overline{\mathbf{x}})$, where $\mathcal{S}(\mathbf{x}):=\Pi_{\nu=1}^{N} \mathcal{S}_{\nu}\left(\mathbf{x}^{-\nu}\right)$. There also exist some other approaches, an interesting one being that in [43], where a continuation approach is used. The
main existence result is probably the one established in [5]. The following theorem is a slightly simplified version (see [52]).

Theorem 6. Let a GNEP be given with at least $C^{0} \theta_{\nu}$ functions and suppose that
(a) There exist $N$ nonempty, convex and compact sets $D_{\nu} \in \mathbb{R}^{n_{\nu}}$ such that for every $\mathbf{x} \in \mathbb{R}^{n}$ with $x^{\nu} \in D_{\nu}$ for every $\nu, X_{\nu}\left(\mathbf{x}^{-\nu}\right)$ is nonempty, closed and convex, $X_{\nu}\left(\mathbf{x}^{-\nu}\right) \subseteq D_{\nu}$, and $X_{\nu}$, as a point-to-set map, is both upper and lower semicontinuous.
(b) For every player $\nu$, the function $\theta_{\nu}\left(\bullet, \mathbf{x}^{-\nu}\right)$ is quasi-convex on $X_{\nu}\left(\mathbf{x}^{-\nu}\right)$.

Then a generalized Nash equilibrium exists.
Remark 1. When the sets $X_{\nu}$ are defined by inequality constraints, the lower and upper semicontinuity requirements translate into reasonably mild conditions on the functions $g_{\nu}$. See for example [88].

It is worth pointing out that the relaxation of the assumptions in the previous theorem has been an important topic of research. Relaxations of the continuity assumptions, compactness assumptions and quasi-convexity assumption have all been considered in the literature. The relaxation of the continuity assumption is probably the most interesting and well studied one. Indeed it can be seen that, if the objective functions $\theta_{\nu}$ are assumed to be pseudocontinuous instead of continuous, then the previous result still holds.

In the sequel, also in view of Theorem 6, we shall make the following blanket assumptions, valid for all $\nu=1, \ldots, N$ :
(A1) $\theta_{\nu}: \mathbb{R}^{n} \rightarrow \mathbb{R}$ is continuously differentiable and, as a function of $x^{\nu}$ alone, convex $\forall \mathbf{x}^{-\nu} \in K_{-\nu}:=\prod_{\nu \neq \nu^{\prime}=1}^{N} K_{\nu^{\prime}}$;
(A2) $g^{\nu}: \mathbb{R}^{n} \rightarrow \mathbb{R}^{m_{\nu}}$ and, for $i=1, \ldots, m_{\nu}, g_{i}^{\nu}$ is continuously differentiable and, as a function of $x^{\nu}$ only, convex $\forall \mathbf{x}^{-\nu} \in K_{-\nu}$;
(A3) $K_{\nu}$ is a closed convex and nonempty set.
These assumptions are rather standard, even if many efforts have been made in order to relax them. We do not deal with the issue of relaxing these assumptions here, and simply note that at this stage of the development of solution algorithms, they seem more or less essential in order to propose efficient numerical methods.

It is useful to illustrate the above definitions with a simple example.
Example 1. Consider a game with two players, i.e. $N=2$, with $n_{1}=1$ and $n_{2}=1$, so that each player controls one variable (for simplicity we therefore set $x_{1}^{1}=x^{1}$ and $x_{1}^{2}=x^{2}$ ). Assume that the players' problems are
$\min _{x^{1}} \frac{1}{2}\left(x^{1}-1\right)^{2}$
$\min _{x^{2}}-x^{2}$
s.t. $-x^{1}+x^{2} \leq 1$,
s.t. $\quad-x^{1}+x^{2} \leq 1$.


Fig. 2.1: In bold the solution set $\mathcal{S}(\mathbf{x})$ of Example 1 .

The optimal solution sets are given by

$$
\mathcal{S}_{1}\left(x^{2}\right)=\left\{\begin{array}{ll}
1, & \text { if } x^{2} \leq 2, \\
x^{2}-1, & \text { if } x^{2} \geq 2,
\end{array} \quad \text { and } \quad \mathcal{S}_{2}\left(x^{1}\right)=x^{1}+1\right.
$$

Then it is easy to check that the solutions of this problem are given by $(\alpha, \alpha+1)$ for every $\alpha \geq 1$. Note that the problem has infinitely many solutions and it is an instance of a jointly convex GNEP.
We may distinguish three main classes of GNEPs based on the "nature" of the players' feasible sets:

- General GNEPs
- Jointly convex GNEPs: where the feasible set of each player $\nu$ is defined by $\left(x^{\nu}, \mathbf{x}^{-\nu}\right) \in \mathbf{F}$, with $\mathbf{F} \subseteq \mathbb{R}^{n}$ a nonempty closed convex set shared by all players. It is not difficult to see that it is equivalent to say that $g^{1}=\cdots=g^{N}=g$ and $g$ is a convex function in $\mathbf{x}$
- NEPs: where $g^{1}, \ldots, g^{N}$ vanish and the only constraints left are the private ones, i.e. $x^{\nu} \in K_{\nu}$

The relations between these three classes are very simple and are illustrated in Figure 2.2.
As it is easy to understand, the more general the problem is the more difficult the solution of the problem itself becomes. In the following we make some general considerations on the solution of these three classes of problems.
NEPs are the simplest GNEPs to solve. Indeed it can be shown, see e.g. [32], that under (A1)-(A3) the NEP can be equivalently reformulated and solved as the variational inequality problem $\mathrm{VI}(K, f)$ :

$$
\begin{equation*}
\text { find } \overline{\mathbf{x}} \in K:=\prod_{\nu=1}^{N} K_{\nu} \text { such that } f(\overline{\mathbf{x}})^{T}(y-\overline{\mathbf{x}}) \geq 0, \forall y \in K \tag{2.3}
\end{equation*}
$$



Fig. 2.2: The three classes of GNEP.
where $f(\mathbf{x})=\left(\nabla_{x^{\nu}} \theta_{\nu}(\mathbf{x})\right)_{\nu=1}^{N}$. This problem, under some monotonicity assumptions on $f$, can be efficiently solved by a host of methods (see Chapter 1). However, it is worth mentioning that other globally convergent methods are available for the solution of a NEP that works well under similar conditions [31]. We refer the interested reader to [31, 32, 34] and to the vast literature therein.

Unfortunately, when one considers jointly convex or general GNEPs, the picture immediately becomes more complex. Jointly convex problems as well as general GNEPs can be reformulated as the quasi-variational inequality problem $\operatorname{QVI}(\mathcal{F}(\mathbf{x}), f)$ :

$$
\text { find } \overline{\mathbf{x}} \in \mathcal{F}(\overline{\mathbf{x}}):=\prod_{\nu=1}^{N} \mathcal{F}_{\nu}\left(\overline{\mathbf{x}}^{-\nu}\right) \text { such that } f(\overline{\mathbf{x}})^{T}(y-\overline{\mathbf{x}}) \geq 0, \forall y \in \mathcal{F}(\overline{\mathbf{x}})
$$

Note that the difference with VI (2.3) is that in the QVI case the feasible set depends on $\mathbf{x}$. However, the algorithms for QVIs (see, e.g., Chan and Pang [15], Fukushima [39] and Pang and Fukushima [79]) are far less advanced than those for VIs and, in spite of some interesting and promising recent adavncements (see [62]), no efficient numerical methods based on the QVI reformulation have been developed yet.

Nevertheless in the jointly convex case, some VI techniques can still be employed. Indeed, it can be proved that an equilibrium can be computed by solving the "simple" variational inequality $\operatorname{VI}(\mathbf{F}, f(\mathbf{x}))$, where $\mathbf{F}=\{\mathbf{x} \in K$ : $g(\mathbf{x}) \leq 0\}$, see $[28,46]$.

Theorem 7. Let a jointly convex GNEP be given and suppose that (A1)(A3) hold. Then every solution of the $\operatorname{VI}(\mathbf{F}, f(\mathbf{x}))$ is also a solution of the GNEP.

The particular equilibria that one obtains by solving $\operatorname{VI}(\mathbf{F}, f(\mathbf{x}))$ are called variational equilibria (also known in the literature as normalized equilibria). We remark that the previous result does not say that any solution of a jointly convex GNEP is also a solution of the $\operatorname{VI}(\mathbf{F}, f(\mathbf{x}))$.

We illustrate the previous considerations by the following simple examples.

Example 2. Consider a game with two players, i.e. $N=2$, with $n_{1}=1$ and $n_{2}=1$, so that each player controls one variable (for simplicity we therefore set $x_{1}^{1}=x^{1}$ and $x_{1}^{2}=x^{2}$ ). Assume that the players' problems are

$$
\begin{array}{ll}
\min _{x^{1}} & x^{1} \\
\text { s.t. } & \min _{x^{2}} x^{2} \\
\text { s })^{2}+\left(x^{2}\right)^{2} \leq 1, & \text { s.t. } \quad\left(x^{1}\right)^{2}+\left(x^{2}\right)^{2} \leq 1
\end{array}
$$

The optimal solution sets are given by

$$
\mathcal{S}_{\nu}\left(x^{-\nu}\right)=\left\{\begin{array}{ll}
-\sqrt{1-\left(x^{-\nu}\right)^{2}}, & \text { if }-1 \leq x^{-\nu} \leq 1, \\
\emptyset, & \text { if } x^{-\nu}>1 \text { or } x^{-\nu}<-1
\end{array}, \quad \nu=1,2\right.
$$

Then it is easy to check that the solutions of this problem are given by $\left(x^{1}\right)^{2}+\left(x^{2}\right)^{2}=1$ with $-1 \leq x^{\nu} \leq 0, \nu=1,2$. Note that the problem has infinitely many solutions, see Figure 2.3. Taking into account that


Fig. 2.3: The solution set $\mathcal{S}(\mathbf{x})$ of Example 2.
$\left(\nabla_{x^{1}} \theta_{1}, \nabla_{x^{2}} \theta_{2}\right)=(1,1)$, the variational equilibria are obtained by solving the following variational inequality:
$\left(\begin{array}{ll}1 & 1\end{array}\right)\binom{y^{1}-x^{1}}{y^{2}-x^{2}} \geq 0, \quad \forall\left(y^{1}, y^{2}\right):\left(y^{1}\right)^{2}+\left(y^{2}\right)^{2} \leq 1$,

$$
\text { with }\left(x^{1}, x^{2}\right):\left(x^{1}\right)^{2}+\left(x^{2}\right)^{2} \leq 1
$$

whose only solution is $\left(-\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right)$, see Figure 2.4.
It is important to underline that $\mathcal{F}(\mathbf{x})$ neither includes nor is included in $\mathbf{F}$ in general, see Figure 2.5. Furthermore, in the passage from the GNEP to the VI it is clear that most solutions are lost.


Fig. 2.4: The only variational equilibrium
Fig. 2.5: In Example 2, $\mathbf{x}=(1 / 2,1 / 2)$ and $\mathcal{F}(\overline{\mathbf{x}})=[-\sqrt{3} / 2, \sqrt{3} / 2] \times$ $[-\sqrt{3} / 2, \sqrt{3} / 2]$ (dashed set) neither includes nor is included in $\mathbf{F}$ (gray set)

Example 3. Consider another game with two players, i.e. $N=2$, with $n_{1}=1$ and $n_{2}=1$, so that each player controls one variable (for simplicity we therefore set $x_{1}^{1}=x^{1}$ and $x_{1}^{2}=x^{2}$ ). Assume that the players' problems are

$$
\begin{array}{ll}
\min _{x^{1}} & \frac{1}{2}\left(x^{1}-1\right)^{2} \\
\text { s.t. } & \min _{x^{2}} \frac{1}{2}\left(x^{2}\right)^{2}+\frac{8}{9} x^{2} \\
\left.x^{1}\right)^{2}-x^{2} \leq 0, & \text { s.t. } \quad\left(x^{1}\right)^{2}-x^{2} \leq 0 .
\end{array}
$$

The optimal solution sets are given by

$$
\mathcal{S}_{1}\left(x^{2}\right)=\left\{\begin{array}{ll}
\sqrt{x^{2}}, & \text { if } 0 \leq x^{2} \leq 1, \\
1, & \text { if } x^{2} \geq 1, \\
\emptyset, & \text { if } x^{2}<0
\end{array} \quad \text { and } \quad \mathcal{S}_{2}\left(x^{1}\right)=\left(x^{1}\right)^{2}\right.
$$

Then it is easy to check that the solutions of this problem are given by $\left(x^{1}\right)^{2}-x^{2}=0$ with $0 \leq x^{1} \leq 1$. Note that the problem has infinitely many solutions, see Figure 2.6. Taking into account that $\left(\nabla_{x^{1}} \theta_{1}, \nabla_{x^{2}} \theta_{2}\right)=$ $\left(x^{1}-1, x^{2}+\frac{8}{9}\right)$, the variational equilibria are obtained by solving the following variational inequality:

$$
\begin{aligned}
&\left(x^{1}-1 x^{2}+\frac{8}{9}\right)\binom{y^{1}-x^{1}}{y^{2}-x^{2}} \geq 0, \quad \forall\left(y^{1}, y^{2}\right):\left(y^{1}\right)^{2}-y^{2} \leq 0 \\
& \text { with }\left(x^{1}, x^{2}\right):\left(x^{1}\right)^{2}-x^{2} \leq 0
\end{aligned}
$$

whose only solution is $\left(\frac{1}{3}, \frac{1}{9}\right)$, see Figure 2.6.
The considerations made so far allow us to say that the computation of a solution of a simple NEP and of a variational equilibrium of a jointly convex GNEP is a reasonable task for which well established techniques can be used. Vice versa the calculation of non-variational equilibria of jointly


Fig. 2.6: In bold the set $\mathcal{S}(\mathbf{x})$ of Example 3; in red the variational equilibrium.
convex GNEPs and of solutions of general GNEPs is far less advanced and the subject of current research. Indeed, we remark that jointly convex and general GNEPs usually have a manifold of solutions (that sometimes could be disconnected) which is one of the many peculiarities that make these problems so difficult to analyze and solve.

We mention briefly that the computation of a solution of a general GNEP is just one of the main issues on which recent research has focused. Here we only cite some of the other topics of interest.

- Calculation of a good sampling of the solution set (see e.g. [24, 36, 71])
- Development of distributed algorithm (see e.g. [34, 80, 81])
- Characterization of the solutions to which an algorithm converges (see e.g. [40])
- Development of Newton-type methods (see e.g. [29])

Until very recently there were no provably convergent algorithms for the solution of general GNEPs: here we mention KKT methods (see e.g. [23]), the Nikaido-Isoda reformulation of the GNEP (see e.g. [25]) and other reformulation approaches (see e.g. [31, 33]). Below we consider one of the most promising approach for which we are able to prove convergence results: penalty methods.

This chapter is organized as follows. In the next section we present a brief historical introduction to the GNEPs. In Section 2.3 we discuss existing results in the literature and the main contributions of this work. In Section 2.4 we show how to (partially) penalize the coupling constraints and reduce the original problem to a "simpler" constrained NEP and, in Section 2.5, we analyze the relations between the original GNEP and the transformed problem. In Section 2.6 we propose a way to algorithmically update the penalty parameters to the "right" values while Section 2.7 deals with the solution of the non differentiable penalized NEP (for fixed values of the penalty parameters):
in this latter section finally we are able to state conditions for our algorithm to converge.

### 2.2 Brief historical background

The Nash equilibrium problem (NEP), where $\mathcal{F}_{\nu}\left(\mathbf{x}^{-\nu}\right)=\mathcal{F}_{\nu}$ for all $\nu=$ $1, \ldots, N$, was formally introduced by Nash in his 1950/1 papers [73, 74]. Nash 1950 paper [73] in the Proceedings of the National Academy of Sciences contains the main results of his 28 pages long PhD thesis and was submitted just fourteen months after he started his graduate studies! Obviously, as usually happens, ideas related to what we currently call game theory can be found in the work of much older researchers. Cournot [20], in the first half of the XIX century, proposed, in the context of an oligopolistic economy, a model that can now be interpreted as a two-players game and whose solution is a Nash equilibrium. But it was only at the beginning of the XX century that major developments occured, with the work of von Neumann [101] and von Neumann and Morgenstern, who published in 1944 the book Theory of Games and Economic Behavior [102] that essentially deals with zero-sum, two players-games, and that established game theory as an important field of study. The notion of Nash equilibrium introduced in [73] expanded enormously the scope of game theory, that had previously been essentially limited to zero-sum, two-players games, and proved to be the fundamental springboard for all successive developments. Nash papers [73, 74] are a landmark in the scientific history of the twentieth century and the notion of Nash equilibrium has extensively proved to be powerful, flexible, and rich of consequences.

In spite of all its importance, the need of an extension of the NEP, where the players interact also at the level of the feasible sets, soon emerged as necessary. The GNEP was first formally introduced in 1952 by Debreu in [21], where the term social equilibrium was coined. This paper was actually intended to be just a mathematical preparation for the famous 1954 Arrow and Debreu paper [5] about economic equilibria. In this latter paper, Arrow and Debreu termed the GNEP "an abstract economy" and explicitly note that "... In a game, the pay-off to each player depends upon the strategies chosen by all, but the domain from which strategies are to be chosen is given to each player independently of the strategies chosen by other players. An abstract economy, then, may be characterized as a generalization of a game in which the choice of an action by one agent affects both the pay-off and the domain of actions of other agents", see [5, p. 273]. It is safe to say that [5] and the subsequent book [22] provided the rigorous foundation for the contemporary development of mathematical economics.

The mathematical-economic origin of the GNEP explains why the GNEP has long been (let us say up to the beginning of the nineties) the almost
exclusive domain of economists and game-theory experts. In truth, it must also be noted that in this community some reserves have been advanced on GNEPs, on the grounds that a GNEP is not a game. For example, Ichiishi states, in his influential 1983 book [52, p. 60], "It should be emphasized, however, that an abstract economy is not a game, ... since player $j$ must know the others' strategies in order to know his own feasible strategy set ..., but the others cannot determine their feasible strategies without knowing $j$ 's strategy. Thus an abstract economy is a pseudo-game and it is useful only as a mathematical tool to establish existence theorems in various applied contexts."

The point here is that one cannot imagine a game where the players make their choices simultaneously and then, for some reason, it happens that the constraints are satisfied. But indeed, this point of view appears to be rather limited, and severely undervalues
(a) the descriptive and explanatory power of the GNEP model;
(b) its normative value, i.e., the possibility to use GNEPs to design rules and protocols, set taxes and so forth, in order to achieve certain goals, a point of view that has been central to recent applications of GNEPs outside the economic field;
(c) the fact that in any case different paradigms for games can and have been adopted, where it is possible to imagine that, although in a noncooperative setting, there are mechanisms that make the satisfactions of the constraints possible.
Following the founding paper [5], researchers dedicated most of their energies to the study of the existence of equilibria under weaker and weaker assumptions and to the analysis of some structural properties of the solutions (for example uniqueness or local uniqueness). However, and with few exceptions, it was not until the beginning of the 1990s that applications of the GNEP outside the economic field started to be considered along with algorithms for calculation of equilibria. In this respect, possibly one of the early contributions was given by Robinson in 1993 in [86, 87]. In these twin papers, Robinson considers the problem of measuring effectiveness in optimizationbased combat models, and gives several formulations that are nothing else but, in our terminlogy, GNEPs. For some of these GNEPs, Robinson provides both existence results and computational procedures.

More or less at the same time, Scotti, see [90] and references therein, introduced GNEPs in the study and solution of complex structural design problems as an evolution of the more standard use of nonlinear programming techniques promoted by Schmit in the 1960s (see [89] for a review) and motivated by some early suggestions in the previous decade, see [85, 100].

After these pioneering contributions, in the last decade the GNEP became a relatively common paradigm, used to model problems from many different fields. In fact GNEPs arise quite naturally from standard NEPs if the players share some common resource (a communication link, an electrical transmis-
sion line, a transportation link etc.) or limitations (for example a common limit on the total pollution in a certain area). More in general the ongoing process of liberalization of many markets (electricity, gas, telecommunications, transportation and others) naturally leads to GNEPs. But GNEPs have also been employed to model more technical problems that do not fit any of the categories listed above, and it just seems likely that now that the model is winning more and more popularity, many other applications will be uncovered in the near future. It is impossible to list here all relevant references for these applications; we limit ourselves to a few that, in our view, are either particularly interesting or good entry points to the literature $[2,3,4,5,6,7,12,17,26,27,41,42,46,47,49,50,51,56,59,60,61,76,80$, $81,84,94,95,96,103,107]$. In very recent years GNEPs seem to have gained an especially important place in two fields: Computer Science and Telecommunications. The book Algorithmic Game Theory [75] and the special issue of the IEEE Magazine in Signal Processing [58] constitute fundamental and rich sources for these exciting developments. We finally mention that the recent reference [34] presents a rather comprehensive and sophisticated treatment of the GNEP based on a variational inequality approach that allows to establish some interesting new results.

### 2.3 Penalty methods for GNEPs: the state of the art

One of the main difficulties of the GNEP (2.2) stems from the variability of the feasible sets $\mathcal{F}_{\nu}\left(\mathbf{x}^{-\nu}\right)$. Thus a suitable penalization of the (difficult) coupling constraints seems to be a quite intuitive and natural approach: the main idea is to reduce the original GNEP to a conceptually simpler NEP. In this way, as we will see, difficulties arising form the variability of sets $\mathcal{F}_{\nu}\left(\mathbf{x}^{-\nu}\right)$ are traded with the difficulty of a more "complex" objective function.

The aim is then to solve the GNEP by solving a NEP for which each player's minimization problem is given by

$$
\begin{gather*}
\operatorname{minimize}_{x^{\nu}} \\
\theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)+\rho_{\nu} \psi^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)  \tag{2.4}\\
x^{\nu} \in K_{\nu}
\end{gather*}
$$

where $\rho^{\nu}$ is a positive penalty parameter and $\psi^{\nu}$ is the penalty term: it must be noted that one can employ both differentiable and nondifferentiable penalty terms (see Chapter 1). Functions $\psi^{\nu}$ must be properly chosen so that no helpful property of the original objective function is lost: typical examples are $\left\|\max \left\{0, g^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\}\right\|$ and $\left\|\max \left\{0, g^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\}\right\|^{2}$ where in function $\max \left\{0, g^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\}$ the maximum is taken componentwise. To the solution of this latter problem, we can then apply methods based mainly on the optimization or variational inequalities techniques. In particular, the VI
formulation provides a strong theoretical and numerical framework for the solution of the classical NEP.

Two different approaches could be distinguished: in a complete penalization scheme no private constraints are left in (2.4) and an unconstrained NEP is obtained. In a partial penalization scheme only the difficult coupling constraints are penalized and, if this is the case, useful properties (like the compactness, for example) of the private constraints' set can be kept in (2.4).

Furthermore, in the literature, both sequential and exact penalty approaches have been considered. The biggest challenge of the penalty method is the non-trivial solution of the penalized conceptually simpler subproblems: indeed problem (2.4) could be very difficult to solve in practice (for example because of the non differentiability of the penalized objective functions) both in a sequential and in an exact approach.

It should be remarked that the generalization of the penalization technique for constrained optimization to the case of GNEPs is not straightforward. Weaker theoretical results can be expected due to the variable nature of $\mathcal{F}_{\nu}\left(\mathbf{x}^{-\nu}\right)$ : in a GNEP framework, each player deals with a different optimization problem in correspondence of each rivals' choice and some of these problems may have no solutions or even no feasible points.

In [79] Fukushima and Pang made the first attempt to solve GNEPs by means of a suitable sequential partial penalty method. Based on the equivalent reformulation of a GNEP as a QVI, they proposed a sequential augmented Lagrangian-type approach in which an infinite sequence of differentiable penalized problems is solved. A numerical experiment is presented in which the penalized subproblems are converted as linear complementarity problem and solved by the MATLAB code PATHLCP.m. However it should be noted that no conditions are established for the solvability of the Nash subproblems.

In [40] another exact penalty method is analyzed by Fukushima. In this approach, at each iteration, a nondifferentiable penalized NEP has to be solved exactly.

Under some assumptions, based on suitable Lagrange multipliers, an incremental penalty algorithm is proposed. Despite the convergence of the scheme, no theoretical results for the solution of the difficult nondifferentiable subproblems (the penalized NEP) are given. In the same work, for jointly convex GNEPs, in order to identify in the whole set of equilibria (possibly infinitely many) those which are of significance in some sense, a new solution class, called restricted equilibria, is introduced. The class of restricted equilibria contains as special case the variational equilibrium and can be characterized by certain properties enjoyed by the palyers' Lagrange multipliers associated with the shared constraints. Furthermore, an heuristic method able to find such equilibria is designed and numerically tested.

Finally, in the same line of research of [79], inspired by [33], an exact penalty method has been put forward recently by Facchinei and Kanzow [30]. Although the method proposed by Facchinei and Kanzow showed a good
numerical behavior for a host of GNEPs of different nature, [30] left open the question of determining classes of problems that can be provably solved by the penalization approach: in particular the critical question is whether we are able to solve the penalized unconstrained NEP. In order to fill this gap and overcome some theoretical difficulties of that method (the existence of a solution of the penalized NEP subproblems and the boundedness of the sequence generated by the algorithm could not be guaranteed), in the following sections we propose a partial exact penalization scheme. Hence, the main contribution of this work is that, for the first time, classes of GNEPs are exhibited for which the penalty approach is guaranteed to converge to a solution: in addition to classical cases (NEP), the algorithm, in particular, is proved to be able to find any solution, also the non-variational ones, of a jointly convex game. Furthermore, under some other technical conditions this method can provably find solutions of general GNEPs in presence of "separable" constraints.

### 2.4 The partial penalization approach

Penalty approaches to the solution of GNEPs are based on the usual penalization idea: recover a solution of the original game by finding a solution of a "simpler" penalized problem. We then propose to penalize only the difficult coupling constraints by keeping the private ones (with their possibly useful properties) and thus reducing the original problem (2.1) to the following pure NEP:

$$
\begin{gather*}
\operatorname{minimize}_{x^{\nu}} \theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)+\rho_{\nu}\left\|g_{+}^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\|_{\gamma} \\
x^{\nu} \in K_{\nu} \tag{2.5}
\end{gather*}
$$

where for a vector $\mathbf{x}$ we denote by $\|\mathbf{x}\|_{\gamma}$ the $\gamma$-norm for some fixed $\gamma \in(1, \infty)$, the $\rho_{\nu}$ are positive penalty parameters and $g_{+}(\mathbf{x}):=\max \{0, g(\mathbf{x})\}$ with the maximum taken componentwise.
By setting

$$
P_{\nu}\left(\mathbf{x}, \rho_{\nu}\right):=\theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)+\rho_{\nu}\left\|g_{+}^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\|_{\gamma}
$$

problem (2.5) can be rewritten as a (standard) Nash equilibrium problem, where each player's problem is given by

$$
\begin{align*}
\min _{x^{\nu}} & P_{\nu}\left(\mathbf{x}, \rho_{\nu}\right)  \tag{2.6}\\
& x^{\nu} \in K_{\nu} .
\end{align*}
$$

We refer to problem (2.5) or, equivalently, (2.6) as the Penalized Nash Equilibrium Problem with parameters $\rho_{\nu}, \mathrm{PNEP}_{\boldsymbol{\rho}}$, where with $\rho$ we denote the vector of parameters $\boldsymbol{\rho}:=\left(\rho_{1}, \cdots, \rho_{N}\right)^{T}$. Our basic idea, that will be substantiated in the next sections, is that we can solve the original GNEP by finding a solution of the $\operatorname{PNEP}_{\rho}$ for suitable values of the penalty parameters
$\rho_{\nu}$. Since the $\operatorname{PNEP}_{\rho}$ is a NEP, we also hope that we are actually able to solve the $\mathrm{PNEP}_{\boldsymbol{\rho}}$. However we point out from the outset that the $\mathrm{PNEP}_{\boldsymbol{\rho}}$ has nondifferentiable objective functions, a fact that, we will see, causes some troubles. In a sense we traded the difficulty of the variable coupling sets with the difficulty given by a nondifferentiable objective function.

Lastly we introduce the notation used throughout the following sections. $J g(\mathbf{x})$ is the Jacobian of $g$ at $\mathbf{x}$, while $\nabla g(\mathbf{x})$ denotes the transposed Jacobian of that mapping. $J_{x^{\nu}} g(\mathbf{x})\left(\nabla_{x^{\nu}} g(\mathbf{x})\right)$ is the partial (transposed) Jacobian with the partial derivatives taken only with respect to the components belonging to the subvector $x^{\nu}$ of $\mathbf{x} . \mathbf{K}:=\prod_{\nu} K_{\nu}$ is the Cartesian product of the $K_{\nu}$ sets. With $T_{K_{\nu}}\left(\bar{x}^{\nu}\right)$ and $N_{K_{\nu}}\left(\bar{x}^{\nu}\right)$, we denote respectively the tangent and the normal cone to $K_{\nu}$ in $\bar{x}^{\nu}$. Finally we denote by $M_{m, n}\left(M_{n}\right)$ the space of $m \times n(n \times n)$ real matrices.

### 2.5 The original problem and the penalized problem

As we have seen in the previous section, we aim at solving the GNEP by finding a solution of the $\mathrm{PNEP}_{\boldsymbol{\rho}}$. Therefore, our first task is to investigate the relations between the solutions of problem (2.2) and those of (2.5). To this end, we introduce a (non standard) constraint qualification. Let us denote by $\partial_{x^{\nu}}^{>}\left\|g_{+}^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\|_{\gamma}$ the set

$$
\begin{aligned}
\partial_{x^{\nu}}^{>}\left\|g_{+}^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\|_{\gamma}:= & \left\{\xi^{\nu} \in \mathbb{R}^{n_{\nu}} \mid \exists\left\{\mathbf{y}^{k}\right\} \text { with } y^{k} \in \mathbf{K}\right. \\
& \text { and }\left\|g_{+}^{\nu}\left(\left(y^{k}\right)^{\nu},\left(\mathbf{y}^{k}\right)^{-\nu}\right)\right\|_{\gamma}>0: \\
& \left.\left\{\mathbf{y}^{k}\right\} \rightarrow \mathbf{x} \text { and } \nabla_{x^{\nu}}\left\|g_{+}^{\nu}\left(\left(y^{k}\right)^{\nu},\left(\mathbf{y}^{k}\right)^{-\nu}\right)\right\|_{\gamma} \rightarrow \xi^{\nu}\right\} .
\end{aligned}
$$

Roughly speaking, for every player $\nu, \partial_{x^{\nu}}^{>}\left\|g_{+}^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\|_{\gamma}$ is the set of subgradients of $\left\|g_{+}^{\nu}(\mathbf{x})\right\|_{\gamma}$ arising from unfeasible points for that player. This set should capture the geometry of $\left\|g_{+}^{\nu}(\mathbf{x})\right\|_{\gamma}$ at unfeasible points (see [16]) and can also be seen as a generalization of a similar definition used in [54] to study metric regularity. Note that, since in the definition we assume that $\left\|g_{+}^{\nu}\left(\mathbf{y}^{k}\right)\right\|_{\gamma}>0,\left\|g_{+}^{\nu}\left(\mathbf{y}^{k}\right)\right\|_{\gamma}$ is continuously differentiable at $\mathbf{y}^{k}$, so that the definition above is well posed.

Definition 5. We say that the GNEP (2.2) satisfies the constraint qualification $\mathrm{CQ}_{\gamma}$ at a point $\overline{\mathbf{x}} \in \mathbf{K}$ if, for every player $\nu=1, \ldots, N$,

$$
\begin{equation*}
\xi^{\nu} \in \partial_{x^{\nu}}^{>}\left\|g_{+}^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)\right\|_{\gamma} \Rightarrow\left(-\xi^{\nu}\right) \notin N_{K_{\nu}}\left(\bar{x}^{\nu}\right) . \tag{2.7}
\end{equation*}
$$

If $\overline{\mathbf{x}}$ is such that $g^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)<0$ for every player, then $\mathrm{CQ}_{\gamma}$ holds because $\partial_{x^{\nu}}^{>}\left\|g_{+}^{\nu}(\overline{\mathbf{x}})\right\|_{\gamma}=\emptyset$ for all $\nu$. We illustrate the definition with a simple example.

Example 4. We consider a game with two players, where the first player controls the variable $x^{1} \in \mathbb{R}$ and the second $x^{2} \in \mathbb{R}$ :

$$
\begin{array}{cl}
\min _{x^{1}}\left(x^{1}\right)^{2}-2 x^{1} x^{2} & \min _{x^{2}} \\
\frac{1}{2}\left(x^{2}\right)^{2}+\left(x^{1}-1\right) x^{2} \\
0 \leq x^{1} \leq 1 & 0 \leq x^{2} \leq 1
\end{array}
$$

In this case $K_{1}=[0,1]$ and $K_{2}=[0,1]$. For the second player, condition (2.7) holds trivially. For the first player, in view of the observation right after Definition 5 , it suffices to show that (2.7) holds for each point $\left(\tilde{x}^{1}, \tilde{x}^{2}\right)$ such that $\tilde{x}^{1} \in K_{1}, \tilde{x}^{2} \in K_{2}$ and $\tilde{x}^{1}+\tilde{x}^{2} \geq \frac{3}{2}$. Note that these conditions imply that $\tilde{x}^{1} \in\left[\frac{1}{2}, 1\right]$. For such points we have $\partial^{>}{ }_{x^{1}}\left\|g_{+}\left(\tilde{x}^{1}, \tilde{x}^{2}\right)\right\|_{\gamma}=\{1\}$. Furthermore, we also have $N_{K_{1}}\left(\tilde{x}^{1}\right)=0$ for every $\tilde{x}^{1} \in\left[\frac{1}{2}, 1\right)$ while $N_{K_{1}}(1)=\mathbb{R}_{+}$. Therefore, it is immediate to see that $-1 \notin N_{K_{1}}\left(\tilde{x}^{1}\right)$ so that (2.7) holds at every point of $\mathbf{K}$.

Furthermore, if $\mathrm{CQ}_{\gamma}$ holds we can prove the following theorem.
Theorem 8. Suppose $C Q_{\gamma}$ holds at every point of $\mathbf{K}$, then there exists a $\bar{\rho}$ such that, for every $\boldsymbol{\rho}$ with $\rho_{\nu} \geq \bar{\rho}$ for all $\nu$, every solution of the $P N E P_{\boldsymbol{\rho}}$ (2.5) is a solution of the GNEP (2.2).

Proof. Note first that it is enough to show that for all penalty parameters sufficiently large, every solution of the $\mathrm{PNEP}_{\rho}$ is feasible for the original GNEP. Assume then by contradiction that $\rho_{\nu}^{k} \rightarrow \infty$ for every $\nu$ and there exists a sequence $\left\{\mathbf{x}^{k}\right\}$ of solutions of the $\mathrm{PNEP}_{\boldsymbol{\rho}^{k}}$ that are unfeasible for the GNEP (note that, since the sets $K_{\nu}$ are convex and compact and the objective functions $P_{\nu}\left(\mathbf{x}, \rho_{\nu}\right)$ are continuous and convex with respect to the players' variables, the $\mathrm{PNEP}_{\boldsymbol{\rho}^{k}}$ always has a solution). There exists a $\nu$ for which we can assume, without loss of generality, that $\left\|g_{+}^{\nu}\left(\left(x^{k}\right)^{\nu},\left(x^{k}\right)^{-\nu}\right)\right\|_{\gamma}>0$ for all $k$. Furthermore, subsequencing if necessary, by the compactness of $\mathbf{K}$, we may assume that the entire sequence $\mathbf{x}^{k}$ converges to $\overline{\mathbf{x}} \in \mathbf{K}$. The function $\left\|g_{+}^{\nu}\left(x^{\nu}, \mathbf{x}^{\nu}\right)\right\|_{\gamma}$ is continuously differentiable at $\mathbf{x}^{k}$, so that (see Chapter 1 )

$$
\begin{equation*}
-\left[\frac{\nabla_{x^{\nu}} \theta_{\nu}\left(\left(x^{k}\right)^{\nu},\left(x^{k}\right)^{-\nu}\right)}{\rho_{\nu}^{k}}+\nabla_{x^{\nu}}\left\|g_{+}^{\nu}\left(\left(x^{k}\right)^{\nu},\left(x^{k}\right)^{-\nu}\right)\right\|_{\gamma}\right] \in N_{K_{\nu}}\left(\left(x^{k}\right)^{\nu}\right) . \tag{2.9}
\end{equation*}
$$

Taking the limit of (2.9) and taking into account the outer semicontinuity of the mapping $N_{K_{\nu}}\left(x^{\nu}\right)$ (see [88], Proposition 6.6), the boundedness of the (partial sub)gradient of $\left\|g_{+}^{\nu}\left(x^{\nu}, \mathbf{x}^{\nu}\right)\right\|_{\gamma}$ (see Chapter 1) on bounded sets and its closedness, we then get

$$
-\bar{\xi}^{\nu} \in N_{K_{\nu}}\left(\bar{x}^{\nu}\right)
$$

for some $\bar{\xi}^{\nu} \in \partial_{x^{\nu}}^{>}\left\|g_{+}^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)\right\|_{\gamma}$. This obviously contradicts the definition of $\mathrm{CQ}_{\gamma}$ and concludes the proof.

The following example clarifies Theorem 8.
Example 4 (continued) Consider again Example 4. One can see that the solution set $S_{1}\left(x^{2}\right)$ of the first player's problem (for a given $x^{2}$ ) is

$$
S_{1}\left(x^{2}\right)=\left\{\begin{array}{lll}
x^{2} & \text { if } & x^{2} \in\left[0, \frac{3}{4}\right]  \tag{2.10}\\
\frac{3}{2}-x^{2} & \text { if } & x^{2} \in\left[\frac{3}{4}, 1\right]
\end{array}\right.
$$

while $S_{2}\left(x^{1}\right)=1-x^{1}$ is the solution set of the second player's problem (for a given $x^{1}$ ). Since a solution is by definition a point $\left(x^{1}, x^{2}\right)$ such that $x^{1} \in S_{1}\left(x^{2}\right)$ and $x^{2} \in S_{2}\left(x^{1}\right)$, it is easy to see that the game has a unique solution $\left(\bar{x}^{1}, \bar{x}^{2}\right)=\left(\frac{1}{2}, \frac{1}{2}\right)$. Consider now the penalized version of the game (see (2.5)):

$$
\begin{array}{cc}
\min _{x}^{1}\left(x^{1}\right)^{2}-2 x^{1} x^{2}+3\left(x^{1}+x^{2}-\frac{3}{2}\right)_{+} & \min _{x}^{2} \\
& \frac{1}{2}\left(x^{2}\right)^{2}+\left(x^{1}-1\right) x^{2}  \tag{2.11}\\
0 \leq x^{1} \leq 1 & 0 \leq x^{2} \leq 1
\end{array}
$$

where we took $\rho=3$ (note that since there is only one constraint being penalized, the value of $\gamma$ is immaterial). We now show by contradiction that all the solutions of Nash game (2.11) are feasible for the original problem. Note that the Nash game (2.11) has a solution since its feasible set is compact. Assume then that we have a solution $\left(\tilde{x}^{1}, \tilde{x}^{2}\right)$ of $(2.11)$ for which $\tilde{x}^{1}+\tilde{x}^{2}>\frac{3}{2}$. Since $\tilde{x}^{2} \in[0,1]$, this implies that $\tilde{x}^{1} \in\left(\frac{1}{2}, 1\right]$. Since $\tilde{x}^{1}+\tilde{x}^{2}-\frac{3}{2}>0$, the objective function of the first player in $\left(\tilde{x}^{2}, \tilde{x}^{2}\right)$ is given by $\left(\tilde{x}^{1}\right)^{2}-2 \tilde{x}^{1} \tilde{x}^{2}+$ $3\left(\tilde{x}^{1}+\tilde{x}^{2}-\frac{3}{2}\right)$ so that the gradient of the first player's objective function is $2 \tilde{x}^{1}-2 \tilde{x}^{2}+3$ and, in view of $\tilde{x}^{1} \in\left(\frac{1}{2}, 1\right]$ and $\tilde{x}^{2} \in[0,1]$, we have $2 \tilde{x}^{1}-2 \tilde{x}^{2}+3>$ 0 . By the minimum principle this implies that $\tilde{x}^{1}=0$ which is a contradiction to $\tilde{x}^{1} \in\left(\frac{1}{2}, 1\right]$ so that all the solutions of (2.11) are also solutions of (2.8). Observe also that it is easy to see that the same result holds for any $\rho>3$.

Theorem 8 obviously shows that indeed we can hope to obtain a solution of the original GNEP by solving the $\mathrm{PNEP}_{\rho}$. However, before analyzing this issue more in detail, we deepen our analysis of the relations between the GNEP (2.2) and the $\operatorname{PNEP}_{\boldsymbol{\rho}}(2.5)$ and we study some consequences of the $\mathrm{CQ}_{\gamma}$.
Theorem 9. Suppose the $C Q_{\gamma}$ holds at every point of $\mathbf{K}$. Then:
(a) The feasible regions of all players are non empty for every $\mathbf{x} \in \mathbf{K}$, i.e. for each $\mathbf{x} \in \mathbf{K}, \mathcal{F}_{\nu}\left(\mathbf{x}^{-\nu}\right)=K_{\nu} \cap X_{\nu}\left(\mathbf{x}^{-\nu}\right) \neq \emptyset$ for every player $\nu=1, \ldots, N ;$
(b) The GNEP (2.2) has a solution.

Proof. Suppose, by contradiction, that $\overline{\mathbf{x}} \in \mathbf{K}$ is such that there exists a player $\nu \in\{1, \ldots, N\}$ with $K_{\nu} \cap X_{\nu}\left(\overline{\mathbf{x}}^{-\nu}\right)=\emptyset$. Therefore, $\left\|g_{+}^{\nu}\left(x^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)\right\|_{\gamma}>$
$0, \forall x^{\nu} \in K_{\nu}$ and function $\left\|g_{+}^{\nu}\left(x^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)\right\|_{\gamma}$ is continuously differentiable with respect to $x^{\nu}$. We consider the feasibility problem

$$
\begin{gather*}
\min _{x^{\nu}}\left\|g_{+}^{\nu}\left(x^{\nu}, \bar{x}^{-\nu}\right)\right\|_{\gamma}  \tag{2.12}\\
x^{\nu} \in K_{\nu} .
\end{gather*}
$$

This is a convex optimization problem with a continuously differentiable objective function and a compact feasible set. Therefore, it admits a global solution $\tilde{x}^{\nu} \in K_{\nu}$ and

$$
-\nabla_{x^{\nu}}\left\|g_{+}^{\nu}\left(\tilde{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)\right\|_{\gamma} \in N_{K_{\nu}}\left(\tilde{x}^{\nu}\right)
$$

Since in this case we have $\partial_{x^{\nu}}^{>}\left\|g_{+}^{\nu}\left(\tilde{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)\right\|_{\gamma}=\left\{\nabla_{x^{\nu}}\left\|g_{+}^{\nu}\left(\tilde{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)\right\|_{\gamma}\right\}$, we get a contradiction to the $\mathrm{CQ}_{\gamma}$ and therefore, $\forall \mathbf{x} \in \mathbf{K}, K_{\nu} \cap X_{\nu}\left(\mathbf{x}^{-\nu}\right) \neq \emptyset$ for every player.

To prove part (b) it is enough to observe that for any $\boldsymbol{\rho}>0$ the $\operatorname{PNEP}_{\boldsymbol{\rho}}$ has a solution in $\mathbf{K}$. By Theorem 8, if all the penalty parameters are large enough, this solution will also be a solution of the GNEP (2.2).

Remark 2. We recall that Therem 6 rests on some non emptiness and continuity assumptions of the point-to-set mappings $\mathcal{F}_{\nu}(\bullet)$. These assumptions are, however, usually difficult to check in practice. Therefore, Theorem 9 might be useful in applications and can be seen as a practical tool for establishing the existence of a solution of GNEP (2.2).

We end this section with some considerations on constraint qualifications for games. Firstly we relate the $\mathrm{CQ}_{\gamma}$ to a more classical constraint qualification: the Extended Mangasarian-Fromovitz Constraint Qualification (EMFCQ).

Definition 6. We say that the GNEP (2.2) satisfies the EMFCQ at a point $\overline{\mathbf{x}} \in \mathbf{K}$ if, for every player $\nu=1, \ldots, N$, there exists a vector $d^{\nu} \in T_{K_{\nu}}\left(\overline{\mathbf{x}}^{\nu}\right)$ such that

$$
\begin{equation*}
\nabla_{x^{\nu}} g_{i}^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)^{T} d^{\nu}<0 \quad \forall i \in I_{+}^{\nu}(\overline{\mathbf{x}}) \tag{2.13}
\end{equation*}
$$

where $I_{+}^{\nu}(\overline{\mathbf{x}}):=\left\{i \in\left\{1, \ldots, m_{\nu}\right\} \mid g_{i}^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right) \geq 0\right\}$ is the index set of all active and violated constraints at $\overline{\mathbf{x}}$.

By standard reasonings, it is easily seen that Condition (2.13) is equivalently satisfied if, for every player $\nu$, there is no nonzero vector $v^{\nu} \in \mathbb{R}_{+}^{m_{\nu}}$, with $v_{i}^{\nu}=0$ if $g_{i}^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)<0$, such that:

$$
\begin{equation*}
-\left[v_{1}^{\nu} \nabla_{x^{\nu}} g_{1}^{\nu}(\overline{\mathbf{x}})+\ldots+v_{m_{\nu}}^{\nu}(\overline{\mathbf{x}}) \nabla_{x^{\nu}} g_{m_{\nu}}^{\nu}(\overline{\mathbf{x}})\right] \in N_{K_{\nu}}\left(\bar{x}^{\nu}\right) \tag{2.14}
\end{equation*}
$$

Remark 3. We show for completeness the equivalence of (2.13) and (2.14). Assume that (2.13) holds and suppose, by contradiction, that there is $\nu \in$ $\{1, \ldots, N\}$ and a nonzero vector $v^{\nu} \in \mathbb{R}_{+}^{m_{\nu}}$, with $v_{i}^{\nu}=0$ if $g_{i}^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)<0$, such that

$$
\begin{align*}
0 & \leq\left(d^{\nu}\right)^{T} \sum_{i=1}^{m_{\nu}} v_{i}^{\nu} \nabla_{x^{\nu}} g_{i}^{\nu}(\overline{\mathbf{x}})  \tag{2.15}\\
& =\left(d^{\nu}\right)^{T} \sum_{i \in I_{+}^{\nu}(\overline{\mathbf{x}})} v_{i}^{\nu} \nabla_{x^{\nu}} g_{i}^{\nu}(\overline{\mathbf{x}})<0
\end{align*}
$$

which is a contradiction.
For the converse, by (2.14), for every player and for every nonzero vector $v^{\nu} \in \mathbb{R}_{+}^{m_{\nu}}$, such that $v_{i}^{\nu}=0$ if $g_{i}^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)<0$, there exists $\bar{d}^{\nu} \in T_{K_{\nu}}\left(\bar{x}^{\nu}\right)$ such that

$$
\begin{equation*}
\left[\sum_{i \in I_{+}^{\nu}(\overline{\mathbf{x}})} v_{i}^{\nu} \nabla_{x^{\nu}} g_{i}^{\nu}(\overline{\mathbf{x}})\right]^{T} \bar{d}^{\nu}<0 \tag{2.16}
\end{equation*}
$$

Relation (2.16) must hold for every such vector $v^{\nu}$ and this fact in turn implies the EMFCQ.

We observe that it is classical to show that if the EMFCQ holds at a solution of the game, then the KKT conditions are satisfied for each player at this point.

The EMFCQ for optimization problems has been used often in analyzing solution algorithms. Our definition of the EMFCQ for games is the natural extension of the constraint qualification from optimization problems to games.
We note that, in Example 4, the EMFCQ holds at every point of $K$. This fact does not occur by chance: the following theorem proves that there is a relationship between the EMFCQ and $\mathrm{CQ}_{\gamma}$.

Proposition 8. If the $E M F C Q$ holds at $\overline{\mathbf{x}} \in \mathbf{K}$, then the $C Q_{\gamma}$ holds at $\overline{\mathbf{x}}$. The vice versa does not hold in general.

Proof. Suppose that EMFCQ holds at $\overline{\mathbf{x}}$, but, for some $\nu$, there is a $\xi^{\nu} \in$ $\partial_{x^{\nu}}^{>}\left\|g_{+}^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)\right\|_{\gamma}$ such that $-\xi^{\nu} \in N_{K_{\nu}}\left(\bar{x}^{\nu}\right)$. By definition there is a sequence $\left\{\mathbf{y}^{k}\right\} \rightarrow \overline{\mathbf{x}}$ with $\mathbf{y}^{k} \in \mathbf{K}$ and $\left\|g_{+}^{\nu}\left(\left(y^{k}\right)^{\nu},\left(\mathbf{y}^{k}\right)^{-\nu}\right)\right\|_{\gamma}>0$ such that

$$
\left\{\nabla_{x^{\nu}}\left\|g_{+}^{\nu}\left(\left(y^{k}\right)^{\nu},\left(\mathbf{y}^{k}\right)^{-\nu}\right)\right\|_{\gamma}\right\} \rightarrow \xi^{\nu}
$$

We have

$$
\nabla_{x^{\nu}}\left\|g_{+}^{\nu}\left(\left(y^{k}\right)^{\nu},\left(\mathbf{y}^{k}\right)^{-\nu}\right)\right\|_{\gamma}=\nabla_{x^{\nu}} g^{\nu}\left(\left(y^{k}\right)^{\nu},\left(\mathbf{y}^{k}\right)^{-\nu}\right)\left(\lambda^{k}\right)^{\nu}
$$

where $\left(\lambda^{k}\right)^{\nu}:=\frac{g_{+}^{\nu}\left(\left(y^{k}\right)^{\nu},\left(\mathbf{y}^{k}\right)^{-\nu}\right)^{\gamma-1}}{\left\|g_{+}^{\nu}\left(\left(y^{k}\right)^{\nu},\left(\mathbf{y}^{k}\right)^{-\nu}\right)\right\|_{\gamma}^{\gamma-1}}$.
It is not difficult to see that there are constants $0<\alpha_{1} \leq \alpha_{2}$ such that $\alpha_{1} \leq\left\|\left(\lambda^{k}\right)^{\nu}\right\|_{\gamma} \leq \alpha_{2}$ for all $k$. Hence, we may assume that there is a limiting
nonzero vector $\bar{\lambda}^{\nu} \in \mathbb{R}_{+}^{m_{\nu}}$ and, by definition of $\left(\lambda^{k}\right)^{\nu}, \bar{\lambda}_{i}^{\nu}=0$ for $g_{i}^{\nu}\left(\overline{\mathbf{x}}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)<$ 0 . Therefore,

$$
\begin{equation*}
\nabla_{x^{\nu}}\left\|g_{+}^{\nu}\left(\left(y^{k}\right)^{\nu},\left(\mathbf{y}^{k}\right)^{-\nu}\right)\right\|_{\gamma} \rightarrow \nabla_{x^{\nu}} g^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right) \bar{\lambda}^{\nu}=\xi^{\nu} \tag{2.17}
\end{equation*}
$$

with $\bar{\lambda}^{\nu} \in \mathbb{R}_{+}^{m_{\nu}}, \bar{\lambda}^{\nu} \neq 0$ and $\bar{\lambda}_{j}^{\nu} g_{j}^{\nu}(\overline{\mathbf{x}})=0$. Recalling that $-\xi^{\nu} \in N_{K_{\nu}}\left(\bar{x}^{\nu}\right)$, we see that (2.17) contradicts (2.14).

To show the second assertion, it is enough to give an example. Indeed, consider a game with one player whose feasible set in $\mathbb{R}$ is defined by $x^{3} \leq 0$, $x \leq 0$ and $K=[-1,1]$. EMFCQ does not hold in $\bar{x}=0$ since the gradient of $x^{3}$ at the origin is zero. On the other hand, straightforward calculations show that $\partial^{>}\left\|g_{+}(0)\right\|_{2}=\{1\}$ so that the $\mathrm{CQ}_{\gamma}$ holds at the origin.

If we assume that EMFCQ holds instead of the $\mathrm{CQ}_{\gamma}$, we can improve on Theorem 8 and get a complete equivalence between the GNEP (2.2) and the $\mathrm{PNEP}_{\boldsymbol{\rho}}$ (2.5).

Theorem 10. Suppose that EMFCQ holds at every point of $\mathbf{K}$; then there exists a $\bar{\rho}$ such that, for every $\rho$ with $\rho_{\nu} \geq \bar{\rho}$ for all $\nu$, the solutions of the $P_{N E P}^{\boldsymbol{\rho}}$ (2.5) coincide with the solution of the GNEP (2.2).

Proof. Because of Theorem 8 and Proposition 8 it is enough to show that every solution of the GNEP (2.2) is a solution of the $\mathrm{PNEP}_{\boldsymbol{\rho}}$ (2.5), if the penalty parameters are large enough. To this end we first prove a simple Lemma.

Lemma 1. Consider the minimization problem

$$
\begin{gather*}
\min f(z) \\
v(z) \leq 0  \tag{2.18}\\
z \in Z
\end{gather*}
$$

where $f: \mathbb{R}^{n} \rightarrow \mathbb{R}$ and $v: \mathbb{R}^{n} \rightarrow \mathbb{R}^{m}$ are continuously differentiable and convex and $Z \subseteq \mathbb{R}^{n}$ is a closed, convex set. Suppose that the feasible set of (2.18) is nonempty. Let $\bar{z}$ be a solution of (2.18) and assume that the set $M$ of multipliers is non empty. Let $\lambda$ be any multiplier in $M$ and denote by $\|\bullet\|_{\gamma}^{*}$ the dual $\gamma$-norm. Then, for every $\sigma>\|\lambda\|_{\gamma}^{*}$, the solutions sets of (2.18) and

$$
\begin{align*}
& \min f(z)+\sigma\left\|v_{+}(z)\right\|_{\gamma} \\
& z \in Z . \tag{2.19}
\end{align*}
$$

coincide.
Proof. First of all, recall that the set of multipliers of the convex problem (2.18) does not depend on the optimal solution we are considering (see e.g. the remark on page 354 in [10]). Therefore, if $\tilde{z}$ is a solution of (2.18) we have

$$
(z-\tilde{z})^{T}\left[\nabla f(\tilde{z})+\lambda_{1} \nabla v_{1}(\tilde{z})+\ldots+\lambda_{m} \nabla v_{m}(\tilde{z})\right] \geq 0, \quad \forall z \in Z
$$

Hence, $\tilde{z}$ also minimizes the Lagrangian $L(z, \lambda):=f(z)+\sum_{j=1}^{m} \lambda_{j} g_{j}(z)$ over the set $Z$ so that

$$
\begin{equation*}
f(z)+\lambda^{T} v(z) \geq f(\tilde{z})+\lambda^{T} v(\tilde{z}), \quad \forall z \in Z \tag{2.20}
\end{equation*}
$$

Recalling that $\sigma>\|\lambda\|_{\gamma}^{*}$

$$
\begin{aligned}
f(z)+\sigma\left\|v_{+}(z)\right\|_{\gamma} & \geq f(z)+\|\lambda\|_{\gamma}^{*}\left\|v_{+}(z)\right\|_{\gamma} \\
& \geq f(z)+\lambda^{T} v_{+}(z) \\
& \geq f(z)+\lambda^{T} v(z) \\
& \geq f(\tilde{z})+\lambda^{T} v(\tilde{z}) \\
& =f(\tilde{z})+\sigma\left\|v_{+}(\tilde{z})\right\|_{\gamma}
\end{aligned}
$$

where we have used, in sequence, $\sigma>\|\lambda\|_{\gamma}^{*}$, the Cauchy-Schwarz inequality, $\lambda \geq 0,(2.20), \lambda^{T} v(\tilde{z})=0$ and $v_{+}(\tilde{z})=0$. Therefore, $\tilde{z}$ is a minimum point of the penalty function on $Z$.
For the converse, suppose that $\tilde{z} \in Z$ is a minimum point of the penalty function $Q(z, \sigma)=f(z)+\sigma\left\|v_{+}(z)\right\|_{\gamma}$ : it is sufficient to verify that $\tilde{z}$ is feasible to (2.18). If this were not the case,

$$
\begin{aligned}
f(\tilde{z})+\sigma\left\|v_{+}(\tilde{z})\right\|_{\gamma} & >f(\tilde{z})+\|\lambda\|_{\gamma}^{*}\left\|v_{+}(\tilde{z})\right\|_{\gamma} \\
& \geq f(\tilde{z})+\lambda^{T} v_{+}(\tilde{z}) \\
& \geq f(\bar{z})+\lambda^{T} v(\bar{z}) \\
& =f(\bar{z}) \\
& =f(\bar{z})+\sigma\left\|v_{+}(\bar{z})\right\|_{\gamma}
\end{aligned}
$$

a contradiction to the fact that $\tilde{z}$ is a global minimum of the penalty function; hence $\tilde{z}$ must be feasible.

Back to the proof of Theorem 10. Taking into account that, by the EMFCQ, multipliers exist at any solution of the GNEP, this lemma easily implies that, if the set of all multipliers at any solution of the GNEP is bounded above by a common constant $r$, then, for $\rho_{\nu}>r$ for every $\nu$, every solution of the GNEP is a solution of $\operatorname{PNEP}_{\rho}$ (note that the reverse does not derive from the lemma because in certain points the feasible sets of some players could be empty). Assume then by contradiction that we can find a sequence $\left\{\mathbf{x}^{k}\right\}$ of solutions of the GNEP and a sequence of corresponding multipliers $\left\{\left(\mu^{k}\right)^{\nu}\right\}$ such that at least for one $\nu$ we have $\left\|\left(\mu^{k}\right)^{\nu}\right\| \rightarrow \infty$. We also assume, without loss of generality, that $\mathbf{x}^{k} \rightarrow \overline{\mathbf{x}} \in \mathbf{K}$. Then we can write, by the EMFCQ, the KKT conditions (see [88]) for player $\nu$ :

$$
\begin{gather*}
-\left[\nabla_{x^{\nu}} \theta_{\nu}\left(\left(x^{k}\right)^{\nu},\left(\mathbf{x}^{k}\right)^{-\nu}\right)+\sum_{j=1}^{m_{\nu}}\left(\mu_{j}^{k}\right)^{\nu} \nabla_{x^{\nu}} g_{j}^{\nu}\left(\left(x^{k}\right)^{\nu},\left(\mathbf{x}^{k}\right)^{-\nu}\right)\right] \in N_{K_{\nu}}\left(\left(x^{k}\right)^{\nu}\right) \\
0 \leq\left(\mu^{k}\right)^{\nu} \perp g^{\nu}\left(\left(x^{k}\right)^{\nu},\left(\mathbf{x}^{k}\right)^{-\nu}\right) \leq 0 . \tag{2.21}
\end{gather*}
$$

Dividing the inclusion on the first line in (2.21) and the first inequality on the second line by $\left\|\left(\mu^{k}\right)^{\nu}\right\|$ and taking the limit, thanks to the outer semicontinuity of the mapping $N_{K_{\nu}}\left(x^{\nu}\right)$ (see [88]), we have

$$
\begin{gather*}
-\nabla_{x^{\nu}} g^{\nu}(\overline{\mathbf{x}}) \bar{\mu}^{\nu} \in N_{K_{\nu}}\left(\bar{x}^{\nu}\right) \\
0 \leq \bar{\mu}^{\nu} \perp g^{\nu}(\overline{\mathbf{x}}) \leq 0, \quad \bar{\mu}^{\nu} \neq 0 . \tag{2.22}
\end{gather*}
$$

Since $\bar{\mu}^{\nu} \neq 0$ we have a contradiction to the EMFCQ (see (2.14)).

### 2.6 How to update the penalty parameters

Based on Theorem 8, we can state that, if we knew the value of the parameter $\bar{\rho}$ in Theorem 8, our only task would be that of studying an algorithm for the solution of the $\operatorname{PNEP}_{\boldsymbol{\rho}}$ (2.5), which has the peculiarity of having non differentiable objective functions. Unfortunately, in general, the value of $\bar{\rho}$ is not known in advance and so we see that our task is twofold: on the one hand we have to (iteratively) find the "right" values for the penalty parameters and, on the other hand, we must also be able to solve the non differentiable penalized Nash game. We will discuss these two main issues separately: in this section we present a way to algorithmically update the penalty parameters to an appropriate value, while the next section is devoted to the study of an algorithm for the solution of the $\mathrm{PNEP}_{\boldsymbol{\rho}}$. For the time being, in this section we assume that we have an algorithm for the solution of the penalized Nash $\operatorname{PNEP}_{\boldsymbol{\rho}}$ (2.5) for given fixed values of the penalty parameters $\rho_{\nu}$. More precisely, we suppose an iterative algorithm $\mathcal{A}$ is available that, given a point $\mathbf{x}^{k}$, generates a new point $\mathbf{x}^{k+1}:=\mathcal{A}\left[\mathbf{x}^{k}\right]$ such that $\mathbf{x}^{k} \in \mathbf{K}$ for all $k$. We suppose that $\mathcal{A}$ enjoys the following property.

Property 1. For every $\mathbf{x}^{0} \in \mathbf{K}$, the sequence $\left\{\mathbf{x}^{k}\right\}$ obtained by setting $\mathbf{x}^{k+1}=$ $\mathcal{A}\left[\mathbf{x}^{k}\right]$ is such that $x^{k} \in \mathbf{K}$ for all $k$ and every limit point is a solution of (2.5).

The meaning of the property is plain: Algorithm $\mathcal{A}$ solves the $\mathrm{PNEP}_{\boldsymbol{\rho}}$ when the $\rho_{\nu}$ are fixed. As we already said, a suitable algorithm $\mathcal{A}$ will be described in the next section. Having this algorithm $\mathcal{A}$ at hand, the scheme we propose for the updating of the penalty parameters is described in Fig. 2.7.


Fig. 2.7: The algorithmic scheme

At each iteration two steps are performed: first we test whether we should update the penalty parameters and then we perform a single step of algorithm $\mathcal{A}$. The hope is that eventually the outcome of the test on the updating of the penalty parameters is always not to update, so that the scheme reduces to the application of algorithm $\mathcal{A}$ to the $\operatorname{PNEP}_{\rho}$ with fixed values of penalty parameters. What we must then be sure of, is that the value of the penalty parameters is the "right" one. In order to achieve this and also to understand the rationale behind the updating test, we note that if, for any value of the penalty parameters, we find a solution of $\mathrm{PNEP}_{\boldsymbol{\rho}}(2.5)$ that is feasible for the original problem (2.2), then this solution solves problem (2.2) itself. Indeed, if $\overline{\mathbf{x}}$ is a Nash equilibrium of $\operatorname{PNEP}_{\boldsymbol{\rho}}(2.5)$, this means that $\bar{x}^{\nu} \in K_{\nu}$ is the optimal solution of (2.5). But on the set $X_{\nu}\left(\overline{\mathbf{x}}^{-\nu}\right)$, we have $\theta_{\nu}\left(x^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)=$ $P_{\nu}\left(x^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)$; therefore $\overline{\mathbf{x}}$ is also a solution of the GNEP. So the updating rule for the penalty parameters should aim at avoiding convergence of algorithm $\mathcal{A}$ to solutions that are unfeasible for the original GNEP by increasing the penalty parameters if this "dangerous" situation seems to be occurring. To this end, let us consider a solution $\overline{\mathbf{x}}$ of (2.5) infeasible for the GNEP. This means that there exists a player $\nu \in\{1, \cdots, N\}$ such that $\left\|g_{+}^{\nu}\left(\overline{\mathbf{x}}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)\right\|_{\gamma}>$ 0 . By the closedness and the convexity of the feasible set $K_{\nu}$ and by the continuous differentiability of $P_{\nu}\left(x^{\nu}, x^{-\nu}\right)$ in $\overline{\mathbf{x}}$, we have, in view of the minimum principle (see also Proposition 6 for the equivalent VI formulation)

$$
\begin{equation*}
\bar{x}^{\nu}=\operatorname{Proj}_{K_{\nu}}\left[\bar{x}^{\nu}-\nabla_{x^{\nu}} P_{\nu}\left(\overline{\mathbf{x}}, \rho_{\nu}\right)\right] . \tag{2.23}
\end{equation*}
$$

where $\operatorname{Proj}_{K_{\nu}}(\bullet)$ is the Euclidean projector on the closed convex set $K_{\nu}$. Our updating rule tries to detect when condition (2.23) is nearly satisfied (see
test (2.24) below), and in such case, tries to force feasibility by increasing the penalty parameters.

$$
\begin{align*}
& \text { Algorithm 2: Penalty Updating Scheme for GNEPs } \\
& \hline \text { (S.0) : } \mathbf{x}^{0} \in \mathbf{K}, \rho_{\nu}^{0}>0 \text { and } c_{\nu} \in(0,1) \text { for all } \nu \mathrm{s} \text {. Set } k:=0 . \\
& \text { (S.1) : If } \mathbf{x}^{k} \text { is a solution of the GNEP }(2.2) \text { : STOP } \\
& \text { (S.2) : Let } I^{k}:=\left\{\nu \mid\left(x^{k}\right)^{\nu} \notin X_{\nu}\left(\left(\mathbf{x}^{k}\right)^{-\nu}\right)\right\} \text {. For every } \nu \in I^{k} \text {, if } \\
& \qquad\left\|\left(x^{k}\right)^{\nu}-\operatorname{Proj}_{K_{\nu}}\left[\left(x^{k}\right)^{\nu}-\nabla_{x^{\nu}} P_{\nu}\left(\mathbf{x}^{k}, \rho_{\nu}^{k}\right)\right]\right\| \leq \frac{c_{\nu}}{\rho_{\nu}^{k}},  \tag{2.24}\\
& \text { then double the penalty parameters } \rho_{\nu}^{k} . \\
& \text { (S.3) : Compute } \mathbf{x}^{k+1}=\mathcal{A}\left[\mathbf{x}^{k}\right] \text {, set } k \leftarrow k+1 \text {, and go to (S.1). }
\end{align*}
$$

Based on Theorem 8, we know that the $\mathrm{CQ}_{\gamma}$ is sufficient to guarantee the existence of suitable values of the penalty parameters that allow to recover a solution of the GNEP from a solution of the $\mathrm{PNEP}_{\boldsymbol{\rho}}$. We would therefore expect that $\mathrm{CQ}_{\gamma}$ should be all we need to have Algorithm 2 work properly. Luckily, this is the case as shown by the following theorem.

Theorem 11. Let a GNEP be given satisfying the $C Q_{\gamma}$ at every point in $\mathbf{K}$. Let $\left\{\mathbf{x}^{k}\right\}$ be the sequence generated by Algorithm 2. The following two assertions hold:
(a) The penalty parameters are updated a finite number of times only;
(b) Every limit point $\overline{\mathbf{x}}$ of the sequence $\left\{\mathbf{x}^{k}\right\}$ is a solution of the GNEP.

Proof. (a) Suppose by contradiction that there is $\nu$ such that $\rho_{\nu}^{k} \rightarrow \infty$. Subsequencing if necessary, we may assume that $\rho_{\nu}^{k}$ is updated at every iteration and $K_{\nu} \ni\left(x^{k}\right)^{\nu} \rightarrow \bar{x}^{\nu}$. Furthermore, recalling that $P_{\nu}\left(\mathbf{x}^{k}, \rho_{\nu}^{k}\right)$ is continuously differentiable at $\mathbf{x}^{k}$ and taking into account test (2.24), we also have $\operatorname{Proj}_{K_{\nu}}\left[\left(x^{k}\right)^{\nu}-\nabla_{x^{\nu}} P_{\nu}\left(\mathrm{x}^{k}, \rho_{\nu}^{k}\right)\right] \rightarrow \bar{x}^{\nu}$. Then we define $\left(y^{k}\right)^{\nu}:=$ $\operatorname{Proj}_{K_{\nu}}\left[\left(x^{k}\right)^{\nu}-\nabla_{x^{\nu}} P_{\nu}\left(\mathbf{x}^{k}, \rho_{\nu}^{k}\right)\right]:$ therefore $\left(y^{k}\right)^{\nu} \rightarrow \bar{x}^{\nu}$. By the projection characterization property we have

$$
\begin{equation*}
\left[w^{\nu}-\left(y^{k}\right)^{\nu}\right]^{T}\left[\left(y^{k}\right)^{\nu}-\left(x^{k}\right)^{\nu}+\nabla_{x^{\nu}} \theta\left(\mathbf{x}^{k}\right)+\rho_{\nu}^{k} \nabla_{x^{\nu}}\left\|g_{+}^{\nu}\left(\left(x^{k}\right)^{\nu},\left(x^{k}\right)^{-\nu}\right)\right\|_{\gamma}\right] \geq 0 \tag{2.25}
\end{equation*}
$$

for all $w^{\nu} \in K_{\nu}$. Dividing both sides of (2.25) by $\rho_{\nu}^{k}$ and considering the boundedness of the (partial sub)gradient of $\left\|g_{+}^{\nu}\left(x^{\nu}, \mathbf{x}^{\nu}\right)\right\|_{\gamma}$ on bounded sets and its closedness, we get

$$
\left(w^{\nu}-\bar{x}^{\nu}\right)^{T} \xi^{\nu} \geq 0
$$

for all $w^{\nu} \in K_{\nu}$ with $\xi^{\nu} \in \partial_{x^{\nu}}^{>}\left\|g_{+}^{\nu}\left(\bar{x}^{\nu}, \overline{\mathbf{x}}^{-\nu}\right)\right\|_{\gamma}$. So we have $\left(-\xi^{\nu}\right) \in N_{K_{\nu}}\left(\bar{x}^{\nu}\right)$, a contradiction to (2.7). Hence, all penalty parameters are updated a finite number of times only.
(b) We have shown that, thanks to $\mathrm{CQ}_{\gamma}$, the penalty parameters are updated a finite number of times only. Now we prove that (a) implies (b). In order to show that a limit point $\overline{\mathbf{x}}$ is a Nash equilibrium of the original GNEP, it is enough to show that $\overline{\mathbf{x}}$ is feasible. Suppose then, by contradiction, that $\overline{\mathbf{x}}$ is not feasible so that there is a $\nu$ such that $\left\|g_{+}^{\nu}(\overline{\mathbf{x}})\right\|_{\gamma}>0$. Suppose also, without loss of generality, that the penalty parameters are not updated for any $k$ and, as usual, $\mathbf{x}^{k} \rightarrow \overline{\mathbf{x}}$. Since $\left\|g_{+}^{\nu}\left(\mathbf{x}^{k}\right)\right\|_{\gamma}>0, P_{\nu}$ is continuously differentiable in a neighborhood of $\overline{\mathbf{x}}$. By the convexity and compactness of $K_{\nu}$ and by simple continuity arguments, this shows that the test (2.24) must be satisfied eventually for all $\nu$ such that $\left\|g_{+}^{\nu}\left(\mathbf{x}^{k}\right)\right\|_{\gamma}>0$ and for all $k$ sufficiently large. Hence, the corresponding penalty parameters $\rho_{\nu}^{k}$ are updated. This contradiction shows that $\overline{\mathbf{x}}$ is feasible.

If the $\mathrm{CQ}_{\gamma}$ is not satisfied everywhere, it could happen that the penalty parameters are updated an infinite number of times. In this case, we could still prove, along the same lines used in [30], that the limit points of the sequence $\left\{\mathbf{x}^{k}\right\}$ are still meaningful to the GNEP. However, this analysis is somewhat involved and, in the end, of little practical interest, since, if some penalty parameters go to infinite, the algorithm is anyway becoming numerically unstable. We therefore omit this kind of analysis here and refer the interested reader to [30] for details.

### 2.7 Algorithm $\mathcal{A}$

Theorem 11 says that the updating scheme will increase the penalty parameter only a finite number of times, so that eventually Algorithm 2 reduces, as we had hoped, to the application of Algorithm $\mathcal{A}$ to the $\mathrm{PNEP}_{\rho}$ for a fixed value of the penalty parameters. The overall convergence properties of Algorithm 2 therefore are based on those of Algorithm $\mathcal{A}$ and in particular on Property 1. In this section we show that, under reasonable assumptions, we can develop an algorithm $\mathcal{A}$ that enjoys Property 1. We stress once more that, based on the results of the previous section, we are only interested in algorithms for the solution of $\mathrm{PNEP}_{\rho}$ for fixed values of the penalty parameters. Therefore, in this section the penalty parameters are assumed to be fixed throughout.

Problem (2.6) is a standard Nash equilibrium problem whose solution is problematic because of the nondifferentiability of the objective functions $P_{\nu}\left(\mathbf{x}, \rho_{\nu}\right)$. We propose to deal with this difficulty by using smoothing techniques. We recall that the objective function of player $\nu$ in (2.6) is given by:

$$
\begin{aligned}
P_{\nu}\left(\mathbf{x}, \rho_{\nu}\right) & =P_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}, \rho_{\nu}\right) \\
& =\theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)+\rho_{\nu}\left\|g_{+}^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\|_{\gamma} \\
& =\theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)+\rho_{\nu}\left(\sum_{j=1}^{m_{\nu}} \max \left\{0, g_{j}^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\}^{\gamma}\right)^{1 / \gamma} .
\end{aligned}
$$

We can approximate these functions by the smooth mappings:

$$
\begin{aligned}
& \tilde{P}_{\nu}\left(\mathbf{x}, \rho_{\nu}, \varepsilon\right):=\tilde{P}_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}, \rho_{\nu}, \varepsilon\right) \\
& \quad:=\theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)+\rho_{\nu}\left(\sum_{j=1}^{m_{\nu}} \max \left\{0, g_{j}^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\}^{\gamma}+\varepsilon\right)^{1 / \gamma}+\frac{\varepsilon}{2}\left\|x^{\nu}\right\|^{2}
\end{aligned}
$$

where $\varepsilon>0$ is a given parameter which makes the root mapping smooth. So far, $\gamma$ was taken arbitrarily from the open interval $(1, \infty)$. From now on we assume that
$-\gamma>2$ holds;

- $\theta_{\nu}$ and $g^{\nu}$ are twice continuously differentiable on $\mathbf{K}$.

The two assumptions above, make the max-term twice continuously differentiable so that $\tilde{P}_{\nu}$ itself is twice continuously differentiable. We are therefore naturally led to define a smooth "approximation" of the $\operatorname{PNEP}_{\boldsymbol{\rho}}(2.6)$, namely the $\operatorname{PNEP}_{\rho}(\varepsilon)$ where the problem of player $\nu$ is minimizing the function $\tilde{P}_{\nu}$ on the set $K_{\nu}$ :

$$
\begin{gathered}
\min _{x^{\nu}} \\
\tilde{P}_{\nu}\left(\mathbf{x}, \rho_{\nu}, \varepsilon\right) \\
x^{\nu} \in K_{\nu}
\end{gathered}
$$

We remark that the presence of the regularization term $(\varepsilon / 2)\left\|x^{\nu}\right\|^{2}$ guarantees that $\tilde{P}_{\nu}\left(\mathbf{x}, \rho_{\nu}, \varepsilon\right)$ is uniformly convex as a function of $x^{\nu}$.

We observe that the $\operatorname{PNEP}_{\rho}(\varepsilon)$ is a game where each player's problem is a continuously differentiable, constrained (the feasible sets $K_{\nu}$ are compact for all $\nu$ ), convex optimization problem. It is well know, see Section 2.1 and, for example, [32], that the game $\operatorname{PNEP}_{\rho}(\varepsilon)$ can equivalently be cast and solved as a variational inequality $\operatorname{VI}\left(\mathbf{K}, F_{\varepsilon}(\mathbf{x})\right)$, with

$$
F_{\varepsilon}(\mathbf{x}):=\left(\begin{array}{c}
\nabla_{x^{1}} \tilde{P}_{1}\left(x^{1}, \mathbf{x}^{-1}, \rho_{1}, \varepsilon\right)  \tag{2.26}\\
\vdots \\
\nabla_{x^{N}} \tilde{P}_{N}\left(x^{N}, \mathbf{x}^{-N}, \rho_{N}, \varepsilon\right)
\end{array}\right)
$$

We remark that thanks to the assumptions made, $F_{\varepsilon}(\mathbf{x})$ is continuously differentiable. The simple idea, common to all smoothing methods, is now to solve $\mathrm{PNEP}_{\rho}$ by solving, inaccurately, but with increasing accuracy, a sequence of $\operatorname{VI}\left(\mathbf{K}, F_{\varepsilon_{i}}(\mathbf{x})\right)$ for values of the smoothing parameter going to zero
(we use the parameter $i$ to denote iterations here in order have no confusion with the iteration counter $k$ in the previous section). We can then rely on the rich literature about solution methods for VIs.

In the next section we then show that indeed, a sequence of inaccurate solutions of $\operatorname{VI}\left(\mathbf{K}, F_{\varepsilon_{i}}(\mathbf{x})\right)$ converges to a solution of the $\mathrm{PNEP}_{\boldsymbol{\rho}}$. The analysis here is similar to that carried out in [30]. The following section, instead, is the core of this work and presents novel and interesting results showing that under suitable condition we are actually able to find a(n inaccurate) solution of $\operatorname{VI}\left(\mathbf{K}, F_{\varepsilon_{i}}(\mathbf{x})\right)$. Putting together the results of this and the previous section we will finally show that Algorithm 2 is implementable and guaranteed to find a solution of the GNEP under conditions that cover both known and novel classes of problems.

### 2.8 Convergence of the smoothing procedure

In order to analyze the behavior of solutions of $\operatorname{VI}\left(\mathbf{K}, F_{\varepsilon_{i}}(\mathbf{x})\right)$ when $\varepsilon_{i}$ goes to zero, we recall a technical result from [35].

Lemma 2. Let $f: \mathbb{R}^{s} \times \mathbb{R}^{t} \rightarrow \mathbb{R}$ be given and assume that $f$ is locally Lipschitz continuous around a point $(\bar{u}, \bar{v}) \in \mathbb{R}^{s} \times \mathbb{R}^{t}$ and such that $f(\bullet, v)$ is convex for every $v$ in a neighborhood of $\bar{v}$. Let $\left\{\left(u^{i}, v^{i}\right)\right\}$ be a sequence of points converging to $(\bar{u}, \bar{v})$ and let $\left\{\xi^{i}\right\}$, with $\xi^{i} \in \partial_{u} f\left(u^{i}, v^{i}\right)$, be a sequence of (Clarke's) partial generalized gradients. Then, every limit point $\bar{\xi}$ of this sequence (and there is at least one such limit point) belongs to $\partial_{u} f(\bar{u}, \bar{v})$.

We are now ready to show convergence of the smoothing procedure. We recall that the left-hand side of (2.27) below is a measure of how accurate we have solved the $\operatorname{VI}\left(\mathbf{K}, F_{\varepsilon_{i}}(\mathbf{x})\right)$, see [32]. If $\eta_{i}=0$ we have an exact solution, otherwise an inaccurate one, the inaccuracy increasing when $\eta_{i}$ increases.

Proposition 9. Let $\left\{\varepsilon_{i}\right\}$ and $\left\{\eta_{i}\right\}$ be two sequences of positive numbers converging to 0 and, for every $i$, let $\mathbf{x}\left(\varepsilon_{i}\right) \in \mathbf{K}$ be a point such that

$$
\begin{equation*}
\left\|\mathbf{x}\left(\varepsilon_{i}\right)-\operatorname{Proj}_{\mathbf{K}}\left[\mathbf{x}\left(\varepsilon_{i}\right)-F_{\varepsilon_{i}}\left(\mathbf{x}\left(\varepsilon_{i}\right)\right)\right]\right\| \leq \eta_{i} \tag{2.27}
\end{equation*}
$$

Then every limit point of the sequence $\mathbf{x}\left(\varepsilon_{i}\right)$ is a solution of the $P N E P_{\boldsymbol{\rho}}(2.6)$.
Proof. We assume without loss of generality that $\mathbf{x}\left(\varepsilon_{i}\right) \rightarrow \overline{\mathbf{x}} \in \mathbf{K}$. In order to show that $\overline{\mathbf{x}}$ is a solution of the $\operatorname{PNEP}_{\rho}$, it is sufficient to show that, for all $\nu$, there is a $\bar{\xi}^{\nu} \in \partial_{x^{\nu}} P_{\nu}\left(\overline{\mathbf{x}}, \rho_{\nu}\right)$ such that

$$
\begin{equation*}
-\bar{\xi}^{\nu} \in N_{K_{\nu}}\left(\bar{x}^{\nu}\right) \tag{2.28}
\end{equation*}
$$

To this end, and for a fixed $\nu$, define the function
$\tilde{\tilde{P}}_{\nu}(\mathbf{x}, \delta):=\theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)+\rho_{\nu}\left(\sum_{j=1}^{m_{\nu}} \max \left\{0, g_{j}^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right\}^{\gamma}+|\delta|^{\gamma}\right)^{1 / \gamma}+\frac{\delta^{\gamma}}{2}\left\|x^{\nu}\right\|^{2}$
(remember that $\rho_{\nu}$ is a fixed quantity in the present setting), and note that $\tilde{\tilde{P}}_{\nu}$ is everywhere locally Lipschitz as a function of $(\bar{x}, \delta)$ (note that, for this observation to be correct, we had to replace $\varepsilon$ by $|\delta|^{\gamma}$ in the definition of $\left.\tilde{\tilde{P}}_{\nu}\right)$. Note furthermore that $P_{\nu}\left(\mathbf{x}, \rho_{\nu}\right)=\tilde{P}_{\nu}\left(\mathbf{x}, \rho_{\nu}, 0\right)=\tilde{\tilde{P}}_{\nu}(\mathbf{x}, 0)$ and that $\partial_{x^{\nu}} \tilde{P}_{\nu}\left(\mathbf{x}, \rho_{\nu}, \varepsilon_{i}\right)=\partial_{x^{\nu}} \tilde{\tilde{P}}\left(\mathbf{x},\left(\varepsilon_{i}\right)^{1 / \gamma}\right)$.
Let us define $\mathbf{y}\left(\varepsilon_{i}\right):=\operatorname{Proj}_{\mathbf{K}}\left(\mathbf{x}\left(\varepsilon_{i}\right)-F_{\delta_{i}^{\gamma}}\left(\mathbf{x}\left(\varepsilon_{i}\right)\right)\right)$. By the projection characterization property (see Theorem 1) we have

$$
\left(\mathbf{z}-\mathbf{y}\left(\varepsilon_{i}\right)\right)^{T}\left(\mathbf{y}\left(\varepsilon_{i}\right)-\mathbf{x}\left(\varepsilon_{i}\right)+F_{\delta_{i}^{\gamma}}\left(\mathbf{x}\left(\varepsilon_{i}\right)\right)\right) \geq 0, \quad \forall i, \quad \forall \mathbf{z} \in \mathbf{K}
$$

This latter property implies the following relation for every player $\nu$ :

$$
\begin{equation*}
\left(z^{\nu}-y^{\nu}\left(\varepsilon_{i}\right)\right)^{T}\left(\left(y^{\nu}\left(\varepsilon_{i}\right)-x^{\nu}\left(\varepsilon_{i}\right)+\nabla_{x^{\nu}} \tilde{\tilde{P}}_{\nu}\left(\mathbf{x}\left(\varepsilon_{i}\right), \delta_{i}^{\gamma}\right)\right) \geq 0, \quad \forall i, \forall z^{\nu} \in K_{\nu}\right. \tag{2.29}
\end{equation*}
$$

We now set

$$
\begin{aligned}
\delta_{i}:=\left(\varepsilon_{i}\right)^{1 / \gamma}, \quad u^{i} & :=x^{\nu}\left(\varepsilon_{i}\right), \quad v^{i}:=\left(\mathbf{x}^{-\nu}\left(\varepsilon_{i}\right), \delta_{i}\right), \\
\left(\xi^{\nu}\right)^{i} & :=\nabla_{x^{\nu}} \tilde{\tilde{P}}_{\nu}\left(\left(x^{\nu}\right)^{i},\left(\mathbf{x}^{-\nu}\right)^{i}, \delta_{i}\right) .
\end{aligned}
$$

Therefore, taking the limit of an appropriate subsequence of both sides of (2.29), considering that, for (2.27), $\mathbf{y}\left(\varepsilon_{i}\right) \rightarrow \overline{\mathbf{x}}$ and taking (2.26) into account, Lemma 2 implies

$$
-\bar{\xi}^{\nu} \in N_{K_{\nu}}\left(\bar{x}^{\nu}\right), \quad \forall \nu
$$

where we have assumed, without loss of generality, that $\left(\xi^{\nu}\right)^{i} \rightarrow \bar{\xi}^{\nu}$.
We see from the previous proposition that we can solve the PNEP $_{\rho}$ by finding a sequence of inaccurate solutions of $\operatorname{VI}\left(\mathbf{K}, F_{\varepsilon_{i}}(\mathbf{x})\right)$ for values of $\varepsilon_{i}$ and $\eta_{i}$ going to zero. All we are then left with, to complete the analysis, is to study under what conditions we are able to calculate such approximate solutions. This is the topic of the next subsection.

## $2.9 \mathrm{VI}\left(\mathrm{K}, \boldsymbol{F}_{\varepsilon_{i}}(\mathrm{x})\right)$

As we have seen in the previous sections, problem $\mathrm{PNEP}_{\rho}$ can be equivalently recast as the variational inequality $\mathrm{VI}\left(\mathbf{K}, F_{\varepsilon}(\mathbf{x})\right)$. Then our task is now to
solve, for fixed $\varepsilon, \operatorname{VI}\left(\mathbf{K}, F_{\varepsilon}(\mathbf{x})\right.$ ) (since throughout this subsection $\varepsilon_{i}$ is fixed, we drop the subscritp $i$ ). We recall that $\mathbf{K}=\prod_{\nu} K_{\nu}$ and note that we can write

$$
F_{\varepsilon}(\mathbf{x}):=\left(\begin{array}{c}
\nabla_{x^{1}} \tilde{P}_{1}\left(x^{1}, \mathbf{x}^{-1}, \rho_{1}, \varepsilon\right)  \tag{2.30}\\
\vdots \\
\nabla_{x^{N}} \tilde{P}_{N}\left(x^{N}, \mathbf{x}^{-N}, \rho_{N}, \varepsilon\right)
\end{array}\right)=F(\mathbf{x})+V(\mathbf{x}, \boldsymbol{\rho}, \varepsilon)+\varepsilon \mathbf{x}
$$

with

$$
\begin{gathered}
F(\mathbf{x}):=\left(\begin{array}{c}
\nabla_{x^{1}} \theta_{1}(\mathbf{x}) \\
\vdots \\
\nabla_{x^{N}} \theta_{N}(\mathbf{x})
\end{array}\right), \quad V(\mathbf{x}, \boldsymbol{\rho}, \varepsilon):=\left(\begin{array}{c}
V^{1}\left(\mathbf{x}, \rho_{1}, \varepsilon\right) \\
\vdots \\
V^{N}\left(\mathbf{x}, \rho_{N}, \varepsilon\right)
\end{array}\right) \\
V^{\nu}(\mathbf{x}):=\frac{\rho_{\nu}}{\left\|\left(g_{+}^{\nu}(\mathbf{x}), \varepsilon^{\frac{1}{\gamma}}\right)\right\|_{\gamma}^{\gamma-1}} \nabla_{x^{\nu}} g^{\nu}(\mathbf{x})\left[g_{+}^{\nu}(\mathbf{x})\right]^{\gamma-1}
\end{gathered}
$$

and where $\left[g_{+}^{\nu}(\mathbf{x})\right]^{\gamma-1}$ means that we apply the exponent $\gamma-1$ to each component of the vector $g_{+}^{\nu}(\mathbf{x})$.
Since $\mathbf{K}$ is a compact convex set and $F_{\varepsilon}$ is continuous, $\operatorname{VI}\left(\mathbf{K}, F_{\varepsilon}(\mathbf{x})\right)$ has a solution.
There exist in literature many, well known and efficient methods for the solution of variational inequalities but the conditions for convergence are somewhat restrictive (see Subsection 1.3.3): one of the weakest conditions under which global convergence to a solution of the $\operatorname{VI}\left(\mathbf{K}, F_{\varepsilon}(\mathbf{x})\right)$ can be guaranteed is the monotonicity of $F_{\varepsilon}(\mathbf{x})$ on $\mathbf{K}$. Suffices to say here, that monotonicity of $F_{\varepsilon}(\mathbf{x})$ is enough to design convergent algorithms for the solution of the $\mathrm{VI}\left(\mathbf{K}, F_{\varepsilon}(\mathbf{x})\right)$ (see [32]): the monotonicity of the defining function is a standard and well accepted assumption in the VI theory and it is satisfied in many practical applications. With regard to the specific structure of the defining function, it can be seen that the monotonicity assumption implies some connections among the $\tilde{P}_{\nu}$ s that cannot be expected to hold in general. If, indeed, the defining function $F_{\varepsilon}$ is continuously differentiable, then it is well known that $F_{\varepsilon}$ is monotone if and only if its Jacobian is positive semidefinite: this fact let us understand that some assumptions on the structure of $F_{\varepsilon}$ and relations between $\tilde{P}_{\nu}$ s must hold.

Nevertheless, for the purpose of this work it could be said that our task reduces to the study of conditions guaranteeing the monotonicity of $F_{\varepsilon}(\mathbf{x})$. In turn, since $F_{\varepsilon}(\mathbf{x})$ is $C^{1}$ under the assumptions we made, it will be enough, to check monotonicity, to verify the positive semidefiniteness of the Jacobian of $F_{\varepsilon}(\mathbf{x})$ (see Proposition 1).

To this end let us introduce some additional notations. First of all we denote by $\mathbf{g}$ the vector of all constraints: $\mathbf{g}=\left(g^{\nu}\right)_{\nu=1}^{N}$. In the sequel we will
also often need to partition $n \times n$ (or $m \times n$ ) matrices quite naturally as

$$
T=\left(\begin{array}{cccc}
T_{11} & T_{12} & \cdots & T_{1 N} \\
\vdots & \vdots & \ddots & \vdots \\
T_{N 1} & T_{N 2} & \cdots & T_{N N}
\end{array}\right)
$$

where block $T_{\nu \mu} \in M_{n_{\nu}, n_{\mu}}$ (or $\in M_{m_{\nu}, n_{\mu}}$ ). For example, $[J V(\mathbf{x}, \boldsymbol{\rho}, \varepsilon)]_{\nu \mu}=$ $J_{x^{\mu}} V^{\nu}(\mathbf{x}, \boldsymbol{\rho}, \varepsilon) \in M_{n_{\nu}, n_{\mu}}$, while $[J \mathbf{g}(\mathbf{x})]_{\nu \mu}=J_{x^{\mu}} g^{\nu}(\mathbf{x}) \in M_{m_{\nu}, n_{\mu}}$. It is now useful to define the following matrices:

$$
\begin{array}{r}
\operatorname{diag} J \mathbf{g}(\mathbf{x}):=\left(\begin{array}{cccc}
J_{x^{1}} g^{1}(\mathbf{x}) & 0_{m_{1} n_{2}} & \cdots & 0_{m_{1} n_{N}} \\
0_{m_{2} n_{1}} & J_{x^{2}} g^{2}(\mathbf{x}) & \cdots & 0_{m_{2} n_{N}} \\
\vdots & & & \vdots \\
0_{m_{N} n_{1}} & 0_{m_{N} n_{2}} & \cdots & J_{x^{N}} g^{N}(\mathbf{x})
\end{array}\right), \\
\operatorname{off} J \mathbf{g}(\mathbf{x}):=\left(\begin{array}{cccc}
0_{m_{1} n_{1}} & J_{x^{2}} g^{1}(\mathbf{x}) & \cdots & J_{x^{N}} g^{1}(\mathbf{x}) \\
J_{x^{1}} g^{2}(\mathbf{x}) & 0_{m_{2} n_{2}} & \cdots & J_{x^{N}} g^{2}(\mathbf{x}) \\
\vdots & & & \vdots \\
J_{x^{1}} g^{N}(\mathbf{x}) & J_{x^{2}} g^{N}(\mathbf{x}) & \cdots & 0_{m_{N} n_{N}}
\end{array}\right)
\end{array}
$$

and

$$
\bar{M}(\mathbf{x}, \boldsymbol{\rho})=\left(\begin{array}{ccc}
\rho_{1} M_{11}(\mathbf{x}) & \cdots & 0 \\
0 & \ddots & 0 \\
0 & \cdots & \rho_{N} M_{N N}(\mathbf{x})
\end{array}\right)
$$

where the diagonal blocks $M_{\nu \nu}(\mathbf{x}) \in M_{m_{\nu}}$ are given by

$$
M_{\nu \nu}(\mathbf{x}):=\frac{\left[\operatorname{diag}\left\{g_{+}^{\nu}(\mathbf{x})\right\}\right]^{\gamma-2}}{\left\|\left(g_{+}^{\nu}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{\gamma-1}}-\frac{\left[g_{+}^{\nu}(\mathbf{x})\right]^{\gamma-1}\left(\left[g_{+}^{\nu}(\mathbf{x})\right]^{\gamma-1}\right)^{T}}{\left\|\left(g_{+}^{\nu}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{2 \gamma-1}}
$$

Proposition 10. In the setting above, and for fixed positive $\varepsilon$ and $\rho_{\nu}$, the Jacobian $J V(\mathbf{x})$ is equal to the following expression:

$$
\left(\begin{array}{ccc}
\rho_{1} \frac{\sum_{i=1}^{m_{1}}\left(g_{i}^{1}\right)_{+}^{\gamma-1}(\mathbf{x})\left[\nabla_{x^{1}}^{2}\left(g_{i}^{1}\right)(\mathbf{x})\right]^{T}}{\left\|\left(g_{+}^{1}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{\gamma-1}} & \cdots & \rho_{1} \frac{\sum_{i=1}^{m_{1}}\left(g_{i}^{1}\right)_{+}^{\gamma-1}(\mathbf{x})\left[\nabla_{x^{N}}\left(\nabla_{x^{1}}\left(g_{i}^{1}\right)(\mathbf{x})\right)\right]^{T}}{\left\|\left(g_{+}^{1}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{\gamma-1}} \\
\vdots & \cdots & \vdots  \tag{2.31}\\
\rho_{N} \frac{\sum_{i=1}^{m_{N}}\left(g_{i}^{N}\right)_{+}^{\gamma-1}(\mathbf{x})\left[\nabla_{x^{1}}\left(\nabla_{x^{N}}\left(g_{i}^{N}\right)(\mathbf{x})\right)\right]^{T}}{\left\|\left(g_{+}^{N}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{\gamma-1}} \cdots & \rho_{N} \frac{\sum_{i=1}^{m_{N}}\left(g_{i}^{N}\right)_{+}^{\gamma-1}(\mathbf{x})\left[\nabla_{x^{N}}^{2}\left(g_{i}^{N}\right)(\mathbf{x})\right]^{T}}{\left\|\left(g_{+}^{N}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{\gamma-1}}
\end{array}\right)
$$

Proof. The calculation of the Jacobian of $V$ is straightforward, even if complicated. It is convenient to calculate the $[J V(\mathbf{x})]_{\nu \mu}$ block separately. Direct calculation gives

$$
\begin{align*}
& {[J V(\mathbf{x})]_{\nu \mu}=\frac{\rho_{\nu}}{\left\|\left(g_{+}^{\nu}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{\gamma-1}} \sum_{j=1}^{m_{\nu}}\left\{\left(g_{j}^{\nu}\right)_{+}^{\gamma-1}(\mathbf{x})\left[\nabla_{x^{\nu} x^{\mu}}^{2} g_{j}^{\nu}(\mathbf{x})\right]^{T}\right.} \\
& \left.\quad+(\gamma-1)\left(g_{j}^{\nu}\right)_{+}^{\gamma-2} \nabla_{x^{\nu}} g_{j}^{\nu}(\mathbf{x})\left[\nabla_{x^{\mu}} g_{j}^{\nu}(\mathbf{x})\right]^{T}\right\} \\
& \quad+\frac{\rho_{\nu}(1-\gamma)}{\left\|\left(g_{+}^{\nu}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{2 \gamma-1}}\left[\nabla_{x^{\nu}} g^{\nu}(\mathbf{x})\left(g_{+}^{\nu}\right)^{\gamma-1}(\mathbf{x})\right]\left[\nabla_{x^{\mu}}\left(g^{\nu}\right)(\mathbf{x})\left(g_{+}^{\nu}\right)^{\gamma-1}(\mathbf{x})\right]^{T} \\
& \quad=\frac{\rho_{\nu}}{\left\|\left(g_{+}^{\nu}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{\gamma-1}} \sum_{j=1}^{m_{\nu}}\left(g_{j}^{\nu}\right)_{+}^{\gamma-1}(\mathbf{x})\left[\nabla_{x^{\nu} x^{\mu}}^{2} g_{i}^{\nu}(\mathbf{x})\right]^{T} \\
& \quad+\rho_{\nu}(\gamma-1) \nabla_{x^{\nu}} g^{\nu}(\mathbf{x}) M_{\nu \nu}(\mathbf{x})\left[\nabla_{x^{\mu}} g^{\nu}(x)\right]^{T} \tag{2.32}
\end{align*}
$$

from which the thesis easily follows.
With the result in mind, the following Proposition 12 allows us to individuate classes of problems for which the Jacobian of $F_{\varepsilon}(\mathbf{x})$ is positive definite under suitable assumptions. In order to prove Theorem 12, we need some other definitions and a preliminary result.

First we show that the $M_{\nu \nu}(\mathbf{x})$ block is positive semidefinite.
Lemma 3. The $M_{\nu \nu}(\mathbf{x})$ block is positive semidefinite for every $\mathbf{x} \in \mathbb{R}^{n}$ and for every $\nu=1, \ldots, N$.

Proof. For sake of simplicity, let us consider the generic symmetric matrix

$$
M(\mathbf{x})=\left\{\frac{\left[\operatorname{diag}\left\{g_{+}(\mathbf{x})\right\}\right]^{\gamma-2}}{\left\|\left(g_{+}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{\gamma-1}}-\frac{\left[g_{+}(\mathbf{x})\right]^{\gamma-1}\left(\left[g_{+}(\mathbf{x})\right]^{\gamma-1}\right)^{T}}{\left\|\left(g_{+}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{2 \gamma-1}}\right\}
$$

and $M(\mathbf{x}) \in M_{m}$.

It suffices to show the semidefinite positiveness of $\tilde{M}(\mathbf{x})$, where

$$
\tilde{M}(\mathbf{x})=\left\|\left(g_{+}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{\gamma}\left[\operatorname{diag}\left\{g_{+}(\mathbf{x})\right\}\right]^{\gamma-2}-\left[g_{+}(\mathbf{x})\right]^{\gamma-1}\left(\left[g_{+}(\mathbf{x})\right]^{\gamma-1}\right)^{T}
$$

and $M(\mathbf{x})=\frac{1}{\left\|\left(g_{+}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{2 \gamma-1}} \tilde{M}(\mathbf{x})$.
Since $\varepsilon^{1 / \gamma}>0$ we have, for $z \in \mathbb{R}^{m}$,

$$
\begin{array}{r}
z^{T} \tilde{M} z \geq\left(g_{1}^{\gamma}+\cdots+g_{m}^{\gamma}{ }_{+}^{\gamma}\right)\left(z_{1} \cdots z_{m}\right)\left(\begin{array}{cccc}
g_{1} \\
{ }_{+}^{\gamma-2} & 0 & \cdots & 0 \\
0 & g_{2}{ }_{+}^{\gamma-2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & g_{m}{ }_{+}^{\gamma-2}
\end{array}\right)\left(\begin{array}{c}
z_{1} \\
\vdots \\
z_{m}
\end{array}\right) \\
-\left(z_{1} \cdots z_{m}\right)\left(\begin{array}{c}
g_{1}{ }_{+}^{\gamma-1} \\
\vdots \\
g_{m} \\
\vdots-1
\end{array}\right)\left(\begin{array}{lll}
g_{1}{ }_{+}^{\gamma-1} & \cdots & g_{m}{ }_{+}^{\gamma-1}
\end{array}\right)\left(\begin{array}{c}
z_{1} \\
\vdots \\
z_{m}
\end{array}\right) .
\end{array}
$$

The first term in the right-hand side above, i.e. $\left(g_{1}{ }_{+}^{\gamma}+\cdots+g_{m}{ }_{+}^{\gamma}\right)\left(g_{1}{ }_{+}^{\gamma-2} z_{1}^{2}+\right.$ $\left.\cdots+g_{m}{ }_{+}^{\gamma-2} z_{m}^{2}\right)$, is constituted by $m^{2}$ addenda, the second, i.e. $\left(g_{1}{ }_{+}^{\gamma-1} z_{1}+\right.$ $\left.\cdots+g_{m}{ }_{+}^{\gamma-1} z_{m}\right)^{2}$, by $\frac{m(m+1)}{2}$ addenda.
Therefore, we can write:

$$
\begin{align*}
& \left(g_{1}{ }_{+}^{\gamma}+\cdots+g_{m} \stackrel{\gamma}{\gamma}\right)\left(g_{1}{ }_{+}^{\gamma-2} z_{1}^{2}+\cdots+g_{m}{ }_{+}^{\gamma-2} z_{m}^{2}\right)-\left(g_{1}{ }_{+}^{\gamma-1} z_{1}+\cdots+g_{m}{ }_{+}^{\gamma-1} z_{m}\right)^{2} \\
& =\left(g_{1}^{\gamma}+\cdots+g_{m}^{\gamma}\right) g_{1}^{\gamma-2} z_{1}^{2}+\cdots+\left(g_{1}^{\gamma}+\cdots+g_{m}^{\gamma}\right) g_{m}{ }_{+}^{\gamma-2} z_{m}^{2} \\
& \quad-g_{1}^{2 \gamma-2} z_{1}^{2}-g_{2}^{\gamma-2} z_{2}^{2}-\cdots-g_{m}{ }_{+}^{\gamma-2} z_{m}^{2}-2 g_{1}^{\gamma-1} g_{2}^{\gamma-1}{ }_{+}^{\gamma-1} z_{1} z_{2}-\cdots \\
& \quad-2 g_{1}^{\gamma-1} g_{m}^{\gamma-1} z_{1} z_{m}-\cdots-2 g_{m-1}^{\gamma-1} g_{m}{ }_{+}^{\gamma-1} z_{m-1} z_{m} . \tag{2.33}
\end{align*}
$$

We note that the $m$ addenda $g_{i+}^{2 \gamma-2} z_{i}^{2}, i=1, \ldots, m$ of the first term (see the second line of (2.33)) are equal to those of the second term in the third line of the expression (2.33). So we can erase them. The remaining part of (2.33) is constituted by the sum of the trinomials $g_{k}{ }_{+}^{\gamma-2} g_{l}{ }_{+}^{\gamma} z_{k}^{2}+g_{k}{ }_{+}^{\gamma} g_{l}^{\gamma-2} z_{l}^{2}-$ $2 g_{k}{ }_{+}^{\gamma-1} g_{l}{ }_{+}^{\gamma-1} z_{k} z_{l}=g_{k}{ }_{+}^{\gamma-2} g_{l_{+}}^{\gamma-2}\left(g_{l_{+}} z_{k}-g_{k+} z_{l}\right)^{2} \geq 0, k \neq l ; k, l=1, \ldots, m$.
We conclude that $z^{T} \tilde{M}(\mathbf{x}) z \geq 0$ for all $\mathbf{x} \in \mathbb{R}^{n}$ and $z \in \mathbb{R}^{m}$.
Now we are able to individuate classes of GNEPs for which our penalty approach is guaranteed to converge to a solution of the original problem.

First we can prove that, in NEP and jointly convex cases, $J F_{\varepsilon}(\mathbf{x})$ is positive definite and our algorithm is provably convergent to a solution of the original problem.

Theorem 12. In the setting of this section and assuming $F(\mathbf{x})$ to be monotone, if the original GNEP (2.2) reduces to either a NEP or to a jointly convex $G N E P$, then $F_{\varepsilon}(\mathbf{x})$ is strongly monotone.

Proof. We recall that

$$
F_{\varepsilon}(\mathbf{x})=F(\mathbf{x})+V(\mathbf{x}, \boldsymbol{\rho}, \varepsilon)+\varepsilon \mathbf{x}
$$

Therefore, by the monotonicity of $F(\mathbf{x})$, it is sufficient to show the semidefinite positiveness of $J V(\mathbf{x})$. We distinguish two cases.
(a) In the NEP case, $[J V(\mathbf{x})]_{\nu \mu}=0$ for $\nu \neq \mu$. Thanks to the positive semidefiniteness of $M_{\nu \nu}(\mathbf{x})$ for every $\nu=1, \ldots, N$ and by the convexity of $g^{\nu}(\mathbf{x})$ as a function of $x^{\nu}$ only, (2.32) implies the thesis.
(b) In our settings, jointly convex case arises when $g^{1}(\mathbf{x})=\ldots=g^{N}(\mathbf{x})=$ $g(\mathbf{x})$ and $g$ is convex with respect to the whole vector $\mathbf{x}$. Therefore, $M_{11}(\mathbf{x})=$ $\ldots=M_{N N}(\mathbf{x})=M(\mathbf{x})$. For sake of simplicity and without loss of generality, for this proof we assume $\rho=\rho_{1}=\rho_{2}=\cdots=\rho_{N}$; then, by (2.32), it is not hard to see that

$$
\begin{aligned}
J V(\mathbf{x}, \boldsymbol{\rho}, \varepsilon)= & \frac{\rho}{\left\|\left(g_{+}(\mathbf{x}), \varepsilon^{1 / \gamma}\right)\right\|_{\gamma}^{\gamma-1}} \sum_{j=1}^{m}\left(g_{j}\right)_{+}^{\gamma-1}(\mathbf{x}) \nabla^{2} g_{j}(\mathbf{x}) \\
& +\rho(\gamma-1)[J g(\mathbf{x})]^{T} M(\mathbf{x}) J g(\mathbf{x})
\end{aligned}
$$

which in turn is positive semidefinite.
The above proposition shows that the two major subclasses of the GNEP (2.1) for which solution algorithms are well known, see [30], can also be handled by penalty methods, since they give rise to monotone $\mathrm{VI}\left(\mathbf{K}, F_{\varepsilon}(\mathbf{x})\right)$. We conclude by showing that penalty methods are actually able to tackle also some new classes of problems for which no previous convergent algorithms are known.

Let us suppose that constraints in (2.2) have the following form:

$$
\begin{equation*}
g^{\nu}(\mathbf{x})=h^{\nu}\left(x^{\nu}\right)+l^{\nu}\left(\mathbf{x}^{-\nu}\right) \forall \nu \tag{2.34}
\end{equation*}
$$

where $h^{\nu}: \mathbb{R}^{n_{\nu}} \rightarrow \mathbb{R}^{m_{\nu}}$ is a convex function, $l^{\nu}: \mathbb{R}^{n-n_{\nu}} \rightarrow \mathbb{R}^{m_{\nu}}$ and both are twice continuously differentiable.
Linear constraints, i.e. $g^{\nu}(\mathbf{x})=A^{\nu 1} x^{1}+\cdots+A^{\nu N} x^{N} \forall \nu$, with $A^{\nu i} \in M_{m_{\nu}, n_{i}}$, are a particular kind of this class of constraints.

Theorem 13. In the setting of this section, for a GNEP with constraints defined in (2.34), if there exists a constant $\alpha>0$ such that

$$
d^{T} J F(\mathbf{x}) d \geq \begin{cases}0 & \forall d \in\{\operatorname{ker}(\operatorname{diag} J \mathbf{g}(\mathbf{x})) \cup \operatorname{ker}(\mathbf{o f f} J \mathbf{g}(\mathbf{x}))\}  \tag{2.35}\\ \alpha\|d\|^{2} & \text { otherwise }\end{cases}
$$

for all $\mathbf{x} \in \mathbf{K}$, with

$$
\begin{equation*}
\alpha \geq(\gamma-1)\|\operatorname{diag} J \mathbf{g}(\mathbf{x})\|\|\bar{M}(\mathbf{x}, \boldsymbol{\rho})\|\|\mathbf{o f f} J \mathbf{g}(\mathbf{x})\| \tag{2.36}
\end{equation*}
$$

then $F_{\varepsilon}(\mathbf{x})$ is strongly monotone.
Proof. It is enough to prove that $J F_{\varepsilon}(\mathbf{x})$ is uniformly positive definite.
Observe that first matrix in expression (2.31) has null off-diagonal blocks. By the convexity of $h^{\nu}\left(x^{\nu}\right)$ this term is therefore positive semidefinite. For the second term in (2.31), we have
$(\operatorname{diag} J \mathbf{g}(\mathbf{x}))^{T} \bar{M}(\mathbf{x}, \boldsymbol{\rho}) J \mathbf{g}(\mathbf{x})=(\operatorname{diag} J \mathbf{g}(\mathbf{x}))^{T} \bar{M}(\mathbf{x}, \boldsymbol{\rho})[\operatorname{diag} J \mathbf{g}(\mathbf{x})+\mathbf{o f f} J \mathbf{g}(\mathbf{x})]$.
Tacking all these facts into account, we distinguish three cases.
(a) If $d \in \operatorname{ker}(\operatorname{diag} J \mathbf{g}(\mathbf{x}))$ then $d^{T} J F_{\varepsilon}(\mathbf{x}) d \geq d^{T} J F(\mathbf{x}) d+\varepsilon\|d\|^{2} \geq \varepsilon\|d\|^{2}$.
(b) If $d \in \operatorname{ker}(\mathbf{o f f} J \mathbf{g}(\mathbf{x}))$ then

$$
d^{T} J F_{\varepsilon}(\mathbf{x}) d \geq d^{T} J F(\mathbf{x}) d+d^{T}(\operatorname{diag} J \mathbf{g}(\mathbf{x}))^{T} \bar{M}(\mathbf{x}, \boldsymbol{\rho}) \operatorname{diag} J \mathbf{g}(\mathbf{x}) d+\varepsilon\|d\|^{2}
$$

$$
\geq \varepsilon\|d\|^{2}
$$

(c) Otherwise we have, by (2.35) and (2.36),

$$
\begin{aligned}
& d^{T}\left[J F(\mathbf{x})+(\gamma-1)(\operatorname{diag} J \mathbf{g}(\mathbf{x}))^{T} \bar{M}(\mathbf{x}, \boldsymbol{\rho}) J \mathbf{g}(\mathbf{x})\right] d \geq \alpha\|d\|^{2} \\
& -(\gamma-1)\|\operatorname{diag} J \mathbf{g}(\mathbf{x})\|\|\bar{M}(\mathbf{x}, \boldsymbol{\rho})\|\|\mathbf{o f f} J \mathbf{g}(\mathbf{x})\|\|d\|^{2}
\end{aligned}
$$

for suitable matrix norms, which in turn is non negative and the rest of the proof is straightforward.

Note that, supposing that $F$ is monotone, the assumption of the previous theorem are automatically satisfied if the game reduces to a pure NEP (or equivalently if, for all players, $l^{\nu}\left(\mathbf{x}^{-\nu}\right)$ in (2.34) is a constant). What Theorem 13 tells us then, is that if the coupling in the constraints is "compensated by sufficient curvature of $F$ in the right subspaces", then we may still be able to solve the resulting GNEP. Although these conditions are strong, they are quite natural, and are, as far as we are aware of, the first breakthrough towards the solution of a significant class of non jointly convex GNEPs.

Note that there is a conceptual difference between the requirements made in Theorem 12 and those in Theorem 13 in that the latter theorem requires a condition that depends on the penalty parameters $\rho_{\nu}$, while the former theorem has no such dependence. Conceptually, this means that while under the conditions given in Theorem 12 we can ensure that the overall penalty scheme of Algorithm 2 will converge to a solution, we can draw the same conclusions under the condition of Theorem 13 only provided that (2.36) holds for every penalty parameters generated by the algorithm. On the positive side, we can add that, under the $\mathrm{CQ}_{\gamma}$, we know in advance the algorithm will generate
penalty parameters that have a common upper bound. This means that condition (2.36) must only be verified for this upper bound. Although the upper bound is not known in general, and so condition (2.36) is not checkable $a$ priori, the theoretical value of Theorem 13 is not to be underestimated.

## Chapter 3 Solution algorithms for VI-constrained hemivariational inequalities

### 3.1 Preliminaries

In a standard constrained optimization problem: $\min _{x \in S} \phi(x)$, with $S$ being a closed convex subset $\mathbb{R}^{n}$ and $\phi$ a continuous function defined on $S^{1}$, it is often assumed that the feasible set $S$ is explicitly defined by a system of finitely many equalities and inequalities. In a hierarchical optimization problem, the set $S$ could be implicitly defined as the solution set of a (lower-level) optimization problem, or more generally, a variational inequality:

$$
\begin{align*}
& \underset{x}{\operatorname{minimize}} \phi(x)  \tag{3.1}\\
& \text { subject to } x \in \operatorname{argmin}\{f(z) \mid z \in K\}
\end{align*}
$$

where $f: K \rightarrow \mathbb{R}$ is the lower-level objective function and $K \subseteq \mathbb{R}^{n}$ is the feasible set of the lower-level optimization problem. This kind of problem is usually referred to as a bilevel program and its practical solution cannot be achieved by standard optimization methods, which invariably require an explicit representation of the feasible set in terms of differentiable equations and inequalities. We consider a generalization of (3.1), which we call Variational Inequality-Constrained HemiVariational Inequality (VI-C HVI). We preliminarily recall that the Hemivariational Inequality problem HVI $(X, \Phi, h)$, where $X$ is a closed convex subset of $\mathbb{R}^{n}, \Phi$ is a continuous functions from $X$ into $\mathbb{R}^{n}$, and $h: X \rightarrow \mathbb{R}$ is a convex function that is not necessarily differentiable, is the problem of finding a vector $x \in X$ such that

$$
\Phi(x)^{T}(y-x)+h(y)-h(x) \geq 0, \quad \forall y \in X
$$

[^0]Hemivariational inequalities (also known as variational inequalities of the second kind) are a powerful modeling tool that encompasses both (convex) optimization, when $\Phi:=0$, and VIs, when $h:=0$, as particular instances. In their full generality, HVIs have been mainly considered in infinite-dimensional settings, see e.g. [69, 72, 77]; nevertheless, finite-dimensional HVIs have recently attracted attention in the mathematical programming literature; see e.g. $[1,57,68]$.


Fig. 3.1: HVI problem: we remark that, in case of continuously differentiable objective functions, the convex optimization set is actually a subset of the VI problem set.

If the set $X$ is implicitly defined as the solution set $\operatorname{SOL}(K, F)$ of a lowerlevel Variational Inequality VI $(K, F)$, with $K \subseteq \mathbb{R}^{n}$ closed and convex and $F: K \rightarrow \mathbb{R}^{n}$ continuous, the HVI $(\operatorname{SOL}(K, F), \Phi, h)$ becomes the VI-C HVI defined by the tuple ( $K, F, \Phi, h$ ), which is the problem of finding a vector $x \in \operatorname{SOL}(K, F)$ such that

$$
\Phi(x)^{T}(y-x)+h(y)-h(x) \geq 0, \quad \forall y \in \operatorname{SOL}(K, F)
$$

where

$$
\operatorname{SOL}(K, F) \triangleq\left\{z \in K \mid F(z)^{T}(y-z) \geq 0, \forall y \in K\right\}
$$

Denoting the same problem, the notations: HVI (SOL $(K, F), \Phi, h)$ and VI-C HVI ( $K, F, \Phi, h$ ), will be used interchangeably. A particularly interesting case that often arises in applications is the VI-Constrained Variational Inequality, which is the problem HVI $(\operatorname{SOL}(K, F), \Phi, 0)$, i.e. VI-C HVI $(K, F, \Phi, 0)$, which we also write as VI-C VI $(K, F, \Phi)$. As far as we know, the HVI where the feasible set is defined implicitly as the solution set of a monotone VI is a novel problem.

Our interest in studying the VI-C HVI is several-fold. First, with $F$ being the gradient of the scalar-valued function $f$, it is well known that $\operatorname{SOL}(K, F)$ is the set of stationary solutions of the lower-level optimization problem in the constraint of (3.1); thus, assuming $\phi$ to be convex, the VI-C HVI ( $K, \nabla f, 0, \phi$ )
is intimately related to the bilevel program (3.1); in fact, the two problems are equivalent if the lower-level objective function $f$ is convex.


Fig. 3.2: VI-C HVI problem.

Second, with $F$ being a monotone map and $\Phi \triangleq \nabla \psi$ being the gradient of the (possibly nonconvex) scalar-valued function $\psi$, the VI-C HVI $(K, F, \nabla \psi, 0)$ is the first-order variational formulation for the optimization problem with VI constraints:

$$
\begin{equation*}
\underset{x \in \operatorname{SOL}(K, F)}{\operatorname{minimize}} \psi(x), \tag{3.2}
\end{equation*}
$$

which in turn is a special case of a mathematical program with equilibrium constraints (MPEC) [66] wherein there is no (upper-level) "design variable". Furthermore, since the VI is known [32] to provide a broad mathematical framework for a host of economic equilibrium and game-theoretic problems, (3.2) is a natural formulation for the problem of selecting a particular equilibrium solution to optimize an auxiliary objective function $\psi$.

By using a penalization approach and taking $h \triangleq \rho \operatorname{dist}_{\Omega}(\bullet)$ to be a (sufficiently large) positive multiple $\rho$ of the distance function (in the Euclidean norm) to the closed convex set $\Omega \subseteq \mathbb{R}^{n}$, the VI-C HVI ( $K, F, \nabla \psi, h$ ) will allow us to treat, for example, an extended version of (3.2) wherein the variable $x$, in addition to being a solution of the VI $(K, F)$, is required to belong to the side feasible set $\Omega$, i.e., the problem

$$
\begin{equation*}
\underset{x \in \Omega \cap \mathrm{SOL}(K, F)}{\operatorname{minimize}} \psi(x) \tag{3.3}
\end{equation*}
$$

It should be noted that

$$
\operatorname{minimum}_{x \in \Omega \cap \operatorname{SOL}(K, F)} \psi(x) \geq \operatorname{minimum}_{x \in \operatorname{SOL}(K \cap \Omega, F)} \psi(x)
$$

Since equality does not necessarily hold, the two problems (3.2) and (3.3) are in general not the same.

The main contributions of the next sections can be summarized as follows:

- we establish an exact penalization result that reduces a HVI with VI constraints and side constraints to a VI-C HVI (without side constraints);
- we present a centralized solution method for solving the VI-C HVI and establish its convergence;
- we present a distributed algorithm for solving a "partitioned" VI-C HVI, i.e., the case where the pair $(K, h)$ has a certain partitioned structure;
- we present an iterative descent framework for computing a stationary point of a VI constrained minimization problem, whose objective function is not necessarily convex;
- we apply the developed framework to solve a new power control problem in ad-hoc networks.

To the best of our knowledge, these contributions are new and expand considerably existing results. The proposed distributed algorithm, in which we are interested as motivated by applications in non-cooperative game problems, see [34, Chapter 12], is novel even for a hierarchical optimization problem. Furthermore, the power control problem analyzed in the sequel is new and our results expand considerably the applicability and flexibility of gametheoretic models in ad-hoc networks and also bring considerable gains over existing techniques.

This chapter is organized as follows. In the next section we discuss existing results in the literature. In Section 3.3 we show how to penalize side constraints and reduce an HVI problem with side constraints to one without side constraints. In Section 3.4 we propose the centralized algorithm, considering both the exact and approximate versions, while in Section 3.5 we present the results of the numerical experiments conducted on the centralized version of the algorithm. Section 3.6 describes a broad decomposition scheme for a partitioned HVI that enables the development of distributed versions of the algorithm described in Section 3.4. Section 3.8 deals specifically with VI-constrained minimization problems and shows how it is possible to use the results developed in the previous sections for computing stationary solutions to nonconvex VI constrained minimization problems. Finally, Section 3.9 introduces a new power control problem in ad-hoc networks which gives the original motivation for our interest in VI-C HVIs. The experimental results reported amply demonstrate the significance of the algorithms developed in this work.

### 3.2 Background Results

For the monotonicity definition we refer to Section 1.1. Here we need an extension of the definition of monotonicity (plus) to point-to-set mappings. Specifically, a set-valued map $\mathcal{G}: X \rightarrow \mathbb{R}^{n}$ is monotone on $X$ if for all $x$
and $y$ in $X$ and $\zeta \in \mathcal{G}(x)$ and $\xi \in \mathcal{G}(y)$ it holds that $(\xi-\zeta)^{T}(y-x) \geq 0$. If in addition $(\xi-\zeta)^{T}(y-x)=0$ implies $\zeta \in \mathcal{G}(y)$ and $\xi \in \mathcal{G}(x)$, then $\mathcal{G}$ is termed monotone plus. It is well-known that the subdifferential of a convex function is a monotone plus set-valued map [55]. Subsequently, in Lemma 4, we present a new class of monotone plus functions that seemingly is fairly natural and yet we have not seen this result in the literature. Along with the survey given below, this lemma justifies the blanket assumption (D) that we impose on the VI-C HVI $(K, F, \Phi, h)$ :
(A) $K$ is a closed and convex set in $\mathbb{R}^{n}$,
(B) $F: K \rightarrow \mathbb{R}^{n}$ is a continuous monotone map,
(C) $\operatorname{SOL}(K, F) \neq \emptyset$.
(D) $\Phi: K \rightarrow \mathbb{R}^{n}$ is continuous and monotone plus, and
(E) $h: K \rightarrow \mathbb{R}^{n} \rightarrow \mathbb{R}$ is convex and continuous.

Under assumptions $(\mathrm{A}-\mathrm{C}), \operatorname{SOL}(K, F)$ is a nonempty closed convex set. To motivate our analysis, it is useful to view HVIs and, in particular, the VI-C HVI $(K, F, \Phi, h)$ as a particular case of the Generalized Variational Inequality (GVI) [15]. Defined by the pair $(X, T)$, where $X$ is a closed convex set in $\mathbb{R}^{n}$ and $T$ is a set-valued map defined on $X$ with images $T(x)$ being closed sets in $\mathbb{R}^{n}$, is to find a vector $x \in X$ and a vector $y \in T(x)$ such that $\left(x^{\prime}-x\right)^{T} y \geq 0$ for all $x^{\prime} \in X$. It is not difficult to see that the $\operatorname{HVI}(X, \Phi, h)$ is equivalent to the GVI $(X, T)$, where $T \triangleq \Phi+\partial h$, i.e., the problem of finding a vector $x \in X$ and a subgradient $\zeta \in \partial h(x)$ such that

$$
\begin{equation*}
(\Phi(x)+\zeta)^{T}(y-x) \geq 0, \quad \forall y \in X \tag{3.4}
\end{equation*}
$$

a fact that we will freely use in subsequent developments. Likewise, the VI-C HVI $(K, F, \Phi, h)$ is equivalent to the GVI $(\operatorname{SOL}(K, F), T)$.

Even if the map $T$ is monotone and the set $\operatorname{SOL}(K, F)$ is closed and convex, the GVI $(\operatorname{SOL}(K, F), T)$, and thus the VI-C HVI $(K, F, \Phi, h)$ is not guaranteed to have a solution. For this to be true, there are two classical sufficient conditions: either $T$ is strongly monotone, or $\operatorname{SOL}(K, F)$ is compact. In this work, we assume neither. Instead, we rely on the well-known Tikhonov process (see Subsection 1.3.3) to regularize the map $T$. For a historical account of the application of the Tikhonov process to VIs, we refer the reader to the Notes and Comments in [32, Section 12.9]; in particular, it was noted there that Browder [13] (and more recently, Tseng [99]) showed that given a single-valued, monotone VI $(S, T)$ one can compute a solution of the VI $(\operatorname{SOL}(S, T), G)$, where $G$ is a strongly monotone operator, by solving a sequence of "regularized" problems of the form VI $\left(S, T+\varepsilon_{k} G\right)$, where $\left\{\varepsilon_{k}\right\}$ is a sequence of positive scalars converging to zero. Each regularized VI $\left(S, T+\varepsilon_{k} G\right)$ has a unique solution $x\left(\varepsilon_{k}\right)$ and as $\varepsilon_{k}$ goes to zero, $\left\{x\left(\varepsilon_{k}\right)\right\}$ converge to the unique solution of VI $(\operatorname{SOL}(S, T), G)$.

One key feature of the above regularization method is that it requires $G$ to be strongly monotone. In order to weaken this assumption, it has been proposed to combine Tikhonov's regularization with the proximal-point method
as we have seen in Subsection 1.3.3. One line of research in this direction has been carried out in the context of the hierarchical optimization problem (3.1). For this problem, the regularization step at iteration $k$ amounts to solving the following strongly convex optimization problem:

$$
\underset{x \in K}{\operatorname{minimize}} f(x)+\varepsilon_{k} \phi(x)+\frac{\alpha}{2}\left\|x-x\left(\varepsilon_{k-1}\right)\right\|^{2},
$$

where $\alpha>0$ is a positive parameter and $x\left(\varepsilon_{k-1}\right)$ is the (unique) optimal solution of the above optimization problem at the previous iteration $k-1$. Thanks to the presence of the strongly convex term $\alpha\left\|x-x\left(\varepsilon_{k-1}\right)\right\|^{2}$, it is not necessary to require that $\phi$ be strongly convex. Studies related to this approach have been carried out mainly in the the French mathematical school. Note that actually, in many of these studies, the main focus was not on the solution of hierarchical problems. Rather, the authors were investigating the possibility of obtaining strong convergence to the least norm solution of an infinite dimensional monotone problem by using a combination of Tikhonov and proximal regularization, so as to couple in a single method the good sides of the two approaches. The more advanced results in analyzing this approach in relation to hierarchical problems are probably those obtained by Cabot [14], whose proof techniques have influenced our approach in Section 3.4. Essentially the idea is to show that if $\varepsilon_{k}$ goes to zero "slowly", then $x\left(\varepsilon_{k}\right)$ converges to a solution of (3.1). Extension of this approach to more general problems have been attempted recently, see [53, 70]. However, in order to establish convergence results, in these latter papers, assumptions are made on the behavior of the algorithm, rather than on the defining functions of the problem.

Among the assumptions (A)-(E), the monotone plus property of $\Phi$ requires an explanation. This property can be traced to the existing literature where many authors have considered solving VIs of the type $(\operatorname{Fix}(U), I-V)$, where $\operatorname{Fix}(U)$ denotes the set of fixed points of a nonexpansive map U and $V$ is another nonexpansive map. It turns out that the map $I-V$ is monotone plus if $V$ is nonexpansive.

Lemma 4. If $V: \mathbb{R}^{n} \rightarrow \mathbb{R}^{n}$ is nonexpansive, then $I-V$ is monotone plus.
Proof. For any two vectors $u$ and $v$, we have

$$
(u-v)^{T}[u-v-(V(u)-V(v))] \geq 0
$$

Moreover, if equality holds, then we have

$$
(u-v)^{T}(u-v)=(u-v)^{T}(V(u)-V(v))
$$

which implies

$$
0 \leq\|V(u)-V(v)-(u-v)\|^{2}=\|V(u)-V(v)\|^{2}-\|u-v\|^{2} \leq 0
$$

Thus equality holds throughout and we deduce $u-V(u)=v-V(v)$, which is the plus property of the map $I-V$.

The line of research on the VI $(\operatorname{Fix}(U), I-V)$ was initiated by [104] and we cite the interesting papers $[63,67]$ for recent results and bibliographic references. While we (see 6) can always write the solution set of the VI ( $K, F$ ) as the set of fixed points of the natural map: $F_{K}^{\text {nat }}(x) \triangleq \operatorname{Proj}_{K}(x-\tau F(x))$, where $\operatorname{Proj}_{K}$ denotes the (Euclidean) projection on $K$ and $\tau$ is a positive constant, if $F$ is simply monotone one can not guarantee that $U \triangleq F_{K}^{\text {nat }}$ be nonexpansive. For example, it suffices to consider the univariate VI ( $\mathbb{R}, e^{x}$ ) whose natural map $F_{K}^{\text {nat }}=x-\tau e^{x}$ cannot be nonexpansive. In general, probably the weakest condition that guarantees $F_{K}^{\text {nat }}(x)$ to be nonexpansive (for $\tau$ sufficiently small) is that $F$ be co-coercive; see [32, Lemma 12.1.7]. Similarly, while any map of the type $I-V$, with $V$ nonexpansive, is monotone plus, as shown by Lemma 4, clearly not all monotone plus maps can be put into this form. For example take $G(x)=e^{x}$ which is strictly monotone; if we write $x-V(x)=e^{x}$ we have $V(x)=x-e^{x}$ which is not nonexpansive. Therefore, when applied to VI-C VIs, the setting originally introduced by Yamada imposes rather strong limitations. Nevertheless, the algorithms obtained by considering this fixed-point structure are rather interesting and simple to implement. It is worth pointing out that, if $U$ and $V$ are nonexpansive, by taking $K=\mathbb{R}^{n}$ and $F=I-U$, then the VI $(\operatorname{Fix}(U), I-V)$ becomes the VI ( $\operatorname{SOL}(K, F), I-V)$ with $F$ and $I-V$ monotone. The upshot of this review is that even for the VI-C VI, let alone the HVI, our setting extends the ones in the existing literature. By considering the HVI, we are able to deal with the VI-C VI with additional side constraints on the lower-level VI, a feature that is not included in any of the cited papers on this topic. Indeed, in the next section, we will prove to be able to cope with side constraints via a penalty technique.

In addition to the main algorithm, another contribution of this work is the introduction of a distributed solution method for solving the VI-C HVI with partitioned structure. This is a novelty by itself as there is so far no such algorithms even for structured hierarchical optimization problems. Our main convergence result for such a distributed algorithm generalizes those in [34] for non-cooperative Nash games. Besides being applicable to the HVI and not only to games, the main departure of the convergence result derived in the present work from the ones in [34] is that twice differentiability is not required.

Historically, distributed algorithms for solving VIs can be traced back to the original work in [78] for partitioned problems; see also [11]. Our interest in algorithms of the distributed type for solving the VI-C HVI is very much motivated by the recent surge of interests in the signal processing area on game-theoretic problems wherein centralized algorithms are not realistic for practical implementation; see the series of papers [53, 64, 80, 81, 93, 92, 105], dealing with game-theoretic formulations of power control problems in ad-
hoc, CDMA, or Cognitive Radio networks. Several distributed algorithms along with their convergence properties have been proposed in these papers. However, all these solution methods have a common drawback, which strongly limits their applicability in practical scenarios: they are guaranteed to converge only if the considered power control Nash games have a unique solution, which is not the case in many applications. In the presence of multiple solutions, the distributed computation of even a single Nash equilibrium becomes a complex and unsolved task. We overcome this limitation and propose a novel distributed algorithm that solves the game-theoretic multi-channel power control problem addressed in [64, 93, 92, 105] even when there are multiple solutions; our main contribution is twofold: i) our algorithm converges under milder sufficient conditions that do not imply the uniqueness of the solution of the game; and ii) in the presence of multiple solutions, we can control the quality of the reached solution by guaranteeing the convergence to the "best" NE, according to some prescribed criterion, while keeping the distributed implementation of the algorithm. The latter feature makes our algorithm very appealing for designing practical telecommunication systems, while algorithms with unpredictable performance (like [64, 93, 92, 105], when multiple solutions are present) are not acceptable since a control on the achievable performance is required. A bi-level optimization approach based on solving a variational inequality problem over the fixed point set of a nonexpansive mapping has been proposed in [53] to solve a scalar power control problem in CDMA data networks; such a problem falls in the class of so-called scalar games, modeling narrower and simpler scenarios than those considered in this work (e.g., in [53], each users is assumed to have a scalar variable to optimize rather than a vector as in this work). Hence, theoretical results in [53] cannot be applied to design multi-channel networks, as considered in this work.

### 3.3 Penalization of HVIs with Side Constraints

This section shows that, provided some weak assumptions are satisfied, the hemivariational inequality HVI $(\operatorname{SOL}(K, F) \cap \Omega, \Phi, h)$ with VI $(K, F)$ and side constraints $\Omega$ is equivalent to the penalized HVI without side constraints: VI-C HVI $\left(\operatorname{SOL}(K, F), \Phi, h+\rho \operatorname{dist}_{\Omega}(\bullet)\right)$ for all suitably large values of the penalty parameter $\rho>0$. This result is of independent interest and extends classical results in optimization. Our specific motivation for studying the penalization of side constraints lies in the development in the following section, where we show that Algorithm 3 successfully solves problems of the form HVI (SOL $(K, F), \Phi, h)$. However, the employed proof technique cannot directly be extended to handle the presence of the side constraint set $\Omega$. The main result in this section, Theorem 14, allows us to apply Algorithm 3 to
the penalized problem HVI $\left(\operatorname{SOL}(K, F), \Phi, h+\rho \operatorname{dist}_{\Omega}(\bullet)\right)$ as a way of solving the HVI $(\operatorname{SOL}(K, F) \cap \Omega, \Phi, h)$.

Since in this section the particular structure of the set $\operatorname{SOL}(K, F)$ will play no role, we derive the desired penalty results in a slightly more general framework than is necessary for our purposes. Specifically, we consider the HVI ( $X \cap \Omega, \Phi, h$ ), where $X$ and $\Omega$ are closed convex sets and $\Phi$ and $h$ satisfy the blanket assumptions (D) and (E). We need three preliminary lemmas that are reminiscent of similar results in optimization. The first lemma says that, in the setting above, the values of $\Phi$ over the solution set of the HVI are a constant. This extends the notable fact valid for pseudomonotone VIs (see Subsection 1.3.2). The noteworthy point here is that, while the sum of the two monotone plus mappings $\Phi$ and $\partial h$ is surely monotone, the plus property does not necessarily hold for the sum. So the following lemma does not readily follow from known results such as the cited corollary.

Lemma 5. Let an HVI $(X, \Phi, h)$ be given, where $X \subseteq \mathbb{R}^{n}$ is a closed convex set, $\Phi: X \rightarrow \mathbb{R}^{n}$ is continuous and monotone plus on $X$ and $h: X \rightarrow \mathbb{R}$ is convex. Let $x$ and $\bar{x}$ be two solutions of the HVI, so that (see (3.4)), for suitable $\zeta \in \partial h(x)$ and $\bar{\zeta} \in \partial h(\bar{x})$,

$$
\begin{equation*}
(\Phi(x)+\zeta)^{T}(z-x) \geq 0 \quad \text { and } \quad(\Phi(\bar{x})+\bar{\zeta})^{T}(w-\bar{x}) \geq 0 \tag{3.5}
\end{equation*}
$$

for all $z$ and $w$ in $X$. Then $\Phi(x)=\Phi(\bar{x})$, i.e. $\Phi$ is constant on the solution set; moreover, $\zeta \in \partial h(\bar{x})$ and $\bar{\zeta} \in \partial h(x)$.

Proof. Summing the two inequalities in (3.5), with $z=\bar{x}$ and $w=x$, we get

$$
[(\Phi(x)+\zeta)-(\Phi(\bar{x})+\bar{\zeta})]^{T}(\bar{x}-x) \geq 0
$$

The monotonicity of $\Phi+\partial h$ shows that the inequality above is actually an equality so that, rearranging terms, we get

$$
\begin{equation*}
[\Phi(x)-\Phi(\bar{x})]^{T}(\bar{x}-x)+(\zeta-\bar{\zeta})^{T}(\bar{x}-x)=0 \tag{3.6}
\end{equation*}
$$

Monotonicity of $\Phi$ and $h$ then shows that

$$
[\Phi(x)-\Phi(\bar{x})]^{T}(x-\bar{x})=0=(\zeta-\bar{\zeta})^{T}(x-\bar{x})
$$

and this, in turn, by the plus property of $\Phi$ and $\partial h$, completes the proof.
Consider the HVI ( $X \cap \Omega, \Phi, h$ ). By (3.4), a vector $\bar{x} \in X \cap \Omega$ is a solution of this HVI if and only if there exists $\bar{\zeta} \in \partial h(\bar{x})$ such that

$$
(y-\bar{x})^{T}(\Phi(\bar{x})+\bar{\zeta}) \geq 0, \quad \forall y \in X \cap \Omega
$$

or equivalently, $-(\Phi(\bar{x})+\bar{\zeta}) \in N_{X \cap \Omega}(\bar{x})$, where $N_{S}(x)$ denotes the normal cone of the closed convex set $S$ at $x \in S$. Under the constraint qualification that

$$
\begin{equation*}
N_{X \cap \Omega}(\bar{x})=N_{X}(\bar{x})+N_{\Omega}(\bar{x}) \tag{3.7}
\end{equation*}
$$

it follows that a vector $\bar{x} \in X \cap \Omega$ is a solution of HVI ( $X \cap \Omega, \Phi, h$ ) if and only if there exist $\bar{\zeta} \in \partial h(\bar{x})$ and $\bar{\eta} \in N_{\Omega}(\bar{x})$ such that $-(\Phi(\bar{x})+\bar{\zeta}+\bar{\eta}) \in N_{X}(\bar{x})$, or equivalently, that

$$
\begin{equation*}
(y-\bar{x})^{T}(\Phi(\bar{x})+\bar{\zeta}+\bar{\eta}) \geq 0, \quad \forall y \in X \tag{3.8}
\end{equation*}
$$

Let $\mathcal{E}(\bar{x})$ be the set of such vectors $\bar{\eta}$ associated with the solution $\bar{x}$; i.e., $\bar{\eta} \in \mathcal{E}(\bar{x})$ if and only if $\bar{\eta} \in N_{\Omega}(\bar{x})$ and there exists $\bar{\zeta} \in \partial h(\bar{x})$ such that (3.8) holds. Note that (3.7) provides a sufficient condition for $\mathcal{E}(\bar{x})$ to be nonempty. Equality (3.7) is a rather weak requirement that is implied by the condition ri $X \cap \operatorname{ri} \Omega \neq \emptyset$, where ri $S$ denotes the relative interior of the convex set $S$. The latter relative condition, in turn, can be further relaxed if, as it is often the case in applications, $\Omega$ is polyhedral, in which case (3.7) is implied by $\Omega \cap \operatorname{ri} X \neq \emptyset$.

Assuming that $\Phi$ is monotone plus, the next lemma establishes that the set $\mathcal{E}(\bar{x})$ is independent of the solution $\bar{x}$.

Lemma 6. Let an HVI ( $X \cap \Omega, \Phi, h$ ) be given, where $X$ and $\Omega$ are closed convex subsets of $\mathbb{R}^{n}, \Phi: X \cap \Omega \rightarrow \mathbb{R}^{n}$ is continuous and monotone plus on $X$, and $h: X \cap \Omega \rightarrow \mathbb{R}$ is continuous and convex. Let $\widehat{x}$ be a solution of this HVI such that $\mathcal{E}(\widehat{x})$ is nonempty. Then, if $\widetilde{x}$ is any other solution of the HVI, $\mathcal{E}(\widetilde{x})$ is nonempty and $\mathcal{E}(\widetilde{x})=\mathcal{E}(\widehat{x})$.

Proof. We first show that $\mathcal{E}(\widetilde{x})$ is nonempty. Suppose by contradiction that $\mathcal{E}(\widetilde{x})=\emptyset$. We know that

$$
\begin{equation*}
(\Phi(\widetilde{x})+\widetilde{\zeta})^{T}(w-\widetilde{x}) \geq 0, \quad \forall w \in X \cap \Omega \tag{3.9}
\end{equation*}
$$

for some $\widetilde{\zeta} \in \partial h(\widetilde{x})$. Let $\widehat{\eta} \in \mathcal{E}(\widehat{x})$ be arbitrary, by definition there exists $\widehat{\zeta} \in \partial h(\widehat{x})$ such that

$$
\begin{equation*}
(v-\widehat{x})^{T}(\Phi(\widehat{x})+\widehat{\zeta}+\widehat{\eta}) \geq 0, \quad \forall v \in X \tag{3.10}
\end{equation*}
$$

Substituting $w=\widehat{x}$ in the former inequality and $v=\widetilde{x}$ in the latter inequality, and adding, we obtain

$$
(\widetilde{x}-\widehat{x})^{T}(\Phi(\widehat{x})-\Phi(\widetilde{x}))+(\widetilde{x}-\widehat{x})^{T}(\widehat{\zeta}-\widetilde{\zeta})+(\widetilde{x}-\widehat{x})^{T} \widehat{\eta} \geq 0
$$

But the left-hand side is also non-positive by the monotonicity of $\Phi$ and $\partial h$ and the fact that $\widetilde{x} \in \Omega, \widehat{x} \in \Omega, \widehat{\eta} \in N_{\Omega}(\widehat{x})$. Hence, the three addends are all zero and, in particular, it holds $(\widetilde{x}-\widehat{x})^{T} \widehat{\eta}=0$, which easily implies that $\widehat{\eta} \in N_{\Omega}(\widetilde{x})$. Moreover, since the sum of the two inequalities (3.9) and (3.10), with $w=\widehat{x}$ in the former and $v=\widetilde{x}$ in the latter, is equal to zero, we have

$$
\begin{equation*}
(\widetilde{x}-\widehat{x})^{T}(\Phi(\widehat{x})+\widehat{\zeta}+\widehat{\eta})=0 \tag{3.11}
\end{equation*}
$$

Since we assumed $\mathcal{E}(\widetilde{x})=\emptyset$, there exists $\widetilde{y} \in X$ such that

$$
\begin{equation*}
(\widetilde{y}-\widetilde{x})^{T}(\Phi(\widehat{x})+\widehat{\zeta}+\widehat{\eta})<0 \tag{3.12}
\end{equation*}
$$

because $\Phi(\widetilde{x})=\Phi(\widehat{x}), \widehat{\zeta} \in \partial h(\widetilde{x})$ by Lemma 5 and we have just established that $\widehat{\eta} \in N_{\Omega}(\widetilde{x})$. We get, by (3.10), (3.11) and (3.12),

$$
0 \leq(\widetilde{y}-\widehat{x})^{T}(\Phi(\widehat{x})+\widehat{\zeta}+\widehat{\eta})=(\widetilde{y}-\widetilde{x}+\widetilde{x}-\widehat{x})^{T}(\Phi(\widehat{x})+\widehat{\zeta}+\widehat{\eta})<0
$$

which is impossible so that $\mathcal{E}(\widetilde{x})$ is nonempty.
Let now $\widehat{\eta} \in \mathcal{E}(\widehat{x})$ and $\widetilde{\eta} \in \mathcal{E}(\widetilde{x})$ be arbitrary. There exist $\widehat{\zeta} \in \partial h(\widehat{x})$ and $\widetilde{\zeta} \in \partial h(\widetilde{x})$ such that for all $y \in X$,

$$
\begin{equation*}
(y-\widehat{x})^{T}(\Phi(\widehat{x})+\widehat{\zeta}+\widehat{\eta}) \geq 0, \quad \text { and } \quad(y-\widetilde{x})^{T}(\Phi(\widetilde{x})+\widetilde{\zeta}+\widetilde{\eta}) \geq 0 \tag{3.13}
\end{equation*}
$$

Substituting $y=\widehat{x}$ in the former inequality and $y=\widetilde{x}$ in the latter inequality, and adding, we obtain

$$
(\widetilde{x}-\widehat{x})^{T}[(\Phi(\widehat{x})-\Phi(\widetilde{x}))+(\widehat{\zeta}-\widetilde{\zeta})]+(\widetilde{x}-\widehat{x})^{T} \widehat{\eta}+(\widehat{x}-\widetilde{x})^{T} \widetilde{\eta} \geq 0
$$

but the left-hand side is also non-positive because all its addends are nonpositive by the monotonicity of $\Phi$ and $\partial h$ and the fact that $\widetilde{x} \in \Omega, \widehat{x} \in \Omega$, $\widehat{\eta} \in N_{\Omega}(\widehat{x})$, and $\widetilde{\eta} \in N_{\Omega}(\widetilde{x})$. Hence, all the addends above are zero and, in particular, it holds $(\widetilde{x}-\widehat{x})^{T} \widehat{\eta}=0=(\widehat{x}-\widetilde{x})^{T} \widetilde{\eta}$, which easily implies $\widetilde{\eta} \in N_{\Omega}(\widehat{x})$ and $\widehat{\eta} \in N_{\Omega}(\widetilde{x})$.

Moreover, since the sum of the two inequalities in (3.13), with $y=\widetilde{x}$ in the former and $y=\widehat{x}$ the latter, is equal to zero, we have

$$
(\widetilde{x}-\widehat{x})^{T}(\Phi(\widehat{x})+\widehat{\zeta}+\widehat{\eta})=0=(\widehat{x}-\widetilde{x})^{T}(\Phi(\widetilde{x})+\widetilde{\zeta}+\widetilde{\eta})
$$

Therefore, recalling that by Lemma $5 \Phi(\widetilde{x})=\Phi(\widehat{x}), \widetilde{\zeta} \in \partial h(\widehat{x})$, and $\widehat{\zeta} \in \partial h(\widetilde{x})$, we have for any $y \in X$,

$$
\begin{aligned}
(y & -\widetilde{x})^{T}(\Phi(\widetilde{x})+\widehat{\zeta}+\widehat{\eta}) \\
& =(y-\widehat{x})^{T}(\Phi(\widehat{x})+\widehat{\zeta}+\widehat{\eta})+(\widehat{x}-\widetilde{x})^{T}(\Phi(\widehat{x})+\widehat{\zeta}+\widehat{\eta}) \\
& =(y-\widehat{x})^{T}(\Phi(\widehat{x})+\widehat{\zeta}+\widehat{\eta}) \geq 0
\end{aligned}
$$

Thus $\widehat{\eta} \in \mathcal{E}(\widetilde{x})$. Similarly, we also have

$$
(y-\widehat{x})^{T}(\Phi(\widehat{x})+\widetilde{\zeta}+\widetilde{\eta}) \geq 0
$$

which implies $\widetilde{\eta} \in \mathcal{E}(\widehat{x})$. This is enough to show that $\mathcal{E}(\widehat{x})=\mathcal{E}(\widetilde{x})$.
Remark 4. Lemma 6 is reminiscent of the fact that for a convex program with non-unique optimal solutions, the set of optimal Lagrange multipliers does
not depend on the optimal solutions; see e.g. the remark on page 354 in [10].

The next lemma deals with the minimization problem:

$$
\begin{equation*}
\underset{x \in \Omega \cap X}{\operatorname{minimize}} f(x) \tag{3.14}
\end{equation*}
$$

where $f: X \cap \Omega \rightarrow \mathbb{R}$ is a continuously differentiable convex function and $X$ and $\Omega$ are closed convex sets in $\mathbb{R}^{n}$.

The following result is strongly related to Lemma 1 . We remark that, in that case, set $\Omega$ is explicitly defined by inequalities while Lemma 7 copes with an implicitly defined $\Omega$ set.

Lemma 7. Let $\widehat{x}$ be an optimal solution of (3.14) for which there exist $\lambda \geq 0$ and $\xi \in \partial \operatorname{dist}_{\Omega}(\widehat{x})$ such that

$$
\begin{equation*}
(\nabla f(\widehat{x})+\lambda \xi)^{T}(y-\widehat{x}) \geq 0, \quad \forall y \in X \tag{3.15}
\end{equation*}
$$

Then, for every $\rho>\lambda$ the optimal solution sets of (3.14) and the penalized problem

$$
\begin{equation*}
\underset{x \in X}{\operatorname{minimize}}\left[f(x)+\rho \operatorname{dist}_{\Omega}(x)\right] \tag{3.16}
\end{equation*}
$$

coincide.
Proof. It follows from the inequality (3.15) that $\widehat{x}$ is a minimizer on $X$ of the convex function $f+\lambda \operatorname{dist}_{\Omega}(\bullet)$. Let $\bar{x}$ be a solution of (3.14) and let $\rho>\lambda$. We have, for any $y \in X$,

$$
\begin{aligned}
f(y)+\rho \operatorname{dist}_{\Omega}(y) & \geq f(y)+\lambda \operatorname{dist}_{\Omega}(y) \\
& \geq f(\widehat{x})+\lambda \operatorname{dist}_{\Omega}(\widehat{x}) \\
& =f(\widehat{x})=f(\bar{x})+\rho \operatorname{dist}_{\Omega}(\bar{x})
\end{aligned}
$$

thus showing that every solution of (3.14) is also a solution of (3.16).
Conversely, suppose that $\bar{x}$ is a solution of (3.16). If $\bar{x}$ belongs to $\Omega$, it is obvious that $\bar{x}$ is a also a solution of (3.14). Assume then that $\bar{x} \notin \Omega$, or equivalently, $\operatorname{dist}_{\Omega}(\bar{x})>0$. We can write

$$
\begin{aligned}
f(\bar{x})+\rho \operatorname{dist}_{\Omega}(\bar{x}) & >f(\bar{x})+\lambda \operatorname{dist}_{\Omega}(\bar{x}) \\
& \geq f(\widehat{x})+\lambda \operatorname{dist}_{\Omega}(\widehat{x})=f(\widehat{x})+\rho \operatorname{dist}_{\Omega}(\widehat{x})
\end{aligned}
$$

where the second inequality follows from the fact that $\widehat{x}$ is a global minimizer of $f+\lambda \operatorname{dist}_{\Omega}(\widehat{x})$ on $X$ and the third from $\operatorname{dist}_{\Omega}(\widehat{x})=0$. But this chain of inequalities contradicts the optimality of $\bar{x}$. Thus $\bar{x} \in \Omega$ and so $\bar{x}$ is also a solution of (3.14).

The following is the main result of this section; it shows that, under mild assumptions, a VI-C VI with side constraints can be converted to a HVI.

Theorem 14. Let $\widehat{x}$ be a solution of the HVI $(X \cap \Omega, \Phi, h)$ with $\mathcal{E}(\widehat{x}) \neq \emptyset$, where $\Phi$ is continuous and monotone plus, $h$ is continuous and convex on the feasible set $X \cap \Omega$, and $X$ and $\Omega$ are both closed and convex. Let $\eta \in \mathcal{E}(\widehat{x})$ be arbitrary. Then for every $\rho>\|\eta\|$, the solution set of the HVI $(X \cap \Omega, \Phi, h)$ and that of the penalized HVI $\left(X, \Phi, h+\rho \operatorname{dist}_{\Omega}(\bullet)\right)$ coincide.

Proof. Let $\bar{x}$ be any solution of HVI ( $X \cap \Omega, \Phi, h$ ). By Lemma 6, it follows that $\eta \in \mathcal{E}(\bar{x})$. Let $\lambda \triangleq\|\eta\|$ and

$$
\bar{\eta} \triangleq \begin{cases}\text { any element in } \partial \operatorname{dist}_{\Omega}(\bar{x}) & \text { if } \eta=0 \\ \frac{\eta}{\|\eta\|} & \text { if } \eta \neq 0\end{cases}
$$

Since $\partial \operatorname{dist}_{\Omega}(\bar{x})=\mathbb{B} \cap N_{\Omega}(\bar{x})$, where $\mathbb{B}$ is the Euclidean unit ball, it follows that $\bar{\eta} \in \partial \operatorname{dist}_{\Omega}(\bar{x})$. Moreover, by definition of the elements of the set $\mathcal{E}(\bar{x})$, $\bar{\zeta} \in \partial h(\bar{x})$ exists such that

$$
\bar{x} \in \underset{y \in X}{\operatorname{argmin}}[\Phi(\bar{x})+\bar{\zeta}+\lambda \bar{\eta}]^{T}(y-\bar{x})
$$

Lemma 7 tells us that $\bar{x}$ is also a minimizer of

$$
\begin{equation*}
\underset{y \in X}{\operatorname{minimize}}\left[(\Phi(\bar{x})+\bar{\zeta})^{T}(y-\bar{x})+\rho \operatorname{dist}_{\Omega}(y)\right] \tag{3.17}
\end{equation*}
$$

if $\rho>\lambda$. Applying the minimum principle we immediately get that $\bar{x}$ is a solution of the penalized HVI $\left(X, \Phi, h+\rho \operatorname{dist}_{\Omega}(\bullet)\right)$.

Conversely, suppose that $\bar{x}$ is a solution of HVI $\left(X, \Phi, h+\rho \operatorname{dist}_{\Omega}(\bullet)\right)$ for some $\rho>\lambda$. We only need to show that $\bar{x} \in \Omega$. Suppose the contrary, so that $\operatorname{dist}_{\Omega}(\bar{x})>0$. We note that by the first part of the theorem we know that also $\widehat{x}$ is a solution of the same HVI, so that by Lemma 5 we have $\Phi(\widehat{x})=\Phi(\bar{x})$. Let now $\theta$ be a number in $(\lambda, \rho)$. Since by the first part of the theorem $\widehat{x}$ is still a solution of HVI $\left(X, \Phi, h+\theta \operatorname{dist}_{\Omega}(\bullet)\right)$ we can write, noting that $\bar{x}$ belongs to $X$ and recalling that $\operatorname{dist}_{\Omega}(\widehat{x})=0$,

$$
\Phi(\widehat{x})^{T}(\bar{x}-\widehat{x})+h(\bar{x})-h(\widehat{x})+\theta \operatorname{dist}_{\Omega}(\bar{x}) \geq 0
$$

By $\Phi(\widehat{x})=\Phi(\bar{x})$ this immediately gives, recalling again that $\operatorname{dist}_{\Omega}(\widehat{x})=0$,

$$
\begin{align*}
& \Phi(\bar{x})^{T}(\widehat{x}-\bar{x})+h(\widehat{x})-h(\bar{x})+\rho \operatorname{dist}_{\Omega}(\widehat{x}) \\
& \quad \leq \theta \operatorname{dist}_{\Omega}(\bar{x}) \\
& \quad<\rho \operatorname{dist}_{\Omega}(\bar{x})  \tag{3.18}\\
& \quad \leq \Phi(\bar{x})^{T}(\bar{x}-\bar{x})+h(\bar{x})-h(\bar{x})+\rho \operatorname{dist}_{\Omega}(\bar{x})
\end{align*}
$$

But this contradicts the fact that $\bar{x}$ solves the HVI $\left(X, \Phi, h+\rho \operatorname{dist}_{\Omega}(\bullet)\right)$ which, in turn, implies that $\bar{x}$ minimizes $\Phi(\bar{x})^{T}(y-\bar{x})+h(y)-h(\bar{x})+\rho \operatorname{dist}_{\Omega}(y)$ on $X$.

Remark 5. In the above developments, the norm used was the Euclidean norm. Theorem 14 remains valid if the distance function is in terms of any other norm; it suffices to note that for any norm $\|\bullet\|$ with $\|\bullet\|_{D}$ as its dual norm, $\partial \operatorname{dist}_{\Omega}(x)=N_{\Omega}(x) \cap \mathbb{B}_{D}$, where $\mathbb{B}_{D}$ is the unit ball in the norm $\|\bullet\|_{D}$. Then it is immediate to check that Theorem 14 still holds for $\rho>\|\eta\|_{D}$.

### 3.4 The Main Algorithm: Centralized Version

This section presents a centralized iterative algorithm for the solution of the VI-C HVI $(K, F, \Phi, h)$. Before doing so, we should describe alternative ways to solve this problem by a centralized algorithm that depends on available representations of the set $\operatorname{SOL}(K, F)$. Foremost is the affine case where $K$ is a polyhedron and $F(x)=q+M x$ is an affine mapping with a positive semidefinite (albeit not necessarily symmetric) matrix $M$. In this case, provided that a solution of the affine VI $(K, q, M)$ is known, $\operatorname{SOL}(K, q, M)$ has the polyhedral representation given by [32, Expression 2.5.11]. While the solution of the (A)VI-C HVI ( $K, F, \Phi, h$ ) by a centralized algorithm can therefore be accomplished by any known approach for a linearly constrained monotone HVI, it is not immediately clear that any such algorithm, which is based on the polyhedral representation of the $\operatorname{SOL}(K, q, M)$, would admit a distributed implementation when the problem $(K, F, \Phi, h)$ has the requisite Cartesian structure; see Section 3.6. Another case where $\operatorname{SOL}(K, F)$ has an explicit representation that could be exploited by a centralized algorithm is when $F(\operatorname{SOL}(K, F))$ is a singleton (such as when $F$ is pseudomonotone plus on $K)$. However, the representation [32, Proposition 2.3.12] of $\operatorname{SOL}(K, F)$ in this case is not so easy to take advantage of, unless more details are available on the pair $(K, F)$. Lastly, even though $\operatorname{SOL}(K, F)$ is a closed convex set for a convex-monotone pair $(K, F)$, its representation in general, cf. [32, Expression 2.3.2], is in terms of a semi-infinite system of linear inequalities that is not helpful for practical calculations. While the algorithm that we describe below does not depend on any special representation of the pair $(K, F)$, it requires solving sub-HVIs each defined on the set $K$ and by the mapping $F+\varepsilon_{k}(\Phi+\partial h)+\alpha\left(\bullet-x_{k}\right)$ for positive scalars $\varepsilon_{k}$ and $\alpha$.

Specifically, given the current iteration $x_{k} \in K$, the core step of the algorithm consists of calculating an approximate solution of the

$$
\begin{equation*}
\operatorname{HVI}\left(K, F^{k}, \varepsilon_{k} h\right), \quad \text { where } F^{k} \triangleq F+\varepsilon_{k} \Phi+\alpha\left(\bullet-x_{k}\right) \tag{3.19}
\end{equation*}
$$

We denote by $\bar{x}_{k+1}$ the exact solution of this HVI, which is equivalent to the $\operatorname{GVI}\left(K, F^{k}+\varepsilon_{k} \partial h\right)$; i.e., $\bar{x}_{k+1} \in K$ is such that for some $\bar{\zeta}_{k+1} \in \partial h\left(\bar{x}_{k+1}\right)$ and all $y \in K$,

$$
\begin{equation*}
\left[F\left(\bar{x}_{k+1}\right)+\varepsilon_{k}\left(\Phi\left(\bar{x}_{k+1}\right)+\bar{\zeta}_{k+1}\right)+\alpha\left(\bar{x}_{k+1}-x_{k}\right)\right]^{T}\left(y-\bar{x}_{k+1}\right) \geq 0 \tag{3.20}
\end{equation*}
$$

Since $F^{k}+\varepsilon_{k} \partial h$ is a strongly monotone set-valued mapping because of (B), (D), and (E), the HVI (3.19) has one and only one solution, so that $\bar{x}_{k+1}$ is well defined.

```
Algorithm 3: Inexact Prox-Tikhonov Algorithm
(S.0) : Let \(\left\{e_{k}\right\}\) be a sequence of nonnegative scalars, and \(\left\{\varepsilon_{k}\right\}\)
a sequence of positive scalars, both tending to zero. Let \(\alpha>0\) be
arbitrary. Choose \(x_{0} \in K\) and set \(k=0\).
```

(S.1): If $x_{k}$ is a solution of VI-C HVI $(K, F, \Phi, h)$, STOP.
(S.2) : Find $x_{k+1} \in K$ such that $\left\|x_{k+1}-\bar{x}_{k+1}\right\| \leq e_{k}$, where $\bar{x}_{k+1} \in$ $K$ is an exact solution of (3.19).
(S.3) : Set $k \leftarrow k+1$ and return to (S.1).

Note that if $e_{k}=0$ we solve (3.19) exactly at iteration $k$, while if $e_{k}>0$ we permit inaccurate solution of the same problem. To establish the convergence of the algorithm, we assume, in addition to the blanket setting (A-E), the following condition:
(F) the solution set of the VI-C HVI $(K, F, \Phi, h)$, denoted $\operatorname{SOL}(K, F, \Phi, h)$, is non-empty and bounded; moreover, the set $L$ defined below is bounded:

$$
\begin{aligned}
& L \triangleq\{x \in K \mid \exists y \in \operatorname{SOL}(K, F, \Phi, h) \\
& \left.\quad \text { such that } \Phi(x)^{T}(y-x)+h(y)-h(x)>0\right\} .
\end{aligned}
$$

Since $\operatorname{SOL}(K, F, \Phi, h)$ is the solution set of the monotone $\operatorname{GVI}(\operatorname{SOL}(K, F), \Phi+$ $\partial h)$, it follows that $\operatorname{SOL}(K, F, \Phi, h)$ is a non-empty, compact, convex set. The boundedness requirement of the two sets, $\operatorname{SOL}(K, F, \Phi, h)$ and $L$ is automatically satisfied if $K$ is bounded. The boundedness of the set $L$ is reminiscent of a sufficient condition for the existence of a solution to the HVI; see [32, Exercise 2.9.11]. The following example shows that neither $K$ nor $\operatorname{SOL}(K, F)$ need to be bounded under all the assumptions made herein.

Example 5. Take $K=[-1,+\infty), F(x) \triangleq 0, \Phi(x)=\max (0, x)$ and $h \triangleq 0$. It is clear that $F$ and $\Phi$ are both monotone plus. Obviously, since $F$ is identically zero, $\operatorname{SOL}(K, F)$ coincides with $K$, which is unbounded. It is easy to check that $\operatorname{SOL}(K, F ; \Phi, h)=[-1,0]$ and $L=\emptyset$.

We also need to impose two technical conditions on the sequences of scalars $\left\{e_{k}\right\}$ and $\left\{\varepsilon_{k}\right\}$. Note that these are not assumptions on the problem, but merely conditions we enforce in the algorithmic scheme.

C1 $\sum_{k=0}^{\infty} \varepsilon_{k}=\infty$;
$\mathrm{C} 2\left\{e_{k} / \varepsilon_{k}\right\} \rightarrow 0$.
Condition C 1 states that $\varepsilon_{k}$ cannot go to zero too fast. The reason for this requirement is rather intuitive, for if $\varepsilon_{k}$ becomes "too small, too soon", then the term $\alpha\left(\bullet-x_{k}\right)$ dominates the term $\varepsilon_{k} \Phi$ and the sequence $\left\{x_{k}\right\}$ "collapses" to the sequence generated by the proximal-point algorithm; the role of $\Phi$ then becomes negligible so that we cannot guarantee that we find anything more than a point in $\operatorname{SOL}(K, F)$. By the same token, Condition C2 says that the inexact solution we employ in place of the exact solution of the GVI (3.19) should not deviate too far from the exact solution and that the smaller $\varepsilon_{k}$ is the more precise $x_{k+1}$ should be to $\bar{x}_{k+1}$, so much so as to guarantee $\left\{e_{k} / \varepsilon_{k}\right\} \rightarrow 0$; if this were not so, once again the influence of $\Phi$ would be lost in the solution process.

In order to prove convergence of Algorithm 3 we need a preliminary lemma.
Lemma 8. Let $X \subseteq \mathbb{R}^{n}$ be a closed convex set, $\Phi: X \rightarrow \mathbb{R}^{n}$ a monotone plus and continuous function, and $h: X \rightarrow \mathbb{R}$ a convex function. If a solution $y$ to the HVI $(X, \Phi, h)$ and a point $x \in X$ exist such that

$$
\begin{equation*}
\Phi(x)^{T}(y-x)+h(y)-h(x) \geq 0 \tag{3.21}
\end{equation*}
$$

then $x$ is also a solution to the HVI $(X, \Phi, h)$.
Proof. Since $y$ is a solution we can write

$$
\Phi(y)^{T}(z-y)+h(z)-h(y) \geq 0 \quad \forall z \in X
$$

In particular, by taking $z=x$ we get

$$
\begin{equation*}
\Phi(y)^{T}(x-y)+h(x)-h(y) \geq 0 . \tag{3.22}
\end{equation*}
$$

Summing (3.21) and (3.22) we have

$$
(\Phi(x)-\Phi(y))^{T}(y-x) \geq 0
$$

which implies, by monotonicity,

$$
(\Phi(x)-\Phi(y))^{T}(x-y)=0
$$

By the plus property, we obtain

$$
\Phi(x)=\Phi(y)
$$

Therefore, we can write, for any $z \in X$,

$$
\begin{aligned}
& \Phi(x)^{T}(z-x)+h(z)-h(x) \\
& \quad=\Phi(y)^{T}(z-y)+\Phi(x)^{T}(y-x)+h(z)-h(y)+h(y)-h(x) \geq 0
\end{aligned}
$$

thus showing that $x$ is a solution of the HVI $(X, \Phi, h)$.

Remark 6. If $h:=0$, it is easy to see that Lemma 8 holds if we only require $\Phi$ to be pseudo monotone plus on $X$ instead of monotone plus.


Fig. 3.3: Algorithm 3 scheme

The next theorem is the main result of this section. In this theorem we use a global Lispchitz continuity assumption that, however, is only required if inexactness is allowed in the solution of the sub-HVI in (3.19). This Lipschitz condition is certainly satisfied if, for example, $K$ is bounded and $F$, $\Phi$ are locally Lipschitz. We recall that since $h$ is required to be convex and continuous, it is automatically locally Lipschitz.

Theorem 15. Assume that (A-F), and C1 and C2 hold. Assume further that either $e_{k}=0$ eventually, or $K$ is bounded and $F, \Phi$ and $h$ are Lipschitz on $K$. Then Algorithm 3 is well defined and produces a bounded sequence $\left\{x_{k}\right\}$ such that each of its limit points is a solution of VI-C HVI ( $K, F, \Phi, h$ ).

Proof. Write $S_{\Phi}^{h} \triangleq \operatorname{SOL}(K, F, \Phi, h)$, which is assumed to be bounded. To prove the theorem it is enough to show that the sequence $\left\{\delta_{k}\right\}$ converges to zero, where

$$
\delta_{k} \triangleq \frac{1}{2} \operatorname{dist}_{S_{\Phi}^{h}}\left(x^{k}\right)=\frac{1}{2}\left\|x_{k}-\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)\right\|^{2}
$$

By (3.20), we get, for $y \in K$,

$$
\begin{align*}
& {\left[F\left(\bar{x}_{k+1}\right)+\alpha\left(\bar{x}_{k+1}-x_{k}\right)\right]^{T}\left(y-\bar{x}_{k+1}\right)}  \tag{3.23}\\
& \quad \geq \varepsilon_{k}\left[\left(h\left(\bar{x}_{k+1}\right)-h(y)\right)+\Phi\left(\bar{x}_{k+1}\right)^{T}\left(\bar{x}_{k+1}-y\right)\right]
\end{align*}
$$

We have for every $y \in K$

$$
\begin{align*}
& \alpha\left(x_{k}-x_{k+1}\right)^{T}\left(y-x_{k+1}\right) \\
& \leq \alpha\left[\left(x_{k}-\bar{x}_{k+1}\right)^{T}\left(y-\bar{x}_{k+1}\right)+\left(\bar{x}_{k+1}-x_{k+1}\right)^{T}\left(y-x_{k+1}\right)+\right. \\
& \left.\quad\left(x_{k}-\bar{x}_{k+1}\right)^{T}\left(\bar{x}_{k+1}-x_{k+1}\right)\right] \\
& \leq \alpha\left(\bar{x}_{k+1}-x_{k+1}\right)^{T}\left(y-x_{k+1}\right)+F\left(\bar{x}_{k+1}\right)^{T}\left(y-\bar{x}_{k+1}\right)+ \\
& \alpha\left(x_{k}-\bar{x}_{k+1}\right)^{T}\left(\bar{x}_{k+1}-x_{k+1}\right)+ \\
& \varepsilon_{k}\left[h(y)-h\left(\bar{x}_{k+1}\right)+\Phi\left(\bar{x}_{k+1}\right)^{T}\left(y-\bar{x}_{k+1}\right)\right] \\
& \leq \eta_{k}+F\left(x_{k+1}\right)^{T}\left(y-x_{k+1}\right)+ \\
& \varepsilon_{k}\left[\left(h(y)-h\left(x_{k+1}\right)\right)+\Phi\left(x_{k+1}\right)^{T}\left(y-x_{k+1}\right)\right] \tag{3.24}
\end{align*}
$$

where $\eta_{k}$ is equal to zero if $e_{k}=0$, and is a positive scalar satisfying $\eta_{k} \leq M e_{k}$ for some constant $M>0$ if $e_{k}>0$. More precisely, if we denote by $D$ the diameter of $K$, by $L$ a (common) Lipschitz constant for $F, \Phi$, and $h$ on $K$, by $U$ a (common) upper bound for $\|F(x)\|$ and $\|\Phi(x)\|$ on $K$, and by $\bar{\varepsilon}$ a constant such that $\varepsilon_{k} \leq \bar{\varepsilon}$ for all $k$, we can take

$$
M=2 \alpha D+\bar{\varepsilon}(L+U+D)+U+L D
$$

We can write

$$
\begin{align*}
& \delta_{k+1}-\delta_{k}=\frac{1}{2}\left\|x_{k+1}-\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k+1}\right)\right\|^{2}-\frac{1}{2}\left\|x_{k}-\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)\right\|^{2} \\
& \leq \frac{1}{2}\left\|x_{k+1}-\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)\right\|^{2}-\frac{1}{2}\left\|x_{k}-\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)\right\|^{2} \\
& =-\frac{1}{2}\left\|x_{k+1}-x_{k}\right\|^{2}+\left(x_{k}-x_{k+1}\right)^{T}\left(\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)-x_{k+1}\right) \\
& \leq-\frac{1}{2}\left\|x_{k+1}-x_{k}\right\|^{2}+\frac{1}{\alpha} F\left(x_{k+1}\right)^{T}\left(\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)-x_{k+1}\right)+ \\
& \quad \frac{\varepsilon_{k}}{\alpha}\left[\left(h\left(\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)\right)-h\left(x_{k+1}\right)\right)+\Phi\left(x_{k+1}\right)^{T}\left(\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)-x_{k+1}\right)\right]+\frac{\eta_{k}}{\alpha} \\
& \leq-\frac{1}{2}\left\|x_{k+1}-x_{k}\right\|^{2}+ \\
& \quad \frac{\varepsilon_{k}}{\alpha}\left[\left(h\left(\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)\right)-h\left(x_{k+1}\right)\right)+\Phi\left(x_{k+1}\right)^{T}\left(\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)-x_{k+1}\right)\right] \tag{3.25}
\end{align*} \frac{\eta_{k}}{\alpha}
$$

The first inequality is obvious by the definition of projection; the second is obtained from (3.24) evaluated at $y=\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right) \in K$. The third inequality can be obtained by observing that since $\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right) \in S_{\Phi}^{h} \subseteq \operatorname{SOL}(K, F)$, we have $F\left(\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)\right)^{T}\left(x_{k+1}-\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)\right) \geq 0$, which yields in turn, by the monotonicity of $F, F\left(x_{k+1}\right)^{T}\left(\operatorname{Proj}_{S_{\Phi}^{h}}\left(x_{k}\right)-x_{k+1}\right) \leq 0$. We now consider three cases.

Case 1: Eventually, $\varepsilon_{k} V_{k+1}+\eta_{k} \leq 0$.
In this case the nonnegative sequence $\left\{\delta_{k}\right\}$ is (eventually) non-increasing and therefore convergent. Since $S_{\Phi}^{h}$ is bounded, this implies that also $\left\{x_{k}\right\}$ is bounded. Furthermore, $\left\{\delta_{k+1}-\delta_{k}\right\}$ converges to zero. From (3.25), we have

$$
\delta_{k+1}-\delta_{k} \leq-\frac{1}{2}\left\|x_{k+1}-x_{k}\right\|^{2}
$$

which shows that

$$
\begin{equation*}
\lim _{k \rightarrow \infty}\left\|x_{k+1}-x_{k}\right\|=0 \tag{3.26}
\end{equation*}
$$

Summing over $i$ from $k_{0}$ to $k-1$ we get from (3.25)

$$
\begin{equation*}
\delta_{k}-\delta_{k_{0}} \leq \frac{1}{\alpha} \sum_{i=k_{0}}^{k-1} \varepsilon_{i} V_{i+1}+\frac{1}{\alpha} \sum_{i=k_{0}}^{k-1} \eta_{i} \tag{3.27}
\end{equation*}
$$

Since $\varepsilon_{k-1} V_{k} \leq \varepsilon_{k-1} V_{k}+\eta_{k-1} \leq 0$, all $V_{k}$ are non positive. We show that $\lim \sup V_{k}=0$. Suppose by contradiction this is not the case, so that a positive $\underset{\substack{k \rightarrow \infty \\ \text { constant } \\ V}}{ }$ exists such that $V_{k} \leq-\bar{V}<0$ for all $k$. This implies that the right-hand side in (3.27) goes to $-\infty$. In fact, tacking into account C2, we can write $\eta_{i} \leq M e_{i} \leq c_{i} M \varepsilon_{i}$, for some sequence $\left\{c_{i}\right\}$ of positive constants converging to zero. Assuming, without loss of generality, that all $c_{i}$ are such that $c_{i} M \leq \bar{V} / 2$ we can write, by (3.27),

$$
\delta_{k}-\delta_{k_{0}} \leq \frac{1}{\alpha} \sum_{i=k_{0}}^{k-1} \varepsilon_{i} V_{i+1}+\frac{1}{\alpha} \sum_{i=k_{0}}^{k-1} \eta_{i} \leq-\frac{\bar{V}}{2 \alpha} \sum_{i=k_{0}}^{k-1} \varepsilon_{i}
$$

Since $\left\{\delta_{k}\right\}$ converges, we get a contradiction and have thus proved that $\limsup _{k \rightarrow \infty} V_{k}=0$. Therefore, a subsequence $J$ exists such that

$$
\begin{equation*}
\lim _{\substack{k \in J \\ k \rightarrow \infty}} V_{k}=0 \tag{3.28}
\end{equation*}
$$

Since $\left\{x_{k}\right\}$ is bounded, we may assume, without loss of generality, that

$$
\lim _{\substack{k \in J \\ k \rightarrow \infty}} x_{k}=\widetilde{x}
$$

Note that since $K$ is closed, $\widetilde{x} \in K$. We show that actually $\tilde{x} \in \operatorname{SOL}(K, F)$. If this is not so, there exists a point $y \in K$ such that $F(\tilde{x})^{T}(y-\tilde{x})<0$. By (3.24) we can write

$$
\begin{align*}
& {\left[F\left(x_{k}\right)+\alpha\left(x_{k}-x_{k-1}\right)\right]^{T}\left(y-x_{k}\right)+}  \tag{3.29}\\
& \quad \varepsilon_{k-1}\left[h(y)-h\left(x_{k}\right)+\Phi\left(x_{k}\right)^{T}\left(y-x_{k}\right)\right] \geq-\eta_{k-1}
\end{align*}
$$

By continuity, the definition of $y$, the boundedness of $\left\{x_{k}\right\}$ and (3.26), we have, without loss of generality (after a suitable renumeration),

$$
\lim _{\substack{k \in J \\ k \rightarrow \infty}} F\left(x_{k}\right)^{T}\left(y-x_{k}\right)<0, \quad \lim _{\substack{k \in J \\ k \rightarrow \infty}} \alpha\left(x_{k}-x_{k-1}\right)^{T}\left(y-x_{k}\right)=0
$$

and

$$
\lim _{\substack{k \in J \\ k \rightarrow \infty}} \varepsilon_{k-1}\left[h(y)-h\left(x_{k}\right)+\Phi\left(x_{k}\right)^{T}\left(y-x_{k}\right)\right]=0
$$

Then we get a contradiction to (3.29) since $\left\{\eta_{k}\right\} \downarrow 0$. Therefore, $\tilde{x} \in$ $\operatorname{SOL}(K, F)$. Thanks to (3.26) we have $\lim _{k \in J, k \rightarrow \infty} x_{k-1}=\widetilde{x}$. Therefore, by (3.28) and continuity, we get $h\left(\operatorname{Proj}_{S_{\Phi}^{h}}(\widetilde{x})\right)-h(\widetilde{x})+\Phi(\widetilde{x})^{T}\left(\operatorname{Proj}_{S_{\Phi}^{h}}(\widetilde{x})-\widetilde{x}\right)=0$. By Lemma 8, it follows that $\widetilde{x} \in S_{\Phi}^{h}$. Therefore, we get

$$
\lim _{\substack{k \in J \\ k \rightarrow \infty}} \delta_{k}=0
$$

But since the whole sequence $\left\{\delta_{k}\right\}$ is convergent, this implies that the entire sequence $\left\{\delta_{k}\right\} \downarrow 0$, thus concluding the analysis of Case 1 .

Case 2: The two index sets $J$ and $\bar{J}$ are both infinite, where

$$
J \triangleq\left\{k \mid \varepsilon_{k-1} V_{k}+\eta_{k-1}>0\right\}
$$

and

$$
\bar{J} \triangleq\left\{k \in J \left\lvert\,-\frac{1}{2}\left\|x_{k}-x_{k-1}\right\|^{2}+\frac{\varepsilon_{k-1}}{\alpha} V_{k}+\frac{\eta_{k-1}}{\alpha}>0\right.\right\}
$$

By (3.25), if $k \in \bar{J}$ it might happen that $\delta_{k}>\delta_{k-1}$, while if $k \notin \bar{J}$ then necessarily $\delta_{k} \leq \delta_{k-1}$. Therefore, since $\bar{J}$ is infinite, to prove that $\left\{\delta_{k}\right\}$ goes to zero it is enough to show that the subsequence $\left\{\delta_{k}\right\}_{\bar{J}}$ converges to zero. To this end, first observe that for every $k \in \bar{J}$ it holds that

$$
\begin{equation*}
\varepsilon_{k-1} V_{k}+\eta_{k-1}>\frac{\alpha}{2}\left\|x_{k}-x_{k-1}\right\|^{2} \tag{3.30}
\end{equation*}
$$

The sequence $\left\{x_{k}\right\}_{\bar{J}}$ is bounded. In fact, if $\eta_{k}=0$ eventually (exact solution of the subproblems), then $x_{k}$ belongs to $L$ which is bounded by assumption. Otherwise, we have $\operatorname{dist}_{K}\left(x_{k+1}\right) \leq\left\|x_{k+1}-\bar{x}_{k+1}\right\| \leq e_{k}$, from which the boundedness of $\left\{x_{k}\right\}$ follows from that of $K$. By continuity, also $\left\{V_{k}\right\}_{\bar{J}}$ is bounded. Hence, (3.30) yields

$$
\begin{equation*}
\lim _{\substack{k \in J \\ k \rightarrow \infty}}\left\|x_{k}-x_{k-1}\right\|=0 \tag{3.31}
\end{equation*}
$$

Since $\left\{x_{k}\right\}_{\bar{J}}$ is bounded, it has limit points. Let $\tilde{J} \subseteq \bar{J}$ be a subsequence such that

$$
\lim _{\substack{k \in \tilde{J} \\ k \rightarrow \infty}}=\widetilde{x}
$$

Reasoning exactly as in Case 1, (the only difference is that instead of (3.26) we use (3.31)), we may deduce that $\widetilde{x} \in \operatorname{SOL}(K, F)$. By continuity, recalling that $\tilde{J} \subseteq \bar{J} \subseteq J$ and Condition (C2), which implies $\left\{\eta_{k} / \varepsilon_{k}\right\} \rightarrow 0$, we get $h\left(\operatorname{Proj}_{S_{\Phi}^{h}}(\widetilde{x})\right)-h(\widetilde{x})+\Phi(\widetilde{x})^{T}\left(\operatorname{Proj}_{S_{\Phi}^{h}}(\widetilde{x})-\widetilde{x}\right) \geq 0$. Thus $\widetilde{x} \in S_{\Phi}^{h}$; hence $\lim _{\substack{k \in \tilde{J} \\ k \rightarrow \infty}} \delta_{k}=$ 0 . Since this reasoning can be repeated for every convergent subsequence of $\left\{x_{k}\right\}_{\bar{J}}$, we may conclude that $\lim _{k \in J} \delta_{k}=0$, thus concluding the analysis of this case.

Case 3: The index set $J$ is infinite while $\bar{J}$ is finite.
In this case, the sequence $\left\{\delta_{k}\right\}$ is non-increasing eventually. Therefore, $\left\{\delta_{k}\right\}$ converges, implying that $\left\{x_{k}\right\}$ is bounded, $\left\{\delta_{k+1}-\delta_{k}\right\}$ converges to zero and therefore, by (3.25), also (3.26) holds. At this point, we can proceed exactly as in Case 1 and Case 2 to prove that $\left\{\delta_{k}\right\}$ converges to zero, thus concluding the proof of the theorem.

Remark 7. It should be clear that all the results in Theorem 15 still hold if we change $\alpha$ at each iteration, provided that these $\alpha$ s are bounded away from zero and bounded from above. Furthermore, it is interesting to observe that if we are dealing with VI-C VIs, i.e. if $h:=0$, it can be checked, by going through the proof and taking into account Remark 6 , that Theorem 15 still holds with the monotonicity plus assumption on $\Phi$ replaced by a weaker pseudo monotonicity plus assumption.

Some observations are in order. The first is that Algorithm 3 would not be implementable if we were not able to solve the subproblems at Step 2. Since these subproblems are instances of strongly monotone generalized VIs, a host of methods is available for their solution; the fact that we only have to deal with strongly monotone subproblems is one of the advantages of our approach. Another issue worth mentioning is that if we are implementing the inexact version of Algorithm 3, i.e., if $e_{k}>0$, we need to translate the theoretical criterion $\left\|x_{k+1}-\bar{x}_{k+1}\right\| \leq e_{k}$ at Step 2 into something computable, since the exact solution $\bar{x}_{k+1}$ of the sub-HVI (3.19) is not known. Here, again, the strong monotonicity of the subproblems at Step 2 comes into help. Consider for example and for simplicity the case of a VI-C VI $(h:=0)$. In this case a rather complete theory of "error bounds" is available, see [32, Chapter 6], and can be used. For example, by [32, Proposition 6.3.1] it follows that at each iteration we have, for some constant $C>0$

$$
\left\|x_{k+1}-\bar{x}_{k+1}\right\| \leq \frac{C+1}{\alpha}\left\|x_{k+1}-P_{K}\left(x_{k+1}-F^{k}\left(x_{k+1}\right)\right)\right\|
$$

with the right-hand norm being a practically computable quantity. Therefore, in a practical implementation of Algorithm 3 we could simply stop when
$\left\|x_{k+1}-P_{K}\left(x_{k+1}-F^{k}\left(x_{k+1}\right)\right)\right\| \leq e_{k}$ without using any of the theoretical properties in Theorem 15. Other possibilities are available and a choice should be made taking into account the particular problem at hand.

### 3.5 Numerical experiments

In this section we present the results of the numerical experiments conducted on the centralized (thus employing a centralized scheme for the solution of the inner problem (3.19)) version of Algorithm 3.

### 3.5.1 Examined problems

We tested the centralized Proximal point-Tikhonov Algorithm 3 on the VI-C $\mathrm{VI}(K, F, \Phi)$ problem.

We recall that $K$ must be a closed and convex set, $F$ a continuous monotone map, $\operatorname{SOL}(K, F)$ a non empty set and $\Phi$ a continuous and monotone plus function. In the examples that we have considered, the $K$ set is a compact polyhedron. We remark that the compactness of $K$ guarantees that $\operatorname{SOL}(K, F)$ is not empty. Furthermore, in some cases, the choice of a compact polyhedral $K$ set allows the explicit calculation of the set $\operatorname{SOL}(K, F)$, by simple theoretical considerations (see Appendix B for more details); the knowledge of the set $\operatorname{SOL}(K, F)$ may be useful in checking the accuracy that is achieved by the algorithm. In every problem, function $F$ is monotone but not strictly monotone. This choice is crucial in order to get a set $\operatorname{SOL}(K, F)$ which possibly is not a singleton; indeed, if this were not the case, problem VI-C $\operatorname{VI}(K, F, \Phi)$ obviously would be less interesting. Furthermore, $F$ is not required to be symmetric. Finally, function $\Phi$, not necessarily symmetric, is chosen to be strongly monotone on $K$ : we recall that a Lipschitz continuous, strongly monotone function is monotone plus. Moreover, all the function that we have considered are continuously differentiable. We remark that, by considering both defining functions with a symmetric Jacobian and defining functions without a symmetric Jacobian, the problems examined are instances of hierarchical optimization problems, equilibrium selection problems and general VI-C VIs.

Two main classes of problems are proposed. Problems for which $\operatorname{SOL}(K, F)$ certainly is not a singleton and is known in advance: in this case it is possible to evaluate the "correctness" of the solution found by our method. And problems for which $\operatorname{SOL}(K, F)$ is not easily computable a priori. Each problem is an instance of $\operatorname{HVI}(K, F, \Phi)$ : thus, each problem is "individuated" by the choice for a particular feasible set, a particular lower and a particular upper function. We have identified each feasible set, each lower function $F$ and each
upper function $\Phi$ with an alphanumeric tag. Then, the test problem name is a suitable combination of three tags: for example, problem K1F11 $\Phi 1$ is the $\operatorname{VI}-\mathrm{C} \operatorname{VI}(K 1, F 11, \Phi 1)$.

The complete collection of feasible sets and functions that we have considered, the resulting test problems and further details can be found in Appendix B.

### 3.5.2 The implemented algorithm

We implemented the centralized version of Algorithm 3 in MATLAB (Release 2009 b). Firstly, we have analyzed the algorithm performances by considering different choices for the fundamental parameters. Initially, we have tested the exact version of the scheme: we recall that, in this case, the strongly monotone inner problem

$$
\begin{equation*}
\mathrm{VI}\left(K, F^{k}\right), \quad \text { where } \quad F^{k}:=F+\varepsilon_{k} \Phi+\alpha\left(\bullet-x_{k}\right) \tag{3.32}
\end{equation*}
$$

is solved exactly at each "outer" iteration $k$. We evaluated the number of outer and inner iterations performed by the algorithm and (if possible) the accuracy of the solution found by our method, by varying the values of $\alpha$ and $\varepsilon_{0}$. Our goal is twofold: to study the behavior of the scheme for different values of $\alpha$ and $\varepsilon_{0}$ and to identify a reasonable choice for the values of these two parameters.

Once $\alpha$ and $\varepsilon_{0}$ are "properly" set, the exact version is further examined and the inexact version is tested by considering different updating rules for the parameter $e_{k}$ and for fixed values of $\alpha$ and $\varepsilon_{0}$.

## Choice of the initial point

The starting point $x_{0}$ is obtained by solving the feasibility problem

$$
\min _{x \in K} 0_{n, 1}^{T} x
$$

by means of the MATLAB function linprog, thanks to the fact that $K$ is a polyhedral set.

## Parameters $\varepsilon_{k}$ and $\alpha$

We recall that the scalar $\varepsilon_{k}$ must be properly set in order to guarantee that conditions C 1 and C 2 hold: $\varepsilon_{k}$ could be, for example, chosen so that $\varepsilon_{k}=$ $\varepsilon_{0} /(k+1)$. However, in some experiments, due to the good behavior of the


Fig. 3.4: The centralized algorithmic scheme tested
algorithm, we noticed that convergence is obtained also for faster updating rules which not necessarily are such that condition C 1 still holds (for example, one can consider $\varepsilon_{k}=\varepsilon_{0} / k!$ ). It stands to reason that also the choice of $\varepsilon_{0}$ (see the discussion about the stopping criterion) plays a role: indeed, we would like to prevent the term $\varepsilon_{k} \Phi$ in function $F^{k}$ to vanish too rapidly. In the same line of reasoning of above, the "right" setting of the scalar $\alpha$ is not an easy task: on the one hand, we do not have to let the regularization term $\alpha\left(\bullet-x_{k}\right)$ prevail against term $\varepsilon_{k} \Phi$. On the other hand, a sufficiently strongly monotone function $F^{k}$ is desirable and the parameter $\alpha$ is fundamental for the convergence of the overall scheme. We would also like to recall that the convergence still holds if $\alpha$ is changed at each iteration provided that these $\alpha \mathrm{s}$ remain bounded away from zero and bounded from above.

We notice that, if we consider different values for $\varepsilon_{0}$ and $\alpha$, we get different behaviors of the algorithm: we aim at proving the robustness of the scheme with respect to the variations of these parameters and at properly setting their values. In order to achieve these goals, in the practical implementation, firstly, we have fixed different values for $\varepsilon_{0}$; then, for each choice for $\varepsilon_{0}$, we have tested the exact version of the algorithm with different values of parameter $\alpha$. After some preliminary tests (the results are not reported here), we have chosen the values $\varepsilon_{0}^{1}=0.4, \varepsilon_{0}^{2}=0.8, \varepsilon_{0}^{3}=1.6, \varepsilon_{0}^{4}=3.2$ and $\varepsilon_{0}^{5}=6.4$ : we remark that $\varepsilon_{0}^{j+1}=4 \varepsilon_{0}^{j}$, for $j=1, \ldots, 4$. In a like manner, for each fixed $\varepsilon_{0}^{j}$, we have considered five different values for $\alpha$, namely $\alpha^{1}=0.1, \alpha^{2}=0.3$, $\alpha^{3}=0.9, \alpha^{4}=2.7$ and $\alpha^{5}=8.1$, thus following the rule $\alpha^{j+1}=3 \alpha^{j}$, for $j=1, \ldots, 4$ and with $\alpha^{1}=0.1$. Then, by keeping track of the results obtained
and considering the most sound choices for $\varepsilon_{0}$ and $\alpha$, we have tested once more the exact version on further problems and, finally, the inexact version.

## A stopping criterion

We remark that the choice of a stopping criterion for problems with an implicitly defined feasible set is difficult: in such cases, the definition of suitable and efficiently implementable optimality conditions (whose violation is a measure of the "correctness" of the calculated solution) is problematic. Indeed, we can not easily use the KKT conditions; furthermore, in order to resort to the minimum principle, we have to cope with the calculation of a projection onto an implicitly defined feasible set: this operation is, in general, not easier than the solution of the overall original problem. For these reasons, we propose an heuristic approach for which two different stopping criteria are considered for step (S.1) of Algorithm 3: the algorithm stops whenever both criteria are satisfied.

Firstly, the difference between two successive iterations $\left\|x_{k}-x_{k-1}\right\|$ must be, in Euclidean norm, smaller than $10^{-5} \cdot \sqrt{n}$, where $n$ is the number of variables of the examined problem. Secondly, the value of $\varepsilon_{k}$ is required to be smaller than 0.01 . The rationale of this latter stopping criterion is that we want our algorithm to iterate a "sufficient" number of times in order to get a value of the scalar $\varepsilon_{k}$ as close as possible to zero; indeed, the convergence of the method is guaranteed for values of $\varepsilon_{k}$ that go to zero. Lastly we have fixed a maximum number of outer iterations for the whole scheme equal to 3000: if this amount of iterations is reached, we declare a failure.

We emphasize that, actually, the combined action of $\varepsilon_{k}$ (which goes to $0)$ and $\alpha$ is desired in order to push the convergence of the whole scheme; nevertheless, we have to make sure that the algorithm is converging to the right point. Indeed, we point out that, if $\varepsilon_{k} \approx \varepsilon_{k-1}$, problem $(\mathrm{H}) \mathrm{VI}\left(K, F^{k}\right)$ is very similar to $(\mathrm{H}) \mathrm{VI}\left(K, F^{k-1}\right)$ and $x_{k}$ should be "not too different" from the previous $x_{k-1}$ : the algorithm, eventually, does not make any further progress, even though it is not converging to a solution of the original problem. Then, in order to get our scheme working properly, whenever $\left\|x_{k}-x_{k-1}\right\|<10^{-5} \cdot \sqrt{n}$, we check further if the difference between two successive outer iterations is small "enough" only because the difference $\left|\varepsilon_{k}-\varepsilon_{k-1}\right|$ becomes negligible and $\alpha$ is large. If this is the case, the outer iterations may be close to each other not necessarily because the whole scheme is converging to a solution. In order to try to avoid this situation, in the practical implementation (see Figure 3.5), when, for the first time, we get $\left\|x_{k}-x_{k-1}\right\|<10^{-5} \cdot \sqrt{n}$, we set, for the next iteration, $\varepsilon_{k+1}=\varepsilon_{0}$; thus, we reset the updating process of $\varepsilon_{k}$ by starting, at the same time, with a (hopefully) "more promising" point $x_{k}$. It is obvious that, if $\left\|x_{k+1}-x_{k}\right\|$ is still "small", the solution previously calculated was probably the "right" one. Hence, this approach, on the one hand, allows us to check, in a more reliable way, if our method has found the
solution of the original problem; on the other hand, it let the algorithm "move on" if the scheme gets stuck in a point which is not the solution sought.

## The inner algorithm

It is worth mentioning that any convergent algorithm can be employed for the solution of the strongly monotone problem (3.32) at each step of the outer scheme: for every single choice of a solution method at step S.2, we get a "different" overall algorithm. Although we believe that an appropriate choice of such method is crucial to the numerical success of the whole scheme, it is outside the scope of this study to analyze and compare different algorithms for the solution of the strongly monotone problem (3.32). In this section, we therefore consider a single choice for the inner method, a choice that we think is effective in practice, and then proceed to numerical testings in order to have a feel for the efficiency of our approach. We recall that the inner algorithm must be able to solve, at each outer iteration, thus for fixed values of the parameter $\varepsilon_{k}$, the strongly monotone $(\mathrm{H}) \mathrm{VI}\left(K, F^{k}\right)$. In order to solve the latter problem at step S. 2 of Algorithm 3, we have chosen the Basic Projection Algorithm (BPA) (see Algorithm 1 and the discussion therein). The choice of a projection method is quite natural in order to fully take advantage of the strong monotonicity property of the defining function $F^{k}$. Furthermore, the projection algorithms do not require the use of the derivatives of $F^{k}$ and do not involve any complex computation besides the projection on $K$ : thus, for sets $K$ such that the projection can be easily carried out (for example, the polyhedral $K$ sets that we have chosen for our experiments), projection methods can be applied to the solution of large problems because of their simplicity. On the negative side, the use of no derivative information could prevent these methods from being fast.

We remark that the choice of an inner solution method influences the performances of the whole scheme. Amongst other (namely BPA, PAVS, EgA and HPA) projection methods, we decided for the basic one, due to its "simplicity"; this choice allows us to better control the "impact" of the inner algorithm's behavior on the overall algorithm in order to "purify" as much as possible our analysis from the influence of the particular choice for the inner scheme. Furthermore, the BPA seems to be an effective option thanks to the minor computational cost of the implemented method. Indeed, we recall that, while BPA requires only one projection per iteration, PAVS, although it needs a single projection per iteration, has to cope with the right tuning of the variable step size $\tau_{k}, \mathrm{EgA}$ executes at least two projections per iteration and HPA three (and a line search). We recall that, within a basic projection approach, in order to obtain the convergence of the method, the defining function must be strongly monotone and the knowledge of the strong monotonicity constant and Lipschitz modulus is required in principle. Nevertheless, we note that function $F^{k}$ is strongly monotone by construction,
with constant $\alpha$ known a priori, and continuous; moreover, if the feasible set $K$ is compact, $F^{k}$ is Lipschitz continuous. Hence, once again, BPA seems to match well with the properties of problem (3.32). Then, in order to design a convergent scheme and, thus, a suitable projection step size $\tau$, the only task will be the knowledge of the Lipschitz modulus of $F^{k}$. In the implemented scheme, a preliminary "worst" estimate of the Lipschitz continuity constant of $F^{k}$ is evaluated by considering the following problem

$$
\max _{K}\left(\|J F(x)\|+\varepsilon_{0}\|J \Phi(x)\|+\alpha\right)=\hat{L}_{\varepsilon_{0}, \alpha}
$$

Then the step size $\tau$, in view of Theorem 3, is properly chosen by considering the value of the following expression:

$$
c_{\varepsilon_{0}, \alpha}=\frac{2 \alpha}{\hat{L}_{\varepsilon_{0}, \alpha}^{2}}
$$

Then again, we remark that, when we study the behavior of the whole scheme by varying $\varepsilon_{0}$ and $\alpha$, the following estimate of the Lipschitz continuity constant of $F^{k}$ is instead considered

$$
\hat{\hat{L}}_{\alpha}=\max _{K}\left(\|J F(x)\|+\varepsilon_{0 \max }\|J \Phi(x)\|+\alpha\right)
$$

where $\varepsilon_{0 \text { max }}$ is the greatest parameter $\varepsilon_{0}$ considered. In this case, the step size $\tau$ is set by considering the value of the following scalar:

$$
c_{\tilde{\alpha}}=\frac{2 \tilde{\alpha}}{\hat{\hat{L}}_{\tilde{\alpha}}^{2}}
$$

where $\tilde{\alpha}$ is the value of parameter $\alpha$, between those considered, for which $c_{\alpha}$ is minimum. In order to keep as small as possible the influence of the inner algorithm on the whole scheme, the value of $\tau$ (which certainly plays an important role) is kept constant for every test performed by varying $\alpha$ and $\varepsilon_{0}$.

In order to establish a stopping criterion for the inner algorithm, for every outer iteration $k$, first $\operatorname{Proj}_{K}\left(x_{i}-\tau F^{k}\left(x_{i}\right)\right)$ is calculated by means of the MATLAB function quadprog, thus solving the quadratic problem:

$$
\min _{y \in K} \frac{1}{2}\left\|y-\left[x_{i}-\tau F^{k}\left(x_{i}\right)\right]\right\|^{2}
$$

In view of the equivalent reformulation of $\operatorname{VI}\left(K, F^{k}\right)$ (see (1.11)) with the natural map $\mathbf{F}_{K}^{\mathbf{k}^{n a t}}$, if the difference between the solution of this quadratic problem and $x_{i}$ is smaller than $e_{k} \cdot \sqrt{n}$ (we remark here that, if $e_{k}=10^{-5}$, the inner VI is considered to be solved exactly) then the execution of the inner BPA stops and $x_{k+1}=x_{i}$. Otherwise, the next inner iteration is calculated: $x_{i+1}=\operatorname{Proj}_{K}\left(x_{i}-\tau F^{k}\left(x_{i}\right)\right)$.

The maximum number of inner iterations allowed at the $k$ th step is equal to 1000 .


Fig. 3.5: The centralized algorithmic scheme implemented

### 3.5.3 Numerical results and comments

## Choice of $\varepsilon_{0}$ and $\alpha$

We ran many preliminary experiments in order to set properly the values of parameters $\varepsilon_{0}$ and $\alpha$. For the sake of brevity only few results are reported in Appendix C (see page 128 and the following pages): in fact, the behavior of the scheme is more or less the same for all the experiments performed. We evaluated the number of outer and inner iterations and the exactness of the solution found by our method by considering different values of $\alpha$ and $\varepsilon_{0}$. Initially, we have considered the exact version of the scheme: thus, the strongly monotone inner problem has been solved exactly at each "outer" iteration $k$ (as we have seen in the previous section).

First of all, we underline the expected reduction of the correctness of the solution found by our method and reached by considering larger values for
$\alpha$ : we recall that if we let the regularization term $\alpha\left(\bullet-x_{k}\right)$ prevail against term $\varepsilon_{k} \Phi$, the solution calculated, in some sense, may lack significance with respect to the original problem. Thus, in the same line of reasoning, a "not too small" value of $\varepsilon_{0}$ is desirable. For these reasons, the different values that we have considered for $\varepsilon_{0}$ and $\alpha$ (see Section 3.5.2) seem to be proper.

The results obtained, for fixed values of $\varepsilon_{0}$, although they depend on the particular functions involved and on the initial point $x_{0}$ calculated, showed, in general, the growth of the number of outer iterations by increasing $\alpha$ in the interval considered. Indeed, this is quite a natural consequence of the fact that, by increasing $\alpha$, we do not let the actual outer iteration $x_{k}$ move far away from the previous one $x_{k-1}$. A similar reasoning holds true if the difference between two consecutive values of $\varepsilon_{k}$, i.e. $\left|\varepsilon_{k}-\varepsilon_{k-1}\right|$ becomes too small too soon. We remark here that, in order to get a more complete picture of the behavior of the algorithm, whenever the inner algorithm eventually stops updating $x_{k}$, even though $\varepsilon_{k}$ is not smaller than epsilon_tol, we stop counting the outer iterations: in fact, at that point, the algorithm does not make any further progress.

A clear trend analysis of the total number of inner iterations is a more challenging task. Let $b_{k}$ be the number of inner iterations performed by the BPA at the $k$ th outer iteration. The total number of inner iterations will be

$$
\# \text { inn.it }:=\sum_{k} b_{k} \text {. }
$$

The number of inner iterations is more deeply related to the features of the chosen inner algorithm and to the particular problem considered. Furthermore, the total amount of the inner iterations obviously depends also on the number of outer iterations. We can say, with some cautions, that if $\alpha$ is "sufficiently" larger than $\varepsilon_{0}$, we get less inner iterations per single outer iteration. If, indeed, the regularization term dominates the term $\varepsilon_{k} \Phi$, the solution of the actual inner strongly monotone $\mathrm{VI}\left(K, F^{k}\right)$ should be not too "distant" from the previous outer iteration which is also the actual starting point for the solution algorithm for $\mathrm{VI}\left(K, F^{k}\right)$.

To sum up, we can say that, although the parameter $\alpha$ is crucial for the convergence of the whole scheme, as it "freezes" the progresses of the algorithm, one has to pay attention to the right setting of its value with respect to $\varepsilon_{0}$ : in order to make the algorithm work well, we should let the presence of $\alpha$ become "significant" at the "right time". In other words, an "artificial" convergence (that is, mostly forced by the improper action of $\varepsilon_{k}$ and $\alpha$ ) is not desirable.

Finally, by following these guidelines and considering the results that we have obtained, we have chosen $\varepsilon_{0}$ equal to 6.4 and $\alpha$ equal to 0.3 .

## Numerical results obtained for the centralized exact version

Once we have set suitable values for $\varepsilon_{0}$ and for $\alpha$ (respectively 6.4 and 0.3 ), we have tested the exact version of the algorithm on further test problems of various dimensions; the results are reported in Table 3.1. Note that, with the exception of $\operatorname{K} 1 \mathrm{Fr} \Phi 5$, for this set of test problems we know the exact solution: thus, we could easily compare this exact solution to the computed ones. The last column of Table 3.1 confirms the good quality of the solution calculated by our method.

| Problem | n | \# out.it. | \# inn.it. | error |
| :---: | :---: | :---: | :---: | :---: |
| K1F7 $\Phi 7$ | 500 | 8 | 162 | $2 \mathrm{e}-5$ |
| K1F11 $\Phi 4$ | 200 | 69 | 389 | $6 \mathrm{e}-16$ |
| K1F13 $\Phi 7$ | 1000 | 15 | 542 | $7 \mathrm{e}-5$ |
| K1F13 $\Phi 6$ | 100 | 226 | 18955 | $7 \mathrm{e}-17$ |
| K1Fr $\Phi 5$ | 200 | 1 | 1 | n.a. |
| K2F21 $\Phi 6$ | 100 | 353 | 4702 | $1 \mathrm{e}-6$ |
| K5F51 $\Phi 1$ | 500 | 13 | 2468 | $6 \mathrm{e}-6$ |
| K5F52 $\Phi$ r | 100 | 13 | 2455 | $7 \mathrm{e}-5$ |
| K5F7 $\Phi$ r | 100 | 2 | 20 | $9 \mathrm{e}-11$ |
| K6F61 $\Phi 3$ | 100 | 6 | 87 | $7 \mathrm{e}-7$ |

Table 3.1: Centralized exact version with fixed values of $\varepsilon_{0}=6.4$ and $\alpha=0.3$ : numerical results in terms of number of inner and outer iterations and solution accuracy

Some observations are in order.
(i) The centralized version of the Proximal point-Tikhonov Algorithm 3 proved in general simple, reliable, accurate and rapid. Indeed, the scheme proposed was able to solve all the test problems examined and reach high accuracy in a small amount of outer iterations.
(ii) For the most part of the tests that we have conducted, the whole scheme converged to the solution after few outer iterations. Few other iterations are needed for some refinements but, in most cases, any significant updating of the current iterate $x_{k}$ ceases long before $\varepsilon_{k}$ becomes smaller than epsilon_tol. Furthermore, we noticed that after the updating process of $\varepsilon_{k}$ is reset (whenever, we recall, for the first time, we get $\left\|x_{k}-x_{k-1}\right\|<10^{-5} \cdot \sqrt{n}$, see the discussion about the stopping criterion) the convergence is, in general, reached rapidly. This is another confirmation of the good behavior of the algorithm.

## Numerical results obtained for the centralized inexact version

The next step of our analysis has been the evaluation, always with fixed values for $\varepsilon_{0}$ and $\alpha$ (respectively 6.4 and 0.3 ), of the behavior of the inexact version
by considering different updating rules for the parameter $e_{k}$. We recall that $e_{k}$ express the degree of accuracy with which we solve the inner problem

$$
\mathrm{VI}\left(K, F^{k}\right), \quad \text { where } \quad F^{k}:=F+\varepsilon_{k} \Phi+\alpha\left(\bullet-x_{k}\right) .
$$

We remind that the inner algorithm stops whenever $\| x_{k+1}-P_{K}\left(x_{k+1}-\right.$ $\left.F^{k}\left(x_{k+1}\right)\right) \| \leq e_{k}$; we let $e_{k}$ vary at each outer iteration following different updating rules such that Condition C2 is satisfied. We have considered (see Figure 3.6) five different rates for $e_{k}$ to go to 0 . In the practical implementation, in fact, the error tolerated for the solution of the inner VI is set equal to

$$
\begin{equation*}
e_{k}^{i}=10^{-5}+\frac{\varepsilon_{0}}{a^{i}+f_{k}^{i}}, \quad i=1, \ldots, 5 \tag{3.33}
\end{equation*}
$$

where $f_{k}^{i}$ for $i$ that goes from 1 to 5 is equal to $k^{6}, \exp (k), 10 k^{3}, 100 k^{2}$ and $10 k^{2}$, respectively, and the the term $a^{i}$ ensures that $e_{0}^{i}$ is equal to $10^{-2}$ for every $i$. The constant $10^{-5}$ in expression (3.33) does not permit $e_{k}$ to become smaller than $10^{-5}$ : indeed, we recall that, if $e_{k}=10^{-5}$, VI (3.32) is considered to be solved exactly. In Appendix C (see page 138 and the


Fig. 3.6: Different updating rules for $e_{k}: e_{k}^{1} \propto 1 / k^{6}$ (magenta -o- line), $e_{k}^{2} \propto 1 / \exp (k)$ (red $-\triangleright$ - line), $e_{k}^{3} \propto 1 / 10 k^{3}$ (ciano - $\square$ - line), $e_{k}^{4} \propto 1 / 100 k^{2}$ (blue -*- line) and $e_{k}^{5} \propto 1 / 10 k^{2}$ (green $-\triangleleft-$ line)
following pages) we show the results of the simulations in terms of number of outer and inner iterations and solutions error. The interpretation of these data is not straightforward. Indeed, by letting $e_{k}$ be different from 0 for each $k$, the consecutive iterations solution accuracy achieved by the algorithm is not easily controllable. Furthermore, like before, the results of the tests show some reliances on the particular problem considered. Once again, in general, one can say that the behavior of the scheme is strictly linked to the moment in which the action of $\alpha$ becomes significant (as it "freezes" the progresses of
the algorithm) with respect to the precision reached by the algorithm at that time. Hence, the results here obtained confirm, one more time, the central role played by the parameter $\alpha$. Moreover, although more numerical tests must be conducted, we can say that, unlike many cases of optimization algorithms, the inexact version of our scheme does not seem to undoubtedly improve on the performances of the exact counterpart. This fact probably depends on the not easily predictable behavior of the solution accuracy when we let $e_{k}$ vary at each outer iteration.

### 3.6 Distributed Solution of the Hemivariational Inequality

This section discusses a distributed algorithm for solving a HVI on Cartesian product of sets. Our development extends the scope and considerably improves on the treatment in [34, Chapter 12] for non-cooperative games. The results in this section also provide further motivation for the choice of a combined Tikhonov and proximal methods for the solution of the HVI $(K, \Phi, h)$.

First, before we continue with our analysis, we briefly recall some relevant results about block decomposition methods. One of the main motivations of such approaches is that, if we fix some variables, one or more subproblems of special structure in the remaining variables could be obtained (see [11, 44] and the references therein). Furthermore, most of these algorithms are well suited for parallel and distributed computing, thus leading to an increased computing power. The best known among these approaches are probably the Jacobi and the Gauss-Seidel methods: the minimization versions of these algorithms are based on successive global minimizations with respect to each block component. The simple and successful idea is to fix all the block components of $x$ to some value, except for the $\nu$ th block component, and to minimize iteratively the objective function with respect to $x^{\nu}$. In the Jacobi method, the minimizations with respect to the different block components $x^{\nu}$ are carried out simultaneously; in the Gauss-Seidel algorithm, the minimizations are carried out successively for each component.

Let us consider the following problem:

$$
\begin{align*}
& \underset{x}{\operatorname{minimize}} f(x)  \tag{3.34}\\
& \text { subject to } x \in X
\end{align*}
$$

and suppose that $X$ is a Cartesian product of sets $X_{\nu}$ where each $X_{\nu}$ is a closed convex set. Accordingly, we represent vector $x \in \mathbb{R}^{n}$ in the form $x=\left(x^{1}, \ldots, x^{N}\right)$ where each $x^{\nu}$ is an element of $\mathbb{R}^{n_{\nu}}$.

The nonlinear Jacobi iteration is defined by:

$$
\begin{equation*}
x_{k+1}^{\nu}=\underset{x^{\nu} \in X_{\nu}}{\operatorname{argmin}} f\left(x_{k}^{1}, \ldots, x_{k}^{\nu-1}, x^{\nu}, x_{k}^{\nu+1}, \ldots, x_{k}^{N}\right) . \tag{3.35}
\end{equation*}
$$

The nonlinear Gauss-Seidel iteration is defined by:

$$
\begin{equation*}
x_{k+1}^{\nu}=\underset{x^{\nu} \in X_{\nu}}{\operatorname{argmin}} f\left(x_{k+1}^{1}, \ldots, x_{k+1}^{\nu-1}, x^{\nu}, x_{k}^{\nu+1}, \ldots, x_{k}^{N}\right) . \tag{3.36}
\end{equation*}
$$

The following classical result holds (for more recent results see also [45]).
Proposition 11. Let $f: \mathbb{R}^{n} \rightarrow \mathbb{R}$ be continuously differentiable, let $\gamma$ be a positive scalar and suppose that the mapping $R: X \rightarrow \mathbb{R}^{n}$, defined by $R(x)=x-\gamma \nabla f(x)$ is a contraction with respect to block-maximum norm $\|x\|=\left\|\left(x^{1}, \ldots, x^{N}\right)\right\|=\max _{\nu}\left\|x^{\nu}\right\|_{\nu} / w_{\nu}$, where each $\|\bullet\|_{\nu}$ is the Euclidean norm on $\mathbb{R}^{n_{\nu}}$ and each $w_{\nu}$ is a positive scalar. Then, there exists a unique vector $x^{*}$ which minimizes $f$ over $X$. Furthermore, the nonlinear Jacobi and Gauss-Seidel algorithms are well defined and the sequence $\left\{x_{k}\right\}$ generated by either of these algorithms converges to $x^{*}$.

The nonlinear Jacobi algorithm can be parallelized by assigning a separate processor to each block component $x^{\nu}$.

With the previous results in mind, consider now the HVI ( $K, G, h$ ) defined by the product set $K \triangleq \prod_{\nu=1}^{N} K_{\nu}$ with $\sum_{\nu=1}^{N} n_{\nu}=n$, where each $K_{\nu}$ is a closed convex subset of $\mathbb{R}^{n_{\nu}}$, by the continuous function $G: K \rightarrow \mathbb{R}^{n}$, and the separable function $h: K \rightarrow \mathbb{R}^{n}$ such that $h(x)=\sum_{\nu=1}^{N} h_{\nu}\left(x^{\nu}\right)$, where $x=$ $\left(x^{\nu}\right)_{\nu=1}^{N}$ with each $x^{\nu} \in K_{\nu}$ and $h_{\nu}$ is a convex continuous function defined on $K_{\nu}$. This setting is applicable to the side-constrained VI-C VI $\operatorname{(SOL}(K, F) \cap$ $\Omega, \Phi)$ via its penalized formulation, provided that $\Omega$ has the same Cartesian structure as $K$, i.e., $\Omega \triangleq \prod_{\nu=1}^{N} \Omega_{\nu}$. By Remark 5, the latter VI, which is defined on the intersection $\operatorname{SOL}(K, F) \cap \Omega$, is equivalent to the HVI $(K, \Phi, \rho h)$ which is defined on the set $K$ for all $\rho>0$ sufficiently large, where $h(x) \triangleq$ $\sum_{\nu=1}^{N} \operatorname{dist}_{K_{\nu}}\left(x^{\nu}\right)$.

With the pair $(K, h)$ having the above partitioned structure, we partition $G$ accordingly; i.e., $G(x)=\left(G_{\nu}(x)\right)_{\nu=1}^{N}$ with $G_{\nu}(x) \in \mathbb{R}^{n_{\nu}}$ for all $\nu=1, \ldots, N$. In what follows if $x=\left(x^{\nu}\right)_{\nu=1}^{N}$, we denote by $x^{-\nu}$ the sub-vector of $x$ with the $\nu$-th block omitted: $x^{-\nu} \triangleq\left(x^{\mu}\right)_{\mu=1, \mu \neq \nu}^{N}$, and by $K_{-\nu}$ the corresponding set: $K_{-\nu} \triangleq \prod_{\mu \neq \nu} K_{\mu}$ with $K_{\nu}$ removed.

We now present a synchronous Jacobi scheme. The analysis can be extended to synchronous Gauss-Seidel schemes and to asynchronous versions of both the Jacobi and Gauss-Seidel methods.

```
Algorithm 4: Distributed Algorithm for HVIs
(S.0) : Choose \(x_{0} \in K\) and set \(k=0\).
(S.1) : If \(x_{k}\) is a solution of \(\operatorname{HVI}(K, G, h)\) stop.
(S.2) : For \(\nu=1, \ldots, N\) set
        \(x_{k+1}^{\nu} \triangleq\) solution of \(\operatorname{HVI}\left(K_{\nu}, G_{\nu}\left(\bullet, x_{k}^{-\nu}\right), h_{\nu}(\bullet)\right)\)
(S.3) : Set \(x_{k+1}=\left(x_{k+1}^{\nu}\right)_{\nu=1}^{N}, \quad k \leftarrow k+1\) and go to (S.1).
\[
\begin{equation*}
\text { (S.3) : Set } x_{k+1}=\left(x_{k+1}^{\nu}\right)_{\nu=1}^{N}, k \leftarrow k+1 \text { and go to (S.1). } \tag{3.37}
\end{equation*}
\]
```

The defining equation (3.37) implicitly assumes that the HVI in (3.37) has one and only one solution. As we will see, this is true under the assumptions we make below. Our aim is to determine when the sequence $\left\{x_{k}\right\}$ produced by Algorithm 4 is well defined and converges to a solution of HVI ( $K, G, h$ ). This is done by showing that, under appropriate assumptions, Algorithm 4 is nothing else but a fixed-point iteration for a certain contractive mapping, whose unique fixed point is the desired solution of HVI ( $K, G, h$ ). The key assumption we need is stated next; it requires a certain $N \times N$ matrix to be of the $P$ kind. We recall that a (not necessarily symmetric) square matrix is $P$ if all its principal minors are positive. If $H: X \subseteq \mathbb{R}^{n} \rightarrow \mathbb{R}^{m}$ is Lipschitz continuous, we denote by $\beta(X, H)$ its Lipschitz constant defined by

$$
\beta(X, H) \triangleq \sup _{z, z^{\prime} \in X, z \neq z^{\prime}} \frac{\left\|H(z)-H\left(z^{\prime}\right)\right\|}{\left\|z-z^{\prime}\right\|}
$$

If furthermore $X$ is convex and $H: X \rightarrow \mathbb{R}^{n}$ is strongly monotone, we define the modulus of strong monotonicity by

$$
\gamma(X, H) \triangleq \inf _{z, z^{\prime} \in X, z \neq z^{\prime}} \frac{\left(H(z)-H\left(z^{\prime}\right)\right)^{T}\left(z-z^{\prime}\right)}{\left\|z-z^{\prime}\right\|^{2}}
$$

The two quantities $\beta(X, H)$ and $\gamma(X, H)$ have simple expressions if, for example, $H$ is continuously differentiable with Jacobian matrix at $z$ denoted by $J H(z)$. The following proposition is elementary and does not need a proof. For a real square matrix $M$, we let $\lambda_{m}^{S}(M)$ denotes the minimum eigenvalue of the symmetric part of $M$.

Proposition 12. If $H: X \rightarrow \mathbb{R}^{m}$ is continuously differentiable on an open set containing the convex set $X \subseteq \mathbb{R}^{n}$ and $\|J H(z)\|$ is bounded on $X$, then $\beta(X, H)=\sup _{z \in X}\|J H(z)\|$ is finite. Furthermore, if $m=n$ and $H$ is strongly monotone on $X$, then $\gamma(X, H)=\inf _{z \in X} \lambda_{m}^{S}(J H(z))$ is finite and positive.

The following is the key assumption necessary to prove convergence of Algorithm 4 for solving the $\operatorname{HVI}(K, G, h)$.

Assumption $\mathbf{P}$ Each function $G_{\nu}\left(\bullet, z^{-\nu}\right)$ is strongly monotone on $K_{\nu}$ with a uniformly positive strong monotonicity modulus for all $z^{-\nu} \in K_{-\nu}$; i.e.,

$$
\gamma_{\nu} \triangleq \inf _{z^{-\nu} \in K_{-\nu}} \gamma\left(K_{\nu}, G_{\nu}\left(\bullet, z^{-\nu}\right)\right)>0, \quad \forall \nu=1, \ldots, N
$$

For each $z^{\nu} \in K_{\nu}$, the function $G_{\nu}\left(z^{\nu}, \bullet\right)$ is Lipschitz continuous on $K_{-\nu}$ with a Lipschitz modulus that is independent of $z^{\nu} \in K_{\nu}$ : thus positive constants $\beta_{\nu \mu}$ exist for all $\mu \neq \nu$ such that for all $z^{\nu} \in K_{\nu}$ and all $u^{-\nu}$ and $v^{-\nu}$ in $K_{-\nu}$,

$$
\left\|G_{\nu}\left(z^{\nu}, u^{-\nu}\right)-G_{\nu}\left(z^{\nu}, v^{-\nu}\right)\right\| \leq \sum_{\mu \neq \nu} \beta_{\nu \mu}\left\|u^{\mu}-v^{\mu}\right\|
$$

Most importantly, the $N \times N$ matrix $\Upsilon$ defined below is $P$

$$
\Upsilon \triangleq\left[\begin{array}{cccc}
\gamma_{1} & -\beta_{12} & \cdots & -\beta_{1 N} \\
-\beta_{21} & \gamma_{2} & \cdots & -\beta_{2 N} \\
\vdots & & \ddots & \vdots \\
-\beta_{N 1} & -\beta_{N 2} & \cdots & \gamma_{N}
\end{array}\right]
$$

[The matrix $\Upsilon$ has the "Z-property", i.e., its off-diagonal entries are all nonpositive.]

Remark 8. The constant $\beta_{\nu \mu}$ is given by

$$
\begin{aligned}
& \sup _{z^{\nu} \in K_{\nu}} \sup _{\substack{\mu \\
u^{\mu}, v^{\mu} \in K_{\mu} \\
u^{\mu} \neq v^{\mu}}} \sup \left\{\left.\frac{G_{\nu}\left(z^{\nu}, y, u^{\mu}\right)-G_{\nu}\left(z^{\nu}, y, v^{\mu}\right)}{\left\|u^{\mu}-v^{\mu}\right\|} \right\rvert\, y \in \prod_{\nu^{\prime} \neq \nu, \mu} K_{\nu^{\prime}}\right\} \\
& =\sup _{z^{\nu} \in K_{\nu}} \sup \left\{\beta\left(K_{\mu}, G_{\nu}\left(z^{\nu}, y, \bullet\right)\right) \mid y \in \prod_{\nu^{\prime} \neq \nu, \mu} K_{\nu^{\prime}}\right\}
\end{aligned}
$$

By Proposition 12 the condition of the finiteness of the $\beta$ 's and $\gamma$ 's translates, in the case of a continuously differentiable $G$, into simple requirements on $J G$. However, the meaning of the $\beta$ 's and $\gamma$ 's is independent of the differentiability of $G$. Note also that if all the $\beta$ 's were 0 , HVI $(K, G, h)$ would decompose in $N$ uncoupled HVIs. In an informal sense, then, the terms $\beta_{\nu \mu}$ can be seen as a measure of the coupling of the problem. In this vein, the finiteness of the $\beta$ 's can be read as a requirement that if we want to solve HVI ( $K, G, h$ ) in a distributed way, the "interactions" between the sub-HVIs in which we decompose the original HVI should be limited. But this alone would not be sufficient for Algorithm 4 to work properly, and we also need the matrix $\Upsilon$ to be of the P-kind. In this latter requirement a role is played also by the quantities $\gamma_{\nu}$. These quantities essentially measure the degree of "uniform strong monotonicity" of the HVIs in which we decompose the problem. If the $\gamma_{\nu}$ are positive, this means that all the HVI $\left(K_{\nu}, G_{\nu}\left(\bullet, x^{-\nu}\right), h_{\nu}\right)$ in which we decompose the original HVI ( $K, G, h$ ) (see also Step 2 of Algorithm 4) are uniformly strongly monotone with a common strong monotonicity constant.

From this point of view, requiring the matrix $\Upsilon$ to be $P$ means, in a very loose sense, that the HVIs $\left(K_{\nu}, G_{\nu}\left(\bullet, x^{-\nu}\right), h_{\nu}\right)$ are "strongly monotone enough to dominate the interactions from the other blocks". We note that if $\Upsilon$ is $P$, it must hold that all the $\gamma_{\nu}$ are positive. In turn this implies that for any $z \in K$ and any $\nu$, the HVI $\left(K_{\nu}, G_{\nu}\left(\bullet, z^{-\nu}\right), h_{\nu}\right)$ is strongly monotone and therefore has one and only one solution, which we denote by $x^{\nu}(z)$; we then let $x(z) \triangleq\left(x^{\nu}(z)\right)_{\nu=1}^{N}$. This is a self-map from the set $K$ into itself; Algorithm 4 is easily seen to be the fixed-point iteration: $x_{k+1}=x\left(x_{k}\right)$.

We are now ready to state the main result of this section. Besides treating the HVI, this result extends the treatment in [34, Section 12.6] for a Nash game in that $G$ is not required to be twice continuously differentiable here.

Theorem 16. Let the HVI $(K, G, h)$ be given with the Cartesian product structure set forth at the beginning of this section. Suppose that Assumption P holds. Algorithm 4 generates a well-defined sequence $\left\{x_{k}\right\}$ that converges to the unique solution of $\operatorname{HVI}(K, G, h)$.

Proof. It suffices to show that $x(z)$ is a contraction on $K$. Let $\bar{z}$ and $\tilde{z}$ be two points in $K$; for a generic $\nu$ we can write,

$$
\begin{aligned}
& G_{\nu}\left(x^{\nu}(\bar{z}), \bar{z}^{-\nu}\right)^{T}\left(y^{\nu}-x^{\nu}(\bar{z})\right)+h_{\nu}\left(y^{\nu}\right)-h_{\nu}\left(x^{\nu}(\bar{z})\right) \geq 0, \quad \forall y^{\nu} \in K_{\nu} \\
& G_{\nu}\left(x^{\nu}(\tilde{z}), \tilde{z}^{-\nu}\right)^{T}\left(y^{\nu}-x^{\nu}(\tilde{z})\right)+h_{\nu}\left(y^{\nu}\right)-h_{\nu}\left(x^{\nu}(\tilde{z})\right) \geq 0, \quad \forall y^{\nu} \in K_{\nu} .
\end{aligned}
$$

By taking $y^{\nu}=x^{\nu}(\tilde{z})$ in the former inequality and $y^{\nu}=x^{\nu}(\bar{z})$ in the latter and summing, we deduce

$$
\begin{equation*}
\left[G_{\nu}\left(x^{\nu}(\bar{z}), \bar{z}^{-\nu}\right)-G_{\nu}\left(x^{\nu}(\tilde{z}), \tilde{z}^{-\nu}\right)\right]^{T}\left(x^{\nu}(\tilde{z})-x^{\nu}(\bar{z})\right) \geq 0 \tag{3.38}
\end{equation*}
$$

Adding and subtracting $G_{\nu}\left(x^{\nu}(\bar{z}), \tilde{z}^{-\nu}\right)$ in the square parenthesis and using the uniform monotonicity hypothesis in Assumption P, we get

$$
\gamma_{\nu}\left\|x^{\nu}(\bar{z})-x^{\nu}(\tilde{z})\right\|^{2} \leq\left[G_{\nu}\left(x^{\nu}(\bar{z}), \bar{z}^{-\nu}\right)-G_{\nu}\left(x^{\nu}(\bar{z}), \tilde{z}^{-\nu}\right)\right]^{T}\left(x^{\nu}(\tilde{z})-x^{\nu}(\bar{z})\right)
$$

from which we deduce

$$
\left\|x^{\nu}(\bar{z})-x^{\nu}(\tilde{z})\right\| \leq \frac{1}{\gamma_{\nu}}\left\|G_{\nu}\left(x^{\nu}(\bar{z}), \bar{z}^{-\nu}\right)-G_{\nu}\left(x^{\nu}(\bar{z}), \tilde{z}^{-\nu}\right)\right\| .
$$

In turn, the uniform Lipschitzian hypothesis in Assumption P permits us to write

$$
\left\|G_{\nu}\left(x^{\nu}(\bar{z}), \bar{z}^{-\nu}\right)-G_{\nu}\left(x^{\nu}(\bar{z}), \tilde{z}^{-\nu}\right)\right\| \leq \sum_{\mu \neq \nu} \beta_{\nu \mu}\left\|\bar{z}^{\mu}-\tilde{z}^{\mu}\right\|
$$

The last two inequalities immediately give

$$
\left\|x^{\nu}(\bar{z})-x^{\nu}(\tilde{z})\right\| \leq \frac{1}{\gamma_{\nu}} \sum_{\mu \neq \nu} \beta_{\nu \mu}\left\|\bar{z}^{\mu}-\tilde{z}^{\mu}\right\|
$$

Since this relation is valid for all $\nu$, we readily get

$$
\left(\begin{array}{c}
\left\|x^{1}(\bar{z})-x^{1}(\tilde{z})\right\|  \tag{3.39}\\
\vdots \\
\left\|x^{N}(\bar{z})-x^{N}(\tilde{z})\right\|
\end{array}\right) \leq \Gamma\left(\begin{array}{c}
\left\|\bar{z}^{1}-\tilde{z}^{1}\right\| \\
\vdots \\
\left\|\bar{z}^{N}-\tilde{z}^{N}\right\|
\end{array}\right)
$$

where

$$
\Gamma \triangleq\left[\begin{array}{ccccc}
0 & \frac{\beta_{12}}{\gamma_{1}} & \cdots & \frac{\beta_{1 N-1}}{\gamma_{1}} & \frac{\beta_{1 N}}{\gamma_{1}} \\
\frac{\beta_{21}}{\gamma_{2}} & 0 & \cdots & \frac{\beta_{2 N-1}}{\gamma_{2}} & \frac{\beta_{2 N}}{\gamma_{2}} \\
\vdots & & & \vdots \\
\frac{\beta_{N 1}}{\gamma_{N}} & \frac{\beta_{N 2}}{\gamma_{N}} & \cdots & \frac{\beta_{N N-1}}{\gamma_{N}} & 0
\end{array}\right]
$$

Taking into account Assumption P, [18, Lemma 5.13.14] implies that $\Gamma$ has spectral radius less than 1 (see [34, Proposition 12.7] for details) thus concluding the proof.

The significance of the previous result is that it fits very well with the results in Section 3.4. In fact, note that the parameter $\alpha$ can be chosen large enough so that the HVI $\left(K, F^{k}, \varepsilon_{k} h\right)$ in Step 2 of Algorithm 3, where $F^{k} \triangleq F+$ $\varepsilon_{k} \Phi+\alpha\left(\bullet-x^{k}\right)$, can be solved by the distributed algorithm described in this section, provided that the pair $(K, h)$ has the required Cartesian structure. The upshot is that we can finally obtain a desired distributed algorithm for the solution of HVIs and, on the basis of the results in Section 3.3, also of VI-C HVIs with side constraints.

Corollary 1. Consider the $\operatorname{HVI}\left(K, F+\varepsilon_{k} \Phi+\alpha\left(\bullet-x_{k}\right), \varepsilon_{k} h\right)$ (arising from Step 2 of Algorithm 3) and suppose that $F$ and $\Phi$ are Lipschitz continuous on $K$ with moduli $L_{F}$ and $L_{\Phi}$ respectively, and that all $\gamma_{\nu}$ are non negative. Let $\bar{\varepsilon}$ be such that $\varepsilon_{k} \leq \bar{\varepsilon}$ for all $k$ and let $\alpha$ be a positive constant such that

$$
\begin{equation*}
\alpha>\bar{\alpha} \triangleq\left(\max _{1 \leq \nu \leq N} \sum_{\mu \neq \nu}\left(\beta_{\nu \mu}+\bar{\varepsilon} L_{\Phi}\right)\right)-\min _{1 \leq \nu \leq N} \gamma_{\nu} \tag{3.40}
\end{equation*}
$$

where $\beta_{\nu \mu}$ and $\gamma_{\nu}$ are the constants defined in Assumption P, with $G=F$. Then all subproblems appearing at Step 2 of Algorithm 3 can be solved by the distributed Algorithm 4.

Proof. According to Theorem 16 we only need to show that the matrix $\Upsilon$ associated with a generic HVI $\left(K, F+\varepsilon_{k} \Phi+\alpha\left(\bullet-x_{k}\right), \varepsilon_{k} h\right)$ has the P-property. Invoking [18, Lemma 5.13.14] this is equivalent to the following matrix

$$
\left[\begin{array}{cccc}
0 & \frac{\beta_{12}+\bar{\varepsilon} L_{\Phi}}{\gamma_{1}+\alpha} & \cdots & \frac{\beta_{1 N}+\bar{\varepsilon} L_{\Phi}}{\gamma_{1}+\alpha} \\
\vdots & & & \vdots \\
\frac{\beta_{N 1}+\bar{\varepsilon} L_{\Phi}}{\gamma_{N}+\alpha} & \cdots & \frac{\beta_{N(N-1)}+\bar{\varepsilon} L_{\Phi}}{\gamma_{N}+\alpha} & 0
\end{array}\right]
$$

having spectral radius less than 1 . It can now be checked that (3.40) guarantees this latter condition by Gershgorin circle theorem.

Note that this Corollary clarifies very well the role of $\alpha$, at least from the point of view of the distributed implementation of the algorithm: it ensures that enough strong monotonicity is present; and in fact the larger the $\gamma_{\nu}$, the smaller $\alpha$ can be.


Fig. 3.7: The distributed version of the algorithm; in particular, the inner Jacobi scheme is given in detail (red boxes).

### 3.7 Numerical experiments: centralized vs distributed algorithm

In this section we present a comparison in terms of numerical performances between the centralized and the distributed versions of Algorithm 3. We recall that, in order to solve the inner problem (3.19), within a centralized approach a centralized scheme is employed, within a distributed approach one resorts to a distributed scheme.

### 3.7.1 Examined problems

We tested both versions of the scheme on the VI-C VI $(K, F, \Phi)$ problem by considering as lower feasible set $K$ the set $K 2$ which is the $n$-dimensional box (see Appendix B). Indeed, the $K 2$ set has a Cartesian product structure such that, once $n$ is fixed, we can consider every possible (with $\sum_{\nu=1}^{N} n_{\nu}=n$ and $\prod_{\nu=1}^{N} K_{\nu}$ ) decomposition for the distributed algorithm to be evaluated. Hence, this choice allows us to study the behavior of the distributed version of the scheme (in comparison with that of the centralized one) for fixed $n$ and by varying $N$ and the $n_{\nu}$ s, thus referring to different decomposition "strategies".

We recall that, besides the assumptions of Theorem 15 (the considerations and the assumptions made in Section 3.5.1 still hold in the following analysis), in order to get problem (3.32), i.e. $\mathrm{VI}\left(K, F^{k}:=F+\varepsilon_{k} \Phi+\alpha\left(\bullet-x_{k}\right)\right)$, solved exactly by the inner distributed algorithm at each outer iteration $k$ (see Corollary 1), functions $F$ and $\Phi$ must be Lipschitz continuous on $K$ and function $F_{\nu}\left(\bullet, z^{-\nu}\right)$ must be at least monotone on $K_{\nu}$ and Lipschitz continuous on $K_{-\nu}$ with a Lipschitz modulus that is independent of $z^{\nu} \in K_{\nu}$ : thus positive constants $\beta_{\nu \mu}$ exist for all $\mu \neq \nu$ such that for all $z^{\nu} \in K_{\nu}$ and all $u^{-\nu}$ and $v^{-\nu}$ in $K_{-\nu}$,

$$
\left\|F_{\nu}\left(z^{\nu}, u^{-\nu}\right)-F_{\nu}\left(z^{\nu}, v^{-\nu}\right)\right\| \leq \sum_{\mu \neq \nu} \beta_{\nu \mu}\left\|u^{\mu}-v^{\mu}\right\| .
$$

We notice, firstly, that, by considering as lower function $F$ the monotone functions F31, F21, F61 and Fr (see Appendix B), in view of Proposition 12, the previous condition on $\beta_{\nu \mu}$ is fulfilled. Secondly, thanks to the monotonicity of $F$, functions $F_{\nu}\left(\bullet, z^{-\nu}\right)$ are monotone on $K_{\nu}$.

### 3.7.2 The implemented algorithm

We have examined and compared the number of outer and inner iterations performed by the two versions of the algorithm and the time spent in order
to calculate a solution. We recall that, while the outer scheme is the same, the two algorithms differ in the way they solve the strongly monotone inner problem (3.32), i.e.,

$$
\mathrm{VI}\left(K, F^{k}\right), \quad \text { where } \quad F^{k}:=F+\varepsilon_{k} \Phi+\alpha\left(\bullet-x_{k}\right)
$$

On the one hand, the centralized version of the scheme is characterized by the presence of a centralized algorithm (we have chosen for our analysis the BPA) for the solution of (3.32); on the other hand, the distributed version resort to a decomposition method (here we have considered the Jacobi algorithm) in order to compute a solution of the same problem (3.32), thus allowing parallel and distributed computing. For all the considerations about the outer scheme, one can refer to Section 3.5. Here we remark that, while in the centralized case the scheme converges for every $\alpha>0$, when Algorithm 4 is employed, one has to choose $\alpha$ such that condition (3.40) holds; besides, the minimum value of $\alpha$ that guarantees the convergence of the distributed scheme depend on the function $F$ and $\Phi$ involved, on the particular decomposition strategy adopted and on $\bar{\varepsilon}:=\varepsilon_{0}$ (which certainly is larger than $\varepsilon_{k}$ for all $k$ ). In the experiments that we have conducted we have chosen $\varepsilon_{0}=3.2$ and, for every $k$, the inner problem (3.32) is solved exactly, i.e. $e_{k}=10^{-5}$.

For every test problem (thus for every choice of functions $F$ and $\Phi$ ), we have considered six different decomposition strategies (hence different choices for the number of decomposed problems $N$ and for their dimensions $n_{\nu} \mathrm{s}$ ) $Q=1, \ldots, 6$. For every $Q$, we get a different value of the lower bound $\bar{\alpha}:=\bar{\alpha}_{Q}$. We have evaluated the behavior of the decomposed version for every decomposition considered, and we have chosen for each test on a particular decomposition $Q$ a value of the parameter $\alpha$, namely $\alpha_{Q}$, equal to the corresponding lower bound $\bar{\alpha}_{Q}$. The centralized version is examined with the same values of $\alpha$. Moreover, we have also studied, for every decomposition strategy adopted, the decomposed algorithm with a "risky" (thus not guaranteeing in principle the convergence of the whole scheme) value of $\alpha_{Q}$ equal to $\alpha_{R}$ such that $\alpha_{R}<\bar{\alpha}_{Q}$, relying on the fact that $\bar{\alpha}_{Q}$ is a "conservative" lower bound.

As inner scheme we have used the synchronous Jacobi Algorithm 4. For each $\nu=1, \ldots, N$, the $\nu$-th decomposed problem

$$
\mathrm{VI}\left(K_{\nu}, F_{\nu}^{k}\right), \quad \text { where } \quad F_{\nu}^{k}:=F_{\nu}+\varepsilon_{k} \Phi_{\nu}+\alpha\left(\bullet-\left(x_{k}\right)^{\nu}\right)
$$

is solved exactly by means of the BPA. Again, in these simulations, the value of the step size $\tau$, once is suitably set (see Section 3.5), is kept constant for every test performed by considering different values of the parameter $\alpha$ and it is the same for both the centralized and the decomposed versions.

The maximum number of inner Jacobi iterations allowed at the $k$ th step and the maximum number of BPA iterations allowed at the $i$ th Jacobi step are both set equal to 1000 .

We remark that all the choices that we have made are in view of the fact that, for $N=1$ and $n_{1}=n$, we want the two versions of the scheme to behave the same.

### 3.7.3 Numerical results and comments

We recall that, whenever the inner algorithm eventually stops updating $x_{k}$, even though $\varepsilon_{k}$ is not smaller than epsilon_tol, we stop counting the outer iterations.

Let $\left(b_{k}^{i}\right)^{\nu}$ be the number of iterations performed at the $k$ th outer iteration and at the $i$ th Jacobi fixed-point iteration by the very inner BPA in order to solve the $\nu$ th decomposed problem $\operatorname{VI}\left(K_{\nu}, F_{\nu}^{k}\left(\bullet,\left(x_{k}^{i}\right)^{-\nu}\right)\right)$. In view of the parallel and distributed computing allowed by the algorithm of Figure 3.7, the total amount of inner iterations of the decomposed scheme is set equal to

$$
\text { \# inn.it }:=\sum_{k} \sum_{i} \max _{\nu}\left(b_{k}^{i}\right)^{\nu}
$$

Hence, for each fixed $k$ and $i$, we consider only the number of BPA iterations $\left(b_{k}^{i}\right)^{\hat{\nu}}:=\max _{\nu}\left(b_{k}^{i}\right)^{\nu}$ necessary to solve the "bottle-neck" $\hat{\nu}$ th decomposed problem.

In a similar manner, the total amount of the time spent in order to get to a solution is given by the following expression:

$$
\text { time }:=\sum_{k} \sum_{i} \max _{\nu}\left(t_{k}^{i}\right)^{\nu}
$$

where $\left(t_{k}^{i}\right)^{\nu}$ is the time spent at the $k$ th outer iteration and at the $i$ th Jacobi iteration by the very inner BPA in order to solve the $\nu$ th decomposed problem. For the centralized scheme, instead, time is simply equal to $\sum_{k} t_{k}$, where $t_{k}$ is the time spent at the $k$ th outer iteration by the inner BPA.

In Appendix C (see page 141 and the following pages) we show the results of the simulations in terms of number of outer and inner iterations and of time spent by the two versions of the algorithm. In Table 3.2 we report a sample (for $Q=5$ ) of the results that we have obtained when both schemes are tested with the same values of $\alpha$.

Once again, the central role played by the parameter $\alpha$ is evident. No matter if the scheme is centralized or distributed, what really counts is the value of $\alpha$ that is considered.
(i) We notice that, for a fixed decomposition $Q$, in view of the parallel implementation of the distributed algorithm, the worst estimate of the overall execution time spent to solve the $\nu$ th problem (assigning a separate pro-

| $\mathrm{Q}=5$ | Centralized |  |  | Decomposed |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Problem | \# out.it. | \# inn.it. | time(s) | \# out.it. | \# inn.it. | time(s) |
| K2F13 $\Phi 6$ | 1962 | 6575 | 72.55 | 1953 | 6893 | 68.79 |
| K2F21 $\Phi 2$ | 1001 | 4987 | 19.37 | 1001 | 5003 | 12.52 |
| K2F61 $\Phi$ r | 382 | 1626 | 83 | 382 | 1961 | 10.68 |
| K2Fr $\Phi 2$ | 61 | 780 | 40.06 | 61 | 937 | 5.11 |
| K2Fr $\Phi 7$ | 25 | 813 | 53.18 | 25 | 1095 | 14.08 |

Table 3.2: Centralized vs decomposed versions: numerical results in terms of number of inner and outer iterations and execution time for $Q=5$
cessor to each block component $x^{\nu}$ ) is shorter than the time spent by the centralized version in order to get to a solution.
(ii) It is worth noting that with the same values of $\alpha_{Q}$ for the centralized and the decomposed algorithms, the number of outer iterations performed by the whole scheme in both versions is, as expected, the same, except for very little variations due to numerical reasons.
(iii) We recall that the number of inner iterations is more deeply related to the features of the chosen inner algorithm and to the particular problem considered and it obviously depends also on the number of outer iterations. With regard to the distributed version, we recall also that the number of inner iterations \# inn.it, which is a worst case count, clearly depends also on the number of the Jacobi (intermediate) algorithm iterations. With these considerations in mind, we note that, with the same values of $\alpha_{Q}$ for the centralized and the decomposed algorithms, the total amount of inner iterations performed in the two cases is slightly different; in particular, it is greater for the distributed version.
(iv) In general, the centralized version could benefit from an "unconstrained" positive parameter $\alpha$, while the decomposed version, in principle, has to cope with condition (3.40). We have tested the distributed scheme also with "risky" (thus not guaranteeing, in principle, convergence) values for parameter $\alpha_{Q}$ equal to $\alpha_{R}$. It turns out that, if, for the $Q$ th decomposition, $\alpha_{R} \ll \bar{\alpha}_{Q}$, some problems concerning the behavior of the algorithm may occur. Once again, the comparison, for a fixed decomposition $Q$, of the performances of the distributed scheme evaluated with different values for $\alpha_{Q}$ (i.e. $\alpha_{Q}=\bar{\alpha}_{Q}$ and $\alpha_{Q}=\alpha_{R}$ ) point out the significance of this parameter. Here we stress that all the considerations made in Section 3.5.3 about the "right" value of $\alpha$ remain valid in this analysis.

To sum up, we emphasize the good behavior of the distributed scheme (compared to that of the centralized version): it proved reliable and fast. Furthermore, the opportunity of a distributed and parallel implementation of the algorithm does not entail a deterioration of the performances of the overall scheme provided that the decomposition strategy adopted does not lead to an unsuitable (in the light of the considerations made in Section 3.5.3) value for the parameter $\alpha_{Q}$.


Fig. 3.8: The distributed algorithmic scheme implemented

### 3.8 VI-C Minimization Problems

A particularly important case of the general HVI is the hierarchical minimization problem (3.2), whose stationarity condition is precisely the VI-C VI $(K, F, \nabla \psi)$. The developments in the previous sections can certainly be applied to this VI when $\psi$ is a convex function. In this case, a simple level-set condition can be given to ensure the boundedness of the set $L$ in assumption (F). Indeed, we have

$$
\begin{aligned}
L & =\left\{x \in K \mid \exists y \in \underset{z \in \operatorname{SOL}(K, F)}{\operatorname{argmin}} \psi(z) \text { such that } \nabla \psi(x)^{T}(y-x)>0\right\} \\
& \subseteq\{x \in K \mid \psi(x)<\underset{y \in \operatorname{SOL}(K, F)}{\operatorname{minimum}} \psi(y)\}
\end{aligned}
$$

Thus, if $\psi$ has bounded level sets on $K$, then $L$ is bounded.
It turns out that the convex case can be used as the basis for designing iterative feasible descent methods [9] for computing a stationary point of (3.2) when $\psi$ is not convex; such a point is by definition a solution to the VI-C VI $(K, F, \nabla \psi)$. This class of algorithms requires the calculation of a search direction that involves the solution of a VI constrained optimization problem with a linear, or convex quadratic objective function. Algorithm 3 can be used as a "black-box" for computing such a direction. A line-search is then performed. In what follows, we present a scaled projected gradient method as an illustration of the overall iterative procedure for solving the VI-C VI $(K, F, \nabla \psi)$ with a nonconvex $\psi$.

$$
\begin{align*}
& \text { Algorithm 5: A descent method for non-convex } \psi \\
& \hline \text { (S.0) : Choose } x_{0} \in K, \beta \in(0,1), \delta \in(0,1) \text { and set } k=0 . \\
& \text { (S.1) : If } x_{k} \in \mathrm{SOL}(\operatorname{SOL}(K, F), \nabla \psi) \text {, stop. } \\
& \text { (S.2): Use Algorithm } 3 \text { to calculate a solution } y_{k} \text { of the following } \\
& \text { problem: } \\
& \qquad \underset{y \in \operatorname{SOL}(K, F)}{\operatorname{minimize}} \nabla \psi\left(x_{k}\right)^{T} y+\frac{1}{2} y^{T} Q_{k} y, \tag{3.41}
\end{align*}
$$

where $Q_{k}$ is a symmetric positive semidefinite matrix. Set $d_{k} \triangleq y_{k}-$ $x_{k}$. (S.3) : Compute the smallest nonnegative integer $i$ such that

$$
\begin{equation*}
\psi\left(x_{k}+\beta^{i} d_{k}\right) \leq \psi\left(x_{k}\right)+\delta \beta^{i} \nabla \psi\left(x_{k}\right)^{T} d_{k} \tag{3.42}
\end{equation*}
$$

and set $x_{k+1} \triangleq x_{k}+\beta^{i} d_{k}$.
(S.4) : Set $k \leftarrow k+1$ and return to (S.1).

Convergence of Algorithm 5 follows easily from standard results, once one can show that it is well defined.

Theorem 17. Consider the VI-C VI $(K, F, \nabla \psi)$, with $\psi$ continuously differentiable on $K$. Suppose that the assumptions (A), (B), and (C) hold and the matrices $Q_{k}$ are uniformly positive definite, i.e. $m\|d\|^{2} \leq d^{T} Q_{k} d \leq M\|d\|^{2}$ for some positive $m$ and $M$ and for all $d \in \mathbb{R}^{n}$. Then Algorithm 5 is well defined and every limit point of the sequence $\left\{x_{k}\right\}$ it generates is a stationary point of the minimization problem (3.2).

Proof. We only need to show that we are able to solve subproblem (3.41); convergence then follows by well-established results on the conditional gradient method, see for example [9]. Subproblem (3.41) is just the VI-C HVI $\left(K, F, 0, h_{k}\right)$, where $h_{k}(y) \triangleq \nabla \psi\left(x_{k}\right)^{T} y+\frac{1}{2} y^{T} Q_{k} y$; so it is enough to check that (A-F) hold for this subproblem. Conditions (A-C) hold by assumption. Condition (D) obviously holds and (E) is satisfied because $Q_{k}$ is positive definite. Finally (F) holds because of the discussion at the beginning of this section, by noting the level sets of $h_{k}$ are clearly bounded.

### 3.9 An Application to a Rate Maximization Game

In this section, we illustrate the application of the theoretical framework previously developed to a resource allocation problem in wireless networks, namely the power control problem over parallel Gaussian Interference Channels (IC). This problem in fact provided the motivation of our work and, in particular, the need of devising distributed solution methods of monotone VI-constraints (Hemi)VIs. We begin with a very informal description of the problem and then give a more detailed technical account.

System Model. The IC model is of great interest in the Signal Processing and Communications community; it is in fact sufficiently general to encompass many multiuser communication systems of practical interest, such as peer-to-peer networks, digital subscriber lines (DSL), wireless frequency-selective ad-hoc networks, and Orthogonal Frequency-Division Multiplexing (OFDM)/ Time Division Multiple Access (TDMA) single/multicell cellular systems.

In the IC, there are $Q$ transmitter-receiver pairs; each transmitter wants to communicate with its corresponding receiver over a set of $N$ independent noisy channels; these channels may represent either time or frequency physical channels (here, for the sake of terminology and without loss of generality, we consider transmissions over ICs in the frequency domain, termed as frequency-selective ICs). We associate with each of the $Q$ users a nonnegative vector variable $p^{\nu} \triangleq\left(p_{n}^{\nu}\right)_{n=1}^{N} \geq 0$, representing the power allocated over the $N$ channels by the transmission-receiver pair $\nu$. As such, these variables satisfy some bound constraints $0 \leq p_{n}^{\nu} \leq p_{n}^{\nu, \max }, \nu=1, \ldots, Q, n=1, \ldots, N$, where $p_{n}^{\nu, \max }$ are given upper bounds; technically, they represent the so-called "mask constraints", imposed by the regulator to limit the amount of power radiated by user $\nu$ over licensed bands. Furthermore, each transmitter has a
power budget limit denoted by $P^{\nu}$, so that the power vector $p^{\nu}$ is constrained to $\sum_{n=1}^{N} p_{n}^{\nu} \leq P^{\nu}$. The set of power constraints of each user $\nu$ is thus defined as

$$
\begin{equation*}
\mathcal{P}^{\nu} \triangleq\left\{p^{\nu} \in \mathbb{R}_{+}^{N}: \sum_{n=1}^{N} p_{n}^{\nu} \leq P^{\nu}, \quad 0 \leq p^{\nu} \leq p^{\nu, \max } \triangleq\left(p_{n}^{\nu, \max }\right)_{n=1}^{N}\right\} \tag{3.43}
\end{equation*}
$$

The IC is used to model practical multiuser systems that do not have any infrastructure, meaning that there is neither a centralized authority scheduling the transmissions in the network nor coordination among the users. It follows that the communications of the $Q$ pairs may occur simultaneously; this implies that, in addition to the desired signal, each user receives also the signal transmitted by the other $Q-1$ pairs, which is an undesired signal, termed as Multi-User Interference (MUI). Stated in mathematical terms, the quality of the transmission of each pair $\nu$ over the channel $n$ is measured by the Signal-to-Noise-plus-Interference ratio (SINR):

$$
\begin{equation*}
\operatorname{SINR}_{n}^{\nu}\left(p_{n}^{\nu}, p_{n}^{-\nu}\right)=\frac{\left|H_{\nu \nu}(n)\right|^{2} p_{n}^{\nu}}{\sigma_{n}^{\nu 2}+\sum_{\mu \neq \nu}\left|H_{\nu \mu}(n)\right|^{2} p_{n}^{\mu}} \tag{3.44}
\end{equation*}
$$

where $\left|H_{\nu \nu}(n)\right|>0$ is the channel gain of pair $\nu$ over the frequency band $n$, and $\left|H_{\nu \mu}(n)\right| \geq 0$ is the (cross-)channel gain between the transmitter $\mu$ and the receiver $\nu ; \sigma_{n}^{\nu 2}$ is the power spectral density (PSD) of the noise at receiver $\nu$ over the band $n$; and the set of all the users power allocations over the channel $n$, except the $\nu$-th one is denoted with $p_{n}^{-\nu} \triangleq\left(p_{n}^{1}, \ldots, p_{n}^{\nu-1}, p_{n}^{\nu+1}, \ldots, p_{n}^{Q}\right)$. The useful power signal of pair $\nu$ over the channel $n$ is thus $\left|H_{\nu \nu}(n)\right|^{2} p_{n}^{\nu}$, whereas $\sum_{\mu \neq \nu}\left|H_{\nu \mu}(n)\right|^{2} p_{n}^{\mu}$ is the PSD of MUI measured by the receiver $\nu$ over the channel $n$. The overall performance of each transmission $\nu$ is measured in terms of the maximum achievable information rate $r^{\nu}\left(p^{\nu}, p^{-\nu}\right)$ over the set of the $N$ parallel channels, which depends on the power allocation of all the users $\left(p^{\nu}, p^{-\nu}\right)$, with $p^{-\nu} \triangleq\left(p^{1}, \ldots, p^{\nu-1}, p^{\nu+1}, \ldots, p^{Q}\right)$. Under basic information theoretical assumptions (see, e.g., [93, 105]) and given the users' power allocation profile $p^{1}, \ldots, p^{Q}, r^{\nu}\left(p^{\nu}, p^{-\nu}\right)$ is

$$
\begin{equation*}
r^{\nu}\left(p^{\nu}, p^{-\nu}\right)=\sum_{n=1}^{N} \log \left(1+\operatorname{SINR}_{n}^{\nu}\left(p_{n}^{\nu}, p_{n}^{-\nu}\right)\right) \tag{3.45}
\end{equation*}
$$

Problem Formulation. The system design consists of finding the optimal power allocation of the users in order to maximize the information rates of the links, according to some performance metrics. A natural objective function would be the sum-rate of the users $\sum_{\nu=1}^{Q} r^{\nu}(p)$. However, the resulting optimization problem has been showed to be NP hard [65]. Several attempts have been pursued in the literature to deal with the nonconvexity of such a problem; however all the proposed schemes are centralized and computation-
ally expensive, which makes them non-implementable in a network with no infrastructure. Thus, it seems natural to concentrate on decentralized power control solutions, where the users are able to self-enforce the negotiated agreements on the usage of the available spectrum without the intervention of a centralized authority. This motivates the formulation of the system design as a Nash Equilibrium Problem (NEP): the aim of each player (link) $\nu$, given the strategy profile $p^{-\nu}$ of the others, is to choose a feasible power allocation $p^{\nu}$ that maximizes the rate $r^{\nu}\left(p^{\nu}, p^{-\nu}\right)$, i.e.,

$$
\begin{array}{ll}
\underset{p^{\nu}}{\operatorname{maximize}} & r^{\nu}\left(p^{\nu}, p^{-\nu}\right)  \tag{3.46}\\
\text { subject to } & p^{\nu} \in \mathcal{P}^{\nu}
\end{array}
$$

for all $\nu=1, \ldots, Q$, where $\mathcal{P}^{\nu}$ and $r^{\nu}\left(p^{\nu}, p^{-\nu}\right)$ are defined in (3.43) and (3.45), respectively. We will denote the above NEP as $\mathcal{G}=<\mathcal{P},\left(r_{i}\right)_{i=1}^{Q}>$, with $\mathcal{P} \triangleq \prod_{\nu} \mathcal{P}^{\nu}$. In this setting, the design aim becomes the computation of a Nash Equilibrium (NE), i.e., the calculation of power allocation $p^{\star}$ such that $p^{\nu, \star}$ is optimal for (3.46), given $p^{-\nu, \star}$.

Note that, for any fixed $p^{-\nu} \geq 0$, the single-user optimization problem in (3.46) admits a unique solution $p^{\nu, \star}$, given by the well-known waterfilling expression [93, 105]:

$$
\begin{equation*}
p_{n}^{\nu, \star}=\operatorname{wf}_{n}^{\nu}\left(p_{n}^{-\nu}\right) \triangleq\left(\lambda^{\nu}-\frac{\sigma_{n}^{\nu^{2}}+\sum_{\mu \neq \nu}\left|H_{\nu \mu}(n)\right|^{2} p_{n}^{\mu}}{\left|H_{\nu \nu}(n)\right|^{2}}\right)_{+} \tag{3.47}
\end{equation*}
$$

with $n=1, \ldots, N$, where $[x]_{+} \triangleq \max (0, x), p_{n}^{-\nu} \triangleq\left(p_{n}^{1}, \ldots, p_{n}^{\nu-1}, p_{n}^{\nu+1}, \ldots\right.$, $p_{n}^{Q}$ ) and the waterlevel $\lambda^{\nu}$ is chosen to satisfy the transmit power constraint $\sum_{n=1}^{N} p_{n}^{\star \nu}=P^{\nu} ; \lambda^{\nu}$ can be computed very efficiently in at most $N$ extremely simple steps. Interestingly, the best-response (3.47) can be computed locally and distributively by the players, since each user only needs to measure the overall interference-plus-noise PSD $\sigma_{n}^{\nu 2}+\sum_{\mu \neq \nu}\left|H_{\nu \mu}(n)\right|^{2} p_{n}^{\mu}$ and "waterfill" over it. The Nash equilibria $p^{\star}$ of the NEP are thus the fixedpoints of the waterfilling mapping $w f(p) \triangleq\left(\operatorname{wf}^{\nu}\left(p^{-\nu}\right)\right)_{\nu=1}^{Q}$, with wf ${ }^{\nu}\left(p^{-\nu}\right) \triangleq$ $\left(\mathrm{wf}_{n}^{\nu}\left(p_{n}^{-\nu}\right)\right)_{n=1}^{N}$.

Related Work. Since the seminal paper of Yu et al. [105] in 2002 (and the conference version in 2001), the NEP $\mathcal{G}$ has been studied in a number of works during the past nine years for the case of SISO frequency-selective channels or, equivalently, a set of parallel non-interfering scalar channels [64, 80, 81, 93, 92, 105]. Several sufficient conditions have been derived that guarantee the uniqueness of the Nash Equilibrium (NE) and the convergence of alternative distributed waterfilling based algorithms; the state-of-the-art algorithm is the asynchronous iterative waterfilling algorithm (IWFA) [92]. In this algorithm, all the users update their power allocation according to the best-response waterfilling solution (3.47) in a totally asynchronous way (in the sense of [92]), meaning that some users may update their power allocation
more frequently than others and they may even use an outdated measurement of the MUI caused from the others. These features make the asynchronous IWFA appealing for many practical scenarios, either wired or wireless, since it strongly relaxes the constraints on the synchronization of the users updates with respect to those imposed, for example, by the simultaneous or sequential updating schemes.

The main properties of the NEP $\mathcal{G}$ are summarized in Theorem 18 below, where the set $\overline{\mathcal{P}}$ is defined as $\mathcal{P}$ in (3.43) but with the power budget inequalities $\sum_{n} p_{n}^{\nu} \leq P^{\nu}$, replaced by $\sum_{n} p_{n}^{\nu}=P^{\nu}$ for all $\nu=1, \ldots, Q$, and the vector function $F(p) \triangleq\left(F_{\nu}(p)\right)_{\nu=1}^{Q}$ and the matrices $M \triangleq\left(M_{\nu \mu}\right)_{\nu, \mu=1}^{Q} \in$ $\mathbb{R}^{N Q \times N Q}$ and $\Delta \in \mathbb{R}^{Q \times Q}$ are defined as

$$
\begin{equation*}
F_{\nu}(p)=\hat{\sigma}^{\nu}+\sum_{\mu=1}^{Q} M_{\nu \mu} p^{\mu} \tag{3.48}
\end{equation*}
$$

with

$$
\hat{\sigma}^{\nu} \triangleq\left(\frac{\sigma_{n}^{\nu 2}}{\left|H_{\nu \nu}(n)\right|^{2}}\right)_{n=1}^{N} \text { and } M_{\nu \mu} \triangleq \operatorname{diag}\left\{\left(\frac{\left|H_{\nu \mu}(n)\right|^{2}}{\left|H_{\nu \nu}(n)\right|^{2}}\right)_{n=1}^{N}\right\}
$$

and

$$
[\Delta]_{\nu \mu} \triangleq \begin{cases}1, & \text { if } \nu=\mu  \tag{3.49}\\ -\max _{n} \frac{\left|H_{\nu \mu}(n)\right|^{2}}{\left|H_{\nu \nu}(n)\right|^{2}}, & \text { otherwise }\end{cases}
$$

Theorem 18. Given the NEP $\mathcal{G}=<\mathcal{P},\left(r_{i}\right)_{i=1}^{Q}>$, then the following hold.
(a) The NEP is equivalent to the affine $\operatorname{VI}(\overline{\mathcal{P}}, F)$ [64], which has a nonempty and bounded solution set;
(b) If $M$ is positive definite (semidefinite), then the $\mathrm{VI}(\overline{\mathcal{P}}, F)$ is strongly monotone (monotone); therefore, if $M$ is positive definite, $\mathcal{G}$ has a unique NE;
(c) If $\Delta$ is a P-matrix, then the asynchronous IWFA based on the waterfilling best-response (3.47) converges to the unique Nash equilibrium of $\mathcal{G}$ [92].

Theorem 18, which represents the state-of-the-art results on $\mathcal{G}$, provides a satisfactory characterization of $\mathcal{G}$ (namely, conditions for the existence/ uniqueness of the solution and global convergence of distributed algorithms) when $M$ is positive definite. Interestingly, such a condition has also a physical interpretation: it quantifies the maximum level of MUI that can be tolerated in the system for the asynchronous IWFA to converge to the (unique) NE of the game. However, the positive definiteness of $M$ may be too restrictive in practice; there are indeed networks having multiple Nash equilibria and thus for which $M$ cannot be positive definite; roughly speaking, this happens, for example, when the users are located quite close to each other. In such cases, the asynchronous IWFA is not longer guaranteed to converge, and
the calculation of even a single NE becomes a complex task. If $M$ is positive semidefinite, the $\operatorname{VI}(\overline{\mathcal{P}}, F)$ is monotone, implying that one could apply double loop schemes based on Tikhonov [32, Algorithm 12.2.9] or proximal [32, Algorithm 12.3.8] regularization to the $\operatorname{VI}(\overline{\mathcal{P}}, F)$ and still compute a Nash equilibrium of $\mathcal{G}$. The main drawback of these algorithms is that they may converge to any of the solutions of $\mathcal{G}$, meaning that there is no a priori guarantee or control about the quality of the solution they reach and thus the achievable performance of the network. This unpredictable behaviour makes the aforementioned algorithms not applicable to design real systems.

We would like instead to develop distributed solution schemes for the monotone NEP $\mathcal{G}$ that converge to the "best" NE, according to some prescribed criterion, while keeping the same desired feature of the asynchronous IWFA (e.g., distributed implementation, low signalling/coordination among the users, etc...). Up to date, no method has been proposed to select one specific NE, in case of multiple solutions. The framework here developed provides a satisfactory answer to this issue. The first step is to choose a merit function that quantifies the quality of a NE. Different heuristics can be used; as an example, here we focus on the following merit function: given the vector $w \triangleq\left(w^{\nu}\right)_{\nu=1}^{Q} \geq 0$, let

$$
\begin{equation*}
\phi(p) \triangleq \sum_{\nu=1}^{Q} w^{\nu} \sum_{\mu \neq \nu} \sum_{n=1}^{N}\left|H_{\nu \mu}(n)\right|^{2} p_{n}^{\mu} \tag{3.50}
\end{equation*}
$$

This choice is motivated by the intuition that among all the Nash equilibria of $\mathcal{G}$, a good candidate is the one that minimizes the overall interference among the users, resulting in a "higher" value of the sum-rate function $\sum_{\nu=1}^{Q} r^{\nu}(p)$. The NE selection problem based on the merit function $\phi$ can be then formulated as a hierarchical optimization problem with VI constraints:

$$
\begin{equation*}
\underset{p \in \operatorname{SOL}(\overline{\mathcal{P}}, F)}{\operatorname{minimize}} \phi(p), \tag{3.51}
\end{equation*}
$$

which is an instance of the more general VI-C HVI. We can then use the machinery developed in the previous sections to successfully study problem (3.51) and devise distributed iterative algorithms. In particular, (3.51) can be solved using Algorithm 3, where the VI-C HVI reduces to VI-C $\operatorname{HVI}(\overline{\mathcal{P}}, F, \nabla \phi, 0)$ and the sub-HVI at iteration $k$ given $p^{(k)} \in \overline{\mathcal{P}}$ corresponds to the VI

$$
\begin{equation*}
\mathrm{VI}\left(\overline{\mathcal{P}}, F+\varepsilon_{k} \nabla \phi+\alpha\left(\bullet-p^{(k)}\right)\right) \tag{3.52}
\end{equation*}
$$

The convergence of Algorithm 3 applied to (3.51) is guaranteed under the conditions of Theorem 15; note that assumptions (A, C, D, E) in the theorem are readily satisfied by (3.51), and assumption (B) is equivalent to the positive semidefiniteness of matrix $M$.

The last thing left to discuss is how to compute in a distributed way a(n approximate) solution of the sub-VIs in (3.52). We can readily use Algorithm

4; it follows from Theorem 16 indeed that Algorithm 4 globally converges to the unique solution of the VI in (3.52) if the matrix

$$
\Delta_{\alpha} \triangleq \Delta+\alpha I
$$

with $\Delta$ defined in (3.49), is a P-matrix, which is guaranteed for any $\alpha$ sufficiently large (see Corollary 1). Interestingly, such an algorithm can be implemented in a distributed way, without requiring any signalling among the users; indeed, given $\varepsilon_{k}$ and $p^{(k)}$, at each iteration, every user $\nu$ solves the quadratic optimization problem: given $p^{-\nu} \geq 0$,

$$
\begin{array}{cl}
\underset{p^{\nu} \in \overline{\mathcal{P}}^{\nu}}{\operatorname{minimize}} & \frac{1}{2}\left\|p^{\nu}+\left(\hat{\sigma}^{\nu}+\sum_{\mu \neq \nu}^{Q} M_{\nu \mu} p^{\mu}\right)\right\|^{2}+\varepsilon_{k} \gamma^{\nu T} p^{\nu}  \tag{3.53}\\
& +\frac{\alpha}{2}\left\|p^{\nu}-p^{(k), \nu}\right\|^{2}
\end{array}
$$

where $\gamma^{\nu} \triangleq\left(\sum_{\mu \neq \nu} w^{\mu}\left|H_{\mu \nu}(n)\right|^{2}\right)_{n=1}^{N}$. The solution of (3.53), has a similar waterfilling-like expression as (3.47) and thus can be efficiently and locally computed, given the MUI $\hat{\sigma}^{\nu}+\sum_{\mu \neq \nu}^{Q} M_{\nu \mu} p^{\mu}$ measured at each receiver $\nu$, $p^{(k), \nu}$, and $\varepsilon_{k}$. The asynchronous implementation of Algorithm 4 is also possible, whose convergence is guaranteed under the same P property of matrix $\Delta_{\alpha}$.

Remark 9. Overall, in Algorithm 3 applied to (3.51), there are two levels of updates: 1) the computation of the users' optimal power allocations [the (approximate) solution of the sub-VI (3.52)], given $p^{(k), \nu} \in \overline{\mathcal{P}}$ and $\varepsilon_{k}$; and 2) the updates of $p^{(k), \nu}$ and $\varepsilon_{k}$, after checking that the termination condition in Step 2 is satisfied. The former can be performed locally and distributively by the users as previously discussed. The check of the termination condition in the latter can be certainly accomplished but it is a rather technical issue and we refer to [91] for practical implementations of this check, under different level of signalling and supervision. It turns out that Algorithm 3 has the same desired properties as the asynchronous IWFA [92]: it is fairly distributed and requires a limited signalling/coordination among the users, which makes it appealing for a practical implementation in telecommunication networks.

Numerical Results. We compare now the performance of our new algorithm with those of the IWFA, which is the state-of-the-art algorithm proposed in the literature $[64,93,92,91]$ for solving the rate maximization game $\mathcal{G}$.

In Figure 3.9, we plot the sum-rate of the users $\sum_{\nu=1}^{Q} r^{\nu}(p)$ versus the iteration index, achieved by the following algorithms: i) The simultaneous IWFA [93] based on the waterfilling map $\operatorname{wf}(p)$ (dashed-dot line curve); ii) Algorithm 3 applied to (3.51) (solid line curve), with $w=1$; and iii) Algorithm 3 applied to (3.51), where $\phi(p)$ in (3.50) is replaced by $-\phi(p)$ and $w=1$. The solution of each sub-VI (3.52) in the inner loop of Algorithm 3 is computed using Algorithm 4. The choice of the merit function $-\phi(p)$ leads to the selection of the NE solution that maximizes the overall MUI in the
system, which provides a benchmark of the sum-rate variability and an estimate of the worst-case performance over the set of the Nash equilibria of the game $\mathcal{G}$. We count one iteration of Algorithm 3 as one iteration of the inner Jacobi scheme (Algorithm 4), which corresponds to a physical transmission of the users.

We examined the behaviour of the above algorithms under the following setup. We considered an ad-hoc network where there are $Q=25$ active users; the (cross-)channels among the links are simulated as FIR filter of order $\mathrm{L}=$ 10 , where each tap is a zero mean complex Gaussian random variable with variance equal to $1 / L$; the channel transfer functions are the FFT of the corresponding impulse responses over $N=128$ points (carriers). We focused on two scenarios, namely low/medium interference and high interference scenario. Low/medium interference means that the channel realizations are such that the matrix $M$ is positive definite, whereas in the high interference case $M$ is positive semidefinite. The thermal noise variance $\sigma_{n}^{\nu 2}$ is set to one for all $n$ and $\nu$, and the transmit power budget $P^{\nu}$ is chosen so that the Signal-to-Noise-Ratio (SNR) of each user $\mathrm{SNR}^{\nu} \triangleq 10 \log _{10}\left(P^{\nu} / \sigma_{n}^{\nu 2}\right)=5 \mathrm{~dB}$ for all $\nu$ and $n$. All the algorithms are initialized by the same starting point, chosen randomly in the set $\mathcal{P}$, and are terminated when the Euclidean norm of the error in two consecutive iterations becomes smaller than $10^{-6}$. In the inner loop of Algorithm 3, we chose the center $p^{(0)}$ of the regularization randomly in $\overline{\mathcal{P}}, \alpha=3.5$, and $\varepsilon_{k}=\varepsilon_{0} /(1+k)$, where $\varepsilon_{0}=0.5$ and $k$ is the iteration index of the outer loop; the termination criterion of the inner loop is the same as the outer loop. The above choice of the free parameters is the result of some preliminary tests; furthermore we have to remark that these choices are tailored to this specific problem and do not take account of the general considerations made in the previous sections. It is important to underline however that the proposed algorithm has been observed to be robust against the variation of the aforementioned parameters. Finally, note that, in case of multiple solutions, the IWFA is not guaranteed to converge [93]. In fact, in our experiments, we sometimes observed oscillating behaviours of the IWFA. In such cases, following a common approach in signal processing, we just terminated the algorithm after 250 iterations and used as power allocation for the transmission of the users the value obtained in the last iteration.

The following comments are in order from Figure 3.9. In the case of multiple Nash equilibria of the game $\mathcal{G}$ (high interference scenario), the sum-rate performance of the network can vary significantly over the set of the Nash equilibria; the relative sum-rate gap between the"worst" and "best" Nash equilibrium is around $90 \%$. Interestingly, our algorithm is shown to significantly outperform the classical IWFA, which validates our heuristic (3.50) in choosing the Nash equilibrium. When the NE of the game is unique, as expected, both the IWFA and our algorithm converge to the same sumrate solution. Finally, note that even though our algorithm is in principle a double-loop scheme, its convergence speed (measured in terms of number of iterations required to reach the desired error accuracy) is as the same order
as the simultaneous (single loop) IWFA; this is due to the fact that the bestresponse Jacobi scheme used in the inner loop converges quite fast, typically in three/four iterations.

Figure 3.9 refers to a specific channel scenario (realization); nevertheless, it has been observed that the qualitative behaviour of the simulated algorithms is almost independent of the specific channel realization (provided that the matrix $M$ remains positive semidefinite), making the conclusions drawn from Figure 3.9 very general. This is confirmed also by Figure 3.10, where we provide the average performance of the algorithms considered in Figure 3.9. More specifically, in Figure 3.10(a) we plotted the average sumrate versus the $\mathrm{SNR} \triangleq 10 \log _{10}(P)$, with $P^{\nu}=P$ and $\sigma_{n}^{\nu 2}=1$ for all $\nu$ and $k$, achievable at the NE reached by the simultaneous IWFA and our Algorithm 3; for each value of the SNR, the curves are averaged over 5000 random channel realizations, chosen so that the matrix $M$ is positive semidefinite. In Figure 3.10 (b) we plotted the outage sum-rate (which is the probability that $\sum_{\nu=1}^{Q} r^{\nu}(p) \geq \mathrm{sr}$ ) versus sr, achieved by the aforementioned algorithms, for a SNR $\triangleq P=5 \mathrm{~dB}$; as in Figure $3.10(\mathrm{~b})$, the outage probability has been estimated using 5000 channel realizations, chosen so that the matrix $M$ is positive semidefinite. The free parameters are chosen as in Figure 3.9. The outage probability provides a quantitative indication of the dispersion of the sum-rate values (as a function of the random channels) around the mean value: the higher the slope of the outage curves, the less the dispersion of the sum-rate around its mean. We would like to have each sum-rate realization as close as possible to its mean, so that the average performance as given in Figure 3.10(a) are meaningful in practice. Note that the discrepancy in the sum-rate gap between Algorithm 3 and the IWFA as observed in Figure 3.9 and Figure 3.10 is due to the oscillating behaviour of the IWFA experienced for some channel realizations. The following remarks are in order from Figure 3.10. As expected, the performance of all the examined algorithms are almost the same in low SNR regime (roughly speaking when $\mathrm{SNR} \leq-5 \mathrm{~dB}$ ), since in that regime the thermal noise dominates the MUI term at the denominator of the SINR, and thus the pairs behave like decoupled links; for medium/high SNR's (i.e., when the MUI is the dominant factor), the proposed algorithm significantly outperforms the state-of-the-art IWFA, which makes it a good candidate for the design of infrastructureless networks.


Fig. 3.9: Comparison of distributed algorithms solving the game $\mathcal{G}$ : Sum-rate of the users versus the iteration index, achieved by the simultaneous IWFA (dashed-dot line curve), Algorithm 3 based on the outer function $\phi(p)$ in (3.50) (solid line curve), and Algorithm 3 based on the outer function $-\phi(p)$ (dashed line curve), in the low/medium interference regime (i.e., $M$ is positive definite) and high interference regime (i.e., $M$ is positive semidefinite).


Fig. 3.10: Average sum-rate versus the SNR [subplot (a)] and outage sum-rate [subplot(b)] at the Nash equilibria of $\mathcal{G}$ : simultaneous IWFA (dashed-dot line curve), Algorithm 3 based on the outer function $\phi(p)$ in (3.50) (solid line curve), and Algorithm 3 based on the outer function $-\phi(p)$ (dashed line curve).

## Conclusions

In this thesis we have proposed solution methods for Generalized Nash Equilibrium Problems (GNEPs) and Hemivariational Inequalities (HVIs) where the feasible set is given by the intersection of a closed convex set with the solution set of a lower-level monotone Variational Inequality (VI).
The Generalized Nash Equilibrium Problem (GNEP) is a central model in game theory that has been used to model problems from many different fields. It turns out that, until very recently, there were no provably convergent algorithms for the solution of GNEPs. In this work we have considered one of the most promising approaches for which we are able to prove convergence results: penalty methods. Penalty algorithms have originally been developed and studied for the solution of constrained (standard) optimization problems and their generalization to the case of GNEPs is not straightforward. Weaker theoretical results can be expected due to the variable nature of the players' feasible sets.

In order to get rid of the "difficult" coupling constraints, we have considered a "partial" exact penalty method which allows us to reduce the original GNEP to a nonsmooth standard NEP. After having established some relevant general theoretical results, we were finally able, for the first time, to identify suitable and non trivial classes of GNEPs for which the proposed scheme is guaranteed to converge to a solution of the original game.

HemiVariational Inequalities (HVI)s are a powerful modeling tool that encompasses both (convex) optimization and variational inequalities as special cases. In particular, we have considered the Variational InequalityConstrained HemiVariational Inequality (VI-C HVI) with side constraints, thus an HVI for which the feasible set is implicitly defined as the intersection of a closed convex set with the solution set of a lower-level monotone variational inequality; this problem is, in turn, a generalization of a hierarchical optimization problem and includes also, as special case, the problem of selecting a particular equilibrium solution to optimize an auxiliary ("upper") objective function. In this work, we presented centralized and distributed al-
gorithms for the numerical solution of such problems. The algorithms consist of a main loop wherein a sequence of one-level, strongly monotone HVIs is solved that involve the penalization of the non-VI constraint and a combination of proximal and Tikhonov regularization to handle the lower-level VI constraints. Subsequently, the methods developed are used to successfully solve a new power control problem in ad-hoc networks. Finally the analysis is equipped with a series of numerical tests and results which show the outstanding behavior of the proposed approach.

To the best of our knowledge, these contributions are new and considerably expand existing results. The proposed distributed algorithm, in which we are interested as motivated by applications in non-cooperative game problems is novel even for a hierarchical optimization problem. Furthermore, the power control problem that we have analyzed is new and our results expand the applicability and flexibility of game-theoretic models in ad-hoc networks and also bring considerable gains over existing techniques.

## Appendix A Some considerations on constraint qualifications

In this appendix, we recall definitions and properties of some relevant constraint qualifications. Furthermore we briefly investigate their consequences and the relations to each other.

Let us consider, for sake of simplicity, the jointly convex GNEP where each player's problem is the following

$$
\begin{array}{ll}
\min _{x^{\nu}} & \theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right) \\
\text { s.t. } & g\left(x^{\nu}, \mathbf{x}^{-\nu}\right) \leq 0, \\
& x \in K_{\nu} .
\end{array}
$$

Definition 7. The Slater condition holds if there exists a point $\hat{x} \in \mathbf{K}$ such that $g(\hat{x})<0$.

We recall that, in case of optimization problems with convex and differentiable (inequality) constraints, the Mangasarian-Fromovitz Constrain Qualification (MFCQ) and the Slater's one are equivalent. In the GNEP case, instead, the EMFCQ (at feasible points) is stronger than a Slater-type condition. In fact, consider a feasible point $x$ and assume that it satisfies the EMFCQ. By setting $d:=\left(d^{1}, \ldots, d^{N}\right)$, it is clear that we have $\nabla_{x} g(x)^{T} d<0$ and in turn, it is classical to show that this implies Slaters condition. On the contrary, let us consider a GNEP in which there are two players, each one controlling one variable, and assume $m=1$ with $g^{1}(\mathbf{x})=g^{2}(\mathbf{x})=\left(\mathbf{x}^{1}\right)^{2}+\left(\mathbf{x}^{2}\right)^{2}-1$. The Slater-type condition holds (considering the origin for example). On the other hand, at point $(1,0), \nabla_{x^{2}} g(\mathbf{x})=0$ and EMFCQ does not hold.

We recall the KKT conditions for the player $\nu$ 's optimization problem (2.2) (see [88]):

$$
\begin{gather*}
-\left[\nabla_{x^{\nu}} \theta_{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)+\sum_{j=1}^{m_{\nu}} v_{j}^{\nu} \nabla_{x^{\nu}} g_{j}^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right)\right] \in N_{K_{\nu}}\left(x^{\nu}\right)  \tag{A-1}\\
0 \leq v^{\nu} \perp g^{\nu}\left(x^{\nu}, \mathbf{x}^{-\nu}\right) \leq 0
\end{gather*}
$$

where $v^{\nu} \in \mathbb{R}^{m_{\nu}}$ and $x^{\nu} \in K_{\nu}$. We finally recall two more constraint qualifications.

Definition 8. We say that the sequentially bounded constraint qualification (SBCQ) [32] holds at $\overline{\mathbf{x}}$, solution of the GNEP (2.2), if, for every sequence of vectors $\left\{\mathbf{x}^{k}\right\}$ converging to $\overline{\mathbf{x}}$ and such that, for every $\nu,\left(x^{k}\right)^{\nu}$ is a solution of the player $\nu$ 's optimization problem given the other players' strategies $\left(x^{k}\right)^{-\nu}$ for every $k$, there exists a bounded sequence of multipliers $\left\{\left(v^{k}\right)^{\nu}\right\}$ such that $\left\{\left(v^{k}\right)^{\nu}\right\}$ satisfies the KKT (A-1) conditions corresponding to $x^{k}$ for every $k$.

The latter condition is implied by the following constraint qualification.
Definition 9. The Generalized Sequentially Bounded Constraint Qualification (GSBCQ) (see [79]) holds if, for every player $\nu=1, \ldots, N$ and for every bounded sequence $\left\{\mathbf{x}^{k}\right\}$ such that $\left(x^{k}\right)^{\nu}$ is a solution of the player $\nu$ 's optimization problem (2.2) given the other players' strategies $\left(x^{k}\right)^{-\nu}$, there exists a bounded sequence $\left\{\left(v^{k}\right)^{\nu}\right\}$ such that $\left\{\left(v^{k}\right)^{\nu}\right\}$ satisfies the KKT conditions (A-1) corresponding to $\mathrm{x}^{k}$ for every $k$.
Regarding this definition we would like to mention that in general the solution set of a GNEP is not closed. For example, let us consider the following two players game:

$$
\begin{array}{clr}
\min _{x^{1}}-x^{1} & \min _{x^{2}}\left(x^{2}-x^{1}\right)^{2} \\
& 0 \leq x^{1} \leq 1 & 0 \leq x^{2} \leq 1  \tag{A-2}\\
& x^{1} x^{2} \leq\left(x^{2}\right)^{2} &
\end{array}
$$

with $x^{1}, x^{2} \in \mathbb{R}$. Consider the sequence of solutions $z^{k}=\left(\frac{1}{k}, \frac{1}{k}\right) \rightarrow(0,0)$ : $(0,0)$ is not a game's solution. This example shows that, in general the GNEP's solution set is not closed.

We note that, by the GSBCQ, by the KKT conditions (A-1) with $\mathbf{x}^{k} \rightarrow \overline{\mathbf{x}}$ (subsequencing if necessary), we have the closedness of the solutions' set taking into account the outer semicontinuity of the mapping $N_{K_{\nu}}\left(x^{\nu}\right)$ (see [88], Proposition 6.6). Furthermore, by usual reasonings, it can be proven that if the EMFCQ holds then GSBCQ holds too.

Theorem 19. If the EMFCQ holds on $\mathbf{K}$, then $G S B C Q$ holds.
Proof. Suppose by contradiction that there exists a player $\nu$ and a bounded sequence $\left\{\mathbf{x}^{k}\right\}$ of solutions of the game for which every sequence of multipliers $\left\{\left(v^{k}\right)^{\nu}\right\}$ (such that $\left\{\left(v^{k}\right)^{\nu}\right\}$ satisfies KKT conditions) is unbounded, i.e., $\left\|\left(v^{k}\right)^{\nu}\right\| \rightarrow \infty$. For every $k$, KKT conditions hold:

$$
\begin{gather*}
-\left[\nabla_{x^{\nu}} \theta_{\nu}\left(\mathbf{x}^{k}\right)+\nabla_{x^{\nu}} g^{\nu}\left(\mathbf{x}^{k}\right)\left(v^{k}\right)^{\nu}\right] \in N_{K_{\nu}}\left(\left(x^{k}\right)^{\nu}\right)  \tag{A-3}\\
0 \leq\left(v^{k}\right)^{\nu} \perp g^{\nu}\left(\mathbf{x}^{k}\right) \leq 0
\end{gather*}
$$

Subsequencing if necessary, we may assume that the entire sequence $\left\{\mathbf{x}^{k}\right\}$ converges to $\overline{\mathbf{x}}$. Then we have,

$$
\begin{equation*}
\lim _{k}\left\{-\frac{\nabla_{x^{\nu}} \theta_{\nu}\left(\mathbf{x}^{k}\right)}{\left\|\left(v^{k}\right)^{\nu}\right\|}-\frac{\nabla_{x^{\nu}} g^{\nu}\left(\mathbf{x}^{k}\right)\left(v^{k}\right)^{\nu}}{\left\|\left(v^{k}\right)^{\nu}\right\|}\right\}=-\nabla_{x^{\nu}} g^{\nu}(\overline{\mathbf{x}}) \bar{w}^{\nu} \in N_{K_{\nu}}\left(\bar{x}^{\nu}\right) \tag{A-4}
\end{equation*}
$$

with $0 \leq \bar{w}^{\nu} \perp g^{\nu}(\mathbf{x}) \leq 0$ and $\bar{w}^{\nu} \neq 0$, a contradiction to (2.14).
In view of the considerations made in Chapter 2 and here, we resume the properties of and the relations between the analyzed constraint qualifications in the following diagram.


Fig. A.1: Some constraint qualifications' relations.

## Appendix B Test problems

In this appendix, we present the collection of feasible sets and lower and upper functions that we have considered in our simulations. Each problem is an instance of $\operatorname{HVI}(K, F, \Phi)$ : thus each test problem is "individuated" by the choice for a particular feasible set, a particular lower and a particular upper function.

We have identified each feasible set, each lower function $F$ and each upper function $\Phi$ with an alphanumeric tag: for example, the tag K1 refers to the feasible set $K 1$, the tag F11 to the lower function $F 11$, the tag $\Phi 1$ to the upper function $\Phi 1$ and so on. We remark that the lower functions' tag could be formed by two digits; in such cases, the first digit refers to the corresponding feasible set: indeed, we will see in the following sections that some of the lower functions, when considered in combination with a particular feasible set, allows the explicit calculation of the solution set $\operatorname{SOL}(K, F)$. Finally, the test problem name is a suitable combination of the three tags that we have previously introduced. For example, problem $\mathrm{K} 1 \mathrm{~F} 11 \Phi 1$ is the VI-C $\operatorname{VI}(K 1, F 11, \Phi 1)$.

## B. 1 Lower feasible set K of VI(K, F)

Each feasible set $K=\left\{x \in \mathbb{R}^{n}: A x-b \leq 0\right\}$, with $A \in M_{m, n}$ and $b \in \mathbb{R}^{n}$, considered is a compact polyhedron.

## K1

Problem Dimensions:

- $n$
$-m=n+1$


## Constraints matrix and constant terms vector:

$$
-A=\left[\begin{array}{rrrrr}
1 & 1 & \cdots & \cdots & 1 \\
-1 & 0 & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & 0 \\
0 & 0 & \cdots & 0 & -1
\end{array}\right] \quad b=\left[\begin{array}{c}
1 \\
0 \\
\vdots \\
\vdots \\
0
\end{array}\right]
$$

## Further Informations:

- This is the $n$ dimensions simplex:

$$
K 1=\left\{x \in \mathbb{R}^{n}: x_{1}+\cdots+x_{n} \leq 1, x_{i} \geq 0, i=1, \ldots, n\right\}
$$

K2

## Problem Dimensions:

- $n$
$-m=2 n$
Constraints matrix and constant terms vector:
$-A=\left[\begin{array}{rrrrr}1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1\end{array}\right] \quad b=\left[\begin{array}{c}1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 1 \\ 0\end{array}\right]$


## Further Informations:

- In this case the feasible set is the "box" $K 2=\left\{x \in \mathbb{R}^{n}: 0_{n, 1} \leq x \leq e\right\}$, where $0_{n, 1} \in \mathbb{R}^{n}$ is the zero vector and $\mathbb{R}^{n} \ni e=(11 \ldots 1)^{T}$.

Problem Dimensions:

- $n$
$-m=2 n$

Constraints matrix and constant terms vector:
$-A=\left[\begin{array}{rrrrr}1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1\end{array}\right] \quad b=\left[\begin{array}{c}u \\ -l \\ u \\ -l \\ u \\ -l \\ \vdots \\ \vdots \\ u \\ -l\end{array}\right]$

## Further Informations:

- In this case the feasible set K is the "box" $K 3=\left\{x \in \mathbb{R}^{n}: l e \leq x \leq u e\right\}$, with $\mathbb{R}^{n} \ni e=(11 \ldots 1)^{T}$.
K4


## Problem Dimensions:

$-n$
$-m=2 n$

## Constraints matrix and constant terms vector:

$$
-A=\left[\begin{array}{rrrrr}
1 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & \ddots & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & -1
\end{array}\right] \quad b=\left[\begin{array}{c}
u_{1} \\
-l_{1} \\
u_{2} \\
-l_{2} \\
u_{3} \\
-l_{3} \\
\vdots \\
\vdots \\
u_{n} \\
-l_{n}
\end{array}\right]
$$

## Further Informations:

- In this case the feasible set K is the "box" $K 4=\left\{x \in \mathbb{R}^{n}: l \leq x \leq u\right\}$, with $\mathbb{R}^{n} \ni l=\left(l_{1} \ldots l_{n}\right)^{T}$ and $\mathbb{R}^{n} \ni u=\left(u_{1} \ldots u_{n}\right)^{T}$.
$-n$
$-m=n+1$


## Constraints matrix and constant terms vector:

$$
-A=\left[\begin{array}{rrrrr}
2 & 1 & \cdots & \cdots & 1 \\
-1 & 0 & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & 0 \\
0 & 0 & \cdots & 0 & -1
\end{array}\right] \quad b=\left[\begin{array}{c}
2 \\
0 \\
\vdots \\
\vdots \\
0
\end{array}\right]
$$

## Further Informations:

- In this case the feasible set K is the polyhedron:

$$
K 5=\left\{x \in \mathbb{R}^{n}: 2 x_{1}+x_{2}+\cdots+x_{n} \leq 2, x_{i} \geq 0, i=1, \ldots, n\right\}
$$

K6

## Problem Dimensions:

- $n$
$-m=n+1$
Constraints matrix and constant terms vector:
$-A=\left[\begin{array}{rrrrr}a_{1} & a_{2} & \cdots & \cdots & a_{n} \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & -1\end{array}\right] \quad b=\left[\begin{array}{c}b_{1} \\ 0 \\ \vdots \\ \vdots \\ 0\end{array}\right], \quad$ with $a_{i} \geq 0$ and $b_{1} \geq 0$.


## Further Informations:

- In this case the feasible set K is the polyhedron:

$$
K 6=\left\{x \in \mathbb{R}^{n}: a_{1} x_{1}+a_{2} x_{2}+\cdots+a_{n} x_{n} \leq b_{1}, x_{i} \geq 0, i=1, \ldots, n\right\}
$$

## B. 2 Lower function F of $\mathrm{VI}(\mathrm{K}, \mathrm{F})$

Each function $F$ considered is continuous and monotone.

## F11

## Lower function:

$-F 11(x)=\left[\frac{1}{2}\left(x_{1}+\ldots+x_{n}\right)^{2}-1\right]\left[\begin{array}{c}1 \\ \vdots \\ 1\end{array}\right]$

## Derivatives:

$$
-J F 11(x)=\left(x_{1}+\ldots+x_{n}\right)\left[\begin{array}{ccc}
1 & \cdots & 1 \\
\vdots & \ddots & \vdots \\
1 & \cdots & 1
\end{array}\right]
$$

## Further Informations:

- When F11 is considered in combination with the feasible set $K 1$, we have $\operatorname{SOL}(K 1, F 11)=\left\{x \in \mathbb{R}^{n}: x_{1}+\ldots+x_{n}=1, x_{i} \geq 0, i=1, \ldots, n\right\}$ (see F61 for more details).


## F12

## Lower function:

$-F 12(x)=\left[\begin{array}{c}-8 \\ \vdots \\ -8\end{array}\right]$

## Derivatives:

- JF12( $x$ ) $=0_{n, n}$


## Further Informations:

- When F12 is considered in combination with the feasible set $K 1$, we have $\operatorname{SOL}(K 1, F 12)=\left\{x \in \mathbb{R}^{n}: x_{1}+\ldots+x_{n}=1, x_{i} \geq 0, i=1, \ldots, n\right\}$ (see F61 for more details).


## F13

## Lower function:

$$
-F 13(x)=\left[\begin{array}{c}
-x_{2}-x_{3}-\ldots-x_{n} \\
x_{1}-1 \\
\vdots \\
x_{1}-1
\end{array}\right]
$$

## Derivatives:

$-J F 13(x)=\left[\begin{array}{cccc}0 & -1 & \cdots & -1 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & 0 & 0\end{array}\right]$

## Further Informations:

- When F13 is considered in combination with the feasible set $K 1$, we have $\operatorname{SOL}(K 1, F 13)=\left\{x \in \mathbb{R}^{n}: x_{1}+\ldots+x_{n}=1, x_{i} \geq 0, i=1, \ldots, n\right\}$ (see F61 for more details).


## F21

## Lower function:

$$
-F 21(x)=\left[\begin{array}{c}
\frac{1}{2} x_{1}^{2}-1 \\
0 \\
\vdots \\
0
\end{array}\right]
$$

## Derivatives:

$-J F 21(x)=\left[\begin{array}{cccc}x_{1} & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0\end{array}\right]$

## Further Informations:

- When F21 is considered in combination with the feasible set $K 2$, we have $\operatorname{SOL}(K 2, F 21)=\left\{x \in \mathbb{R}^{n}: x_{1}=1,0 \leq x_{j} \leq 1,, j=1, \ldots, n, j \neq 1\right\}$ (see F61 for more details).


## F51

## Lower function:

$-F 51(x)=\left[\begin{array}{c}x_{2}+x_{3}+\ldots+x_{n}-2 \\ -x_{1} \\ \vdots \\ -x_{1}\end{array}\right]$

## Derivatives:

$-J F 51(x)=\left[\begin{array}{cccc}0 & 1 & \cdots & 1 \\ -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ -1 & 0 & 0 & 0\end{array}\right]$

## Further Informations:

- When F51 is considered in combination with the feasible set $K 5$, we have $\operatorname{SOL}(K 5, F 51)=\left\{x \in \mathbb{R}^{n}: 2 x_{1}+x_{2}+\ldots+x_{n}=2, x_{i} \geq 0, i=1, \ldots, n\right\}$ (see F61 for more details).


## F52

## Lower function:

$-F 52(x)=\left[\begin{array}{c}\frac{2}{3}\left(2 x_{1}+\ldots+x_{n}\right)^{3}-12 \\ \frac{1}{3}\left(2 x_{1}+\ldots+x_{n}\right)^{3}-6 \\ \vdots \\ \frac{1}{3}\left(2 x_{1}+\ldots+x_{n}\right)^{3}-6\end{array}\right]$

## Derivatives:

$$
-J F 52(x)=\left[\begin{array}{ccc}
4\left(2 x_{1}+\ldots+x_{n}\right)^{2} & \cdots & 2\left(2 x_{1}+\ldots+x_{n}\right)^{2} \\
2\left(2 x_{1}+\ldots+x_{n}\right)^{2} & \cdots & \left(2 x_{1}+\ldots+x_{n}\right)^{2} \\
\vdots & \ddots & \vdots \\
2\left(2 x_{1}+\ldots+x_{n}\right)^{2} & \cdots & \left(2 x_{1}+\ldots+x_{n}\right)^{2}
\end{array}\right]
$$

## Further Informations:

- When F52 is considered in combination with the feasible set $K 5$, we have $\operatorname{SOL}(K 5, F 52)=\left\{x \in \mathbb{R}^{n}: 2 x_{1}+x_{2}+\ldots+x_{n}=2, x_{i} \geq 0, i=1, \ldots, n\right\}$ (see F61 for more details).


## F61

## Lower function:

$$
-F 61(x)=\left[\begin{array}{c}
a_{2} x_{2}+\ldots+a_{n} x_{n}-b_{1} \\
-a_{2} x_{1} \\
\vdots \\
-a_{n} x_{1}
\end{array}\right]
$$

## Derivatives:

$$
-J F 61(x)=\left[\begin{array}{cccc}
0 & a_{2} & \cdots & a_{n} \\
-a_{2} & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
-a_{n} & 0 & 0 & 0
\end{array}\right]
$$

## Further Informations:

- When F61 is considered in combination with the feasible set $K 6$, we have $\operatorname{SOL}(K 6, F 61)=\left\{x \in \mathbb{R}^{n}: a_{1} x_{1}+\ldots+a_{n} x_{n}=b_{1}, x_{i} \geq 0, i=1, \ldots, n\right\}$. Indeed for $x \in \mathbb{R}^{n}$ such that $x \in \tilde{K} 1=\left\{a_{1} x_{1}+\ldots+a_{n} x_{n}=b_{1}, x_{i} \geq\right.$ $0, i=1, \ldots, n\}$,

$$
F 61(x)=\left[\begin{array}{c}
-a_{1} x_{1} \\
-a_{2} x_{1} \\
\vdots \\
-a_{n} x_{1}
\end{array}\right]
$$

Therefore, for such points, $-F 61(x) \in N_{K}(x)$.

## F7

## Lower function:

$-F 7(x)=\left[\begin{array}{c}0 \\ \vdots \\ 0\end{array}\right]$

## Derivatives:

- $J F 7(x)=0_{n, n}$


## Further Informations:

- We have $\operatorname{SOL}(K, F 7)=K$, for each $K$.


## Fr

## Lower function:

$-\operatorname{Fr}(x)=P x+c$

## Derivatives:

$-\operatorname{JFr}(x)=P$

## Further Informations:

- Fr is a randomly generated monotone linear function.
- The procedure for the random generation of monotone linear function has been proposed in [48]. In this case we have chosen $P=B-B^{T}+D$ with $B \in M_{n}$ upper triangular and $D \in M_{n}$ positive semidefinite diagonal matrix. Thus, if $B$ is not the null matrix, $P$ would not be symmetric. The eigenvalues of the symmetric part of $P$ are the diagonal entries of $D$.
- The diagonal elements of $D$ (at least one is zero in order to obtain nothing more than a monotone function) and the components of the vector $c$ are drawn from the standard uniform distribution on the interval $(0,10]$, those of $B$ on the interval $(0,0.1]$.


## B. 3 Upper function $\Phi$

Each function $\Phi$ considered is continuous and strongly monotone.

```
$1
```


## Upper function:

$-\Phi 1(x)=x$

## Derivatives:

$-J \Phi 1(x)=I_{n, n}$

## Further Informations:

- This function is the gradient of $0.5 x_{1}^{2}+\ldots+0.5 x_{n}^{2}+c$.


## $\Phi 2$

## Upper function:

$-\Phi 2(x)=x-\left[\begin{array}{c}0.1 \\ \vdots \\ 0.1\end{array}\right]$

## Derivatives:

$-J \Phi 2(x)=I_{n, n}$

## Further Informations:

- This function is the gradient of $0.5\left(x_{1}-0.1\right)^{2}+\ldots+0.5\left(x_{n}-0.1\right)^{2}+c$.


## Upper function:

$-\Phi 3(x)=x-\left[\begin{array}{c}0 \\ \vdots \\ 0 \\ 4\end{array}\right]$

## Derivatives:

$-J \Phi 3(x)=I_{n, n}$

## Further Informations:

- This function is the gradient of $0.5 x_{1}^{2}+\ldots+0.5 x_{n-1}^{2}+0.5\left(x_{n}-4\right)^{2}+c$.


## $\Phi 4$

## Upper function:

$-\Phi 4(x)=\left[0.5\left(x_{i}+0.5\right) e^{0.1\left(x_{i}+0.5\right)^{2}}\right]_{i=1}^{n}$
Derivatives:
$-J \Phi 4(x)=\operatorname{diag}\left[0.5\left(1+0.2\left[\left(x_{i}+0.5\right)^{2}\right)\right] e^{0.1\left[\left(x_{i}+0.5\right)^{2}\right]}\right]_{i=1}^{n}$

## Further Informations:

- The Jacobian of this function is symmetric and uniformly positive definite over $K$.

```
$5
```


## Upper function:

$-\Phi 5(x)=\left[0.3 e^{\left(0.15 x_{i}\right)}\right]_{i=1}^{n}$

## Derivatives:

$-J \Phi 5(x)=\operatorname{diag}\left[0.05 e^{\left(0.15 x_{i}\right)}\right]_{i=1}^{n}$

## Further Informations:

- The Jacobian of this function is symmetric and uniformly positive definite over $K$.


## $\Phi 6$

## Upper function:

$$
-\Phi 6(x)=\left[\begin{array}{c}
0.3 e^{0.15 x_{1}}+x_{2}+x_{3}+\ldots+x_{n} \\
0.3 e^{0.15 x_{2}}-x_{1}+x_{3}+\ldots+x_{n} \\
\vdots \\
0.3 e^{0.15 x_{n}}-x_{1}-x_{2}-\ldots-x_{n-1}
\end{array}\right]
$$

## Derivatives:

$-J \Phi 6(x)=\left[\begin{array}{ccccc}0.05 e^{0.15 x_{1}} & 1 & \cdots & \cdots & 1 \\ -1 & 0.05 e^{0.15 x_{2}} & 1 & \cdots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -1 & -1 & \cdots & -1 & 0.05 e^{0.15 x_{n}}\end{array}\right]$

## Further Informations:

- The symmetric part of the Jacobian of this function is the uniformly positive definite (on $K$ ) matrix $\operatorname{diag}\left[0.05 e^{\left(0.15 x_{i}\right)}\right]_{i=1}^{n}$.


## $\Phi 7$

## Upper function:

$$
-\Phi 7(x)=\Phi 1(x)+\left[\begin{array}{c}
x_{1}^{2}\left(x_{2}+\ldots+x_{n}\right) \\
-1 / 3 x_{1}^{3}+x_{2}\left(x_{3}+\ldots+x_{n}\right) \\
-1 / 3 x_{1}^{3}-1 / 2 x_{2}^{2}+x_{3}\left(x_{4}+\ldots+x_{n}\right) \\
\vdots \\
-1 / 3 x_{1}^{3}-1 / 2 x_{2}^{2}-\ldots-1 / 2 x_{n-1}^{2}
\end{array}\right]
$$

## Derivatives:

$$
-J \Phi 7(x)=I_{n, n}+\left[\begin{array}{ccccc}
2 x_{1}\left(x_{2}+\ldots+x_{n}\right) & x_{1}^{2} & \cdots & \cdots & x_{1}^{2} \\
-x_{1}^{2} & \left(x_{3}+\ldots+x_{n}\right) & x_{2} & \cdots & x_{2} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
-x_{1}^{2} & -x_{2} & \cdots & -x_{n-1} & 0
\end{array}\right]
$$

## Further Informations:

- The symmetric part of the Jacobian of this function, if $x \in K$ implies $x \geq 0$, is the uniformly positive definite (on $K$ ) matrix:

$$
\left[\begin{array}{ccccc}
1+2 x_{1}\left(x_{2}+\ldots+x_{n}\right) & 0 & \cdots & \cdots & 0 \\
0 & 1+\left(x_{3}+\ldots+x_{n}\right) & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & 0 & 1
\end{array}\right]
$$

## $\Phi$ r

## Upper function:

$$
-\Phi r(x)=M x+b
$$

## Derivatives:

$-J \Phi r(x)=M$

## Further Informations:

- $\Phi r$ is a randomly generated monotone linear function.
- The procedure for the random generation of monotone linear function has been proposed in [48]. In this case we have chosen $M=A^{T} A+B-B^{T}+D$ with $A \in M_{n}, B \in M_{n}$ upper triangular and $D \in M_{n}$ positive definite diagonal matrix. If $B$ is not the null matrix, $J \Phi r$ would not be symmetric.
- The elements of $A$ and $B$ are drawn from the standard uniform distribution on the open interval $(-5,5)$, those of $D$ on the interval $(0,0.3)$.


# Appendix C <br> Algorithm performances 

## C. 1 Centralized exact version

For each problem we have implemented the algorithm using the parameter $e_{k}=10^{-5}$ for every $k$ (thus solving the strongly monotone inner problem exactly at each outer iteration $k$ ) and for fixed values of $\varepsilon_{0}$, with $\varepsilon_{0}$ equal to $0.4,0.8,1.6,3.2,6.4$ respectively; for each fixed value of $\varepsilon_{0}$ we have considered several values for parameter $\alpha$, namely $0.1,0.3,0.9,2.7,8.1$ which are provided in logarithmic scale for the x axes. We report here the total amount of the outer and inner iterations and the accuracy of the solution found by the algorithm for such values of $\alpha$ and $\varepsilon_{0}$ (see Figures C.1-C.9), by considering three test problems with 50 variables, namely K1F13 $\Phi$ r, K2F21 $\Phi 3$ and K5F52 $\$ 2$.

## C. 2 Centralized inexact version

We report here (see Figures C.10-C.14), respectively, the total number of outer iterations, the total amount of inner iterations and the error of the solution found by the algorithm that we have obtained for different choices for the updating rule for the parameter $e_{k}$ and by considering five test problems with $n=50$ variables, namely K2F7 $\Phi$ r, K5F52 $\Phi 2$, K5F52 $\Phi 4$, K5F52 $\Phi 5$ and K6F61 ${ }^{2}$ 2. Based on the considerations made in Section 3.5.2, we have set preliminarily $\varepsilon_{0}$ equal to 6.4 and $\alpha$ equal to 0.3 .

## C. 3 Centralized version vs distributed version

We have compared the performances of the centralized and the distributed algorithms in terms of number of outer and inner iterations and of time spent (provided in seconds) by the two versions of the scheme. Here we report the results (see Figures C.15-C.19) of our experiments on five test problems with $n=50$ variables, namely K2F13 $\Phi 6, \mathrm{~K} 2 \mathrm{~F} 21 \Phi 2, \mathrm{~K} 2 \mathrm{~F} 61 \Phi \mathrm{r}, \mathrm{K} 2 \mathrm{Fr} \Phi 2, \mathrm{~K} 2 \mathrm{Fr} \Phi 7$. We recall that we have chosen $\varepsilon_{0}=3.2$ and we have considered the following six different decomposition approaches $Q=1, \ldots, 6$ :
$-Q=1: N^{1}=1$ and $n_{1}^{1}=50$
$-Q=2: N^{2}=2$ and $n_{1}^{2}=25, n_{2}^{2}=25$
$-Q=3: N^{3}=3$ and $n_{1}^{3}=20, n_{2}^{3}=15, n_{3}^{3}=15$
$-Q=4: N^{4}=4$ and $n_{1}^{4}=15, n_{2}^{4}=15, n_{3}^{4}=10, n_{4}^{4}=10$
$-Q=5: N^{5}=5$ and $n_{\nu}^{5}=10, \nu=1, \ldots, 5$
$-Q=6: N^{6}=10$ and $n_{\nu}^{6}=5, \nu=1, \ldots, 10$
We note that for $Q=1$ the problem is not decomposed, thus $\bar{\alpha}_{1}=0$; in this case, we have set $\alpha_{1}$ equal to 0.3 (based on the considerations made in Section 3.5.2), no decomposition is considered and there are no differences between the two versions of the scheme tested with the same value for $\alpha_{1}$. In general, for each test problem, each decomposition $Q$ leads to a different value of the lower bound $\bar{\alpha}_{Q}$; for each $Q$, we have tested the distributed version first with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), then with a "risky" value $\alpha_{Q}=\alpha_{R}$ such that $\alpha_{R}<\min _{Q \neq 1} \bar{\alpha}_{Q}$. Then we have tested the centralized version by considering the same values of $\alpha$ equal to $\alpha_{Q}$ for every $Q$.

In each figure, one can find the algorithm performances obtained for different choices of $Q=1, \ldots, 6$ : we report on the x axes the values of the corresponding values of $N^{Q}$, i.e. the number of decomposed problems that we have considered for each $Q$. For each $Q$ (and thus for each $N^{Q}$ ), a triplet of results is reported: the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$.


Fig. C.1: Centralized exact version, problem K1F13Фr, $n=50$. Number of outer iterations.


Fig. C.2: Centralized exact version, problem $\mathrm{K} 1 \mathrm{~F} 13 \Phi \mathrm{r}, n=50$. Number of inner iterations.


Fig. C.3: Centralized exact version, problem $\mathrm{K} 1 \mathrm{~F} 13 \Phi \mathrm{r}, n=50$. Solutions error.


Fig. C.4: Centralized exact version, problem K2F21 $\Phi 3$, $n=50$. Number of outer iterations.


Fig. C.5: Centralized exact version, problem K2F21 $\Phi 3$, $n=50$. Number of inner iterations.


Fig. C.6: Centralized exact version, problem $\mathrm{K} 2 \mathrm{~F} 21 \Phi 3, n=50$. Solutions error.


Fig. C.7: Centralized exact version, problem K5F52 $\Phi 2, n=50$. Number of outer iterations.


Fig. C.8: Centralized exact version, problem K5F52 $\Phi 2$, $n=50$. Number of inner iterations.


Fig. C.9: Centralized exact version, problem K5F52 $22, n=50$. Solutions error.


Fig. C.10: Centralized inexact version, problem $\mathrm{K} 2 \mathrm{~F} 7 \Phi \mathrm{r}, n=50$.


Fig. C.11: Centralized inexact version, problem K5F52 $22, n=50$.


Fig. C.12: Centralized inexact version, problem K5F52 $44, n=50$.


Fig. C.13: Centralized inexact version, problem K5F52 $55, n=50$.


Fig. C.14: Centralized inexact version, problem $\mathrm{K} 6 \mathrm{~F} 61 \Phi 2, n=50$.


Fig. C.15: Problem K2F13 $\Phi 6, n=50$ : for each $Q$, the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$. In this case, $\alpha_{1}=0.3$, $\alpha_{2}=6.0183, \alpha_{3}=9.7825, \alpha_{4}=13.2523, \alpha_{5}=16.7221, \alpha_{6}=29.2889, \alpha_{R}=5$.


Fig. C.16: Problem K2F21 $\Phi 2, n=50$ : for each $Q$, the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$. In this case, $\alpha_{1}=0.3$, $\alpha_{2}=0.8, \alpha_{3}=1.6, \alpha_{4}=2.4, \alpha_{5}=3.2, \alpha_{6}=7.2, \alpha_{R}=0.3$.


Fig. C.17: Problem K2F61 $\Phi$ r, $n=50$ : for each $Q$, the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$. In this case, $\alpha_{1}=0.3$, $\alpha_{2}=4.7438, \alpha_{3}=9.4877, \alpha_{4}=14.2315, \alpha_{5}=18.9753, \alpha_{6}=42.6945, \alpha_{R}=4$.


Fig. C.18: Problem $\operatorname{K2Fr} \Phi 2, n=50$ : for each $Q$, the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$. In this case, $\alpha_{1}=0.3$, $\alpha_{2}=2.093, \alpha_{3}=3.3734, \alpha_{4}=4.4545, \alpha_{5}=5.3132, \alpha_{6}=9.7138, \alpha_{R}=1$.


Fig. C.19: Problem $\operatorname{K2Fr} \Phi 7, n=50$ : for each $Q$, the first bar (in blue) refers to the centralized version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ), the second bar (in grey) refers to the distributed version with $\alpha_{Q}=\bar{\alpha}_{Q}$ (with the exception of $\alpha_{1}=0.3$ ) and the third one (in red) to the decomposed version with a risky value of $\alpha_{Q}=\alpha_{R}$ for $Q=1, \ldots, 6$. In this case, $\alpha_{1}=0.3$, $\alpha_{2}=2.3113, \alpha_{3}=3.8099, \alpha_{4}=5.1092, \alpha_{5}=6.1863, \alpha_{6}=11.6781, \alpha_{R}=0.3$.

## References

1. K. Addi, B. Brogliato, and D. Goeleven. A qualitative mathematical analysis of a class of linear variational inequalities via semi-complementarity problems: application in electronics. Mathematical Programming, (126):31-67, 2011.
2. E. Adida and G. Perakis. Dynamic pricing and inventory control: Uncertainty and competition. Part a: Existence of Nash equilibrium. Technical report, Sloan School of Management, MIT, 2006.
3. E. Adida and G. Perakis. Dynamic pricing and inventory control: Uncertainty and competition. Part b: An algorithm for the normalized Nash equilibrium. Technical report, Sloan School of Management, MIT, 2006.
4. E. Altman and L. Wynter. Equilibrium games, and pricing in transportation and telecommunication networks. Networks and Spatial Economics, 4:7-21, 2004.
5. K.J. Arrow and G. Debreu. Existence of an equilibrium for a competitive economy. Econometrica, 22:265-290, 1954.
6. T. Başar and G.J. Olsder. Dynamic Noncooperative Game Theory. Academic Press, London/New York, 1989. Reprinted in SIAM Series "Classics in Applied Mathematics", 1999.
7. A. Bassanini, A. La Bella, and A. Nastasi. Allocation of railroad capacity under competition: A game theoretic approach to track time pricing, volume Transportation and Networks Analysis: Current Trends, pages 1-17. Kluwer Academic Publisher, Dordrecht, 2002.
8. D. Bertsekas. Constrained Optimization and Lagrange Multiplier Methods. Athena Scientific, 1982.
9. D. Bertsekas. Nonlinear Programming. Athena Scientifica, 1992.
10. D. Bertsekas, A. Nedic, and A.E. Ozdaglar. Convex Analysis and Optimization. Athena Scientifica, 2003.
11. D. Bertsekas and J.N. Tsitziklis. Parallel and Distributed Computation: Numerical Methods. Athena Scientific, 1997.
12. M. Breton, G. Zaccour, and M. Zahaf. A game-theoretic formulation of joint implementation of environmental projects. European Journal of Operational Research, 168:221-239, 2005.
13. F.E. Browder. Existence and approximation of solutions of nonlinear variational inequalities. Proceedings of the National Academy of Sciences, (56):1080-1086, 1966.
14. A. Cabot. Proximal point algorithm controlled by a slowly vanishing term: applications to hierarchical minimization. SIAM Journal on Optimization, (15):555-572, 2005.
15. D. Chan and J.-S. Pang. The generalized quasi-variational inequality problem. Mathematics of Operations Research, 7:211-222, 1982.
16. F.H. Clarke. Optimization and nonsmooth analysis. Wiley, 1983.
17. J. Contreras, M.K. Klusch, and J.-B. Krawczyk. Numerical solution to Nash-Cournot equilibria in coupled constraints electricity markets. IEEE Transactions on Power Systems, 19:195-206, 2004.
18. R.W. Cottle, J.-S. Pang, and R.E. Stone. The Linear Complementarity Problem. Academic Press, 1992.
19. R. Courant. Variational methods for the solution of problems with equilibrium and vibration. Bulletin of the American Mathematical Society, 49:1-23, 1943.
20. A.A. Cournot. Recherches sur les principes mathématiques de la théorie des richesses. Hachette, Paris, 1838.
21. G. Debreu. A social equilibrium existence theorem. Proceedings of the National Academy of Sciences, 38:886-893, 1952.
22. G. Debreu. Theory of Values. Yale University Press, New Haven, USA, 1959.
23. A. Dreves, F. Facchinei, C. Kanzow, and S. Sagratella. On the solution of the KKT conditions of generalized Nash equilibrium problems. Preprint 300, Institute of Mathematics, University of Würzburg, Würzburg, 2010.
24. A. Dreves and C. Kanzow. Nonsmooth optimization reformulations characterizing all solutions of jointly convex generalized Nash equilibrium problems. Computational Optimization and Applications, 50(1):23-48, 2010.
25. A. Dreves, C. Kanzow, and O. Stein. Nonsmooth optimization reformulations of player convex generalized Nash equilibrium problems. Preprint 300, Institute of Mathematics, University of Würzburg, Würzburg, October 2010.
26. L. Drouet, A. Haurie, F. Moresino, J.-P. Vial, M. Vielle, and L. Viguier. An oracle based method to compute a coupled equilibrium in a model of international climate policy. Computational Management Science, 5:119-140, 2008.
27. A. Ehrenmann. Equilibrium problems with equilibrium constraints and their application to electricity markets. Ph.D. Dissertation, Judge Institute of Management, Cambridge, UK, 2004.
28. F. Facchinei, A. Fischer, and V. Piccialli. On generalized Nash games and variational inequalities. Operations Research Letters, (35):159-164, 2007.
29. F. Facchinei, A. Fischer, and V. Piccialli. Generalized Nash equilibrium problems and Newton methods. Mathematical Programming, 35:163-194, 2009.
30. F. Facchinei and C. Kanzow. Penalty methods for the solution of generalized Nash equilibrium problems. SIAM Journal on Optimization, 20(5):2228-2253, 2009.
31. F. Facchinei and C. Kanzow. Generalized Nash equilibrium problems. Annals of Operations Research, 175(1):177-211, 2010.
32. F. Facchinei and J.-S. Pang. Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, 2003.
33. F. Facchinei and J.-S. Pang. Exact penalty functions for generalized Nash problems. G. Di Pillo and M. Roma (eds): Large Scale Nonlinear Optimization, pages 115-126, 2006.
34. F. Facchinei and J.-S. Pang. Nash equilibria: The variational approach, pages 443493. Convex optimization in signal processing and communications. Cambridge University Press, 2010.
35. F. Facchinei, V. Piccialli, and M. Sciandrone. Decomposition algorithms for generalized potential games. Computational Optimization and Applications, 50(2):237-262, 2011.
36. F. Facchinei and S. Sagratella. On the computation of all solutions of jointly convex generalized Nash equilibrium problems. Optimization Letters, 5(3):531-547, 2011.
37. A.V. Fiacco and G.P. Mc Cormick. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. John Wiley \& Sons, 1968.
38. R. E. Fletcher. A class of methods for nonlinear programming with termination and convergence properties. Integer and Nonlinear Programming, pages 157-173, 1970.
39. M. Fukushima. A class of gap functions for quasi-variational inequality problems. Journal of Industrial and Management Optimization, 3:165-171, 2007.
40. M. Fukushima. Restricted generalized Nash equilibria and controlled penalty algorithm. Computational Management Science, 8(3):201-218, 2011.
41. S.A. Gabriel, S. Kiet, and J. Zhuang. A mixed complementarity-based equilibrium model of natural gas markets. Operations Research, 53:799-818, 2005.
42. S.A. Gabriel and Y. Smeers. Complementarity problems in restructured natural gas markets. Lecture Notes in Economics and Mathematical Systems, 563(Part IV):343373, 2006.
43. C.B. Garcia and W.I. Zangwill. Pathways to solutions, fixed points, and equilibria. Prentice-Hall, 1981.
44. L. Grippo and M. Sciandrone. Globally convergent block-coordinate techniques for unconstrained optimization. Optimization Methods and Software, 10(4):587-637, 1999.
45. L. Grippo and M. Sciandrone. On the convergence of the block nonlinear gauss-seidel method under convex constraints. Operations Research Letters, 26(3):127-136, 2000.
46. P.T. Harker. Generalized Nash games and quasivariational inequalities. European Journal of Operational Research, 54:81-94, 1991.
47. P.T. Harker and S. Hong. Pricing of track time in railroad operations: An internal market approach. Transportation Research Part B: Methodological, 28(3):197-212, 1994.
48. P.T. Harker and J.-S. Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. Mathematical Programming, 48:161-220, 1990.
49. A. Haurie and J.-B. Krawczyk. Optimal charges on river effluent from lumped and distributed sources. Environmental Modelling and Assessment, 2:93-106, 1997.
50. B. Hobbs, U. Helman, and J.-S. Pang. Equilibrium market power modeling for large scale power systems. IEEE Power Engineering Society Summer Meeting, 1:558-563, 2001.
51. B. Hobbs and J.-S. Pang. Nash-cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints. Operations Research, 55:113127, 2007.
52. T. Ichiishi. Game theory for economic analysis. Academic Press, New York, 1983.
53. H. Iiduka. Fixed point optmization algorithm and its application to power control in cdma data networks. To appear in Mathematical Programming, 2010.
54. A.D. Ioffe and J.V. Outrata. On metric and calmness qualification conditions in subdifferential calculus. Journal of Set-Valued and Variational Analysis, 16:199-227, 2008.
55. A.N. Iusem. On some properties of paramonotone operators. Journal of Convex Analysis, (5):269-278, 1998.
56. H. Jiang. Network capacity management competition. Technical report, Judge Business School at University of Cambridge, UK, 2007.
57. A. Jofré, R.T. Rockafellar, and J.B. Wets. Variational inequalities and economic equilibrium. Mathematics of Operations research, (32), 32-50.
58. E.A. Jorswieck and E.G. Larsson. Game theory in signal processing and communications. In M. Luise and H.V. Poor, editors, Special issue of the IEEE Signal Processing Magazine, volume 26, 2009.
59. A. Kesselman, S. Leonardi, and V. Bonifaci. Game-theoretic analysis of internet switching with selfish users. Lecture Notes in Computer Science, 3828:236-245, 2005.
60. J.-B. Krawczyk. An open-loop Nash equilibrium in an environmental game with coupled constraints. In Proceedings of the 2000 Symposium of the International Society of Dynamic Games, 2000.
61. J.-B. Krawczyk. Numerical solutions to coupled-constraint (or generalised Nash) equilibrium problems. Computational Management Science, 4:183-204, 2007.
62. K. Kubota and M. Fukushima. Gap function approach to the generalized nash equilibrium problem. Journal of Optimization Theory and Applications, 144(3):511-531, 2010.
63. X. Lu, H.-K. Xu, and X. Yin. Hybrid methods for a class of monotone variational inequalities. Nonlinear Analysis, (71):1032-1041, 2009.
64. T. Luo and J.-S. Pang. Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines. EURASIP Journal on Applied Signal Processing, 2006.
65. T. Luo and S. Zhang. Dynamic spectrum management: Complexity and duality. IEEE journal of selected topics in signal processing, 2(1):57-73, 2008.
66. Z.Q. Luo, J.-S. Pang, and D. Ralph. Mathematical programs with equilibrium constraints. Cambridge University Press, 1996.
67. G. Marino and H.-K. Xu. Explicit hierarchical fixed point approach to varitional inequalities. Journal of Optimization Theory and Applications, (149):61-78, 2011.
68. R.D.C. Monteiro and B.F. Svaiter. Complexity of variants of tseng's modified f-b splitting and korpelevich's method for generalized variational inequalities with applications to saddle point and convex optimization problems. Under review at SIAM Journal on Optimization 2010.
69. D. Motreanu and R.D. Panagiotopoulos. Minimax Theorems and Qualitative Properties of the Solutions of Hemivariational Inequalities. Kluwer Academic Publishers, 1999.
70. A. Moudafi. Proximal methods for a class of bilevel monotone equilibrium problems. Journal of Global Optimization, (47):287-292, 2010.
71. K. Nabetani, P. Tseng, and M. Fukushima. Parametrized variational inequality approaches to generalized Nash equilibrium problems with shared constraints. Computational Optimization and Applications, 48(3):423-452, 2011.
72. Z. Naniewicz and R.D. Panagiotopoulos. Mathematical Theory of Hemivariational Inequalities and Applications. Marcel Dekker, 1995.
73. J.F. Nash. Equilibrium points in n-person games. Proceedings of the National Academy of Sciences, 36(1):48-49, 1950.
74. J.F. Nash. Non-cooperative games. Annals of Mathematics, 54:286-295, 1951.
75. N. Nisan, T. Roughgarden, and E. Tardos. Algorithmic Game Theory. Cambridge University Press, 2007.
76. J.V. Outrata, M. Kocvara, and J. Zowe. Nonsmooth Approach to Optimization Problems with Equilibrium Constraints. Kluwer Academic Publisher, 1998.
77. R.D. Panagiotopoulos. Hemivariational Inequalities: Applications in Mechanics and Engineering. Springer, 1993.
78. J.-S. Pang. Asymmetric variational inequality problems over product sets: Applications and iterative methods. Mathematical Programming, (31):206-219, 1985.
79. J.-S. Pang and M. Fukushima. Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. Computational Management Science, 2:2156, 2009.
80. J.-S. Pang, G. Scutari, F. Facchinei, and C. Wang. Distributed power allocation with rate constraints in gaussian frequency-selective interference channels. IEEE Transactions on Information Theory, 54:3471-3489, 2007.
81. J.-S. Pang, G. Scutari, D.P. Palomar, and F. Facchinei. Design of cognitive radio systems under temperature-interference constraints: A variational inequality approach. IEEE Transactions on Signal Processing, 58:3251-3271, 2010.
82. G. Di Pillo and L. Grippo. A continuously differentiable exact penalty function for nonlinear programming problems with inequality constraints. SIAM Journal on Control and Optimization, 23:72-84, 1985.
83. G. Di Pillo and L. Grippo. Exact penalty functions in constrained optimization. SIAM Journal on Control and Optimization, 27:1333-1360, 1989.
84. J. Puerto, A. Schöbel, and S. Schwarze. The path player game a network game from the point of view of the network providers. Mathematical Methods of Operations Research, 68(1):1-20, 2008.
85. S.S. Rao, V.B. Venkayya, and N.S. Khot. Game theory approach for the integrated design of structures and controls. AIAA Journal, 26:463-469, 1988.
86. S.M. Robinson. Shadow prices for measures of effectiveness. I. Linear model. Operations Research, 41:518-535, 1993.
87. S.M. Robinson. Shadow prices for measures of effectiveness. II. General model. Operations Research, 41:536-548, 1993.
88. R.T. Rockafellar and J.B. Wets. Variational Analysis. Springer, 1998.
89. L.A. Schmit. Structural synthesis - its genesis and development. AIAA Journal, 19:1249-1263, 1981.
90. S.J. Scotti. Structural design using equilibrium programming formulations. NASA Technical Memorandum, 110175, 1995.
91. G. Scutari, F. Facchinei, J.-S. Pang, and D.P. Palomar. Monotone communication games: Theory, algorithms, and methods. Manuscript in preparation.
92. G. Scutari, D.P. Palomar, and S. Barbarossa. Asynchronous iterative water-filling for Gaussian frequency-selective interference channels. IEEE Transaction on Information Theory, (54):2868-2878, 2008.
93. G. Scutari, D.P. Palomar, and S. Barbarossa. Optimal linear precoding strategies for wideband noncooperative systems based on game theory-part I \& II: Nash equilibria \& Algorithms. IEEE Transaction on Signal Processing, (56):1230-1249 \& 1250-1267, 2008.
94. G. Scutari, D.P. Palomar, J.-S. Pang, and F. Facchinei. Flexible design of cognitive radio wireless systems: From game theory to variational inequality theory. IEEE Signal Processing Magazine, 26(5):107-123, 2009.
95. L.-J. Sun and Z.-Y. Gao. An equilibrium model for urban transit assignment based on game theory. European Journal of Operational Research, 181:305-314, 2007.
96. M. Tidball and G. Zaccour. An environmental game with coupling constraints. Environmental Modeling Assessment, 10:153-158, 2005.
97. A.N. Tikhonov. Solution of incorrectly formulated problems and regularization method. Soviet mathematics doklady, (4):1035-1038, 1963.
98. A.N. Tikhonov and V.Y. Arsenin. Solutions of ill-posed problems. John Wiley \& Sons, 1977.
99. P. Tseng. Error bounds for regularized complementarity problems, pages 247-274. Ill-posed variational problems and regularization techniques. Springer, 1999.
100. T.L. Vincent. Game theory as a design tool. ASME Journal of Mechanisms, Transmissions, and Automation in Design, 105:165-170, 1994.
101. J. von Neumann. Zur theorie der gesellschaftsspiele. Mathematische Annalen, 100:295-320, 1928.
102. J. von Neumann and O. Morgenstern. Theory of Games and Economic Behavior. Princeton University Press, Princeton, 1944.
103. J.-Y. Wei and Y. Smeers. Spatial oligopolistic electricity models with cournot generators and regulated transmission prices. Operations Research, 47:102-112, 1999.
104. I. Yamada. The hybrid steepest descent for the variational inequality problems over the intersection of fixed point sets of nonexpansive mappings, volume 8 of Studies in Computational Mathematics, chapter Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, pages 473-504. Elsevier, 2001.
105. W. Yu, G. Ginis, and J.M. Cioffi. Distributed multiuser power control for digital subscriber lines. IEEE Journal on Selected Areas in Communications, (20):11051115, 2002.
106. W.I. Zangwill. Nonlinear programming via penalty functions. Management Science, 13:344-358, 1967.
107. J. Zhou, W.H.K. Lam, and B.G. Heydecker. The generalized Nash equilibrium model for oligopolistic transit market with elastic demand. Transportation Research BMethodological, 39:519-544, 2005.

[^0]:    ${ }^{1}$ Here and in all the paper, when we say that a function is continuous or continuously differentiable on a closed set, we intend that the function is (defined and) continuous or continuously differentiable on an open set containing the closed set

