

# VI-constrained hemivariational inequalities: distributed algorithms and power control in ad-hoc networks

Francisco Facchinei · Jong-Shi Pang ·  
Gesualdo Scutari · Lorenzo Lampariello

Received: 10 July 2011 / Accepted: 22 January 2013 / Published online: 15 February 2013  
© Springer-Verlag Berlin Heidelberg and Mathematical Optimization Society 2013

**Abstract** We consider centralized and distributed algorithms for the numerical solution of a hemivariational inequality (HVI) 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 algorithms consist of a main loop wherein a sequence of one-level, strongly monotone HVIs are 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. Minimization problems, possibly with nonconvex objective functions, over implicitly defined VI constraints are discussed in detail.

---

The work of the first author has been partially supported by MIUR-PRIN 2005 n.2005017083 Research Program “Innovative Problems and Methods in Nonlinear Optimization”. The work of the second author was based on research partially supported by the U.S.A. National Science Foundation grant CMMI 0969600 and by the Air Force Office of Sponsored Research award No. FA9550-09-10329.

---

F. Facchinei (✉) · L. Lampariello  
Department of Computer, Control and Management Engineering Antonio Ruberti,  
University of Rome La Sapienza,  
Via Ariosto 25, 00185 Rome, Italy  
e-mail: facchinei@dis.uniroma1.it

L. Lampariello  
e-mail: lampariello@dis.uniroma1.it

J.-S. Pang  
Department of Industrial and Enterprise Systems Engineering,  
University of Illinois at Urbana–Champaign, Urbana, IL, USA  
e-mail: jspang@illinois.edu

G. Scutari  
Department of Electrical Engineering State University of New York (SUNY) at Buffalo,  
Buffalo, NY, USA  
e-mail: gesualdo@buffalo.edu

The methods developed in the paper are then used to successfully solve a new power control problem in ad-hoc networks.

**Keywords** Hemivariational inequality · Hierarchical optimization · Distributed algorithms · Penalization · Power control · Ad-hoc networks

**Mathematics Subject Classification** 90C33 · 90C90 · 91A10 · 49M27 · 65K15 · 65K10

## 1 Introduction

In a standard constrained optimization problem: minimize  $x \in S \phi(x)$ , with  $S$  being a closed convex subset  $\mathbb{R}^n$  and  $\phi$  a continuous function defined on  $S$ ,<sup>1</sup> it is often assumed that the feasible set  $S$  is explicitly defined by a system of finitely many equalities and inequalities. However, in some cases the set  $S$  is only implicitly defined as the solution set of a second (lower-level) optimization problem, or more generally, a variational inequality. Formally the problem is then defined as

$$\begin{aligned} & \underset{x}{\text{minimize}} \phi(x) \\ & \text{subject to } x \in \operatorname{argmin}\{f(z) \mid z \in K\} \end{aligned} \quad (1)$$

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 (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 function 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

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

Hemivariational inequalities (also known as variational inequalities of the second kind) are a powerful modelling tool that encompass both (convex) optimization, when  $\Phi \equiv 0$ , and VIs, when  $h \equiv 0$ , as particular instances. In its full generality, HVIs have been mainly considered in infinite-dimensional settings, see e.g. [29,31,32]; nevertheless, finite-dimensional HVIs have recently attracted attention in the mathematical programming literature; see e.g. [1,20,28]. If the set  $X$  is implicitly defined as the solution set  $\operatorname{SOL}(K, F)$  of a lower-level Variational Inequality VI  $(K, F)$ , with

<sup>1</sup> 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.

$K \subseteq \mathbb{R}^n$  closed and convex and  $F : K \rightarrow \mathbb{R}^n$  continuous, the HVI  $(\text{SOL}(K, F), \Phi, h)$  becomes the VI-C HVI defined by the tuple  $(K, F, \Phi, h)$ , which is nothing else than the problem of finding a vector  $x \in \text{SOL}(K, F)$  such that

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

where

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

Denoting the same problem, the notations: HVI  $(\text{SOL}(K, F), \Phi, h)$  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  $(\text{SOL}(K, F), \Phi, 0)$ , 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  $f$ , assumed to be continuously differentiable, it is well known that  $\text{SOL}(K, F)$  is the set of stationary solutions of the lower-level optimization problem in the constraint of (1); thus, assuming  $\phi$  to be convex, the VI-C HVI  $(K, \nabla f, 0, \phi)$  is intimately related to the bilevel program (1); in fact, the two problems are equivalent if the lower-level objective function  $f$  is convex. Second, with  $F$  being a monotone map [15, Definition 2.3.1] and  $\Phi \triangleq \nabla\psi$  being the gradient of the (possibly nonconvex) continuously differentiable 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:

$$\underset{x \in \text{SOL}(K, F)}{\text{minimize}} \psi(x), \tag{2}$$

which in turn is a special case of a mathematical program with equilibrium constraints (MPEC) [24] wherein there is no (upper-level) “design variable”. Since the VI is known [15] to provide a broad mathematical framework for a host of economic equilibrium and game-theoretic problems (for a recent survey on the VI approach to the latter problems, see [17]), the optimization problem (2) is a natural formulation for the problem of selecting a particular equilibrium solution to optimize the auxiliary objective function  $\psi$ .

By using a penalization approach and taking  $h \triangleq \rho \text{ dist}(\bullet, \Omega)$  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 (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

$$\underset{x \in \Omega \cap \text{SOL}(K, F)}{\text{minimize}} \psi(x). \tag{3}$$

It should be noted that

$$\underset{x \in \Omega \cap \text{SOL}(K, F)}{\text{minimum}} \psi(x) \geq \underset{x \in \text{SOL}(K \cap \Omega, F)}{\text{minimum}} \psi(x).$$

Since equality does not necessarily hold, the two problems (2) and (3) are in general not the same.

Our main motivation in dealing with a VI-C HVIs comes from the problem of selecting a Nash equilibrium in some power-control games in ad-hoc networks. This application will be considered in Sect. 7; however here we describe more in detail the Nash equilibrium selection problem in full generality, since this gives an interesting, concrete example of the use of our results and also prepares the reader for the developments in Sect. 7.

In a general noncooperative game there are  $N$  players, each associated to a cost function that may depend on the other players' actions and to a given strategy set  $K_\nu$  which is a closed and convex subset of  $\mathbb{R}^{n_\nu}$ . Player  $\nu$ 's cost function  $\theta_\nu(x^\nu, x^{-\nu})$  depends on all players' strategies which are described by the vector  $x \triangleq (x^1, \dots, x^N)$ , where  $x^\nu \in K_\nu$  is the action of the player  $\nu$  and  $x^{-\nu} \triangleq (x^1, \dots, x^{\nu-1}, x^{\nu+1}, \dots, x^N)$  denotes the vector of all players' strategies except that of player  $\nu$ . The joint strategy set of the game is given by  $K \triangleq \prod_{i=1}^N K_i$ . In the Nash Equilibrium Problem (NEP), player  $\nu$ , anticipating the other players' strategies  $x^{-\nu}$ , aims at solving

$$\begin{aligned} & \text{minimize}_{x^\nu} \theta_\nu(x^\nu, x^{-\nu}) \\ & \text{subject to } x^\nu \in K_\nu. \end{aligned} \quad (4)$$

A Nash Equilibrium or, more simply, a solution of the NEP, is a feasible point  $\bar{x}$  such that

$$\theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \leq \theta_\nu(x^\nu, \bar{x}^{-\nu}), \quad \forall x^\nu \in K_\nu \quad (5)$$

holds for each player  $\nu = 1, \dots, N$ ; that is,  $\bar{x}$  is a solution if no *single* player can benefit from a *unilateral* deviation from  $\bar{x}^\nu$ . There is an easy equivalence between a NEP and a suitably defined VI. Given  $K$  as indicated above, let  $F : K \rightarrow \mathbb{R}^n$  (with  $n \triangleq \sum_{\nu=1}^N n_\nu$ ) be defined by  $F(x) = (\nabla_{x^\nu} \theta_\nu(x^\nu, x^{-\nu}))_{\nu=1}^N$ . The following result follows easily from the minimum principle for convex problems and the Cartesian structure of the joint strategy set  $K$ , see for example [17, Prop. 1.4.2].

**Proposition 1** *Given a NEP as described above, suppose that for each player  $\nu$  the following assumptions hold:*

- (i) *The (nonempty) strategy set  $K_\nu$  is closed and convex;*
- (ii) *For every fixed  $x^{-\nu} \in K_{-\nu} \triangleq \prod_{\mu \neq \nu} K_\mu$ , the function  $\theta_\nu(x^\nu, x^{-\nu})$  is convex and continuously differentiable.*

*Then, the game is equivalent to the VI  $(K, F)$ , with  $K$  and  $F$  defined just before the proposition.*

Suppose now that we are dealing with a *monotone game*, i.e. a NEP with a monotone function  $F$ . Then the set of Nash equilibria is the closed convex set  $\text{SOL}(K, F)$ . In

case this set is not a singleton, in some applicative contexts the problem arises (see for example Sect. 7) to select the “best” equilibrium according to some criterion. Having Proposition 1 in mind, it is now clear that this problem can be formulated as in (2), where the function  $\psi$  represents the chosen criterion. If we further want to find an optimal equilibrium which satisfies some other conditions represented by a set  $\Omega \subset \mathbb{R}^n$ , our selection problem can instead be formulated as in (3).

The main contributions of this paper 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. In particular, the proposed distributed algorithm, in which we are particularly interested as motivated by applications in non-cooperative game problems, see [17, Chapter 12] and Sect. 7, is novel even for a hierarchical optimization problem. Furthermore, the power control problem analyzed in Sect. 7 is new and our results expand considerably the applicability and flexibility of game-theoretic models in ad-hoc networks and also bring considerable gains over existing techniques.

The paper is organized as follows. In the next section we discuss existing results in the literature. In Sect. 3 we show how to penalize side constraints and reduce an HVI problem with side constraints to one without side constraints. Section 5 presents and analyzes the main centralized algorithm, considering both the exact and approximate versions. Section 5 describes a broad decomposition scheme for a partitioned HVI that enables the development of distributed versions of the algorithm described in Sect. 4. Section 6 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, Sect. 7 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 paper.

## 2 Background results and related works

The study of the VI-C HVI is based on the theory of the VI; basic concepts and results from the monograph [15] will be used freely in the paper. For ease of reference, we list below several well-known monotonicity and related definitions.

**Definition 1** Let  $X \subseteq \mathbb{R}^n$  be a convex set. A function  $G : X \rightarrow \mathbb{R}^n$  is

– *monotone* on  $X$  if for all  $x$  and  $y$  in  $X$  it holds that

$$[G(y) - G(x)]^T (y - x) \geq 0;$$

– *monotone plus* (or *paramonotone*) on  $X$  if it is monotone on  $X$  and in addition for all  $x$  and  $y$  in  $X$  it holds that

$$[G(y) - G(x)]^T (y - x) = 0 \implies G(x) = G(y);$$

– *pseudo monotone* on  $X$  if for all  $x$  and  $y$  in  $X$  it holds that

$$G(x)^T (y - x) \geq 0 \implies G(y)^T (y - x) \geq 0;$$

– *pseudo monotone plus* on  $X$  if it is pseudo monotone on  $X$  and in addition for all  $x$  and  $y$  in  $X$  it holds that

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

We also 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 [19]. Further conditions implying monotonicity plus are discussed in [15, Chapter 2]. Subsequently, in Lemma 1, 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)  $\text{SOL}(K, F) \neq \emptyset$ .
- (D)  $\Phi : K \rightarrow \mathbb{R}^n$  is continuous and monotone plus, and
- (E)  $h : K \rightarrow \mathbb{R}$  is convex and continuous.

Under assumptions (A–C),  $\text{SOL}(K, F)$  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) [9]. 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  $(x' - x)^T y \geq 0$  for all  $x' \in X$ . It is not difficult to see that the 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

$$(\Phi(x) + \zeta)^T (y - x) \geq 0, \quad \forall y \in X, \quad (6)$$

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

Even if the map  $T$  is monotone and  $\text{SOL}(K, F)$  is closed and convex, the GVI  $(\text{SOL}(K, F), T)$ , 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  $\text{SOL}(X, F)$  is compact. In this paper, we assume neither. Instead, we rely on the well-known Tikhonov process [40,41] 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 [15, Section 12.9]; in particular, it was noted there that Browder [7] (and more recently, Tseng [39]) showed that given a single-valued, monotone VI  $(S, T)$  one can compute a solution of the VI  $(\text{SOL}(S, T), G)$ , where  $G$  is a strongly monotone operator, by solving a sequence of “regularized” problems of the form VI  $(S, T + \varepsilon_k G)$ , where  $\{\varepsilon_k\}$  is a sequence of positive scalars converging to zero. Each regularized VI  $(S, T + \varepsilon_k G)$  has a unique solution  $x(\varepsilon_k)$  and as  $\varepsilon_k$  goes to zero,  $\{x(\varepsilon_k)\}$  converge to the unique solution of VI  $(\text{SOL}(S, T), G)$ .

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. One line of research in this direction has been carried out in the context of the hierarchical optimization problem (1). For this problem, the regularization step at iteration  $k$  amounts to solving the following strongly convex optimization problem:

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

where  $\alpha > 0$  is a positive parameter and  $x(\varepsilon_{k-1})$  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  $\alpha \|x - x(\varepsilon_{k-1})\|^2$ , 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. Early significant results include [2, 6, 10, 21, 22]. The more advanced results in analyzing this approach in relation to hierarchical problems are probably those obtained by Cabot [8], to which we refer the interested reader also for a detailed bibliographical discussion and whose proof techniques have influenced our approach in Sect. 4. Essentially the idea is to show that if  $\varepsilon_k$  goes to zero “slowly”, then  $x(\varepsilon_k)$  converges to a solution of (1). Extension of this approach to more general problems have been attempted recently, see [18, 30]. 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. In other words the flavor of the convergence results is something like: if the algorithm behaves in such and such a way, then it converges. While this certainly gives insights on the convergence of the algorithm itself, results obtained in this way are not totally satisfactory, because

ideally one would like to understand what classes of problems are amenable to the convergence of the iterations.

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  $(\text{Fix}(U), I - V)$ , where  $\text{Fix}(U)$  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 1** *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  $I - V$ .  $\square$

The line of research on the VI  $(\text{Fix}(U), I - V)$  was initiated by [42] and we cite the interesting papers [26,27] for recent results and bibliographic references. While we [15, Proposition 1.5.8] can always write the solution set of the VI  $(K, F)$  as the set of fixed points of the natural map:  $F_K^{\text{nat}}(x) \triangleq P_K(x - \tau F(x))$ , where  $P_K$  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  $F_K^{\text{nat}}(x) = x - \tau e^x$  cannot be nonexpansive. In general, probably the weakest condition that guarantees  $F_K^{\text{nat}}$  to be nonexpansive (for  $\tau$  sufficiently small) is that  $F$  be co-coercive; see [15, Lemma 12.1.7]. Similarly, while any map of the type  $I - V$ , with  $V$  nonexpansive, is monotone plus, as shown by Lemma 1, 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  $(\text{Fix}(U), I - V)$  becomes the VI  $(\text{SOL}(K, F), I - V)$  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.

In addition to the main algorithm, another contribution of this paper 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 [17] 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 [17] is that twice differentiability is not required.

Historically, distributed algorithms for solving VIs can be traced back to the original work in [33] for partitioned problems; see also [5]. 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 [18, 23, 34–37, 43], 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. In this paper we overcome this limitation and propose a novel distributed algorithm that solves the game-theoretic multi-channel power control problem addressed in [23, 36, 37, 43] even when there are multiple solutions; our main contribution is twofold: (1) our algorithm converges under milder sufficient conditions that do not imply the uniqueness of the solution of the game; and (2) 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 [23, 36, 37, 43], 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 [18] to solve a scalar power control problem in CDMA data networks; such a problem falls in the class of so-called *scalar* games, modelling narrower and simpler scenarios than those considered in this paper (e.g., in [18], each user is assumed to have a scalar variable to optimize rather than a vector as in this paper). Hence, theoretical results in [18] cannot be applied to design multi-channel networks, as considered in this paper.

A last but not least contribution of the paper is the treatment of side constraints via a penalty technique. While having its origin in classical constrained optimization, [12, 13, 16], the idea of applying penalization to treat variational problems with side constraints is new and deserves further investigation, which regrettably is beyond the scope of this work.

### 3 Penalization of HVIs with side constraints

This section shows that, provided some weak assumptions are satisfied, the hemivariational inequality HVI  $(\text{SOL}(K, F) \cap \Omega, \Phi, h)$  with VI  $(K, F)$  and side constraints  $\Omega$  is equivalent to the penalized HVI without side constraints: VI-C HVI  $(\text{SOL}(K, F), \Phi, h + \rho \text{dist}(\bullet, \Omega))$  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 1 successfully solves problems of the form HVI  $(\text{SOL}(K, F), \Phi, h)$ . 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 1, allows us to apply Algorithm 1 to the penalized problem HVI  $(\text{SOL}(K, F), \Phi, h + \rho \text{dist}(\bullet, \Omega))$  as a way of solving the HVI  $(\text{SOL}(K, F) \cap \Omega, \Phi, h)$ .

Since in this section the particular structure of the set  $\text{SOL}(K, F)$  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 e.g. [15, Corollary 2.3.10]. 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 2** *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 (6)), for suitable  $\zeta \in \partial h(x)$  and  $\bar{\zeta} \in \partial h(\bar{x})$ ,*

$$(\Phi(x) + \zeta)^T(z - x) \geq 0 \quad \text{and} \quad (\Phi(\bar{x}) + \bar{\zeta})^T(w - \bar{x}) \geq 0, \quad (7)$$

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

*Proof* Summing the two inequalities in (7), 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

$$[\Phi(x) - \Phi(\bar{x})]^T(\bar{x} - x) + (\zeta - \bar{\zeta})^T(\bar{x} - x) = 0. \quad (8)$$

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

$$\mathcal{N}(X \cap \Omega; \bar{x}) = \mathcal{N}(X; \bar{x}) + \mathcal{N}(\Omega; \bar{x}), \tag{9}$$

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  $\bar{\eta} \in \mathcal{N}(\Omega; \bar{x})$  such that  $-(\Phi(\bar{x}) + \bar{\zeta} + \bar{\eta}) \in \mathcal{N}(X; \bar{x})$ , or equivalently, that

$$(y - \bar{x})^T(\Phi(\bar{x}) + \bar{\zeta} + \bar{\eta}) \geq 0, \quad \forall y \in X. \tag{10}$$

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  $\bar{\eta} \in \mathcal{N}(\Omega; \bar{x})$  and there exists  $\bar{\zeta} \in \partial h(\bar{x})$  such that (10) holds. Note that (9) provides a sufficient condition for  $\mathcal{E}(\bar{x})$  to be nonempty. Equality (9) is a rather weak requirement that is implied by the condition  $\text{ri } X \cap \text{ri } \Omega \neq \emptyset$ , where  $\text{ri } S$  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 (9) is implied by  $\Omega \cap \text{ri } X \neq \emptyset$ .

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 3** *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  $\hat{x}$  be a solution of this HVI such that  $\mathcal{E}(\hat{x})$  is nonempty. Then, if  $\tilde{x}$  is any other solution of the HVI,  $\mathcal{E}(\tilde{x})$  is nonempty and  $\mathcal{E}(\tilde{x}) = \mathcal{E}(\hat{x})$ .*

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

$$(\Phi(\tilde{x}) + \tilde{\zeta})^T(w - \tilde{x}) \geq 0, \quad \forall w \in X \cap \Omega \tag{11}$$

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

$$(v - \hat{x})^T(\Phi(\hat{x}) + \hat{\zeta} + \hat{\eta}) \geq 0, \quad \forall v \in X. \tag{12}$$

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  $\widetilde{x} \in \Omega$ ,  $\widehat{x} \in \Omega$ ,  $\widehat{\eta} \in \mathcal{N}(\Omega; \widehat{x})$ . Hence the three addends are all zero and, in particular, it holds  $(\widetilde{x} - \widehat{x})^T \widehat{\eta} = 0$ , which easily implies that  $\widehat{\eta} \in \mathcal{N}(\Omega; \widetilde{x})$ . Moreover, since the sum of the two inequalities (11) and (12), with  $w = \widehat{x}$  in the former and  $v = \widetilde{x}$  in the latter, is equal to zero, we have

$$(\widetilde{x} - \widehat{x})^T (\Phi(\widehat{x}) + \widehat{\zeta} + \widehat{\eta}) = 0. \quad (13)$$

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

$$(\widetilde{y} - \widetilde{x})^T (\Phi(\widehat{x}) + \widehat{\zeta} + \widehat{\eta}) < 0. \quad (14)$$

because  $\Phi(\widetilde{x}) = \Phi(\widehat{x})$ ,  $\widehat{\zeta} \in \partial h(\widetilde{x})$  by Lemma 2 and we have just established that  $\widehat{\eta} \in \mathcal{N}(\Omega; \widetilde{x})$ . We get, by (12), (13) and (14),

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

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  $y \in X$ ,

$$(y - \widehat{x})^T (\Phi(\widehat{x}) + \widehat{\zeta} + \widehat{\eta}) \geq 0, \quad \text{and} \quad (y - \widetilde{x})^T (\Phi(\widetilde{x}) + \widetilde{\zeta} + \widetilde{\eta}) \geq 0. \quad (15)$$

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 non-positive by the monotonicity of  $\Phi$  and  $\partial h$  and the fact that  $\widetilde{x} \in \Omega$ ,  $\widehat{x} \in \Omega$ ,  $\widehat{\eta} \in \mathcal{N}(\Omega; \widehat{x})$ , and  $\widetilde{\eta} \in \mathcal{N}(\Omega; \widetilde{x})$ . 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  $\widehat{\eta} \in \mathcal{N}(\Omega; \widetilde{x})$  and  $\widetilde{\eta} \in \mathcal{N}(\Omega; \widehat{x})$ .

Moreover, since the sum of the two inequalities in (15), with  $y = \widetilde{x}$  in the former and  $y = \widehat{x}$  in 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 2  $\Phi(\tilde{x}) = \Phi(\hat{x})$ ,  $\tilde{\zeta} \in \partial h(\hat{x})$ , and  $\hat{\zeta} \in \partial h(\tilde{x})$ , we have for any  $y \in X$ ,

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

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

$$(y - \hat{x})^T (\Phi(\hat{x}) + \tilde{\zeta} + \tilde{\eta}) \geq 0,$$

which implies  $\tilde{\eta} \in \mathcal{E}(\hat{x})$ . This is enough to show that  $\mathcal{E}(\hat{x}) = \mathcal{E}(\tilde{x})$ . □

*Remark 1* Lemma 3 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 [4].

The next lemma deals with the minimization problem:

$$\underset{x \in \Omega \cap X}{\text{minimize}} f(x), \tag{16}$$

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

**Lemma 4** *Let  $\hat{x}$  be an optimal solution of (16) for which there exist  $\lambda \geq 0$  and  $\xi \in \partial \text{dist}(\hat{x}, \Omega)$  such that*

$$(\nabla f(\hat{x}) + \lambda \xi)^T (y - \hat{x}) \geq 0, \quad \forall y \in X. \tag{17}$$

*Then, for every  $\rho > \lambda$  the optimal solution sets of (16) and the penalized problem*

$$\underset{x \in X}{\text{minimize}} [f(x) + \rho \text{dist}(x, \Omega)] \tag{18}$$

*coincide.*

*Proof* It follows from the inequality (17) that  $\hat{x}$  is a minimizer on  $X$  of the convex function  $f + \lambda \text{dist}(\bullet, \Omega)$ . Let  $\bar{x}$  be a solution of (16) and let  $\rho > \lambda$ . We have, for any  $y \in X$ ,

$$\begin{aligned} f(y) + \rho \text{dist}(y, \Omega) &\geq f(y) + \lambda \text{dist}(y, \Omega) \\ &\geq f(\hat{x}) + \lambda \text{dist}(\hat{x}, \Omega) \\ &= f(\hat{x}) = f(\bar{x}) + \rho \text{dist}(\bar{x}, \Omega), \end{aligned}$$

thus showing that every solution of (16) is also a solution of (18).

Conversely, suppose that  $\bar{x}$  is a solution of (18). If  $\bar{x}$  belongs to  $\Omega$ , it is obvious that  $\bar{x}$  is also a solution of (16). Assume then that  $\bar{x} \notin \Omega$ , or equivalently,  $\text{dist}(\bar{x}, \Omega) > 0$ . We can write

$$\begin{aligned} f(\bar{x}) + \rho \text{dist}(\bar{x}, \Omega) &> f(\bar{x}) + \lambda \text{dist}(\bar{x}, \Omega) \\ &\geq f(\hat{x}) + \lambda \text{dist}(\hat{x}, \Omega) = f(\hat{x}) + \rho \text{dist}(\hat{x}, \Omega), \end{aligned}$$

where the second inequality follows from the fact that  $\hat{x}$  is a global minimizer of  $f + \lambda \text{dist}(\hat{x}, \Omega)$  on  $X$  and the third from  $\text{dist}(\hat{x}, \Omega) = 0$ . 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 (16).  $\square$

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 1** *Let  $\hat{x}$  be a solution of the HVI  $(X \cap \Omega, \Phi, h)$  with  $\mathcal{E}(\hat{x}) \neq \emptyset$ , 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  $\eta \in \mathcal{E}(\hat{x})$  be arbitrary. Then for every  $\rho > \|\eta\|$ , the solution set of the HVI  $(X \cap \Omega, \Phi, h)$  and that of the penalized HVI  $(X, \Phi, h + \rho \text{dist}(\bullet, \Omega))$  coincide.*

*Proof* Let  $\bar{x}$  be any solution of HVI  $(X \cap \Omega, \Phi, h)$ . By Lemma 3, it follows that  $\eta \in \mathcal{E}(\bar{x})$ . Let  $\lambda \triangleq \|\eta\|$  and

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

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

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

Lemma 4 tells us that  $\bar{x}$  is also a minimizer of

$$\underset{y \in X}{\text{minimize}} \left[ \left( \Phi(\bar{x}) + \bar{\zeta} \right)^T (y - \bar{x}) + \rho \text{dist}(y, \Omega) \right] \tag{19}$$

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

Conversely, suppose that  $\bar{x}$  is a solution of HVI  $(X, \Phi, h + \rho \text{dist}(\bullet, \Omega))$  for some  $\rho > \lambda$ . We only need to show that  $\bar{x} \in \Omega$ . Suppose the contrary, so that  $\text{dist}(\bar{x}, \Omega) > 0$ . We first note that by the first part of the theorem we know that also  $\hat{x}$  is a solution of the same HVI, so that by Lemma 2 we have  $\Phi(\hat{x}) = \Phi(\bar{x})$ . Let now  $\theta$  be a number in  $(\lambda, \rho)$ . Since by the first part of the theorem  $\hat{x}$  is still a solution of HVI  $(X, \Phi, h + \theta \text{dist}(\bullet, \Omega))$  we can write, noting that  $\bar{x}$  belongs to  $X$  and recalling that  $\text{dist}(\hat{x}, \Omega) = 0$ ,

$$\Phi(\hat{x})^T (\bar{x} - \hat{x}) + h(\bar{x}) - h(\hat{x}) + \theta \text{dist}(\bar{x}, \Omega) \geq 0.$$

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

$$\begin{aligned} &\Phi(\bar{x})^T(\widehat{x} - \bar{x}) + h(\widehat{x}) - h(\bar{x}) + \rho \text{dist}(\widehat{x}, \Omega) \\ &\leq \theta \text{dist}(\bar{x}, \Omega) \\ &< \rho \text{dist}(\bar{x}, \Omega) \\ &\leq \Phi(\bar{x})^T(\bar{x} - \bar{x}) + h(\bar{x}) - h(\bar{x}) + \rho \text{dist}(\bar{x}, \Omega). \end{aligned}$$

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

*Remark 2* In the above developments, the norm used was the Euclidean norm. Theorem 1 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,  $\partial \text{dist}(x, \Omega) = \mathcal{N}(x; \Omega) \cap \mathbb{B}_D$ , where  $\mathbb{B}_D$  is the unit ball in the norm  $\|\bullet\|_D$ . Then it is immediate to check that Theorem 1 still holds for  $\rho > \|\eta\|_D$ .

#### 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  $\text{SOL}(K, F)$ . Foremost is the affine case where  $K$  is a polyhedron and  $F(x) = q + Mx$  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,  $\text{SOL}(K, q, M)$  has the polyhedral representation given by [15, 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  $\text{SOL}(K, q, M)$ , would admit a distributed implementation when the problem  $(K, F, \Phi, h)$  has the requisite Cartesian structure; see Sect. 5. Another case where  $\text{SOL}(K, F)$  has an explicit representation that could be exploited by a centralized algorithm is when  $F(\text{SOL}(K, F))$  is a singleton (such as when  $F$  is pseudomonotone plus on  $F$ ). However, the representation [15, Proposition 2.3.12] of  $\text{SOL}(K, F)$  in this case is not so easy to take advantage of, unless more details are available on the pair  $(K, F)$ . Lastly, even though  $\text{SOL}(K, F)$  is a closed convex set for a convex-monotone pair  $(K, F)$ , its representation in general, cf. [15, 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  $F + \varepsilon_k(\Phi + \partial h) + \alpha(\bullet - x_k)$  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

$$\text{HVI}(K, F^k, \varepsilon_k h), \quad \text{where } F^k \triangleq F + \varepsilon_k \Phi + \alpha(\bullet - x_k). \tag{20}$$

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

$$[F(\bar{x}_{k+1}) + \varepsilon_k(\Phi(\bar{x}_{k+1}) + \bar{\zeta}_{k+1}) + \alpha(\bar{x}_{k+1} - x_k)]^T (y - \bar{x}_{k+1}) \geq 0. \tag{21}$$

Since  $F^k + \varepsilon_k \partial h$  is a strongly monotone set-valued mapping because of (B), (D), and (E), the HVI (20) has one and only one solution, so that  $\bar{x}_{k+1}$  is well defined.

---

**Algorithm 1: Inexact Prox-Tikhonov Algorithm**

---

- (S.0) : Let  $\{e_k\}$  be a sequence of nonnegative scalars, and  $\{\varepsilon_k\}$  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  $\|x_{k+1} - \bar{x}_{k+1}\| \leq e_k$ , where  $\bar{x}_{k+1} \in K$  is an exact solution of (20).
  - (S.3) : Set  $k \leftarrow k + 1$  and return to Step 1.
- 

Note that if  $e_k = 0$  we solve (20) 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  $SOL(K, F, \Phi, h)$ , is non-empty and bounded; moreover, the set  $L$  defined below is bounded:

$$L \triangleq \{x \in K \mid \exists y \in SOL(K, F, \Phi, h) \text{ such that } \Phi(x)^T(y - x) + h(y) - h(x) > 0\}.$$

Since  $SOL(K, F, \Phi, h)$  is the solution set of the monotone GVI  $(SOL(K, F), \Phi + \partial h)$ ,  $SOL(K, F, \Phi, h)$  is a non-empty, compact, convex set. The boundedness requirement of the two sets,  $SOL(K, F, \Phi, h)$  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 [15, Exercise 2.9.11]. The following example shows that neither  $K$  nor  $SOL(K, F)$  need to be bounded under all the assumptions made herein.

*Example 1* 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,  $SOL(K, F)$  coincides with  $K$ , which is unbounded. It is easy to check that  $SOL(K, F; \Phi, h) = [-1, 0]$  and  $L = \emptyset$ .

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

C1  $\sum_{k=0}^{\infty} \varepsilon_k = \infty$ ;

C2  $\{e_k/\varepsilon_k\} \rightarrow 0$ .

Condition C1 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  $\alpha(\bullet - x_k)$  dominates the term  $\varepsilon_k \Phi$  and the sequence  $\{x_k\}$  “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  $\text{SOL}(K, F)$ . By the same token, Condition C2 says that the inexact solution we employ in place of the exact solution of the GVI (20) 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  $\{e_k/\varepsilon_k\} \rightarrow 0$ ; 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 1 we need a preliminary lemma.

**Lemma 5** *Let  $X \subseteq \mathbb{R}^n$  be a closed convex set,  $G : X \rightarrow \mathbb{R}^n$  a monotone plus and continuous function, and  $f : X \rightarrow \mathbb{R}$  a convex function. If a solution  $y$  to the HVI  $(X, G, f)$  and a point  $x \in X$  exist such that*

$$G(x)^T(y - x) + f(y) - f(x) \geq 0, \tag{22}$$

*then  $x$  is also a solution to the HVI  $(X, G, f)$ .*

*Proof* Since  $y$  is a solution we can write

$$G(y)^T(z - y) + f(z) - f(y) \geq 0 \quad \forall z \in X.$$

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

$$G(y)^T(x - y) + f(x) - f(y) \geq 0. \tag{23}$$

Summing (22) and (23) we have

$$(G(x) - G(y))^T(y - x) \geq 0,$$

which implies, by monotonicity,

$$(G(x) - G(y))^T(x - y) = 0.$$

By the plus property, we obtain

$$G(x) = G(y).$$

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

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

thus showing that  $x$  is a solution of the HVI  $(X, G, f)$ . □

*Remark 3* If  $f \equiv 0$ , it is easy to see that Lemma 5 still holds if we only require  $G$  to be pseudo monotone plus on  $X$  instead of monotone plus.

The next theorem is the main result of this section. In this theorem we use a global Lipschitz continuity assumption that, however, is only required if inexactness is allowed in the solution of the sub-HVI in (20). 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 2** *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 1 is well defined and produces a bounded sequence  $\{x_k\}$  such that each of its limit points is a solution of VI-C HVI  $(K, F, \Phi, h)$ .*

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

$$\delta_k \triangleq \frac{1}{2} \text{dist}(x_k, S_\Phi^h) = \frac{1}{2} \|x_k - P_{S_\Phi^h}(x_k)\|^2.$$

By (21), we get, for  $y \in K$ ,

$$\begin{aligned} & [F(\bar{x}_{k+1}) + \alpha(\bar{x}_{k+1} - x_k)]^T (y - \bar{x}_{k+1}) \\ & \geq \varepsilon_k \left[ (h(\bar{x}_{k+1}) - h(y)) + \Phi(\bar{x}_{k+1})^T (\bar{x}_{k+1} - y) \right]. \end{aligned} \tag{24}$$

We have for every  $y \in K$

$$\begin{aligned} & \alpha(x_k - x_{k+1})^T (y - x_{k+1}) \\ & \leq \alpha \left[ (x_k - \bar{x}_{k+1})^T (y - \bar{x}_{k+1}) + (\bar{x}_{k+1} - x_{k+1})^T (y - x_{k+1}) \right. \\ & \quad \left. + (x_k - \bar{x}_{k+1})^T (\bar{x}_{k+1} - x_{k+1}) \right] \\ & \leq \alpha(\bar{x}_{k+1} - x_{k+1})^T (y - x_{k+1}) + F(\bar{x}_{k+1})^T (y - \bar{x}_{k+1}) \\ & \quad + \alpha(x_k - \bar{x}_{k+1})^T (\bar{x}_{k+1} - x_{k+1}) \\ & \quad + \varepsilon_k \left[ h(y) - h(\bar{x}_{k+1}) + \Phi(\bar{x}_{k+1})^T (y - \bar{x}_{k+1}) \right] \\ & \leq \eta_k + F(x_{k+1})^T (y - x_{k+1}) \\ & \quad + \varepsilon_k \left[ (h(y) - h(x_{k+1})) + \Phi(x_{k+1})^T (y - x_{k+1}) \right], \end{aligned} \tag{25}$$

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

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

We can write

$$\begin{aligned}
 \delta_{k+1} - \delta_k &= \frac{1}{2} \|x_{k+1} - P_{S_\phi^h}(x_{k+1})\|^2 - \frac{1}{2} \|x_k - P_{S_\phi^h}(x_k)\|^2 \\
 &\leq \frac{1}{2} \|x_{k+1} - P_{S_\phi^h}(x_k)\|^2 - \frac{1}{2} \|x_k - P_{S_\phi^h}(x_k)\|^2 \\
 &= -\frac{1}{2} \|x_{k+1} - x_k\|^2 + (x_k - x_{k+1})^T (P_{S_\phi^h}(x_k) - x_{k+1}) \\
 &\leq -\frac{1}{2} \|x_{k+1} - x_k\|^2 + \frac{1}{\alpha} F(x_{k+1})^T (P_{S_\phi^h}(x_k) - x_{k+1}) \\
 &\quad + \frac{\varepsilon_k}{\alpha} \left[ (h(P_{S_\phi^h}(x_k)) - h(x_{k+1})) + \Phi(x_{k+1})^T (P_{S_\phi^h}(x_k) - x_{k+1}) \right] + \frac{\eta_k}{\alpha} \\
 &\leq -\frac{1}{2} \|x_{k+1} - x_k\|^2 \\
 &\quad + \frac{\varepsilon_k}{\alpha} \underbrace{[(h(P_{S_\phi^h}(x_k)) - h(x_{k+1})) + \Phi(x_{k+1})^T (P_{S_\phi^h}(x_k) - x_{k+1}))]}_{V_{k+1}} + \frac{\eta_k}{\alpha}
 \end{aligned}
 \tag{26}$$

The first inequality is obvious by the definition of projection; the second is obtained from (25) evaluated at  $y = P_{S_\phi^h}(x_k) \in K$ . The third inequality can be obtained by observing that since  $P_{S_\phi^h}(x_k) \in S_\phi^h \subseteq \text{SOL}(K, F)$ , we have  $F(P_{S_\phi^h}(x_k))^T(x_{k+1} - P_{S_\phi^h}(x_k)) \geq 0$ , which yields in turn, by the monotonicity of  $F$ ,  $F(x_{k+1})^T(P_{S_\phi^h}(x_k) - x_{k+1}) \leq 0$ . We now consider three cases.

Case 1: Eventually,  $\varepsilon_k V_{k+1} + \eta_k \leq 0$ .

In this case the nonnegative sequence  $\{\delta_k\}$  is (eventually) non-increasing and therefore convergent. Since  $S_\phi^h$  is bounded, this implies that also  $\{x_k\}$  is bounded. Furthermore  $\{\delta_{k+1} - \delta_k\}$  converges to zero. From (26), we have

$$\delta_{k+1} - \delta_k \leq -\frac{1}{2} \|x_{k+1} - x_k\|^2,$$

which shows that

$$\lim_{k \rightarrow \infty} \|x_{k+1} - x_k\| = 0.
 \tag{27}$$

Summing over  $i$  from  $k_0$  to  $k - 1$  we get from (26)

$$\delta_k - \delta_{k_0} \leq \frac{1}{\alpha} \sum_{i=k_0}^{k-1} \varepsilon_i V_{i+1} + \frac{1}{\alpha} \sum_{i=k_0}^{k-1} \eta_i.
 \tag{28}$$

Since  $\varepsilon_{k-1} V_k \leq \varepsilon_{k-1} V_k + \eta_{k-1} \leq 0$  we have that all  $V_k$  are non positive. We show that  $\limsup_{k \rightarrow \infty} V_k = 0$ . Suppose by contradiction this is not the case, so that a positive constant  $\bar{V}$  exists such that  $V_k \leq -\bar{V} < 0$  for all  $k$ . This implies that the

right-hand side in (28) 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  $\{c_i\}$  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 (28),

$$\delta_k - \delta_{k_0} \leq \frac{1}{\alpha} \sum_{i=k_0}^{k-1} \varepsilon_i V_{i+1} + \frac{1}{\alpha} \sum_{i=k_0}^{k-1} \eta_i \leq -\frac{\bar{V}}{2\alpha} \sum_{i=k_0}^{k-1} \varepsilon_i.$$

Since  $\{\delta_k\}$  converges, we get a contradiction to C1 and have thus proved that  $\limsup_{k \rightarrow \infty} V_k = 0$ . Therefore a subsequence  $J$  exists such that

$$\lim_{\substack{k \in J \\ k \rightarrow \infty}} V_k = 0. \tag{29}$$

Since  $\{x_k\}$  is bounded, we may assume, without loss of generality, that

$$\lim_{\substack{k \in J \\ k \rightarrow \infty}} x_k = \tilde{x}.$$

Note that since  $K$  is closed,  $\tilde{x} \in K$ . We show that actually  $\tilde{x} \in \text{SOL}(K, F)$ . If this is not so, there exists a point  $y \in K$  such that  $F(\tilde{x})^T(y - \tilde{x}) < 0$ . By (25) we can write

$$\begin{aligned} & [F(x_k) + \alpha(x_k - x_{k-1})]^T(y - x_k) \\ & + \varepsilon_{k-1} [h(y) - h(x_k) + \Phi(x_k)^T(y - x_k)] \geq -\eta_{k-1}. \end{aligned} \tag{30}$$

By continuity, the definition of  $y$ , the boundedness of  $\{x_k\}$  and (27), we have, without loss of generality (after a suitable renumeration),

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

and

$$\lim_{\substack{k \in J \\ k \rightarrow \infty}} \varepsilon_{k-1} [h(y) - h(x_k) + \Phi(x_k)^T(y - x_k)] = 0,$$

Then we get a contradiction to (30) since  $\{\eta_k\} \downarrow 0$ . Therefore  $\tilde{x} \in \text{SOL}(K, F)$ . Thanks to (27) we have  $\lim_{k \in J, k \rightarrow \infty} x_{k-1} = \tilde{x}$ . Therefore, by (29) and continuity, we get  $h(P_{S_\Phi^h}(\tilde{x})) - h(\tilde{x}) + \Phi(\tilde{x})^T(P_{S_\Phi^h}(\tilde{x}) - \tilde{x}) = 0$ . By Lemma 5, it follows that  $\tilde{x} \in S_\Phi^h$ . Therefore we get

$$\lim_{\substack{k \in J \\ k \rightarrow \infty}} \delta_k = 0.$$

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

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

$$J \triangleq \{k \mid \varepsilon_{k-1} V_k + \eta_{k-1} > 0\}$$

and

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

By (26), 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  $\{\delta_k\}$  goes to zero it is enough to show that the subsequence  $\{\delta_k\}_{\bar{J}}$  converges to zero. To this end, first observe that for every  $k \in \bar{J}$  it holds that

$$\varepsilon_{k-1} V_k + \eta_{k-1} > \frac{\alpha}{2} \|x_k - x_{k-1}\|^2. \tag{31}$$

The sequence  $\{x_k\}_{\bar{J}}$  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  $\text{dist}(x_{k+1}, K) \leq \|x_{k+1} - \bar{x}_{k+1}\| \leq e_k$ , from which the boundedness of  $\{x_k\}$  follows from that of  $K$ . By continuity, also  $\{V_k\}_{\bar{J}}$  is bounded. Hence, (31) yields

$$\lim_{\substack{k \in \bar{J} \\ k \rightarrow \infty}} \|x_k - x_{k-1}\| = 0. \tag{32}$$

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

$$\lim_{\substack{k \in \tilde{J} \\ k \rightarrow \infty}} x_k = \tilde{x}.$$

Reasoning exactly as in Case 1, [(the only difference is that instead of (27) we use (32)], we may deduce that  $\tilde{x} \in \text{SOL}(K, F)$ . By continuity, recalling that  $\tilde{J} \subseteq \bar{J} \subseteq J$  and Condition (C2), which implies  $\{\eta_k/\varepsilon_k\} \rightarrow 0$ , we get  $h(P_{S_\Phi^h}(\tilde{x})) - h(\tilde{x}) + \Phi(\tilde{x})^T (P_{S_\Phi^h}(\tilde{x}) - \tilde{x}) \geq 0$ . Thus  $\tilde{x} \in S_\Phi^h$ ; hence  $\lim_{k \in \tilde{J}, k \rightarrow \infty} \delta_k = 0$ . Since this reasoning can be repeated for every convergent subsequence of  $\{x_k\}_{\bar{J}}$ , we may conclude that  $\lim_{k \in \bar{J}, k \rightarrow \infty} \delta_k = 0$ , 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  $\{\delta_k\}$  is non-increasing eventually. Therefore  $\{\delta_k\}$  converges, implying that  $\{x_k\}$  is bounded. Because  $\{\delta_k\}$  converges we also have that  $\{\delta_{k+1} - \delta_k\}$  converges to zero. Therefore, by (26), the boundedness of  $\{x_k\}$  and taking into account that both  $\{\varepsilon_k\}$  and  $\{\eta_k\}$  converge to zero, we get that (27) holds. Since  $J$  is infinite and  $\{x_k\}$  is bounded we can now consider a convergent subsequence  $\{x_k\}_{\tilde{J}} \rightarrow \tilde{x}$ , with  $\tilde{J} \subseteq J$ . Reasoning as in Case 2 we can prove that  $\tilde{x} \in S_\Phi^h$ ; hence  $\lim_{k \in \tilde{J}, k \rightarrow \infty} \delta_k = 0$ .

But since the whole sequence  $\{\delta_k\}$  converges, this is enough to show that  $\{\delta_k\}$  converges to zero, thus concluding the proof of the theorem.  $\square$

*Remark 4* It should be clear that all the results in Theorem 2 still hold if we change  $\alpha$  at each iteration, provided that these alpha 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  $h \equiv 0$ , it can be checked, going through the proof and taking into account Remark 3, that Theorem 2 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 1 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 are 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 1, i.e., if  $e_k > 0$ , we need to translate the theoretical criterion  $\|x_{k+1} - \bar{x}_{k+1}\| \leq e_k$  at Step 2 into something computable, since the exact solution  $\bar{x}_{k+1}$  of the sub-HVI (20) 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 ( $h \equiv 0$ ). In this case a rather complete theory of “error bounds” is available, see [15, Chapter 6], and can be used. For example, by [15, Proposition 6.3.1] it follows that at each iteration we have, for some constant  $C > 0$

$$\|x_{k+1} - \bar{x}_{k+1}\| \leq \frac{C+1}{\alpha} \|x_{k+1} - P_K(x_{k+1} - \Phi(x_{k+1}))\|,$$

with the right-hand norm being a practically computable quantity. Therefore, in a practical implementation of Algorithm 1 we could simply stop when  $\|x_{k+1} - P_K(x_{k+1} - \Phi(x_{k+1}))\| \leq e_k$  without using any of the theoretical properties in Theorem 2. Other possibilities are available and a choice should be made taking into account the particular problem at hand. We leave these details to the reader. The last section of this paper give a substantial example of how the results in this and in the next section can be used in a practical environment.

## 5 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 [17, 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  $(K, \Phi, h)$ .

The setting is as follows. Consider the HVI  $(K, G, h)$  defined by the product set  $K \triangleq \prod_{v=1}^N K_v$  with  $\sum_{v=1}^N n_v = n$ , where each  $K_v$  is a closed convex subset of  $\mathbb{R}^{n_v}$ , by the continuous function  $G : K \rightarrow \mathbb{R}^n$ , and the separable function  $h : K \rightarrow \mathbb{R}^n$  such that  $h(x) = \sum_{v=1}^N h_v(x^v)$ , where  $x = (x^v)_{v=1}^N$  with each  $x^v \in K_v$  and  $h_v$  is

a convex continuous function defined on  $K_\nu$ . This setting is applicable to the side-constrained VI-C VI  $(\text{SOL}(K, F) \cap \Omega, \Phi)$  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 2, the latter VI, which is defined on the intersection  $\text{SOL}(K, F) \cap \Omega$ , is equivalent to the HVI  $(K, \Phi, \rho h)$  which is defined on the set  $K$  for all  $\rho > 0$  sufficiently large, where  $h(x) \triangleq \sum_{\nu=1}^N \text{dist}(x^\nu, K_\nu)$ .

With the pair  $(K, h)$  having the above partitioned structure, we partition  $G$  accordingly; i.e.,  $G(x) = (G^\nu(x))_{\nu=1}^N$  with  $G^\nu(x) \in \mathbb{R}^{n_\nu}$  for all  $\nu = 1, \dots, N$ . In what follows if  $x = (x^\nu)_{\nu=1}^N$ , we denote by  $x^{-\nu}$  the sub-vector of  $x$  with the  $\nu$ -th block omitted:  $x^{-\nu} \triangleq (x^\mu)_{\mu=1, \mu \neq \nu}^N$ , 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 easily to synchronous Gauss–Seidel schemes and to asynchronous versions of both the Jacobi and Gauss–Seidel methods; we leave these details to the reader.

---

**Algorithm 2: Distributed Algorithm for HVIs**

---

- (S.0) : Choose  $x_0 \in K$  and set  $k = 0$ .
  - (S.1) : If  $x_k$  is a solution of HVI  $(K, G, h)$  stop.
  - (S.2) : For  $\nu = 1, \dots, N$  set (33)
  - $x_{k+1}^\nu \triangleq$  solution of HVI  $(K_\nu, G_\nu(\bullet, x_k^{-\nu}), h_\nu(\bullet))$
  - (S.3) : Set  $x_{k+1} = (x_{k+1}^\nu)_{\nu=1}^N$ ,  $k \leftarrow k + 1$  and go to Step 1.
- 

The defining equation (33) implicitly assumes that the HVI in (33) 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  $\{x_k\}$  produced by Algorithm 2 is well defined and converges to a solution of HVI  $(K, G, h)$ . This is done by showing that, under appropriate assumptions, Algorithm 2 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

$$\beta(X, H) \triangleq \sup_{z, z' \in X, z \neq z'} \frac{\|H(z) - H(z')\|}{\|z - z'\|}.$$

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

$$\gamma(X, H) \triangleq \inf_{z, z' \in X, z \neq z'} \frac{(H(z) - H(z'))^T (z - z')}{\|z - z'\|^2}.$$

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  $JH(z)$ . 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 2** *If  $H : X \rightarrow \mathbb{R}^m$  is continuously differentiable on an open set containing the convex set  $X \subseteq \mathbb{R}^n$  and  $\|JH(z)\|$  is bounded on  $X$ , then  $\beta(X, H) = \sup_{z \in X} \|JH(z)\|$  is finite. Furthermore if  $m = n$  and  $H$  is strongly monotone on  $X$ , then  $\gamma(X, H) = \inf_{z \in X} \lambda_m^S(JH(z))$  is finite and positive.*

The following is the key assumption necessary to prove convergence of Algorithm 2 for solving the HVI  $(K, G, h)$ .

**Assumption P** Each function  $G^v(\bullet, z^{-v})$  is strongly monotone on  $K_v$  with a uniformly positive strong monotonicity modulus for all  $z^{-v} \in K_{-v}$ ; i.e.,

$$\gamma_v \triangleq \inf_{z^{-v} \in K_{-v}} \gamma(K_v, G^v(\bullet, z^{-v})) > 0, \quad \forall v = 1, \dots, N.$$

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

$$\|G^v(z^v, u^{-\mu}) - G^v(z^v, v^{-\mu})\| \leq \sum_{\mu \neq v} \beta_{v\mu} \|u^\mu - v^\mu\|.$$

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

$$\Upsilon \triangleq \begin{bmatrix} \gamma_1 & -\beta_{12} & \cdots & -\beta_{1N} \\ -\beta_{21} & \gamma_2 & \cdots & -\beta_{2N} \\ \vdots & & \ddots & \vdots \\ -\beta_{N1} & -\beta_{N2} & \cdots & \gamma_N \end{bmatrix}.$$

[The matrix  $\Upsilon$  has the ‘‘Z-property’’, i.e., its off-diagonal entries are all non-positive.]

*Remark 5* The constant  $\beta_{v\mu}$  is given by

$$\begin{aligned} & \sup_{z^v \in K_v} \sup_{\substack{u^\mu, v^\mu \in K_\mu \\ u^\mu \neq v^\mu}} \sup \left\{ \frac{G^v(z^v, y, u^\mu) - G^v(z^v, y, v^\mu)}{\|u^\mu - v^\mu\|} \mid y \in \prod_{v' \neq v, \mu} K_{v'} \right\} \\ & = \sup_{z^v \in K_v} \sup \left\{ \beta(K_\mu, G^v(z^v, y, \bullet)) \mid y \in \prod_{v' \neq v, \mu} K_{v'} \right\}. \end{aligned}$$

By Proposition 2 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  $JG$ . 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  $\beta_{v\mu}$  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 2 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  $(K_\nu, G_\nu(\bullet, x^{-\nu}), h_\nu)$  in which we decompose the original HVI  $(K, G, h)$  (see also Step 2 of Algorithm 2) 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  $(K_\nu, G_\nu(\bullet, x^{-\nu}), h_\nu)$  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  $(K_\nu, G_\nu(\bullet, z^{-\nu}), h_\nu)$  is strongly monotone and therefore has one and only one solution, which we denote by  $x^\nu(z)$ ; we then let  $x(z) \triangleq (x^\nu(z))_{\nu=1}^N$ . This is a self-map from the set  $K$  into itself; Algorithm 2 is easily seen to be the fixed-point iteration:  $x_{k+1} = x(x_k)$ .

We are now ready to state the main result of this section. Besides treating the HVI, this result extends the treatment in [17, Section 12.6] for a Nash game in that  $G$  is not required to be twice continuously differentiable here.

**Theorem 3** *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 2 generates a well-defined sequence  $\{x_k\}$  that converges to the unique solution of 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,

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

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

$$\left[ G^\nu(x^\nu(\bar{z}), \bar{z}^{-\nu}) - G^\nu(x^\nu(\tilde{z}), \tilde{z}^{-\nu}) \right]^T (x^\nu(\tilde{z}) - x^\nu(\bar{z})) \geq 0. \tag{34}$$

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

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

from which we deduce

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

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

$$\|G^v(x^v(\bar{z}), \bar{z}^{-v}) - G^v(x^v(\tilde{z}), \tilde{z}^{-v})\| \leq \sum_{\mu \neq v} \beta_{v\mu} \|\bar{z}^\mu - \tilde{z}^\mu\|$$

The last two inequalities immediately give

$$\|x^v(\bar{z}) - x^v(\tilde{z})\| \leq \frac{1}{\gamma_v} \sum_{\mu \neq v} \beta_{v\mu} \|\bar{z}^\mu - \tilde{z}^\mu\|$$

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

$$\begin{pmatrix} \|x^1(\bar{z}) - x^1(\tilde{z})\| \\ \vdots \\ \|x^N(\bar{z}) - x^N(\tilde{z})\| \end{pmatrix} \leq \Gamma \begin{pmatrix} \|\bar{z}^1 - \tilde{z}^1\| \\ \vdots \\ \|\bar{z}^N - \tilde{z}^N\| \end{pmatrix}, \tag{35}$$

where

$$\Gamma \triangleq \begin{bmatrix} 0 & \frac{\beta_{12}}{\gamma_1} & \dots & \frac{\beta_{1N-1}}{\gamma_1} & \frac{\beta_{1N}}{\gamma_1} \\ \frac{\beta_{21}}{\gamma_2} & 0 & \dots & \frac{\beta_{2N-1}}{\gamma_2} & \frac{\beta_{2N}}{\gamma_2} \\ \vdots & & & & \vdots \\ \frac{\beta_{N1}}{\gamma_N} & \frac{\beta_{N2}}{\gamma_N} & \dots & \frac{\beta_{NN-1}}{\gamma_N} & 0 \end{bmatrix}.$$

Taking into account Assumption P, [11, Lemma 5.13.14] implies that  $\Gamma$  has spectral radius less than 1 (see [17, Proposition 12.7] for details) thus concluding the proof.  $\square$

The significance of the previous result is that it fits very well with the results in Sect. 4. In fact, note that the parameter  $\alpha$  can be chosen large enough so that the HVI  $(K, F^k, \varepsilon_k h)$  in Step 2 of Algorithm 1, where  $F^k \triangleq F + \varepsilon_k \Phi + \alpha(\bullet - x^k)$ , 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 Sect. 3, also of VI-C HVIs with side constraints.

**Corollary 1** *Consider the HVI  $(K, F + \varepsilon_k \Phi + \alpha(\bullet - x_k), \varepsilon_k h)$  (arising from Step 2 of Algorithm 1) and suppose that  $F$  and  $\Phi$  are Lipschitz continuous on  $K$  with moduli  $L_F$  and  $L_\Phi$ , respectively, and that all  $\gamma_v$  are nonnegative. Let  $\bar{\varepsilon}$  be such that  $\varepsilon_k \leq \bar{\varepsilon}$  for all  $k$  and let  $\alpha$  be a positive constant such that*

$$\alpha > \bar{\alpha} \triangleq \left( \max_{1 \leq v \leq N} \sum_{\mu \neq v} (\beta_{v\mu} + \bar{\varepsilon} L_\Phi) \right) - \min_{1 \leq v \leq N} \gamma_v, \tag{36}$$

where  $\beta_{v\mu}$  and  $\gamma_v$  are the constants defined in Assumption P, with  $G = F$ . Then all subproblems appearing at Step 2 of Algorithm 1 can be solved by the distributed Algorithm 2.

*Proof* According to Theorem 3 we only need to show that the matrix  $\Upsilon$  associated with a generic HVI  $(K, F + \varepsilon_k \Phi + \alpha(\bullet - x_k), \varepsilon_k h)$  has the P-property. Invoking [11, Lemma 5.13.14] this is equivalent to the following matrix

$$\begin{bmatrix} 0 & \frac{\beta_{12} + \bar{\varepsilon}L\Phi}{\gamma_1 + \alpha} & \cdots & \frac{\beta_{1N} + \bar{\varepsilon}L\Phi}{\gamma_1 + \alpha} \\ \vdots & & & \vdots \\ \frac{\beta_{N1} + \bar{\varepsilon}L\Phi}{\gamma_N + \alpha} & \cdots & \frac{\beta_{N(N-1)} + \bar{\varepsilon}L\Phi}{\gamma_N + \alpha} & 0 \end{bmatrix}$$

having spectral radius less than 1. It can now be checked that (36) 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  $\gamma_v$ , the smaller  $\alpha$  can be.

### 6 VI-C minimization problems

A particularly important case of the general HVI is the hierarchical minimization problem (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

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

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 [3] for computing a stationary point of (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 1 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  $\psi$ .

Convergence of Algorithm 3 follows easily from standard results, once one can show that it is well defined.

---

**Algorithm 3: A descent method for non-convex  $\psi$**

---

- (S. 0) : Choose  $x_0 \in K, \beta \in (0, 1), \delta \in (0, 1)$  and set  $k = 0$ .
  - (S. 1) : If  $x_k \in \text{SOL}(\text{SOL}(K, F), \nabla\psi)$ , stop.
  - (S. 2) : Use Algorithm 1 to calculate a solution  $y_k$  of the following problem: (37)  

$$\underset{y \in \text{SOL}(K, F)}{\text{minimize}} \quad \nabla\psi(x_k)^T y + \frac{1}{2} y^T Q_k y,$$
  - 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 (38)  

$$\psi(x_k + \beta^i d_k) \leq \psi(x_k) + \delta \beta^i \nabla\psi(x_k)^T d_k$$
and set  $x_{k+1} \triangleq x_k + \beta^i d_k$ .
  - (S. 4) : Set  $k \leftarrow k + 1$  and return to Step 1.
- 

**Theorem 4** 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 3 is well defined and every limit point of the sequence  $\{x_k\}$  it generates is a stationary point of the minimization problem (2).

*Proof* We only need to show that we are able to solve subproblem (37); convergence then follows by well-established results on the conditional gradient method, see for example [3]. Subproblem (37) is just the VI-C HVI  $(K, F, 0, h_k)$ , where  $h_k(y) \triangleq \nabla\psi(x_k)^T y + \frac{1}{2} y^T Q_k y$ ; 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.  $\square$

### 7 An application to a rate maximization game

In this section, we illustrate the application of the theoretical framework developed in the paper 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  $p^\nu \triangleq (p_n^\nu)_{n=1}^N \geq 0$ , representing

the power allocated over the  $N$  channels by the transmission-receiver pair  $\nu$ . As such, these variables satisfy some bound constraints  $0 \leq p_n^\nu \leq p_n^{\nu, \max}$ ,  $\nu = 1, \dots, Q$ ,  $n = 1, \dots, N$ , 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

$$P^\nu \triangleq \left\{ p^\nu \in \mathbb{R}_+^N : \sum_{n=1}^N p_n^\nu \leq P^\nu, \quad 0 \leq p^\nu \leq p^{\nu, \max} \triangleq (p_n^{\nu, \max})_{n=1}^N \right\}. \quad (39)$$

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):

$$\text{SINR}_n^\nu(p_n^\nu, p_n^{-\nu}) = \frac{|H_{\nu\nu}(n)|^2 p_n^\nu}{\sigma_n^{\nu 2} + \sum_{\mu \neq \nu} |H_{\nu\mu}(n)|^2 p_n^\mu}, \quad (40)$$

where  $|H_{\nu\nu}(n)| > 0$  is the channel gain of pair  $\nu$  over the frequency band  $n$ , and  $|H_{\nu\mu}(n)| \geq 0$  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  $p_n^{-\nu} \triangleq (p_n^1, \dots, p_n^{\nu-1}, p_n^{\nu+1}, \dots, p_n^Q)$  is the set of all the users power allocations over the channel  $n$ , except the  $\nu$ -th one. The useful power signal of pair  $\nu$  over the channel  $n$  is thus  $|H_{\nu\nu}(n)|^2 p_n^\nu$ , whereas  $\sum_{\mu \neq \nu} |H_{\nu\mu}(n)|^2 p_n^\mu$  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  $r^\nu(p^\nu, p^{-\nu})$  over the set of the  $N$  parallel channels, which depends on the power allocation of all the users  $(p^\nu, p^{-\nu})$ , with  $p^{-\nu} \triangleq (p^1, \dots, p^{\nu-1}, p^{\nu+1}, \dots, p^Q)$ . Under basic information theoretical assumptions (see, e.g., [36,43]) and given the users’ power allocation profile  $p^1, \dots, p^Q$ ,  $r^\nu(p^\nu, p^{-\nu})$  is

$$r^\nu(p^\nu, p^{-\nu}) = \sum_{n=1}^N \log(1 + \text{SINR}_n^\nu(p_n^\nu, p_n^{-\nu})). \quad (41)$$

*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 [25]. 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 computationally 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  $r^\nu(p^\nu, p^{-\nu})$ , i.e.,

$$\begin{aligned} & \underset{p^\nu}{\text{maximize}} \quad r^\nu(p^\nu, p^{-\nu}) \\ & \text{subject to} \quad p^\nu \in \mathcal{P}^\nu, \end{aligned} \quad (42)$$

for all  $\nu = 1, \dots, Q$ , where  $\mathcal{P}^\nu$  and  $r^\nu(p^\nu, p^{-\nu})$  are defined in (39) and (41), respectively. We will denote the above NEP as  $\mathcal{G} = \langle \mathcal{P}, (r_i)_{i=1}^Q \rangle$ , 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  $p^*$  such that  $p^{\nu,*}$  is optimal for (42), given  $p^{-\nu,*}$ .

Note that, for any fixed  $p^{-\nu} \geq 0$ , the single-user optimization problem in (42) admits a unique solution  $p^{\nu,*}$ , given by the well-known waterfilling expression [36,43]:

$$p_n^{\nu,*} = \text{wf}_n^\nu(p_n^{-\nu}) \triangleq \left( \lambda^\nu - \frac{\sigma_n^{\nu 2} + \sum_{\mu \neq \nu} |H_{\nu\mu}(n)|^2 p_n^\mu}{|H_{\nu\nu}(n)|^2} \right)^+, \quad (43)$$

with  $n = 1, \dots, N$ , where  $[x]^+ \triangleq \max(0, x)$ ,  $p_n^{-\nu} \triangleq (p_n^1, \dots, p_n^{\nu-1}, p_n^{\nu+1}, \dots, p_n^Q)$  and the waterlevel  $\lambda^\nu$  is chosen to satisfy the transmit power constraint  $\sum_{n=1}^N p_n^{\nu,*} = P^\nu$ ;  $\lambda^\nu$  can be computed very efficiently in at most  $N$  extremely simple steps. Interestingly, the best-response (43) can be computed locally and distributively by the players, since each user only needs to measure the overall interference-plus-noise PSD  $\sigma_n^{\nu 2} + \sum_{\mu \neq \nu} |H_{\nu\mu}(n)|^2 p_n^\mu$  and “waterfill” over it. The Nash equilibria  $p^*$  of the NEP are thus the fixed-points of the waterfilling mapping  $\text{wf}(p) \triangleq (\text{wf}^\nu(p^{-\nu}))_{\nu=1}^Q$ , with  $\text{wf}^\nu(p^{-\nu}) \triangleq (\text{wf}_n^\nu(p_n^{-\nu}))_{n=1}^N$ .

*Related Work.* Since the seminal paper of Yu et al. [43] 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 [23,34–37,43]. 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) [37]. In this algorithm, all the users update their power allocation according to the best-response waterfilling solution (43) in a totally asynchronous way (in the sense of [37]), 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 5 below, where the set  $\overline{\mathcal{P}}$  is defined as  $\mathcal{P}$  in (39) but with the power budget inequalities  $\sum_n p_n^v \leq P^v$ , replaced by  $\sum_n p_n^v = P^v$  for all  $v = 1, \dots, Q$ , and the vector function  $F(p) \triangleq (F^v(p))_{v=1}^Q$  and the matrices  $M \triangleq (M_{v\mu})_{v,\mu=1}^Q \in \mathbb{R}^{N \times Q \times N \times Q}$  and  $\Delta \in \mathbb{R}^{Q \times Q}$  are defined as

$$F^v(p) = \hat{\sigma}^v + \sum_{\mu=1}^Q M_{v\mu} p^\mu, \tag{44}$$

with

$$\hat{\sigma}^v \triangleq \left( \frac{\sigma_n^{v2}}{|H_{vv}(n)|^2} \right)_{n=1}^N \text{ and } M_{v\mu} \triangleq \text{diag} \left\{ \left( \frac{|H_{v\mu}(n)|^2}{|H_{vv}(n)|^2} \right)_{n=1}^N \right\}.$$

and

$$[\Delta]_{v\mu} \triangleq \begin{cases} 1, & \text{if } v = \mu; \\ -\max_n \frac{|H_{v\mu}(n)|^2}{|H_{vv}(n)|^2}, & \text{otherwise.} \end{cases} \tag{45}$$

**Theorem 5** *Given the NEP  $\mathcal{G} = \langle \mathcal{P}, (r_i)_{i=1}^Q \rangle$ , then the following hold.*

- (a) *The NEP is equivalent to the affine VI( $\overline{\mathcal{P}}, F$ ) [23], which has a nonempty and bounded solution set;*
- (b) *If  $M$  is positive definite (semidefinite), then the 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 (43) converges to the unique Nash equilibrium of  $\mathcal{G}$  [37].*

Theorem 5, 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 VI( $\overline{\mathcal{P}}, F$ ) is monotone, implying that one could apply double loop schemes based on Tikhonov [15, Algorithm 12.2.9] or proximal [15, Algorithm 12.3.8] regularization to

the 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 developed in this paper 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  $w \triangleq (w^v)_{v=1}^Q \geq 0$ , let

$$\phi(p) \triangleq \sum_{v=1}^Q w^v \sum_{\mu \neq v} \sum_{n=1}^N |H_{v\mu}(n)|^2 p_n^\mu. \tag{46}$$

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  $\sum_{v=1}^Q r^v(p)$ . The NE selection problem based on the merit function  $\phi$  can be then formulated as a hierarchical optimization problem with VI constraints:

$$\underset{p \in \text{SOL}(\overline{\mathcal{P}}, F)}{\text{minimize}} \phi(p), \tag{47}$$

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 (47) and devise distributed iterative algorithms. In particular, (47) can be solved using Algorithm 1, where the VI-C HVI reduces to VI-C 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

$$\text{VI} \left( \overline{\mathcal{P}}, F + \varepsilon_k \nabla\phi + \alpha \left( \bullet - p^{(k)} \right) \right). \tag{48}$$

The convergence of Algorithm 1 applied to (47) is guaranteed under the conditions of Theorem 2; note that assumptions (A, C, D, E) in the theorem are readily satisfied by (47), 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 (48). We can readily use Algorithm 2; it follows from Theorem 3 indeed that Algorithm 2 globally converges to the unique solution of the VI in (48) if the matrix

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



with  $\Delta$  defined in (45), 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  $v$  solves the quadratic optimization problem: given  $p^{-v} \geq 0$ ,

$$\underset{p^v \in \overline{\mathcal{P}}^v}{\text{minimize}} \quad \frac{1}{2} \left\| p^v + \left( \hat{\sigma}^v + \sum_{\mu \neq v}^Q M_{v\mu} p^\mu \right) \right\|^2 + \varepsilon_k \gamma^{vT} p^v + \frac{\alpha}{2} \left\| p^v - p^{(k),v} \right\|^2, \tag{49}$$

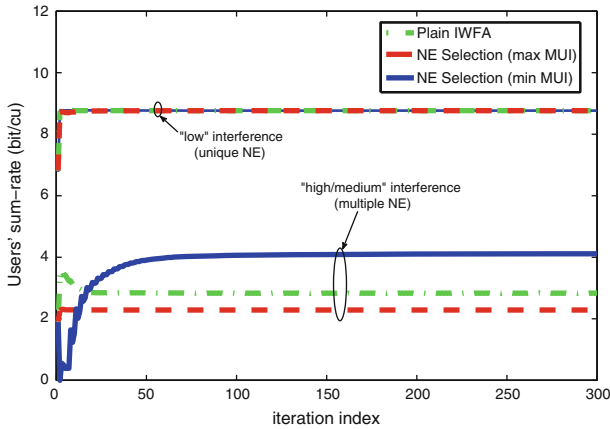
where  $\gamma^v \triangleq (\sum_{\mu \neq v} w^\mu |H_{\mu v}(n)|^2)_{n=1}^N$ . The solution of (49), has a similar waterfilling-like expression as (43) and thus can be efficiently and locally computed, given the MUI  $\hat{\sigma}^v + \sum_{\mu \neq v}^Q M_{v\mu} p^\mu$  measured at each receiver  $v$ ,  $p^{(k),v}$ , and  $\varepsilon_k$ . The asynchronous implementation of Algorithm 2 is also possible, whose convergence is guaranteed under the same P property of matrix  $\Delta_\alpha$ .

*Remark 6* Overall, in Algorithm 1 applied to (47), there are two levels of updates: (1) the computation of the users’ optimal power allocations [the (approximate) solution of the sub-VI (48)], given  $p^{(k),v} \in \overline{\mathcal{P}}$  and  $\varepsilon_k$ ; and (2) the updates of  $p^{(k),v}$  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 [38] for practical implementations of this check, under different level of signalling and supervision. It turns out that Algorithm 1 has the same desired properties as the asynchronous IWFA [37]: 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 [23,36–38] for solving the rate maximization game  $\mathcal{G}$ .

In Fig. 1, we plot the sum-rate of the users  $\sum_{v=1}^Q r^v(p)$  versus the iteration index, achieved by the following algorithms: (1) The simultaneous IWFA [36] based on the waterfilling map  $\text{wf}(p)$  (dashed-dot line curve); (2) Algorithm 1 applied to (47) (solid line curve), with  $w = 1$ ; and (3) Algorithm 1 applied to (47), where  $\phi(p)$  in (46) is replaced by  $-\phi(p)$  and  $w = 1$ . The solution of each sub-VI (48) in the inner loop of Algorithm 1 is computed using Algorithm 2. 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 1 as one iteration of the inner Jacobi scheme (Algorithm 2), which corresponds to a physical transmission of the users.

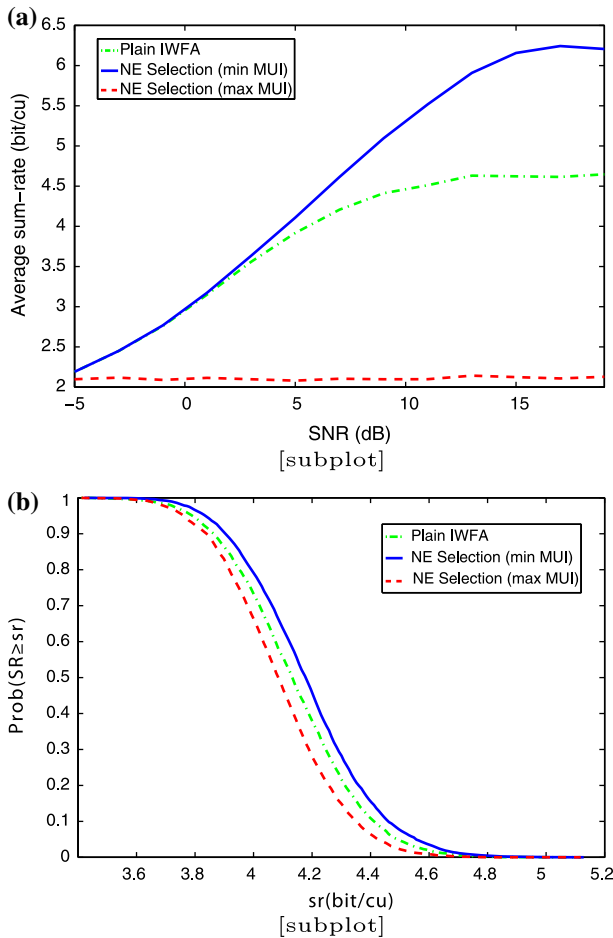
We examined the behaviour of the above algorithms under the following setup. We considered an ad-hoc network where there are  $Q = 25$  active users; the (cross-)channels among the links are simulated as FIR filter of order  $L = 10$ , where each tap is a zero mean complex Gaussian random variable with variance equal to  $1/L$ ;



**Fig. 1** 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 1 based on the outer function  $\phi(p)$  in (46) (*solid line curve*), and Algorithm 1 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)

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  $\sigma_n^{v2}$  is set to one for all  $n$  and  $v$ , and the transmit power budget  $P^v$  is chosen so that the Signal-to-Noise-Ratio (SNR) of each user  $\text{SNR}^v \triangleq 10 \log_{10}(P^v/\sigma_n^{v2}) = 5$  dB for all  $v$  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 1, we chose the center  $p^{(0)}$  of the regularization randomly in  $\bar{\mathcal{P}}$ ,  $\alpha = 3.5$ , 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; it is important to remark 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 [36]. 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 Fig. 1. 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 (46) in choosing the Nash equilibrium. When the NE of the game is



**Fig. 2** 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 1 based on the outer function  $\phi(p)$  in (46) (solid line curve), and Algorithm 1 based on the outer function  $-\phi(p)$  (dashed line curve)

unique, as expected, both the IWFA and our algorithm converge to the same sum-rate 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 best-response Jacobi scheme used in the inner loop converges quite fast, typically in three/four iterations.

Figure 1 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 Fig. 1 very general. This is confirmed also by Fig. 2, where we provide the average performance of the algorithms considered in Fig. 1. More specifically, in Fig. 2a we plotted the average sum-rate

versus the SNR  $\triangleq 10 \log_{10}(P)$ , with  $P^v = P$  and  $\sigma_n^{v^2} = 1$  for all  $v$  and  $k$ , achievable at the NE reached by the simultaneous IWFA and our Algorithm 1; for each value of the SNR, the curves are averaged over 5,000 random channel realizations, chosen so that the matrix  $M$  is positive semidefinite. In Fig. 2b we plotted the outage sum-rate (which is the probability that  $\sum_{v=1}^Q r^v(p) \geq sr$ ) versus  $sr$ , achieved by the aforementioned algorithms, for a SNR  $\triangleq P = 5$  dB; as in Fig. 2b, 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 Fig. 1. 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 Fig. 2a are meaningful in practice. Note that the discrepancy in the sum-rate gap between Algorithm 1 and the IWFA as observed in Figs. 1 and 2 is due to the oscillating behaviour of the IWFA experienced for some channel realizations. The following remarks are in order from Fig. 2. As expected, the performance of all the examined algorithms are almost the same in low SNR regime (roughly speaking when  $\text{SNR} \leq -5$  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.

## 8 Conclusions

In this paper we have studied algorithms for the solution of a hemivariational inequality whose feasible set is given by the intersection of a closed convex set with the solution set of a monotone variational inequality. A complete convergence theory is developed covering both centralized and distributed solution methods. We remark, in particular, that our distributed methods seem to be the first proposal of this type even in the simpler, and relatively well researched field of hierarchical optimization. The main assumptions needed for the centralized algorithm are (A)–(E) discussed in Sect. 2 and the boundedness condition (F) described in Sect. 4. Although these assumptions can possibly be slightly relaxed we believe that it would be more interesting to get a better understanding of the convergence properties of the Distributed Algorithm 2, which currently is based on Assumption P (see Sect. 5). In fact, Assumption P is a somewhat strong assumption and practical experiments show that the Distributed Algorithm 2 often works even when this assumption is violated. Certainly this analysis should be carried out also taking into account realistic settings derived from applications. We have used our results to solve a problem of power control in ad-hoc networks obtaining very good results. In this setting Assumption P becomes a very clear and natural requirement on the interference level that can be tolerated by the system. And indeed, it would seem that telecommunications can provide many settings where the techniques developed in this paper can usefully be used, especially in networking. But, as the discussion in the introduction clearly shows, our results apply more generally to

the Nash equilibrium selection problem, thus paving the way to applications in many different fields, see, for example [14].

## References

1. Addi, K., Brogliato, B., Goeleven, D.: A qualitative mathematical analysis of a class of linear variational inequalities via semi-complementarity problems: application in electronics. *Math. Program.* **126**, 31–67 (2011)
2. Alart, P., Lemaire, B.: Penalization in nonclassical convex programming via variational convergence. *Math. Program.* **51**, 307–331 (1991)
3. Bertsekas, D.: *Nonlinear Programming*. Athena Scientifica (1992).
4. Bertsekas, D., Nedic, A., Ozdaglar, A.E.: *Convex Analysis and Optimization*. Athena Scientifica (2003)
5. Bertsekas, D., Tsitsiklis, J.N.: *Parallel and Distributed Computation: Numerical Methods* Athena Scientific (1997); [Original published by Prentice Hall 1989].
6. Baharaoui, M.A., Lemaire, B.: Convergence of diagonally stationary sequences in convex optimization. *Set-Valued Anal.* **2**, 49–61 (1994)
7. Browder, F.E.: Existence and approximation of solutions of nonlinear variational inequalities. *Proc. Natl. Acad. Sci.* **56**, 1080–1086 (1966)
8. Cabot, A.: Proximal point algorithm controlled by a slowly vanishing term: applications to hierarchical minimization. *SIAM J. Optim.* **15**, 555–572 (2005)
9. Chan, D., Pang, J.S.: The generalized quasi-variational inequality. *Math. Oper. Res.* **7**, 211–224 (1982)
10. Cominetti, R.: Coupling the proximal point algorithm with approximation methods. *J. Optim. Theory Appl.* **95**, 581–600 (1997)
11. Cottle, R.W., Pang, J.S., Stone, R.E.: *The Linear Complementarity Problem*. Cambridge Academic, Cambridge (1992)
12. Demyanov, V.F., Di Pillo, G., Facchinei, F.: Exact penalization via Dini and Hadamard conditional derivatives. *Optim. Methods Softw.* **9**, 19–36 (1998)
13. Facchinei, F., Kanzow, C.: Penalty methods for the solution of generalized Nash equilibrium problems. *SIAM J. Optim.* **20**, 2228–2253 (2010)
14. Facchinei, F., Kanzow, C.: Generalized Nash equilibrium problems. *Ann. Oper. Res.* **175**, 177–211 (2010)
15. Facchinei, F., Pang, J.S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, Berlin (2003)
16. Facchinei, F., Pang, J.S.: Exact penalty function for generalized Nash games. In: Di Pillo, G., Roma, M. (eds.) *Large-scale Nonlinear Optimization. Nonconvex Optimization and Its Applications*, vol. 83, pp. 115–126 (2006).
17. Facchinei, F., Pang, J.S.: Nash equilibria: the variational approach. In: Palomar, D.P., Eldar, Y. (eds.) *Convex Optimization in Signal Processing and Communications*, pp. 443–493. Cambridge University Press, Cambridge (2010)
18. Iiduka, H.: Fixed point optimization algorithm and its application to power control in CDMA data networks. *Mathematical Programming* (to appear)
19. Iusem, A.N.: On some properties of paramonotone operators. *J. Convex Anal.* **5**, 269–278 (1998)
20. Jofré, A., Rockafellar, R.T., Wets, R.J.-B.: Variational inequalities and economic equilibrium. *Math. Oper. Res.* **32**, 32–50 (2007)
21. Ledhili, N., Moudafi, A.: Combining the proximal algorithm and Tikhonov regularization. *Optimization* **37**, 239–252 (1996)
22. Lemaire, B.: On the convergence of some iterative methods for convex minimization. In: *Recent Developments in Optimization. Lectures Notes in Economic and Mathematical Systems*, vol. 429, pp. 252–268. Springer, Berlin (1995).
23. Luo, T., Pang, J.S.: Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines. *EURASIP J. Appl. Signal Process.* (2006); Article ID 24012.
24. Luo, Z.Q., Pang, J.S., Ralph, D.: *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge (1996)
25. Luo, T., Zhang, S.: Spectrum management: complexity and duality. *J. Sel. Top. Signal Process.* **2**, 5772 (2008)

26. Lu, X., Xu, H.-K., Yin, X.: Hybrid methods for a class of monotone variational inequalities. *Nonlinear Anal.* **71**, 1032–1041 (2009)
27. Marino, G., Xu, H.-K.: Explicit hierarchical fixed point approach to variational inequalities. *J. Optim. Theory Appl.* **149**, 61–78 (2011)
28. Monteiro, R.D.C., Svaiter, B.F.: 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. Working paper, School of ISyE, Georgia Tech, USA, June 2010; under review at *SIAM Journal on Optimization*.
29. Motreanu, D., Panagiotopoulos, R.D.: *Minimax Theorems and Qualitative Properties of the Solutions of Hemivariational Inequalities*. Kluwer, Dordrecht (1999)
30. Moudafi, A.: Proximal methods for a class of bilevel monotone equilibrium problems. *J. Glob. Optim.* **47**, 287–292 (2010)
31. Naniewicz, Z., Panagiotopoulos, P.D.: *Mathematical Theory of Hemivariational Inequalities and Applications*. Marcel Dekker, New York (1995)
32. Panagiotopoulos, P.D.: *Hemivariational Inequalities: Applications in Mechanics and Engineering*. Springer, Berlin (1993)
33. Pang, J.S.: Asymmetric variational inequality problems over product sets: applications and iterative methods. *Math. Program.* **31**, 206–219 (1985)
34. Pang, J.S., Scutari, G., Facchinei, F., Wang, C.: Distributed power allocation with rate constraints in Gaussian frequency-selective interference channels. *IEEE Trans. Inf. Theory* **54**, 3471–3489 (2007)
35. Pang, J.S., Scutari, G., Palomar, D.P., Facchinei, F.: Design of cognitive radio systems under temperature-interference constraints: a variational inequality approach. *IEEE Trans. Signal Process.* **58**, 3251–3271 (2010)
36. Scutari, G., Palomar, D.P., Barbarossa, S.: Optimal linear precoding strategies for wideband noncooperative systems based on game theory—Part I & II: Nash equilibria & algorithms. *IEEE Trans. Signal Process.* **56**, 1230–1249 & 1250–1267 (2008).
37. Scutari, G., Palomar, D.P., Barbarossa, S.: Asynchronous iterative water-filling for Gaussian frequency-selective interference channels. *IEEE Trans. Inf. Theory* **54**, 2868–2878 (2008)
38. Scutari, G., Facchinei, F., Pang, J.S., Palomar, D.P.: *Real and Complex Monotone Communication Games*. *IEEE Trans. Inf. Theory* (2013) (Submitted)
39. Tseng, P.: Error bounds for regularized complementarity problems. In: Théra, M., Tichatschke, R. (eds.) *Ill-Posed Variational Problems and Regularization Techniques*. *Lecture Notes in Economics and Mathematical Systems*, vol. 477, pp. 247–274. Springer, Berlin (1999).
40. Tikhonov, A.N.: Solution of incorrectly formulated problems and regularization method. *Sov. Math. Doklady* **4**, 1035–1038 (1963)
41. Tikhonov, A.N., Arsenin, V.Y.: *Solutions of Ill-Posed Problems*. Wiley, New York (1977)
42. Yamada, I.: The hybrid steepest descent for the variational inequality problems over the intersection of fixed point sets of nonexpansive mappings. In: Butnariu, D., Censor, Y., Reich, S. (eds.) *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*. *Studies in Computational Mathematics*, vol. 8, pp. 473–504. Elsevier, Amsterdam (2001).
43. Yu, W., Ginis, G., Cioffi, J.M.: Distributed multiuser power control for digital subscriber lines. *IEEE J. Sel. Areas Commun.* **20**, 1105–1115 (2002)