ORIGINAL ARTICLE

On the solution of monotone nested variational inequalities

Lorenzo Lampariello[1] · Gianluca Priori[2] · Simone Sagratella[2]

Abstract

We study nested variational inequalities, which are variational inequalities whose feasible set is the solution set of another variational inequality. We present a projected averaging Tikhonov algorithm requiring the weakest conditions in the literature to guarantee the convergence to solutions of the nested variational inequality. Specifically, we only need monotonicity of the upper- and the lower-level variational inequalities. Also, we provide the first complexity analysis for nested variational inequalities considering optimality of both the upper- and lower-level.

1 Introduction

We focus on solving (upper-level) variational inequalities whose feasible set is given by the solution set of another (lower-level) variational inequality. These problems are commonly referred to as nested variational inequalities and they represent a flexible modelling tool when it comes to solving, for instance, problems arising in finance (see, e.g., the most recent Lampariello et al. (2021) for an application in the multi-portfolio

✉ Simone Sagratella
   sagratella@diag.uniroma1.it

   Lorenzo Lampariello
   lorenzo.lampariello@uniroma3.it

   Gianluca Priori
   priori@diag.uniroma1.it

[1] Department of Business Studies, Roma Tre University, Rome, Italy

[2] Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome, Rome, Italy

selection context) as well as several other fields (see, e.g., Facchinei et al. 2014; Scutari et al. 2012 for resource allocation problems in communications and networking).

As far as the literature on nested VIs is concerned, it is still in its infancy if compared to the bilevel instance of hierarchical optimization or, more generally, bilevel structures as in Dempe (2002), Lampariello and Sagratella (2017, 2020) and Lampariello et al. (2019). However, there are two main solution methods that are most adopted in the field-related literature: hybrid-like techniques (see, e.g., Lu et al. 2009; Marino and Xu 2011; Yamada 2001) and Tikhonov-type schemes (see, e.g. Lampariello et al. (2020) for the latest developments and Kalashnikov and Kalashinikova (1996) for some earlier developments, as well as Facchinei et al. (2014) and the references therein). It should be pointed out that hybrid-like procedures very often require particularly strong assumptions [(e.g. demanding co-coercivity of the lower-level map as in Lu et al. (2009), Marino and Xu (2011) and Yamada (2001)] in order to ensure convergence or to work properly at all. Hence, we rely on the Tikhonov paradigm drawing from the general schemes proposed in Facchinei et al. (2014) and Lampariello et al. (2020), however asking for less stringent assumptions and combining the Tikhonov approach with a new averaging procedure.

We widen the scope and expand the applicability of nested variational inequalities by showing for the first time in related literature that solutions can be provably computed in the more general framework of simply monotone upper- and lower-level variational inequalities. Specifically, in Facchinei et al. (2014) and Lampariello et al. (2020), where as far as we are aware the most advanced results are obtained, the upper-level map is required to be monotone plus. Relying on a combination of a Tikhonov approach with an averaging procedure, the algorithm we propose is shown to converge provably to a solution of a nested variational inequality where the upper-level map is required to be just monotone.

We also obtain complexity results for our method. Except for Lampariello et al. (2020), not only does this analysis represent the only other complexity study in the literature of nested variational inequalities, it is also the first one in the field dealing with upper-level optimality.

2 Problem definition and motivation

Let us consider the nested variational inequality $\text{VI}\big(G, \text{SOL}(F, Y)\big)$, where $G : \mathbb{R}^n \to \mathbb{R}^n$ is the upper-level map, and $\text{SOL}(F, Y)$ is the solution set of the lower-level $\text{VI}(F, Y)$. We recall that, given a subset $Y \subseteq \mathbb{R}^n$ and a mapping $F : \mathbb{R}^n \to \mathbb{R}^n$, the variational inequality $\text{VI}(F, Y)$ is the problem of computing a vector $x \in \mathbb{R}^n$ such that

$$x \in Y, \quad F(x)^T (y - x) \geq 0, \quad \forall y \in Y. \tag{1}$$

In other words, $\text{VI}\big(G, \text{SOL}(F, Y)\big)$ is the problem of finding $x \in \mathbb{R}^n$ that solves

$$x \in \text{SOL}(F, Y), \quad G(x)^T (y - x) \geq 0, \quad \forall y \in \text{SOL}(F, Y). \tag{2}$$

As it is clear from (2), the feasible set of $\text{VI}\big(G, \text{SOL}(F, Y)\big)$ is implicitly defined as the solution set of the lower-level $\text{VI}(F, Y)$. The nested variational inequality (2) we consider has a purely hierarchical structure in that the lower-level problem (1) is non parametric with respect to the upper-level variables, unlike the more general bilevel structures presented in Lampariello and Sagratella (2020). Under mild conditions, VIs equivalently reformulate NEPs so that in turn, by means of structure (2), we are able to model well-known instances of bilevel optimization and address multi-follower games.

We introduce the following blanket assumptions which are widely adopted in the literature of solution methods for variational inequalities:

(A1) the upper-level map G is Lipschitz continuous with constant L_G and monotone on Y;
(A2) the lower-level map F is Lipschitz continuous with constant L_F and monotone on Y;
(A3) the lower-level feasible set Y is nonempty, convex and compact.

Due to (A2) and (A3), $\text{SOL}(F, Y)$ is a nonempty, convex, compact and not necessarily single-valued set, see e.g. Facchinei and Pang (2003, Section 2.3). As a consequence, the feasible set of the nested variational inequality (2) is not necessarily a singleton. Moreover, thanks to (A1), the solution set of the nested variational inequality (2) can include multiple points.

Notice that assumption (A1) on the upper-level map G is much less demanding than the one required in Facchinei et al. (2014) and Lampariello et al. (2020). Specifically, here we assume G to be only monotone, while in Facchinei et al. (2014) and Lampariello et al. (2020) it must be monotone plus.

For the sake of completeness, we recall that a mapping $G : Y \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone *plus* on Y if both the following conditions hold:

1. G is monotone on Y, i.e. $(G(x) - G(y))^T (x - y) \geq 0 \quad \forall\, x, y \in Y$;
2. $(x - y)^T (G(x) - G(y)) = 0 \Rightarrow G(x) = G(y) \quad \forall\, x, y \in Y$.

Consequently, we are now able to dispose of the monotonicity *plus* assumption on operator G, which inevitably represents a more stringent condition when compared to plain monotonicity, thus hereby asking for G to be simply monotone. Indeed, whenever the upper-level map G is nonsymmetric, requiring G to be monotone *plus* is "slightly less" than assuming G to be strongly monotone (see, e.g. Bigi et al. 2021). In fact, the main objective of this work is to define, for the first time in the field-related literature, an algorithm (Algorithm 1) which is able to compute solutions of the monotone nested variational inequality (2), under the weaker assumptions (A1)-(A3), see the forthcoming Theorem 1.

In addition, we study complexity properties of Algorithm 1 in detail, see Theorems 2 and 3. We highlight that all steps in Algorithm 1 can be readily implemented and no nontrivial computations are required, see e.g. the numerical illustration in Sect. 5.

Summarizing,

– we show that the algorithm we propose is globally (subsequential) convergent to solutions of monotone nested variational inequalities under the weakest conditions in the literature,

– we provide the first complexity analysis for nested variational inequalities considering optimality of both the upper- and lower-level [instead in Lampariello et al. (2020) just lower-level optimality is contemplated].

3 A projected averaging Tikhonov algorithm

For the sake of notation, let us introduce the following operator:

$$\Phi_\tau(x) \triangleq F(x) + \frac{1}{\tau}G(x),$$

which is the classical operator used to define subproblems in Tikhonov-like methods. For any $\tau \in \mathbb{R}_{++}$, by assumptions (A1) and (A2), Φ_τ is monotone and Lipschitz continuous with constant $L_\Phi \triangleq L_F + L_G$ on Y. Moreover, the following finite quantities are useful in the forthcoming analysis:

$$H \triangleq \max_{y \in Y} \|G(y)\|_2, \quad R \triangleq \max_{y \in Y} \|F(y)\|_2, \quad D \triangleq \max_{v, y \in Y} \|v - y\|_2.$$

For the sake of clarity, let us recall that, by definition, the Euclidean projection $P_Y(x)$ of a vector $x \in \mathbb{R}^n$ onto a closed convex subset $Y \subseteq \mathbb{R}^n$ is the unique solution of the strongly convex (in y) problem

$$\underset{y}{\text{minimize}} \quad \frac{1}{2}(y - x)^T(y - x)$$
$$\text{s.t.} \quad y \in Y.$$

The solution of the latter problem is a unique vector $\bar{y} \in Y$ that is closest to x in the Euclidean norm [see, e.g. Facchinei and Pang (2003, Th.1.5.5) for an exhaustive overview on the Euclidean projector and its properties].

In our analysis we rely on approximate solutions of VIs. Specifically, we say that $x \in K$ approximately solves VI(Ψ, K) (with Ψ continuous and K convex and compact) if

$$\Psi(x)^\top(y - x) \geq -\varepsilon, \quad \forall y \in K, \tag{3}$$

where $\varepsilon \geq 0$. Relation (3), for example, when VI(Ψ, K) defines the first-order optimality conditions of a convex problem, guarantees that the problem is solved up to accuracy ε. Relation (3), in view of assumption (A3), is equivalent to

$$\min_{y \in K} \Psi(x)^\top(y - x) \geq -\varepsilon. \tag{4}$$

We remark that $\Psi(x)^\top(y - x)$ is linear in y; moreover, if K is polyhedral (as, e.g., in the multi-portfolio selection context) computing $\min_{y \in K} \Psi(x)^\top(y - x)$ amounts to

solving a linear optimization problem. In any event, we assume this computation to be easy to do in practice.

With the following result we relate approximate solutions of the VI subproblem

$$\Phi_\tau(x)^\top(y - x) \geq -\varepsilon_{\text{sub}}, \quad \forall y \in Y, \tag{5}$$

where $\varepsilon_{\text{sub}} \geq 0$, with approximate solutions of problem (2).

Proposition 1 *Assume conditions (A1)-(A3) to hold, and let $x \in Y$ be a solution of the VI subproblem (5) with $\tau > 0$ and $\varepsilon_{sub} \geq 0$. It holds that*

$$G(x)^\top(y - x) \geq -\varepsilon_{up}, \quad \forall y \in SOL(F, Y), \tag{6}$$

with $\varepsilon_{up} \geq \varepsilon_{sub}\tau$, and

$$F(x)^\top(y - x) \geq -\varepsilon_{low}, \quad \forall y \in Y, \tag{7}$$

with $\varepsilon_{low} \geq \varepsilon_{sub} + \frac{1}{\tau}HD$.

Proof We have for all $y \in SOL(F, Y)$:

$$\begin{aligned} -\varepsilon_{\text{sub}}\tau &\leq [\tau F(x) + G(x)]^\top(y - x) \\ &\leq [\tau F(y) + G(x)]^\top(y - x) \\ &\leq G(x)^\top(y - x), \end{aligned}$$

where the first inequality is due to (5), the second one comes from the monotonicity of F, and the last one is true because $x \in Y$ and then $F(y)^\top(x - y) \geq 0$. That is (6) is true.

Moreover, we have for all $y \in Y$:

$$\begin{aligned} F(x)^\top(y - x) &= \Phi_\tau(x)^\top(y - x) - \frac{1}{\tau}G(x)^\top(y - x) \\ &\geq -\varepsilon_{\text{sub}} - \frac{1}{\tau}HD, \end{aligned}$$

where the inequality is due to (5). Therefore we get (7). $\qquad\square$

Proposition 1 suggests a way to solve, with a good degree of accuracy, the hierarchical problem (2). That is solving the VI subproblem (5) with a big value for τ and an ε_{sub} sufficiently small in order to make $\varepsilon_{\text{sub}}\tau$ small enough. Following this path,

we propose a Projected Averaging Tikhonov Algorithm (PATA), see Algorithm 1, to compute solutions of problem (2).

Algorithm 1: Projected Averaging Tikhonov Algorithm (PATA)

Data: $\{\bar{\varepsilon}^i\} \subseteq \mathbb{R}_+$, $\{\bar{\tau}^i\} \subseteq \mathbb{R}_{++}$, $\gamma^k \in (0, 1]$, $w^0 = z^0 = y^0 \in Y$, $i, l \leftarrow 0$;

for $k = 0, 1, \ldots$ **do**

(S.1) $\quad \varepsilon^k = \bar{\varepsilon}^i, \tau^k = \bar{\tau}^i$;

(S.2) $\quad y^{k+1} = P_Y(y^k - \gamma^{k-l}\Phi_{\tau^k}(y^k))$;

(S.3) $\quad z^{k+1} = \dfrac{\sum_{j=l}^{k+1} \gamma^{j-l} y^j}{\sum_{j=l}^{k+1} \gamma^{j-l}}$;

(S.4) $\quad$ **if** $\min_{y \in Y} \Phi_{\tau^k}(z^{k+1})^\top (y - z^{k+1}) \geq -\varepsilon^k$ **then**

$\quad\quad \mid \quad w^{i+1} = z^{k+1}, i = i + 1, l = k + 1$;

$\quad$ **end**

end

Some comments about PATA are in order. Index i denotes the outer iterations that occur when the condition in step (S.4) is verified, and they correspond to solutions w^{i+1} of the VI subproblems (5) with $\varepsilon_{\text{sub}} = \bar{\varepsilon}^i$ and $\tau = \bar{\tau}^i$. The sequence $\{y^k\}$ is obtained by making classical projection steps with stepsizes γ^k, see step (S.2). The sequence $\{z^k\}$ consists of the inner iterations needed to compute a solution of the VI subproblem (5), and it is obtained by performing a weighted average on the points y^j, see step (S.3). Index l is included in order to let the sequence of the stepsizes $\{\gamma^k\}$ restart at every outer iteration and to consider only the points y^j belonging to the current subproblem to compute z^{k+1}.

We remark that the condition in step (S.4) only requires the solution of an optimization problem with a linear objective function over the convex set Y (see the discussion about inexact solutions of VIs below relation (4)). In Sect. 5 we give a practical implementation of PATA.

In the following section we show that Proposition 1 can be used to prove that PATA effectively computes solutions of problem (2).

4 Main convergence properties

First of all we deal with convergence properties of PATA.

Theorem 1 *Assume conditions (A1)-(A3) to hold, and let conditions*

$$\sum_{k=0}^{\infty} \gamma^k = \infty, \quad \frac{\sum_{k=0}^{\infty}(\gamma^k)^2}{\sum_{k=0}^{\infty} \gamma^k} = 0, \quad \frac{1}{\bar{\tau}^i} \downarrow 0, \quad \sum_{i=0}^{\infty} \frac{1}{\bar{\tau}^i} = \infty, \quad \bar{\varepsilon}^i = \frac{c}{(\bar{\tau}^i)^\beta}, \qquad (8)$$

hold with $\beta > 1$ and $c > 0$. Every limit point of the sequence $\{w^i\}$ generated by PATA is a solution of problem (2).

Proof First of all we show that $i \to \infty$. Assume by contradiction that this is not true, therefore there exists an index $\bar{k}$ such that the condition in step (S.4) is violated for

every $k \geq \bar{k}$, and either $\bar{k} = 0$ or the condition in step (S.4) is satisfied at the iteration $\bar{k} - 1$. We denote $\bar{\tau} = \tau^{\bar{k}}$, and observe that $\tau^k = \bar{\tau}$ for every $k \geq \bar{k}$.

For every $j \in [\bar{k}, k]$, and for any $v \in Y$, we have

$$
\begin{aligned}
\|y^{j+1} - v\|_2^2 &= \|P_Y(y^j - \gamma^{j-\bar{k}} \Phi_{\bar{\tau}}(y^j)) - v\|_2^2 \\
&\leq \|y^j - \gamma^{j-\bar{k}} \Phi_{\bar{\tau}}(y^j) - v\|_2^2 \\
&= \|y^j - v\|_2^2 + (\gamma^{j-\bar{k}})^2 \|\Phi_{\bar{\tau}}(y^j)\|_2^2 - 2\gamma^{j-\bar{k}} \Phi_{\bar{\tau}}(y^j)^\top (y^j - v),
\end{aligned}
$$

and, in turn,

$$
\Phi_{\bar{\tau}}(y^j)^\top (v - y^j) \geq \frac{\|y^{j+1} - v\|_2^2 - \|y^j - v\|_2^2}{2\gamma^{j-\bar{k}}} - \frac{\gamma^{j-\bar{k}}}{2} \|\Phi_{\bar{\tau}}(y^j)\|_2^2.
$$

Summing, we get

$$
\begin{aligned}
\frac{\sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}} \Phi_{\bar{\tau}}(y^j)^\top (v - y^j)}{\sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}}} &\geq \frac{\sum_{j=\bar{k}}^{k} \left(\|y^{j+1} - v\|_2^2 - \|y^j - v\|_2^2 - (\gamma^{j-\bar{k}})^2 \|\Phi_{\bar{\tau}}(y^j)\|_2^2 \right)}{2 \sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}}} \\
&= \frac{\left(\|y^{k+1} - v\|_2^2 - \|y^{\bar{k}} - v\|_2^2 - \sum_{j=\bar{k}}^{k} (\gamma^{j-\bar{k}})^2 \|\Phi_{\bar{\tau}}(y^j)\|_2^2 \right)}{2 \sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}}} \\
&\geq -\frac{\left(\|y^{\bar{k}} - v\|_2^2 + \sum_{j=\bar{k}}^{k} (\gamma^{j-\bar{k}})^2 \|\Phi_{\bar{\tau}}(y^j)\|_2^2 \right)}{2 \sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}}},
\end{aligned} \tag{9}
$$

which implies

$$
\begin{aligned}
\Phi_{\bar{\tau}}(v)^\top (v - z^k) &= \frac{1}{\sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}}} \sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}} \Phi_{\bar{\tau}}(v)^\top (v - y^j) \\
&\geq -\frac{\left(\|y^{\bar{k}} - v\|_2^2 + \sum_{j=\bar{k}}^{k} (\gamma^{j-\bar{k}})^2 \|\Phi_{\bar{\tau}}(y^j)\|_2^2 \right)}{2 \sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}}} \\
&\quad + \frac{1}{\sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}}} \sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}} (\Phi_{\bar{\tau}}(v) - \Phi_{\bar{\tau}}(y^j))^\top (v - y^j) \\
&\geq -\frac{\left(\|y^{\bar{k}} - v\|_2^2 + \sum_{j=\bar{k}}^{k} (\gamma^{j-\bar{k}})^2 \|\Phi_{\bar{\tau}}(y^j)\|_2^2 \right)}{2 \sum_{j=\bar{k}}^{k} \gamma^{j-\bar{k}}},
\end{aligned} \tag{10}
$$

where the last inequality is due to the monotonicity of $\Phi_{\bar{\tau}}$. Denoting by $z \in Y$ any limit point of the sequence $\{z^k\}$, taking the limit in the latter relation and subsequencing, the following inequality holds:

$$
\Phi_{\bar{\tau}}(v)^\top (v - z) \geq -\frac{\left(\|y^{\bar{k}} - v\|_2^2 + \sum_{j=\bar{k}}^{\infty} (\gamma^{j-\bar{k}})^2 \|\Phi_{\bar{\tau}}(y^j)\|_2^2 \right)}{2 \sum_{j=\bar{k}}^{\infty} \gamma^{j-\bar{k}}} = 0,
$$

because $\sum_{j=\bar{k}}^{\infty} \gamma^{j-\bar{k}} = +\infty$ and $\left(\sum_{j=\bar{k}}^{\infty} (\gamma^{j-\bar{k}})^2 \right) / \left(\sum_{j=\bar{k}}^{\infty} \gamma^{j-\bar{k}} \right) = 0$, and then z is a solution of the dual problem

$$
\Phi_{\bar{\tau}}(v)^\top (v - z) \geq 0, \quad \forall v \in Y.
$$

Hence, the sequence $\{z^k\}$ converges to a solution of VI$(Y, \Phi_{\bar{\tau}})$, see e.g. Facchinei and Pang (2003, Theorem 2.3.5) in contradiction to $\min_{y \in Y} \Phi_{\bar{\tau}}(z^{k+1})^\top (y - z^{k+1}) < -\varepsilon^k = -\varepsilon^{\bar{k}}$ for every $k \geq \bar{k}$.

Therefore the algorithm produces an infinite sequence $\{w^i\}$ such that $w^{i+1} \in Y$ and

$$\Phi_{\bar{\tau}^i}(w^{i+1})^\top (y - w^{i+1}) \geq -\frac{c}{(\bar{\tau}^i)^\beta}, \quad \forall y \in Y,$$

that is (5) holds at w^{i+1} with $\varepsilon_{\text{sub}} = \frac{c}{(\bar{\tau}^i)^\beta}$. By Proposition 1, specifically from (6) and (7), we obtain

$$G(w^{i+1})^\top (y - w^{i+1}) \geq -\frac{c}{(\bar{\tau}^i)^{\beta-1}}, \quad \forall y \in \text{SOL}(F, Y), \tag{11}$$

and

$$F(w^{i+1})^\top (y - w^{i+1}) \geq -\frac{c}{(\bar{\tau}^i)^\beta} - \frac{1}{\bar{\tau}^i} H D, \quad \forall y \in Y. \tag{12}$$

Taking the limit $i \to \infty$, and recalling that G and F are continuous and $\beta > 1$, we get the desired convergence property. □

Conditions (8) for the sequence of stepsizes $\{\gamma^k\}$ are satisfied, e.g., if we choose

$$\gamma^k = \min \left\{ 1, \frac{a}{k^\alpha} \right\},$$

with $a > 0$ and $\alpha \in (0, 1]$, see Proposition 4 in the Appendix. Another possible choice of step-size rule satisfying conditions (8), as shown in Facchinei et al. (2015), is

$$\gamma^k = \gamma^{k-1}(1 - \theta \gamma^{k-1}), \quad k = 1, \ldots$$

where $\theta \in (0, 1)$ is a given constant, provided that $\gamma^0 \in (0, 1]$. Both the above step-size rules satisfy conditions (8), needed for Theorem 1 to be valid.

We remark that, even if we require assumptions that are less stringent with respect to related literature, we still obtain the same type of convergence as in related literature, namely *subsequential* convergence to a solution of problem (2). Note that, thanks to assumption (A3), at least a limit point of sequence $\{w^i\}$ generated by PATA exists. As it is common practice when using an iterative algorithm like PATA, referring to (11) and (12), w^{i+1} can be considered an approximate solution of problem (2) as soon as $\frac{c}{(\bar{\tau}^i)^{\beta-1}}$ and $\left(\frac{c}{(\bar{\tau}^i)^\beta} + \frac{1}{\bar{\tau}^i} H D \right)$ are small enough. Clearly, if the upper-level map G is strongly monotone on Y, the whole sequence $\{w^i\}$ converges to the unique solution of problem (2).

We consider the so-called natural residual map for VI(Ψ, K) (with Ψ continuous and K convex and compact)

$$U(x) \triangleq \| P_K(x - \Psi(x)) - x \|. \tag{13}$$

As recalled in Lampariello et al. (2020), the function U is continuous and nonnegative. Moreover, $U(x) = 0$ if and only if $x \in \mathrm{SOL}(\Psi, K)$. Specifically, classes of problems exist for which the value $U(x)$ also gives an actual upper-bound to the distance between x and $\mathrm{SOL}(\Psi, K)$, see Lampariello et al. (2020) and the references therein. Therefore, the following condition

$$U(x) \leq \widehat{\varepsilon}, \tag{14}$$

with $\widehat{\varepsilon} \geq 0$, is alternative to (3). However, in view of the compactness of K, relations (3) and (14) turn out to be related to each other: we show in Appendix (Proposition 3) that if x satisfies (3), then (14) holds with $\widehat{\varepsilon} \geq \sqrt{\varepsilon}$. Vice versa, condition (14) implies (3) with $\varepsilon \geq (\Omega + \Xi)\widehat{\varepsilon}$, where $\Omega \triangleq \max_{v,y \in K} \|v - y\|_2$ and $\Xi \triangleq \max_{y \in K} \|\Psi(y)\|_2$.

We remark that, whenever Y is not compact, e.g. if $Y = \mathbb{R}^n$, the condition in step (S.4) of Algorithm 1 involving the classical gap function for VIs (see, e.g., Facchinei and Pang 2003, page 88) cannot be relied upon and could be replaced by the natural map-related relation $\|P_Y(z^{k+1} - \Phi_{\tau^k}(z^{k+1})) - z^{k+1}\| \leq \varepsilon^k$, at the price of a slightly modified convergence analysis.

In order to deal with the convergence rate analysis of our method, we consider the natural residual map for the lower-level VI(F, Y)

$$V(x) \triangleq \|P_Y(x - F(x)) - x\|. \tag{15}$$

Clearly, the following condition

$$V(x) \leq \widehat{\varepsilon}_{\mathrm{low}}, \tag{16}$$

with $\widehat{\varepsilon}_{\mathrm{low}} \geq 0$, is alternative to (7) to take care of the feasibility of problem (2).

In this context, we underline that the convergence rate we establish is intended to give an upper bound to the number of iterations needed to drive both the upper-level error $\varepsilon_{\mathrm{up}}$, given in (6), and the lower-level error $\widehat{\varepsilon}_{\mathrm{low}}$, given in (16), under some prescribed tolerances δ_{up} and $\widehat{\delta}_{\mathrm{low}}$, respectively.

Theorem 2 *Assume conditions (A1)-(A3) to hold and, without loss of generality, assume $L_\Phi < 1$. Consider PATA. Given some precisions $\delta_{up}, \widehat{\delta}_{low} \in (0, 1)$, set $\gamma^k = \min\left\{1, \frac{1}{2k^{\frac{1}{2}}}\right\}$, $\bar{\tau}^i = \max\{1, i\}$, and $\bar{\varepsilon}^i = \frac{1}{(\bar{\tau}^i)^2}$. Let us define the quantity*

$$I_{\max} \triangleq \left\lceil \max\left\{\frac{1}{\delta_{up}}, \frac{H+1}{\widehat{\delta}_{low}}\right\} \right\rceil.$$

Then, the upper-level approximate problem (6) is solved for $x = z^{k+1}$ with $\varepsilon_{up} \leq \delta_{up}$ and the lower-level approximate problem (16) is solved for $x = z^{k+1}$ with $\widehat{\varepsilon}_{low} \leq \widehat{\delta}_{low}$ and the condition in step (S.4) is satisfied in at most

$$\sigma \triangleq I_{\max}\left\lceil \max\left\{I_{\max}^8 \frac{(D+R)^4}{(1-L_\Phi)^2}C_1, I_{\max}^{\frac{8}{1-2\eta}}\frac{(D+R)^{\frac{4}{1-2\eta}}}{(1-L_\Phi)^{\frac{2}{1-2\eta}}}C_{2,\eta}\right\}\right\rceil,$$

iterations k, where $\eta > 0$ is a small number, and

$$C_1 \triangleq \left(D^2 + \frac{5}{4}(R+H)^2 \right)^2, \quad C_{2,\eta} \triangleq \left(\frac{(R+H)^2}{(4\eta)} \right)^{\frac{2}{1-2\eta}}. \tag{17}$$

Proof First of all we show that if $i \geq I_{\max}$, we reach the desired result. Specifically, about the upper-level problem (6), we obtain

$$\varepsilon_{\text{up}} = \bar{\varepsilon}^i \bar{\tau}^i = \frac{1}{i} \leq \delta_{\text{up}},$$

where the first equality is due to Proposition 1, and the last inequality follows from $i \geq I_{\max} \geq (\delta_{\text{up}})^{-1}$.

About the lower-level problem (16), preliminarily we observe that

$$\left\| P_Y(w^{i+1} - \Phi_{\bar{\tau}_i}(w^{i+1})) - w^{i+1} \right\|_2 \leq \sqrt{\bar{\varepsilon}^i}, \tag{18}$$

because w^{i+1} satisfies the condition in step (S.4) with $\bar{\varepsilon}^i$, see Proposition 3. Moreover, we get

$$\widehat{\varepsilon}_{\text{low}} \leq \frac{1}{\bar{\tau}^i} H + \sqrt{\bar{\varepsilon}^i} = \frac{H+1}{i} \leq \widehat{\delta}_{\text{low}},$$

where the first inequality is due to (26) and (18), and the last inequality follows from $i \geq I_{\max} \geq (\widehat{\delta}_{\text{low}})^{-1}(H+1)$.

Now we consider the number of inner iterations needed to satisfy the condition in step (S.4) with the smallest error $\bar{\varepsilon}^{I_{\max}} = I_{\max}^{-2}$ and for a $\tau > 0$. Without loss of generality, in the following developments we will assume $\bar{k} = 0$, meaning that we are simply computing the number of inner iterations. By (10), the dual subproblem

$$\Phi_\tau(y)^\top(y - z^k) \geq -\varepsilon_{\text{sub\_dual}}^k, \quad \forall y \in Y, \tag{19}$$

is solved for $\varepsilon_{\text{sub\_dual}}^k = \frac{D^2 + \sum_{j=0}^k (\gamma^j)^2 (R+H)^2}{2 \sum_{j=0}^k \gamma^j}$. From Lemma 1, we obtain

$$\sum_{j=0}^k \gamma^j \geq k^{\frac{1}{2}}, \quad \sum_{j=0}^k (\gamma^j)^2 \leq \frac{5}{4} + \frac{1}{4}\ln(k) \leq \frac{5}{4} + \frac{1}{4\eta}k^\eta, \quad \eta > 0,$$

because

$$\ln(k) = \int_1^k t^{-1} dt \leq \int_1^k t^{-1+\eta} dt \leq \frac{1}{\eta}k^\eta.$$

Therefore

$$
\begin{aligned}
\varepsilon_{\text{sub\_dual}}^k &\le \frac{D^2 + \left(\frac{5}{4} + \frac{1}{4\eta} k^\eta\right)(R+H)^2}{2k^{\frac{1}{2}}} \\
&= \frac{D^2 + \frac{5}{4}(R+H)^2}{2k^{\frac{1}{2}}} + \frac{\frac{1}{4\eta} k^\eta (R+H)^2}{2k^{\frac{1}{2}}} \\
&= \frac{C_1^{\frac{1}{2}}}{2k^{\frac{1}{2}}} + \frac{C_{2,\eta}^{\frac{1-2\eta}{2}}}{2k^{\frac{1-2\eta}{2}}} \\
&\le \max\left\{\frac{C_1^{\frac{1}{2}}}{k^{\frac{1}{2}}}, \frac{C_{2,\eta}^{\frac{1-2\eta}{2}}}{k^{\frac{1-2\eta}{2}}}\right\}.
\end{aligned}
\tag{20}
$$

Now we show that

$$
\left\| P_Y(z^k - \Phi_\tau(z^k)) - z^k \right\|_2 \le \frac{\sqrt{\varepsilon_{\text{sub\_dual}}^k}}{\sqrt{1 - L_\Phi}}.
\tag{21}
$$

In fact, taking $y = v^k = P_Y(z^k - \Phi_\tau(z^k)) \in Y$ in (19), we have

$$
\begin{aligned}
\varepsilon_{\text{sub\_dual}}^k &\ge \Phi_\tau(v^k)^\top (z^k - v^k) \\
&= [z^k - v^k - [z^k - v^k - \Phi_\tau(v^k)]]^\top (z^k - v^k) \\
&= \|z^k - v^k\|_2^2 - [z^k - \Phi_\tau(v^k) - v^k]^\top (z^k - v^k) \\
&= \|z^k - v^k\|_2^2 - [z^k - \Phi_\tau(z^k) - v^k]^\top (z^k - v^k) \\
&\quad - [\Phi_\tau(z^k) - \Phi_\tau(v^k)]^\top (z^k - v^k) \\
&\ge (1 - L_\Phi)\|z^k - v^k\|_2^2,
\end{aligned}
$$

where the last inequality follows from the Lipschitz continuity of Φ_τ and the characteristic property of the projection.

From Proposition 3 and inequality (21), we obtain the following error for the subproblem

$$
\Phi_\tau(z^k)^\top (y - z^k) \ge -\frac{D+R}{\sqrt{1 - L_\Phi}}\sqrt{\varepsilon_{\text{sub\_dual}}^k}, \quad \forall\, y \in Y,
\tag{22}
$$

and then, by (20), the desired accuracy for the subproblem is obtained when

$$
I_{\max}^{-2} = \bar{\varepsilon}^{I_{\max}} \ge \frac{D+R}{\sqrt{1 - L_\Phi}} \max\left\{\frac{C_1^{\frac{1}{4}}}{k^{\frac{1}{4}}}, \frac{C_{2,\eta}^{\frac{1-2\eta}{4}}}{k^{\frac{1-2\eta}{4}}}\right\},
$$

that is

$$
k \ge \max\left\{ I_{\max}^8 \frac{(D+R)^4}{(1 - L_\Phi)^2} C_1, \, I_{\max}^{\frac{8}{1-2\eta}} \frac{(D+R)^{\frac{4}{1-2\eta}}}{(1 - L_\Phi)^{\frac{2}{1-2\eta}}} C_{2,\eta} \right\}.
$$

The thesis follows by multiplying the number of outer iterations ($i \geq I_{max}$) for the number of inner ones. $\square$

In order to provide other complexity results for our method, we consider the following proposition, which is the dual counterpart of Proposition 1, and provides a theoretical basis for Theorem 3.

Proposition 2 *Assume conditions (A1)-(A3) to hold, and let $x \in Y$ be an approximate solution of the dual VI subproblem:*

$$\Phi_\tau(y)^\top (y - x) \geq -\varepsilon_{sub\_dual}, \quad \forall \, y \in Y, \tag{23}$$

with $\tau > 0$ and $\varepsilon_{sub\_dual} \geq 0$. It holds that x turns out to be an approximate solution for the dual formulation of problem (2), that is

$$G(y)^\top (y - x) \geq -\varepsilon_{up\_dual}, \quad \forall y \in SOL(F, Y), \tag{24}$$

with $\varepsilon_{up\_dual} \geq \varepsilon_{sub\_dual}\tau$, and

$$F(y)^\top (y - x) \geq -\varepsilon_{low\_dual}, \quad \forall y \in Y, \tag{25}$$

with $\varepsilon_{low\_dual} \geq \varepsilon_{sub\_dual} + \frac{1}{\tau} H D$.

Proof We have for all $y \in SOL(F, Y)$:

$$-\varepsilon_{sub\_dual}\tau \leq [\tau F(y) + G(y)]^\top (y - x)$$
$$\leq G(y)^\top (y - x),$$

where the first inequality is due to (23) and the last one is true because $x \in Y$ and then $F(y)^\top (x - y) \geq 0$. That is (24) is true.

Moreover, we have for all $y \in Y$:

$$F(y)^\top (y - x) = \Phi_\tau(y)^\top (y - x) - \frac{1}{\tau} G(y)^\top (y - x)$$
$$\geq -\varepsilon_{sub\_dual} - \frac{1}{\tau} H D,$$

where the inequality is due to (23). Therefore we get (25). $\square$

The following theorem considers a simplified version of PATA. Specifically, the parameter τ is right away initialized to a value sufficiently large to get the prescribed optimality accuracy. Moreover, approximate optimality for problem (2) is considered only in its dual version. That said, the complexity bound obtained is better than the one given by Theorem 2.

Theorem 3 *Assume conditions (A1)-(A3) to hold. Consider PATA. Given some precision $\delta \in (0, 1)$, set $\gamma^k = \min \left\{ 1, \frac{1}{2k^{\frac{1}{2}}} \right\}$, $\bar{\tau}^0 = \bar{I}_{\max}$, and $\bar{\varepsilon}^0 = 0$ where*

$$\bar{I}_{\max} \triangleq \left\lceil \frac{H+1}{\delta} \right\rceil .$$

Then, the upper-level approximate dual problem (24) is solved for $x = z^{k+1}$ with $\varepsilon_{up\_dual} \leq \delta$ and the lower-level approximate dual problem (25) is solved for $x = z^{k+1}$ with $\varepsilon_{low\_dual} \leq \delta$ in at most

$$\bar{\sigma} \triangleq \left\lceil \max \left\{ \bar{I}_{\max}^4 C_1, \bar{I}_{\max}^{\frac{4}{1-2\eta}} C_{2,\eta} \right\} \right\rceil ,$$

iterations k, where $\eta > 0$ is a small number, and C_1 and $C_{2,\eta}$ are given in (17).

Proof First of all we denote with $\varepsilon_{sub\_dual}^k$ the error with which the current iteration solves the dual subproblem (23). Notice that as soon as $\varepsilon_{sub\_dual}^k \leq \bar{I}_{\max}^{-2}$, the desired accuracy for both the upper- and the lower-level dual problems is reached. In fact, as done in the proof of Theorem 2, and considering Proposition 2, we have:

$$\varepsilon_{up\_dual} = \varepsilon_{sub\_dual}^k \bar{\tau}^0 = \frac{1}{\bar{I}_{\max}} \leq \delta$$

where the first equality is due to Proposition 1, and the last inequality follows from $i \geq \bar{I}_{\max} \geq \delta^{-1}$, and

$$\varepsilon_{low\_dual} \leq \frac{H}{\bar{\tau}^0} + \sqrt{\varepsilon_{sub\_dual}^k} = \frac{H+1}{\bar{I}_{\max}} \leq \delta .$$

where the first inequality is due to (26) and (18), and the last inequality follows from $i \geq \bar{I}_{\max} \geq \delta^{-1}(H+1)$.

By (20) we have

$$\varepsilon_{sub\_dual}^k \leq \max \left\{ \frac{C_1^{\frac{1}{2}}}{k^{\frac{1}{2}}}, \frac{C_{2,\eta}^{\frac{1-2\eta}{2}}}{k^{\frac{1-2\eta}{2}}} \right\} .$$

Therefore, $\varepsilon_{sub\_dual}^k \leq \bar{I}_{\max}^{-2}$ is implied by

$$\bar{I}_{\max}^{-2} \geq \max \left\{ \frac{C_1^{\frac{1}{2}}}{k^{\frac{1}{2}}}, \frac{C_{2,\eta}^{\frac{1-2\eta}{2}}}{k^{\frac{1-2\eta}{2}}} \right\} ,$$

that is

$$k \geq \max \left\{ I_{\max}^4 C_1, \, I_{\max}^{\frac{4}{1-2\eta}} C_{2,\eta} \right\},$$

and the thesis follows. $\qquad\square$

5 Numerical experiments

We now tackle a practical example which is representative of the fact that, under assumptions (A1)-(A3), PATA produces the sequence of points $\{z^k\}$ that is (subsequentially) convergent to a solution of the hierarchical problem, while the sequence $\{y^k\}$ never approaches the solution set. Notice that $\{y^k\}$ coincides with the sequence produced by the Tikhonov methods proposed in Lampariello et al. (2020) when no proximal term is considered.

Let us examine the selection problem (2), where:

$$G(y) = \begin{pmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \; F(y) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \; Y = \mathbb{B}(0,1),$$

where $\mathbb{B}(0,1)$ denotes the unit ball. The unique feasible point and, thus, the unique solution of the problem is $z^* = (0,0)^T$. The assumptions (A1)–(A3) are satisfied, but notice that G does not satisfy convergence conditions of the Tikhonov-like methods proposed in Facchinei et al. (2014) and Lampariello et al. (2020) because it is not monotone plus.

The generic kth iteration of PATA, in this case, should read as reported below:

$$y^{k+1} = P_Y(y^k - \gamma^k [F(y^k) + \frac{1}{\tau^k} G(y^k)]),$$

where we take, for example, but without loss of generality, $\tau^k = \tau \geq 1$ and $\gamma^k = \gamma > 0$. We remark that the unique exact solution of the VI subproblem (5) is the origin, and then every inexact solution, with a reasonably small error, cannot be far from it. For every k it holds that

$$y^{k+1} = P_Y \left(\begin{pmatrix} 1 & \gamma(\frac{1}{2\tau} - 1) \\ \gamma(1 - \frac{1}{2\tau}) & 1 \end{pmatrix} \begin{pmatrix} y_1^k \\ y_2^k \end{pmatrix} \right),$$

hence $\|y^{k+1}\|_2 = \min \left\{ 1, \sqrt{1 + \gamma^2 (\frac{1}{2\tau} - 1)^2} \, \|y^k\|_2 \right\}$. Therefore we consider $\|y^0\|_2 = 1$ and get $\|y^k\|_2 = 1$ for every k, because $\sqrt{1 + \gamma^2 (\frac{1}{2\tau} - 1)^2} > 1$. Therefore, neither does the sequence $\{y^k\}$ produced by PATA lead to the unique solution z^* of problem (2), nor does it approach the inexact solution set of the VI subproblem.

We now consider the sequence $\{z^k\}$ produced by PATA. In order to show that this sequence leads us to the solution of the hierarchical problem, we analyze a numerical implementation of the algorithm. Some further considerations are in order before showing the actual implemented scheme.

– A general rule for the update of the variable z^k is given by the following relation:

$$z^{k+1} = \frac{z^k \gamma^{\text{sum},l,k} + \gamma^{k+1-l} y^{k+1}}{\gamma^{\text{sum},l,k} + \gamma^{k+1-l}},$$

where

$$\gamma^{\text{sum},l,k} \triangleq \sum_{j=l}^{k} \gamma^{j-l},$$

which gives us the expression of z^{k+1} reported in Step (S.3) in PATA. This is done in order to avoid keeping trace of all y^j, $j = l, .., k$, which carries a heavy computational weight. Instead, we only need to know the current value of z^k, the sum $\gamma^{\text{sum},l,k}$, γ^{k+1-l} and, last but not least, the current point y^{k+1}. This allows us to save 4 entities only, which is far more convenient.

– Because the feasible set $Y = \mathbb{B}(0, 1)$ is the unit ball of radius 1, the computation of the projection steps (see Step (S.3)) becomes straightforward, since it is sufficient to divide the argument by its vector norm:

$$P_{\mathbb{B}(0,1)}(w) = \frac{w}{\|w\|_2} \ \forall w : \|w\|_2 \geq 1.$$

Moreover, a closed-form expression for the unique solution u of the minimum problem at Step (S.4):

$$u = \arg \min_{y \in \mathbb{B}(0,1)} [F(z^{k+1}) + \frac{1}{\tau^k} G(z^{k+1})]^\top (y - z^{k+1})$$

is achievable. On the basis that the feasible set $Y = \mathbb{B}(0, 1)$ becomes an active constraint at the optimal solution u, the KKT-multiplier associated to this constraint is strictly positive. We, of course, do not know the value of the multiplier itself, but we can impose that the optimal point has Euclidean norm 1, so that it belongs to the boundary of $\mathbb{B}(0, 1)$:

$$u = -\frac{F(z^{k+1}) + \frac{1}{\tau^k} G(z^{k+1})}{\|F(z^{k+1}) + \frac{1}{\tau^k} G(z^{k+1})\|_2}.$$

We now show the implemented scheme in Algorithm 2.

Algorithm 2: Practical version of PATA

Data: $z^0, y^0 \in \mathbb{B}(0, 1)$, $i \leftarrow 1$, $l \leftarrow 0$, $k^{\max} > 0$, tol > 0, $\gamma^{\text{sum,l,k}} \leftarrow 0$, $a > 0$, $\alpha \in (0, 1]$, $\beta > 1$;

for $k = 1, \ldots, k^{max}$ **do**

(S.1) $\gamma^k = \min \left\{ 1, \frac{a}{(k-l)^\alpha} \right\}$, $\tau^k = i$, $\varepsilon^k = \frac{1}{i^\beta}$;

(S.2) $v = y^k - \gamma^k [F(y^k) + \frac{1}{\tau^k} G(y^k)])$;

(S.3)   **if** $\|v\|_2 \geq 1$ **then**
       $y^{k+1} = \frac{v}{\|v\|_2}$;
    else
       $y^{k+1} = v$;
    end

(S.4) $z^{k+1} = \frac{z^k \gamma^{\text{sum,l,k}} + \gamma^k y^{k+1}}{\gamma^{\text{sum,l,k}} + \gamma^k}$;

(S.5) $u = -\frac{F(z^{k+1}) + \frac{1}{\tau^k} G(z^{k+1})}{\|F(z^{k+1}) + \frac{1}{\tau^k} G(z^{k+1})\|_2}$;

(S.6)   **if** $[F(z^{k+1}) + \frac{1}{\tau^k} G(z^{k+1})]^\top (u - z^{k+1}) \geq -\varepsilon^k$ **then**
       if $\varepsilon^k \leq$ tol **then**
          break
       end
       $i \leftarrow i + 1$;
       $l \leftarrow k + 1$;
       $\gamma^{\text{sum,l,k+1}} = 0$;
    else
       $\gamma^{\text{sum,l,k+1}} = \gamma^{\text{sum,l,k}} + \gamma^k$;
    end

end

return z^{k+1}.

As far as the steps of Algorithm 2 are concerned, (S.2) and (S.3) perform step (S.2) of PATA, while (S.5) and (S.6) fulfil step (S.4) of PATA.

We set the parameters $k^{\max} = 10^6$, tol $= 10^{-3}$, $a = \alpha = \frac{1}{2}$, $\beta = 2$. Table 1 summarizes the results obtained by running Algorithm 2. It is clear to see how $\|z^{k+1}\|_2$ tends to 0 as the number of iterations k grows, which is what we expected, being $z^* = (0, 0)^T$ the unique solution of the problem.

To further reiterate the elements of novelty that PATA displays, we hereby present some numerical experiments in which PATA performs better than Algorithm 1 presented in Lampariello et al. (2020). We do not intend to present a thorough numerical comparison between these solution methods, we just want to show that PATA is a fundamental solution tool when the classical Tikhonov gradient method presented in Lampariello et al. (2020) struggles to converge.

Again, for the sake of simplicity we consider $Y = \mathbb{B}(0, 1)$. This time, we extend the problem to encompass $n = 100$ variables and consider $G(x) = M_G x + b_G$ and $F(x) = M_F$, with

Table 1 Numerical experiment for PATA

i	k	ε^k	$\|z^{k+1}\|_2$
1	1	1.00000	1.00e+00
2	50	0.25000	3.28e−01
3	107	0.11111	1.29e−01
4	165	0.06250	6.78e−02
5	223	0.04000	4.05e−02
6	281	0.02778	2.57e−02
7	339	0.02041	2.01e−02
8	540	0.01562	1.48e−02
9	740	0.01235	1.20e−02
10	1166	0.01000	9.73e−03
⋮	⋮	⋮	⋮
20	17691	0.00250	2.55e−03
21	21952	0.00227	2.32e−03
22	27084	0.00207	2.10e−03
23	33167	0.00189	1.92e−03
24	40281	0.00174	1.77e−03
25	48506	0.00160	1.63e−03
26	59199	0.00148	1.49e−03
27	71242	0.00137	1.39e−03
28	84715	0.00128	1.30e−03
29	99699	0.00119	1.21e−03
30	117950	0.00111	1.13e−03
31	137950	0.00104	1.06e−03
32	161698	0.00098	9.88e−04

$$M_* = \begin{pmatrix} & & & & v_1^{M_*} \\ & & & \iddots & \\ & & v_{\frac{n}{2}}^{M_*} & & \\ & -v_{\frac{n}{2}}^{M_*} & & & \\ & \iddots & & & \\ -v_1^{M_*} & & & & \end{pmatrix} + \zeta u^{M_*}(u^{M_*} + 0.01 w^{M_*})^\top, \quad * = G, F,$$

$b_G = \zeta v^{b_G}$, and $v^{M_G}, u^{M_G}, w^{M_G}, v^{M_F}, u^{M_F}, w^{M_F}$, and v^{b_G} are randomly generated between 0 and 1, $\zeta > 0$. We remark that when $\zeta = 0$ the problem is a generalization of that in the simple example described at the beginning of this section. In our experiments, we consider the cases $\zeta = 0.1$ and 0.01.

As far as PATA parameters are concerned, for the purpose of the implementation we set $k^{\max} = 5 \cdot 10^4$, $a = 1$, $\alpha = \frac{1}{4}$ and $\beta = 2$. As for the Tikhonov scheme proposed in Lampariello et al. (2020), we set $\lambda = 0.1$.

Following Proposition 1, a merit function for the nested variational inequality (2) can be given by

$$\text{optimality measure}(k) \triangleq \max \left\{ \varepsilon_{\text{sub}}^k \tau^k, \varepsilon_{\text{sub}}^k + \frac{1}{\tau^k} \right\}.$$

We generated 3 different instances of the problem and considered 2 values for ζ, for a total of 6 different test problems.

Figure 1 shows the evolution of the optimality measure as the number of inner iterations k grows towards $k^{\max}$, when both PATA and the classical Tikhonov algorithm described in Lampariello et al. (2020) are applied to the 6 test problems.

It is clear to see how the practical implementation for PATA always outperforms the classical Tikhonov in Lampariello et al. (2020), as it needs a significant smaller number of inner iterations k to reach small values of the optimality measure. However, we remark that computing averages, such as in step (S.3) of PATA, can be computationally expensive. Hence, PATA becomes an essential alternative tool when other Tikhonov-like methods, not including averaging steps, either fail to reach small values of the optimality measure, as shown in our practical implementation (see Fig. 1), or do not converge at all.

A similar scenario can be observed also when dealing with nonlinear problems. In Table 2, we report the numerical results that are obtained by running Algorithm 2 to address a nonlinear modification of the problem defined at beginning of this section: in particular, inspired by some of the test problems in Solodov and Svaiter (1999), we take

$$G(y) = \begin{pmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \ F(y) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} \max\{0, y_1\}^2 \\ \max\{0, y_2\}^2 \end{pmatrix},$$
$$Y = \mathbb{B}(0, 1).$$

The unique solution of this problem is again the origin. As for the previous case, PATA seems to work well: it reaches a good approximation level of the exact solution in a fair amount of inner iterations. Note that the nonlinear operator F is monotone on Y but also strongly monotone on $Y \setminus \{(0, 0)\}$: the strong monotonicity modulus vanishes when approaching the origin. This issue is clearly reflected in the behavior of a standard gradient-like version of the algorithm, that is readily obtained neglecting in Algorithm 2 the averaging step, thus taking there $z^{k+1} = y^{k+1}$ (see Table 3). Comparing Tables 2 and 3, one can observe as, for the initial outer iterations i, PATA and the non averaged counterpart perform similarly, while, when approaching the origin, the non averaged gradient-like procedure struggles to converge (see the number of inner iterations that performed for each outer iterate in Table 3).
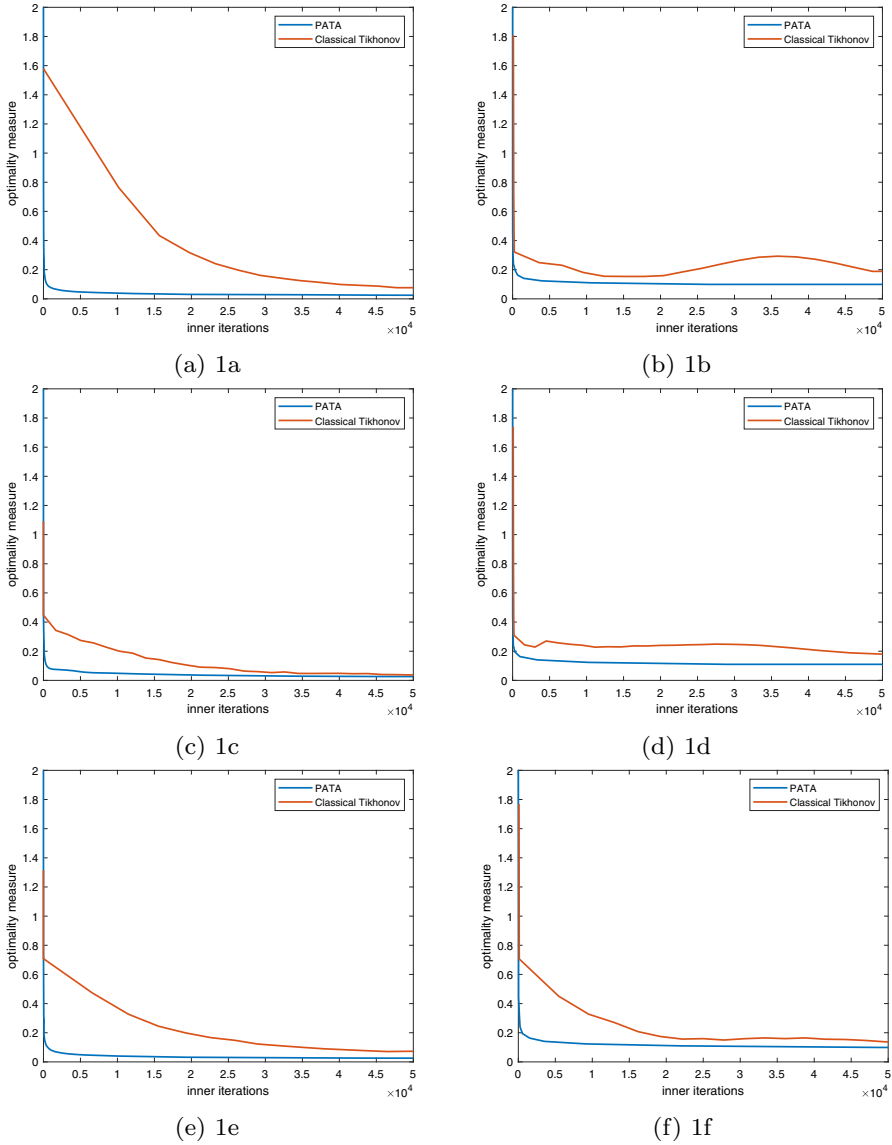
Fig. 1 Plots **a**, **c** and **e** correspond to the value $\zeta = 0.01$; plots **b**, **d** and **f** correspond to $\zeta = 0.1$. Each row is related to a different instance of the problem, namely a different seed for the random generation

Table 2 Numerical experiment (nonlinear case) for PATA

i	k	ε^k	$\|z^{k+1}\|_2$
1	1	1.00000	5.59e−01
2	12	0.25000	3.24e−01
3	32	0.11111	2.60e−01
4	70	0.06250	6.77e−02
5	120	0.04000	4.31e−02
6	174	0.02778	3.03e−02
7	233	0.02041	2.18e−02
8	436	0.01562	1.66e−02
9	489	0.01235	1.27e−02
10	1000	0.01000	1.58e−02
⋮	⋮	⋮	⋮
20	1519	0.00250	5.67e−03
21	1570	0.00227	5.93e−03
22	1621	0.00207	6.05e−03
23	1672	0.00189	6.06e−03
24	1722	0.00174	3.95e−03
25	1772	0.00160	4.49e−03
26	1823	0.00148	6.93e−03
27	1873	0.00137	4.53e−03
28	1923	0.00128	4.85e−03
29	1973	0.00119	5.10e−03
30	2023	0.00111	5.28e−03
31	2073	0.00104	5.42e−03
32	2123	0.00098	5.52e−03

6 Conclusions

We have shown that PATA is (subsequentially) convergent to solutions of monotone nested variational inequalities under the weakest conditions in the literature so far, see Theorem 1. Specifically, besides the standard convexity and monotonicity assumptions, G is required to be just monotone, while all other papers demand the monotonicity plus of G, see Facchinei et al. (2014) and Lampariello et al. (2020).

In addition, PATA enjoys interesting complexity properties, see Theorems 2 and 3. Notice that we have provided the first complexity analysis for nested variational inequalities considering optimality of both the upper- and lower-level. Conversely, authors in Lampariello et al. (2020) only handled lower-level optimality.

Although the convergence and complexity properties of the method are guaranteed by the results in Sect. 4, we plan to test numerically PATA more extensively (also when addressing nonlinear problems) in order to evaluate more accurately its performances from a practical point of view.

Table 3 Numerical experiment (nonlinear case) for the non averaged gradient-like algorithm

i	k	ε^k	$\|z^{k+1}\|_2$
1	1	1.00000	5.59e−01
2	12	0.25000	3.11e−01
3	261	0.11111	1.29e−01
4	1104	0.06250	6.91e−02
5	3390	0.04000	4.40e−02
6	8216	0.02778	3.02e−02
7	17435	0.02041	2.21e−02
8	34201	0.01562	1.66e−02
9	62192	0.01235	1.30e−02
10	106683	0.01000	1.05e−02
11	175259	0.00826	8.68e−03
12	274823	0.00694	7.26e−03
13	415372	0.00592	6.15e−03
14	608135	0.00510	5.28e−03
15	869480	0.00444	4.61e−03
16	1213087	0.00391	4.03e−03
17	1658910	0.00346	3.57e−03
18	2230967	0.00309	3.17e−03
19	2948137	0.00277	2.85e−03
20	3844104	0.00250	2.57e−03
21	4946409	0.00227	2.32e−03

Possible future research may focus on generalizing the problem to consider quasi variational inequalities as well as generalized variational inequalities. The first step would be extending Proposition 1 to encompass these more complex variational problems. We leave this investigation to following works.

Appendix

The following proposition is instrumental for the discussion regarding the natural map V in Sect. 4.

Proposition 3 *Let $x \in K$ satisfy the primal VI approximate optimality condition* (3), *then the natural map approximate optimality condition* (14) *holds with $\widehat{\varepsilon} \geq \sqrt{\varepsilon}$. Vice versa, let $x \in K$ satisfy condition* (14), *then* (3) *holds with $\varepsilon \geq (\Omega + \Xi)\widehat{\varepsilon}$, where $\Omega \triangleq \max_{v,y \in K} \|v - y\|_2$ and $\Xi \triangleq \max_{y \in K} \|\Psi(y)\|_2$.*

Proof Let $z = P_K(x - \Psi(x))$. If (3) holds, then

$$
\begin{aligned}
-\varepsilon &\leq \Psi(x)^\top (z - x) = [x - z - (x - z - \Psi(x))]^\top (z - x) \\
&= -\|x - z\|_2^2 - (x - \Psi(x) - z)^\top (z - x) \leq -\|x - z\|_2^2 = -U(x)^2,
\end{aligned}
$$

where the last inequality is due to the characteristic property of the projection. Therefore, (14) holds with $\widehat{\varepsilon} \geq \sqrt{\varepsilon}$.

Now we consider the case in which (14) holds. Thanks again to the characteristic property of the projection, we have

$$
[z - (x - \Psi(x))]^\top (y - z) \geq 0, \quad \forall y \in K,
$$

and, thus, for all $y \in K$,

$$
\begin{aligned}
\Psi(x)^\top (y - x) &\geq (x - z)^\top (y - z) + \Psi(x)^\top (z - x) \\
&\geq -(\|y - z\|_2 + \|\Psi(x)\|_2)\|x - z\|_2 \\
&\geq -(\Omega + \Xi)U(x) \geq -(\Omega + \Xi)\widehat{\varepsilon}.
\end{aligned}
$$

Therefore, (3) holds with $\varepsilon \geq (\Omega + \Xi)\widehat{\varepsilon}$. $\qquad\square$

The following lemma is helpful to prove Theorems 2 and 3.

Lemma 1 *Let $\{\gamma^k : \gamma^k = \min\{1, \frac{a}{k^\alpha}\}\}$, with $\alpha \in (0, 1]$ and $a > 0$. Setting $K \in \mathbb{N}$, the following upper and lower bounds hold true:*

(i) $\alpha \neq 1 : \sum_{k=0}^{K} \gamma^k \geq \lceil a^{\frac{1}{\alpha}} \rceil + \frac{a}{1-\alpha}[(K+1)^{1-\alpha} - \lceil a^{\frac{1}{\alpha}} \rceil^{1-\alpha}]$;

(ii) $\alpha = 1 : \sum_{k=0}^{K} \gamma^k \geq \lceil a \rceil + \ln\left((\frac{K+1}{\lceil a \rceil})^a\right)$;

(iii) $\alpha \neq \frac{1}{2} : \sum_{k=0}^{K} (\gamma^k)^2 \leq \lceil a^{\frac{1}{\alpha}} \rceil + \frac{a^2}{\lceil a^{\frac{1}{\alpha}} \rceil^{2\alpha}} + \frac{a^2}{1-2\alpha}[K^{1-2\alpha} - \lceil a^{\frac{1}{\alpha}} \rceil^{1-2\alpha}]$;

(iv) $\alpha = \frac{1}{2} : \sum_{k=0}^{K} (\gamma^k)^2 \leq \lceil a^2 \rceil + \frac{a^2}{\lceil a^2 \rceil} + a^2 \ln(\frac{K}{\lceil a^2 \rceil})$.

Proof In cases (i) and (ii):

$$
\begin{aligned}
\sum_{k=0}^{K} \gamma^k &= \sum_{k=0}^{K} \min\{1, \frac{a}{k^\alpha}\} \\
&= \lceil a^{\frac{1}{\alpha}} \rceil + \sum_{k=\lceil a^{\frac{1}{\alpha}} \rceil}^{K} \frac{a}{k^\alpha} \\
&\geq \lceil a^{\frac{1}{\alpha}} \rceil + \int_{\lceil a^{\frac{1}{\alpha}} \rceil}^{K+1} a x^{-\alpha} dx,
\end{aligned}
$$

where the inequality is due to the integral test for Harmonic series. When $\alpha \neq 1$ it follows that:

$$
\begin{aligned}
\sum_{k=0}^{K} \gamma^k &\geq \lceil a^{\frac{1}{\alpha}} \rceil + a \frac{x^{1-\alpha}}{1-\alpha}\Big|_{\lceil a^{\frac{1}{\alpha}} \rceil}^{K+1} \\
&= \lceil a^{\frac{1}{\alpha}} \rceil + \frac{a}{1-\alpha}[(K+1)^{1-\alpha} - \lceil a^{\frac{1}{\alpha}} \rceil^{1-\alpha}],
\end{aligned}
$$

whilst, if $\alpha = 1$, it follows that:

$$\sum_{k=0}^{K} \gamma^k \geq \lceil a \rceil + a \ln|x| \big|_{\lceil a \rceil}^{K+1}$$
$$= \lceil a \rceil + \ln\left(\left(\frac{K+1}{\lceil a \rceil}\right)^a \right).$$

In cases (iii) and (iv):

$$\sum_{k=0}^{K} (\gamma^k)^2 = \sum_{k=0}^{K} (\min\{1, \frac{a}{k^\alpha}\})^2$$
$$= \lceil a^{\frac{1}{\alpha}} \rceil + \sum_{k=\lceil a^{\frac{1}{\alpha}} \rceil}^{K} \frac{a^2}{k^{2\alpha}}$$
$$= \lceil a^{\frac{1}{\alpha}} \rceil + \frac{a^2}{\lceil a^{\frac{1}{\alpha}} \rceil^{2\alpha}} + \sum_{k=\lceil a^{\frac{1}{\alpha}} \rceil+1}^{K} \frac{a^2}{k^{2\alpha}}$$
$$\leq \lceil a^{\frac{1}{\alpha}} \rceil + \frac{a^2}{\lceil a^{\frac{1}{\alpha}} \rceil^{2\alpha}} + \int_{\lceil a^{\frac{1}{\alpha}} \rceil}^{K} a^2 x^{-2\alpha} dx,$$

where, in the third equality, the first term of the series is taken out, whilst the inequality is once again due to the integral test for Harmonic series.

When $\alpha \neq \frac{1}{2}$ it follows that:

$$\sum_{k=0}^{K} (\gamma^k)^2 \leq \lceil a^{\frac{1}{\alpha}} \rceil + \frac{a^2}{\lceil a^{\frac{1}{\alpha}} \rceil^{2\alpha}} + a^2 \frac{x^{1-2\alpha}}{1-2\alpha} \Big|_{\lceil a^{\frac{1}{\alpha}} \rceil}^{K}$$
$$= \lceil a^{\frac{1}{\alpha}} \rceil + \frac{a^2}{\lceil a^{\frac{1}{\alpha}} \rceil^{2\alpha}} + \frac{a^2}{1-2\alpha} [K^{1-2\alpha} - \lceil a^{\frac{1}{\alpha}} \rceil^{1-2\alpha}],$$

while, if $\alpha = \frac{1}{2}$, it follows that:

$$\sum_{k=0}^{K} (\gamma^k)^2 \leq \lceil a^2 \rceil + \frac{a^2}{\lceil a^2 \rceil} + a^2 \ln|x| \big|_{\lceil a^2 \rceil}^{K}$$
$$= \lceil a^2 \rceil + \frac{a^2}{\lceil a^2 \rceil} + a^2 \ln(\frac{K}{\lceil a^2 \rceil}).$$

$$\square$$

The following proposition enables us to use $\gamma^k = \min\{1, \frac{a}{k^\alpha}\}$, which is shown to satisfy conditions (8) in Theorem 1.

Proposition 4 *The sequence of stepsizes $\{\gamma^k : \gamma^k = \min\{1, \frac{a}{k^\alpha}\}\}$, with $\alpha \in (0, 1]$ and $a > 0$, satisfies both sets of hypotheses:*

(i) $\sum_{k=0}^{\infty} \gamma^k = \infty$

(ii) $\frac{\sum_{k=0}^{\infty} (\gamma^k)^2}{\sum_{k=0}^{\infty} \gamma^k} = 0$

needed for Theorem 1 to be valid, see conditions (8).

Proof (i) We need to examine two separate cases, i.e. when $\alpha = 1$ and $\alpha < 1$, when K goes to $+\infty$.

When $\alpha = 1$:

$$\sum_{k=0}^{\infty} \gamma^k \geq \lim_{K \to +\infty} \left(\lceil a \rceil + \ln\left(\left(\frac{K+1}{\lceil a \rceil}\right)^a \right) \right) = +\infty.$$

When, instead, $\alpha < 1$:

$$\sum_{k=0}^{\infty} \gamma^k \geq \lim_{K \to +\infty} \left(\lceil a^{\frac{1}{\alpha}} \rceil + \frac{a}{1-\alpha} [(K+1)^{1-\alpha} - \lceil a^{\frac{1}{\alpha}} \rceil^{1-\alpha}] \right) = +\infty.$$

Notice that, if $\alpha > 1$, the limit would be $\lceil a^{\frac{1}{\alpha}} \rceil - \frac{a}{1-\alpha} \lceil a^{\frac{1}{\alpha}} \rceil^{1-\alpha} \neq +\infty$, hence why it is necessary that $\alpha \leq 1$.

(*ii*) We first show that, if $\alpha \in \left(\frac{1}{2}, 1 \right]$, the following relation holds:

$$\sum_{k=0}^{\infty} (\gamma^k)^2 \leq \lim_{K \to +\infty} \left(\lceil a^{\frac{1}{\alpha}} \rceil + \frac{a^2}{\lceil a^{\frac{1}{\alpha}} \rceil^{2\alpha}} + \frac{a^2}{1-2\alpha} [K^{1-2\alpha} - \lceil a^{\frac{1}{\alpha}} \rceil^{1-2\alpha}] \right)$$

$$= \lceil a^{\frac{1}{\alpha}} \rceil + \frac{a^2}{\lceil a^{\frac{1}{\alpha}} \rceil^{2\alpha}} - \frac{a^2}{1-2\alpha} \lceil a^{\frac{1}{\alpha}} \rceil^{1-2\alpha}$$

$$= C,$$

which, in turn, implies that:

$$\frac{\sum_{k=0}^{\infty} (\gamma^k)^2}{\sum_{k=0}^{\infty} \gamma^k} \leq \frac{C}{+\infty} = 0.$$

When $\alpha = \frac{1}{2}$, it easy to see that:

$$\frac{\sum_{k=0}^{\infty} (\gamma^k)^2}{\sum_{k=0}^{\infty} \gamma^k} \leq \lim_{K \to +\infty} \frac{\lceil a^2 \rceil + \frac{a^2}{\lceil a^2 \rceil} + a^2 \ln \left(\frac{K}{\lceil a^2 \rceil} \right)}{\lceil a^2 \rceil + 2a[\sqrt{K+1} - \lceil a^2 \rceil^{\frac{1}{2}}]}$$

$$= \lim_{K \to +\infty} \frac{\left(\frac{a^2 \lceil a^2 \rceil}{K} \right)}{\left(\frac{2a}{2\sqrt{K+1}} \right)}$$

$$= \lim_{K \to +\infty} \frac{a(\lceil a^2 \rceil)\sqrt{K+1}}{K}$$

$$= \lim_{K \to +\infty} a(\lceil a^2 \rceil)\sqrt{\frac{1}{K} + \frac{1}{K^2}}$$

$$= 0,$$

where the first equality follows from L'Hopital's rule.

Finally, if $\alpha \in \left(0, \frac{1}{2}\right)$, it is once again easy to see that:

$$
\begin{aligned}
\frac{\sum_{k=0}^{\infty}(\gamma^k)^2}{\sum_{k=0}^{\infty}\gamma^k} &\leq \lim_{K \to +\infty} \frac{\lceil a^{\frac{1}{\alpha}} \rceil + \frac{a^2}{\lceil a^{\frac{1}{\alpha}} \rceil^{2\alpha}} + \frac{a^2}{1-2\alpha}[K^{1-2\alpha} - \lceil a^{\frac{1}{\alpha}} \rceil^{1-2\alpha}]}{\lceil a^{\frac{1}{\alpha}} \rceil + \frac{a}{1-\alpha}[(K+1)^{1-\alpha} - \lceil a^{\frac{1}{\alpha}} \rceil^{1-\alpha}]} \\
&= \lim_{K \to +\infty} \frac{\frac{a^2}{1-2\alpha}(1-2\alpha)K^{-2\alpha}}{\frac{a}{1-\alpha}(1-\alpha)(K+1)^{-\alpha}} \\
&= \lim_{K \to +\infty} \frac{aK^{-2\alpha}}{(K+1)^{-\alpha}} \\
&= \lim_{K \to +\infty} \frac{aK^{-\alpha}}{(1+\frac{1}{K})^{-\alpha}} \\
&= 0,
\end{aligned}
$$

where the first equality follows again from L'Hopital's rule, and the last equality is true because $\alpha > 0$. □

This last lemma is again used in the proof of Theorem 2.

Lemma 2 *The following upper bound holds for the lower-level merit function V (see the definition* (15)*) at z for every positive τ:*

$$
V(z) \leq \frac{1}{\tau}\|G(z)\|_2 + \|P_Y(z - \Phi_\tau(z)) - z\|_2. \tag{26}
$$

Proof The claim is a consequence of the following chain of relations:

$$
\begin{aligned}
V(z) &= \|P_Y(z - F(z)) - z\|_2 \\
&\leq \|P_Y(z - F(z)) - P_Y(z - (F(z) + \frac{1}{\tau}G(z)))\|_2 \\
&\quad + \|P_Y(z - (F(z) + \frac{1}{\tau}G(z))) - z\|_2 \\
&\leq \frac{1}{\tau}\|G(z)\|_2 + \|P_Y(z - (F(z) + \frac{1}{\tau}G(z))) - z\|_2,
\end{aligned}
$$

where the last inequality follows from the nonexpansive property of the projection mapping. □

References

Bigi G, Lampariello L, Sagratella S (2021) Combining approximation and exact penalty in hierarchical programming. Optimization 71:1–17

Dempe S (2002) Foundations of bilevel programming. Springer, New York

Facchinei F, Pang JS (2003) Finite-dimensional variational inequalities and complementarity problems. Springer

Facchinei F, Pang JS, Scutari G, Lampariello L (2014) VI-constrained hemivariational inequalities: distributed algorithms and power control in ad-hoc networks. Math Program 145(1–2):59–96

Facchinei F, Scutari G, Sagratella S (2015) Parallel selective algorithms for nonconvex big data optimization. IEEE Trans Signal Process 63(7):1874–1889

Kalashnikov VV, Kalashinikova NI (1996) Solving two-level variational inequality. J Global Optim 8(3):289–294

Lampariello L, Sagratella S (2017) A bridge between bilevel programs and Nash games. J Optim Theory Appl 174(2):613–635

Lampariello L, Sagratella S (2020) Numerically tractable optimistic bilevel problems. Comput Optim Appl 76(2):277–303

Lampariello L, Sagratella S, Stein O (2019) The standard pessimistic bilevel problem. SIAM J Optim 29(2):1634–1656

Lampariello L, Neumann C, Ricci JM, Sagratella S, Stein O (2020) An explicit Tikhonov algorithm for nested variational inequalities. Comput Optim Appl 77(2):335–350

Lampariello L, Neumann C, Ricci JM, Sagratella S, Stein O (2021) Equilibrium selection for multi-portfolio optimization. Eur J Oper Res 295:363–373

Lu X, Xu HK, Yin X (2009) Hybrid methods for a class of monotone variational inequalities. Nonlinear Anal Theory Methods Appl 71(3–4):1032–1041

Marino G, Xu HK (2011) Explicit hierarchical fixed point approach to variational inequalities. J Optim Theory Appl 149(1):61–78

Scutari G, Facchinei F, Pang JS, Lampariello L (2012) Equilibrium selection in power control games on the interference channel. In: 2012 Proceedings IEEE INFOCOM. IEEE, pp 675–683

Solodov M, Svaiter B (1999) A new projection method for variational inequality problems. SIAM J Control Optim 37(3):765–776

Yamada I (2001) The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. Inherently Parallel Algorithms Feasibility Optim Appl 8:473–504