



An explicit Tikhonov algorithm for nested variational inequalities

Lorenzo Lampariello¹ · Christoph Neumann² · Jacopo M. Ricci¹ · Simone Sagratella³ · Oliver Stein²

Received: 16 March 2020 / Published online: 6 July 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

We consider nested variational inequalities consisting in a (upper-level) variational inequality whose feasible set is given by the solution set of another (lower-level) variational inequality. Purely hierarchical convex bilevel optimization problems and certain multi-follower games are particular instances of nested variational inequalities. We present an explicit and ready-to-implement Tikhonov-type solution method for such problems. We give conditions that guarantee the convergence of the proposed method. Moreover, inspired by recent works in the literature, we provide a convergence rate analysis. In particular, for the simple bilevel instance, we are able to obtain enhanced convergence results.

Keywords Nested variational inequality · Purely hierarchical problem · Tikhonov method · Convergence rate

Mathematics Subject Classification 90C33 · 90C25 · 90C30 · 65K15 · 65K10

1 Introduction

This paper presents and analyzes an explicit and ready-to-implement Tikhonov-type solution method for nested Variational Inequalities (VIs), where the feasible set of a (upper-level) VI is given by the solution set of another (lower-level) VI. Under mild

This work was supported by the MIUR-DAAD Joint Mobility Program under Grant 34793 and Project-ID 57396680.

✉ Lorenzo Lampariello
lorenzo.lampariello@uniroma3.it

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

² Institute of Operations Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

³ Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza University of Rome, Rome, Italy

monotonicity and continuity assumptions, this setting does not only model certain convex bilevel optimization problems but also multi-follower games. Note that, differently from the more general bilevel structures [5, 16–18], we focus only on lower-level problems that are *non-parametric* with respect to the upper level variables, but have multiple solutions.

For nested VIs, two main solution approaches have been proposed in the literature: hybrid-like schemes (see, e.g., [19, 20, 28]), and Tikhonov-type techniques (see, e.g., [14] for some early developments, and [12] and the references therein). For the sake of completeness, here we also cite the recent penalty method [4] where an approximated version of the problem is addressed. Since hybrid-like methods need rather strong assumptions to work properly (see Sect. 2), we rely instead on the Tikhonov paradigm, along the lines of the general scheme put forward in [12]. However, as a major departure from [12], our algorithm is explicit and easy-to-implement: no nontrivial problems are requested to be addressed while the method progresses, and it essentially consists of iterative computations of simple projection steps. We provide a convergence analysis, showing that the method is globally (sub-sequential) convergent to a solution of the problem under mild conditions. Furthermore, we give some convergence rate bounds. Concerning the latter topic, we wish to emphasize that, to date, [2, 21] present the only convergence rate results for problems with a hierarchical structure. However, they exclusively consider the instance of simple bilevel problems, that is, the easier case in which G and F are gradients of convex functions. In Sect. 2, see Example 1, we show that the projected-gradient-like schemes of [2, 21] cannot be readily generalized to the framework of nested VIs, and there we also formulate the motivation of our approach in more detail. Section 3 presents our Tikhonov-like algorithm, in Sect. 4 we describe the main convergence properties of our method, while in Sect. 5 we address the specific instance of simple bilevel problems from the point of view of our approach. For the latter simpler case, we are able to show enhanced convergence rate results.

We remark that in [15] our algorithm is applied to a multi-follower game that arises as a model for equilibrium selection in multi-portfolio optimization. There, detailed numerical results illustrate the applicability of our approach.

2 Problem definition and motivation

We address the hierarchical problem

$$\text{VI}(G, \text{SOL}(F, Y)), \quad (1)$$

where $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the upper-level map, and $\text{SOL}(F, Y)$ is the solution set of the lower-level $\text{VI}(F, Y)$, with $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^n$. We recall that the variational inequality (see [8–11, 26]) $\text{VI}(F, Y)$ is the problem of finding $x \in Y$ with

$$F(x)^T(y - x) \geq 0 \quad \forall y \in Y. \quad (2)$$

Hence, the solution set $\text{SOL}(F, Y)$ of the lower-level VI consists of the points $x \in Y$ such that (2) holds. The upper-level VI and its solution set are defined analogously:

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

We impose the following general assumptions:

- (A1) G is Lipschitz continuous with constant L_G and monotone plus on Y , i.e. G is monotone on Y and, for every $x, y \in Y$, $(G(y) - G(x))^T(y - x) = 0$ implies $G(y) = G(x)$;
- (A2) F is Lipschitz continuous with constant L_F and monotone on Y ;
- (A3) Y is nonempty, convex and compact.

Assumptions (A2) and (A3) yield a nonempty, convex, compact and not necessarily single-valued set $\text{SOL}(F, Y)$ (cf. [11, Section 2.3]) and, thus, make the “selection” problem (1) meaningful. Note also that, under the above conditions, problem (1) might turn out to have multiple solutions. The monotone plus condition in (A1) is satisfied, for example (see the results in [11, Section 2.3.1]), if G is symmetric and monotone: a case worth to be considered is the one of G being the gradient of a convex function. We underline that monotonicity plus is also valid whenever G is strongly monotone: hence, our framework covers also certain problems in which the upper-level consists in a uniquely solvable Nash equilibrium problem.

The assumptions on the lower-level problem, on the other hand, do not only allow F to be the gradient of a convex function, but the lower-level VI may also model more general structures, like e.g. Nash games with nonunique equilibria (see [1, 7, 13, 22–25]).

We remark that for the hybrid-like schemes from [19, 20, 28] to work properly for the solution of (1), one has to call for conditions that are stronger than (A2). Basically, in [19, 20, 28], F is required to be co-coercive, a condition that substantially holds in two cases: when the lower-level VI reduces to a convex optimization problem, or when it has a unique solution. On the contrary, here we only require F to be monotone and Lipschitz continuous, allowing for the treatment of more general problems with respect to the approach in [19, 20, 28]. For details we refer the interested reader to [12, Section 2].

Here, we rely on the Tikhonov paradigm, along the lines of the general scheme put forward in [12], whose core iteration k is defined as

$$w^{k+1} \in \text{SOL}\left(F + \frac{1}{\tau^k}G + \lambda(\bullet - w^k), Y\right), \tag{3}$$

where $\frac{1}{\tau^k}$ is the so-called Tikhonov parameter, with $\frac{1}{\tau^k} \downarrow 0$ and $\sum_{k=0}^{\infty} \frac{1}{\tau^k} = \infty$, and $\lambda > 0$ is the proximal constant. Observe that the VI-subproblem $\text{VI}\left(F + \frac{1}{\tau^k}G + \lambda(\bullet - w^k), Y\right)$, to be solved in (3), has a unique solution, in view of the strong monotonicity of its operator. For further aspects regarding Tikhonov-like methods, see [12].

As a major departure from [12], the present article provides an explicit, ready-to-implement algorithm for problem (1), in the sense that it essentially consists of iterative computations of simple projection steps, and then no nontrivial subproblems are requested to be solved. Furthermore, we also give some convergence rate analysis.

To date, [2, 21] present the only converge rate results for problems with a hierarchical structure. However, [2, 21] exclusively consider the simple bilevel instance of problem (1), that is, the case in which G and F are gradients of convex functions. As the following example shows, the projected-gradient-like schemes presented in [2, 21] cannot be readily generalized to the framework of problem (1).

Example 1 For the sake simplicity, we refer to the more recent BiG-SAM in [21]. Consider the selection problem (1) where

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

with the unit ball $\mathbb{B}(0, 1)$. The unique feasible point and, thus, the unique solution of the problem is $y^* = (0, 0)^T$. The map G is symmetric, strongly monotone with constant 1, and Lipschitz continuous with constant $L_G = 1$; F is Lipschitz continuous with constant $L_F = 1$. Therefore, the assumptions (A1)–(A3) are satisfied. Moreover, aside from the fact that F is nonsymmetric, since all the convergence conditions in [21] are satisfied, one could be tempted to apply right away BiG-SAM to the problem at hand. The generic k th iteration of BiG-SAM, in this case, should read as reported below:

$$\begin{aligned} w^k &= P_Y(y^k - F(y^k)); \\ z^k &= y^k - G(y^k); \\ y^{k+1} &= \alpha_k z^k + (1 - \alpha_k)w^k; \end{aligned}$$

where we take, for example, but without loss of generality, $\alpha_k = \frac{\sqrt{2}-1}{\sqrt{2k}}$. Note that all these choices comply with the requirements given in [21]. Taking y^k such that $\|y^k\| \geq \frac{1}{\sqrt{2}}$, we obtain

$$w^k = P_Y\left(\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} y_1^k \\ y_2^k \end{pmatrix}\right),$$

hence $\|w^k\| = \min\{1, \sqrt{2}\|y^k\|\} = 1$,

$$\begin{aligned} z^k &= 0; \\ y^{k+1} &= (1 - \alpha_k)w^k. \end{aligned}$$

Finally $\|y^{k+1}\| = |1 - \alpha_k|\|w^k\| \geq \frac{1}{\sqrt{2}}$. Therefore, considering e.g. $(1, 0)^T$ as starting point, the sequence produced by BiG-SAM does not lead to the unique solution y^* of the problem.

We now show that the Tikhonov approach (3) works well when applied to the same problem as above. In this case, the generic iteration (3), setting for simplicity but without loss of generality $\tau^k = k$, and $\lambda = 1$, reads as:

$$\begin{pmatrix} w_2^{k+1} + \frac{1}{k}w_1^{k+1} + w_1^{k+1} - w_1^k \\ -w_1^{k+1} + \frac{1}{k}w_2^{k+1} + w_2^{k+1} - w_2^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Letting $b = 1 + \frac{1}{k}$, we have

$$\begin{pmatrix} b & 1 \\ -1 & b \end{pmatrix} w^{k+1} = w^k,$$

and, being $b^2 + 1 > 0$, we obtain

$$w^{k+1} = \frac{1}{b^2 + 1} \begin{pmatrix} b & -1 \\ 1 & b \end{pmatrix} w^k.$$

In turn,

$$\|w^{k+1}\| \leq \frac{1}{b^2 + 1} \left\| \begin{pmatrix} b & -1 \\ 1 & b \end{pmatrix} \right\| \|w^k\| = \frac{1}{\sqrt{b^2 + 1}} \|w^k\|,$$

where $\frac{1}{\sqrt{b^2 + 1}} < 1$, and hence, $\|w^k\| \rightarrow 0$, i.e. $w^k \rightarrow y^*$. □

With the above considerations in mind, summarizing, the main goals of this work are:

- to provide an *explicit*, practical version of the general Tikhonov-like scheme proposed in [12] whose generic iteration is given in (3);
- to give, as far as we are aware, the first convergence rate results in the context of general nested VIs. Note that this is much in the spirit of the analysis provided in [2, 21] where, however, only the particular instance of simple bilevel problems is considered.

More specifically, concerning the second point, we believe that our analysis closes a gap in the literature, since, as observed in [21], the major missing part about papers dealing with, inter alia, Tikhonov approaches “is that, while convergence was proven, the convergence rates of these algorithms are unknown”.

3 An explicit projected Tikhonov-prox approach

We introduce an explicit version for the general Tikhonov scheme (3). For a fixed given outer iterate w^k and for a fixed value of the (outer) Tikhonov parameter $1/\tau^k$, the following Algorithm 1 essentially consists in the iterative computation of some (inner) projection steps until the VI-subproblem $\text{VI}\left(F + \frac{1}{\tau^k}G + \lambda(\cdot - w^k), Y\right)$ is solved (up to a prescribed outer accuracy). Then, Tikhonov parameter, current outer iterate and accuracy required in the solution of the VI-subproblem are modified (externally), and so on. Note that, as it will be detailed next, thanks to error bound results, we know how far the current iterate is from the exact solution of the VI-subproblem, even without calculating it. As a result, Algorithm 1 is *explicit*, contrary to the *implicit* nature of the procedure defined by (3): the generic iteration (3) requires in fact a nontrivial VI to be solved (possibly inexactly), whereas in Algorithm 1 no

nontrivial problems are requested to be addressed while the method progresses. Algorithm 1 can be implemented right away and has a very low cost per iterate, especially when projection onto Y can be computed in closed form. In summary, one can say that the definition of Algorithm 1 for nested VIs parallels the introduction of the descent method in [27] for simple bilevel problems. We also underline that Algorithm 1, on the one hand, is not a mere generalization of the one presented in [27], on the other hand, its convergence properties are not straightforward consequences of the ones for the general scheme in [12] (see Theorem 1 below).

Algorithm 1: Explicit method for problem (1)

Data: $\{\bar{\varepsilon}^i\} \subseteq \mathbb{R}_+, \lambda \in \mathbb{R}_+, \{\bar{\tau}^i\} \subseteq \mathbb{R}_{++}, \eta \in [0, 1], \{\gamma^k\}$ s.t.
 $[\eta, 1] \ni \gamma^k \downarrow \eta, y^0 \in Y, w^0 = y^0, i \leftarrow 0;$

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

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

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

(S.3) **if** $\|y^{k+1} - y^k\| \leq \varepsilon^k$ **then**

| $w^{k+1} = y^k, i = i + 1;$

(S.4) **else**

| $w^{k+1} = w^k;$

end

end

Some further comments are in order. We remark that a strictly positive lower bound η for the step size γ^k is mandatory in our general VI framework, in contrast to the standard optimization case where η can be zero, see [11, Theorem 12.1.8] and the developments below (specifically, convergence Theorems 1 and 2).

Let now \mathcal{K} be the set of iteration indices such that the condition in step (S.3) is satisfied; taking $\bar{\varepsilon}^i = 0$ for every i (exact case), the sequence $\{w^k\}_{\mathcal{K}}$ generated by Algorithm 1 is nothing else but the one defined by the general Tikhonov iteration (3): considering any $k \in \mathcal{K}$, being $\gamma^k > 0$, it is standard to show that

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

$$\Updownarrow$$

$$w^{k+1} \in \text{SOL}\left(F + \frac{1}{\tau^k}G + \lambda(\bullet - w^k), Y\right),$$

see [11, Proposition 1.5.8]. However, we observe that solving *exactly* (i.e. for $\bar{\varepsilon}^i = 0$) the VI-subproblem $\text{VI}\left(F + \frac{1}{\tau^k}G + \lambda(\bullet - w^k), Y\right)$ is impractical. For this reason, the possibly inaccurate solution of the VI-subproblem must be taken into account, as done in [12]. But there, the condition on the inexactness is only theoretical. In particular, the inaccurate version of the generic k th iteration (3) requires one to compute w^{k+1} “sufficiently near” to the set $\text{SOL}\left(F + \frac{1}{\tau^k}G + \lambda(\bullet - w^k), Y\right)$, i.e. such that

$$\|y_{\tau^k}(w^k) - w^{k+1}\| \leq \zeta^k, \tag{4}$$

where $y_{\tau^k}(w^k)$ is the unique solution of $\text{VI}\left(F + \frac{1}{\tau^k}G + \lambda(\bullet - w^k), Y\right)$ and ζ^k is a suitable tolerance. Actually, no practical guidelines to calculate w^{k+1} (that avoid the computation of $y_{\tau^k}(w^k)$) are given in [12]. On the contrary, considering strictly positive values for $\bar{\varepsilon}^i$ in Algorithm 1, we propose some practically implementable rules to find w^{k+1} such that (4) is satisfied: specifically, leveraging the (inner) projection step (S.2), for fixed outer iterate w^k , Tikhonov parameter $1/\tau^k$ and accuracy ε^k , the condition in step (S.3) is shown to be verified in a finite number of (inner) iterations. Thanks to error bound results, this makes relation (4), where ζ^k is proportional to ε^k , automatically satisfied without the computation of $y_{\tau^k}(w^k)$, see the proof of Theorem 1.

4 Main convergence properties

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

$$\Phi_\tau(y;w) \triangleq F(y) + \frac{1}{\tau}G(y) + \lambda(y - w).$$

For any $\lambda > 0, \tau \in \mathbb{R}_{++}$ and $w \in \mathbb{R}^n$, $\Phi_\tau(\bullet;w)$ is easily shown to be strongly monotone with constant λ and Lipschitz continuous with constant $L_\Phi \triangleq L_F + L_G + \lambda$, without loss of generality. We also consider, again without loss of generality, $2\lambda \leq L_\Phi^2$. We recall that $y_\tau(w)$ denotes the unique solution of $\text{VI}(\Phi_\tau(\bullet;w), Y)$. The main convergence result for Algorithm 1 is detailed below.

Theorem 1 *Assuming conditions (A1)–(A3) to hold, and setting*

$$\lambda > 0, \eta \in \left(0, \frac{2\lambda}{L_\Phi^2}\right), \frac{1}{\bar{\tau}^i} \downarrow 0 \text{ and } \sum_{i=0}^\infty \frac{1}{\bar{\tau}^i} = \infty, \bar{\varepsilon}^i = \frac{c}{(\bar{\tau}^i)^2},$$

with $\frac{2\lambda}{L_\Phi^2} \leq 1$ and $c > 0$, each limit point of the sequence $\{w^k\}$ generated by Algorithm 1 is a solution of problem (1).

Proof We take the cue from [12, Theorem 2]. Observing that G is monotone plus on Y thanks to (A1), and in view of (A2) and (A3), we only need to prove that:

- (i) $\left\{\frac{1}{\tau^k}\right\} \downarrow 0$ and $\sum_{k \in \mathcal{K}} \frac{1}{\tau^k} = \infty$,
- (ii) for every $k \in \mathcal{K}$, we have $\|y_{\tau^k}(w^k) - w^{k+1}\| \leq \zeta^k$ for some $\zeta^k \in \mathbb{R}_+$ such that $\left\{\frac{\zeta^k}{1/\tau^k}\right\} \rightarrow 0$,

where \mathcal{K} denotes the set of iteration indices such that the condition in step (S.3) is satisfied.

(i) Preliminarily, observe that a necessary condition for the assertion to hold is that \mathcal{K} is an infinite set of indices. Suppose by contradiction that \mathcal{K} is finite, i.e. there

exists an index \bar{k} such that the condition in step (S.3) is violated for every $k \geq \bar{k}$. Since $\eta \in \left(0, \frac{2\lambda}{L_\Phi^2}\right)$, and in view of $\gamma^k \downarrow \eta$, there exists \hat{k} such that

$$0 < \eta = \inf_{k \geq \hat{k}} \gamma^k \leq \sup_{k \geq \hat{k}} \gamma^k < \frac{2\lambda}{L_\Phi^2}.$$

Hence, by [11, Theorem 12.1.8], the sequence $\{y^k\}$ converges to the unique solution $y_{\tau^{\bar{k}}}(w^{\bar{k}})$ of $\text{VI}(Y, \Phi_{\tau^{\bar{k}}}(\cdot; w^{\bar{k}}))$, in contradiction to $\|y^{k+1} - y^k\| > \varepsilon^k = \varepsilon^{\bar{k}}$ for every $k \geq \bar{k}$. As a consequence, we have $\frac{1}{\tau} \downarrow 0$, $\sum_{k \in \mathcal{K}} \frac{1}{\tau^k} = \infty$ and $\varepsilon^k \downarrow 0$.

(ii) We recall that $\frac{1}{\alpha} \|P_Y(y^k - \alpha \Phi_{\tau^k}(y^k; w^k)) - y^k\|$ is monotonically non-increasing with respect to α , for any $\alpha \in \mathbb{R}_{++}$ and $y^k \in Y$ (see [3, Lemma 2.3.1]). As a consequence, we obtain

$$\|P_Y(y^k - \Phi_{\tau^k}(y^k; w^k)) - y^k\| \leq \frac{1}{\gamma^k} \|P_Y(y^k - \gamma^k \Phi_{\tau^k}(y^k; w^k)) - y^k\|, \tag{5}$$

because $\gamma^k \in (0, 1]$. For every $k \in \mathcal{K}$ we obtain the following chain of inequalities:

$$\begin{aligned} \|y_{\tau^k}(w^k) - w^{k+1}\| &= \|y_{\tau^k}(w^k) - y^k\| \\ &\stackrel{(a)}{\leq} \frac{1+L_\Phi}{\lambda} \|P_Y(y^k - \Phi_{\tau^k}(y^k; w^k)) - y^k\| \\ &\stackrel{(b)}{\leq} \frac{1+L_\Phi}{\lambda} \frac{1}{\gamma^k} \|P_Y(y^k - \gamma^k \Phi_{\tau^k}(y^k; w^k)) - y^k\| \\ &= \frac{1+L_\Phi}{\lambda} \frac{1}{\gamma^k} \|y^{k+1} - y^k\| \\ &\stackrel{(c)}{\leq} \frac{1+L_\Phi}{\lambda} \frac{1}{\gamma^k} \varepsilon^k \triangleq \zeta^k, \end{aligned}$$

where (a) follows from the classical error bound result [11, Proposition 6.3.1], (b) is due to (5), and (c) holds in view of the condition in step (S.3). Finally, the choice of $\bar{\varepsilon}^i$ and the updating rules for ε and τ yield $\varepsilon^k \tau^k \rightarrow 0$ and, thus, $\frac{\varepsilon^k}{1/\tau^k} \rightarrow 0$. \square

Clearly, if problem (1) turns out to be uniquely solvable, such as when e.g. G is strictly monotone, the sequence $\{w^k\}$ generated by the algorithm converges strongly to the unique solution of (1).

When it comes to the convergence rate analysis of our method, in the same spirit of [2, 21], we focus on the lower-level problem in (1). But, as opposed to the study in [2, 21] where only the particular instance of simple bilevel problems is considered, we cannot simply rely on some lower-level objective function values, but we must employ a different lower-level merit-function. In particular, we give convergence rate estimates in terms of the following lower-level merit-function:

$$V(y) \triangleq \|P_Y(y - F(y)) - y\|. \tag{6}$$

Note that V is the so-called natural residual map for the lower-level $\text{VI}(F, Y)$, see [11]. We observe that V is a continuous nonnegative measure such that $V(y) = 0$ if and only if $y \in \text{SOL}(F, Y)$. Furthermore, we recall that there exist classes of problems for which the value $V(y)$ also gives an actual upper-bound for the distance

between y and $\text{SOL}(F, Y)$. A noteworthy case is the one of affine VIs where Y is polyhedral and F is affine (see [11, Chapter 6 and, in particular, Proposition 6.3.3]).

The preliminary Lemma 1 is key for the forthcoming developments.

Lemma 1 *Let $\{y^k\}, \{w^k\}$ be the sequences generated by Algorithm 1. The following upper bound holds for the lower-level merit function V (see the definition (6)) at y^k for every k :*

$$V(y^k) \leq \lambda \|y^k - w^k\| + \frac{1}{\tau^k} \|G(y^k)\| + \frac{1}{\gamma^k} \|y^{k+1} - y^k\|. \tag{7}$$

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

$$\begin{aligned} V(y^k) &= \|P_Y(y^k - F(y^k)) - y^k\| \\ &\stackrel{(a)}{\leq} \frac{1}{\gamma^k} \|P_Y(y^k - \gamma^k F(y^k)) - y^k\| \\ &= \frac{1}{\gamma^k} \|P_Y(y^k - \gamma^k F(y^k)) - y^{k+1} + y^{k+1} - y^k\| \\ &\leq \frac{1}{\gamma^k} (\|y^{k+1} - y^k\| + \|P_Y(y^k - \gamma^k F(y^k)) \\ &\quad - P_Y(y^k - \gamma^k [F(y^k) + \frac{1}{\tau^k} G(y^k) + \lambda(y^k - w^k)])\|) \\ &\stackrel{(b)}{\leq} \frac{1}{\gamma^k} \|y^{k+1} - y^k\| + \frac{1}{\tau^k} \|G(y^k)\| + \lambda \|y^k - w^k\|, \end{aligned}$$

where (a) is due to [3, Lemma 2.3.1] recalling that $\gamma^k \in (0, 1]$, and (b) follows from the nonexpansive property of the projection mapping. □

Concerning the upper bound for V in (7), we observe that the last addendum in the right hand side of (7) accounts for the solution of the inner problem $\text{VI}(F + \frac{1}{\tau^k} G + \lambda(\bullet - w^k), Y)$ (consistently, it appears in the condition at step (S.3)). The second term is related to the upper-level operator G weighted by the Tikhonov parameter: as typical in Tikhonov approaches, being G bounded, this addendum eventually vanishes. Finally, if $k \in \mathcal{K}$, being $w^{k+1} = y^k$, the first term measures the distance between two successive outer iterations and, thus roughly speaking, a small such measure means that the outer iterates are barely changing. It is worth noticing that the upper bound in (7) is also useful to establish practical rules to properly set the algorithm parameters, see the following proposition. In this context, we underline that the convergence rate we establish is intended to give an upper bound on the number of iterations needed to drive the lower-level measure V under a prescribed tolerance δ .

Proposition 1 *Assume conditions (A1)–(A3) to hold. Given a precision $\delta \in (0, 1)$, set $\gamma^k = \eta$ for every k , $\lambda = \frac{\delta}{3D}$, where $D \triangleq \max_{v,y \in Y} \|v - y\|$ is the diameter of Y , $\eta = \frac{\lambda}{L_\phi^2}$, $\bar{\tau}^i = \max\{1, i\}$, and $\bar{\varepsilon}^i = \frac{\delta \eta}{(\bar{\tau}^i)^2}$. Then, the lower-level merit function V (see the definition (6)) is driven under δ and the condition in step (S.3) is satisfied in at most*

$$\sigma \triangleq \left\lceil \frac{2 \ln \frac{3(DL_\Phi)^2 \left\lceil \frac{3H}{\delta} \right\rceil^2}{\delta^2}}{\ln \frac{9(DL_\Phi)^2}{9(DL_\Phi)^2 - \delta^2}} \right\rceil \left\lceil \frac{3H}{\delta} \right\rceil$$

iterations, where $H \triangleq \max_{y \in Y} \|G(y)\|$.

Proof Our goal is to show that, in at most σ iterations, each addendum in the right-hand side of (7) is smaller than $\frac{1}{3}\delta$.

We observe that the first addendum $\lambda\|y^k - w^k\|$ is smaller than $\frac{1}{3}\delta$ in view of the value of λ and the definition of D .

In order to have the second term in (7) $\frac{1}{\tau^k}\|G(y^k)\|$ smaller than $\frac{1}{3}\delta$, we need $\tau^k \geq \frac{3H}{\delta}$, and hence at most

$$\bar{i} \triangleq \left\lceil \frac{3H}{\delta} \right\rceil \tag{8}$$

outer iterations.

We now focus on the remaining addendum in (7), $\frac{1}{\eta}\|y^{k+1} - y^k\|$. Preliminarily, we observe that the following chain of relations hold for every j :

$$\begin{aligned} \|y^{k+j+1} - y^{k+j}\|^2 &= \|P_Y(y^{k+j} - \eta\Phi_{\tau^k}(y^{k+j}; w^k)) \\ &\quad - P_Y(y^{k+j-1} - \eta\Phi_{\tau^k}(y^{k+j-1}; w^k))\|^2 \\ &\stackrel{(a)}{\leq} \|y^{k+j} - \eta\Phi_{\tau^k}(y^{k+j}; w^k) - y^{k+j-1} + \eta\Phi_{\tau^k}(y^{k+j-1}; w^k)\|^2 \\ &= \|y^{k+j} - y^{k+j-1}\|^2 \\ &\quad + \eta^2\|(\Phi_{\tau^k}(y^{k+j}; w^k) - \Phi_{\tau^k}(y^{k+j-1}; w^k))\|^2 \\ &\quad - 2\eta(y^{k+j} - y^{k+j-1})^T(\Phi_{\tau^k}(y^{k+j}; w^k) - \Phi_{\tau^k}(y^{k+j-1}; w^k)) \\ &\stackrel{(b)}{\leq} (1 + \eta^2L_\Phi^2 - 2\eta\lambda)\|y^{k+j} - y^{k+j-1}\|^2 \\ &\stackrel{(c)}{\leq} (1 - \eta\lambda)\|y^{k+j} - y^{k+j-1}\|^2, \end{aligned}$$

where (a) follows from the nonexpansive property of the projection mapping, while (b) is due to the strong monotonicity and Lipschitz continuity of $\Phi_{\tau^k}(\bullet; w^k)$, and (c) holds since $\eta \leq \frac{\lambda}{L_\Phi^2}$. In view of (8), we consider only the case $i \leq \bar{i}$. Starting from any iterate k , we count the maximum number of iterations j that are required in order to obtain $\|y^{k+j+1} - y^{k+j}\| \leq \bar{\epsilon}^i \leq \epsilon^k$. Suppose that $\|y^{k+j+1} - y^{k+j}\| > \bar{\epsilon}^i$, for every $j = 1, \dots, \bar{j}$.

We obtain

$$D(1 - \eta\lambda)^{\bar{j}/2} \geq \|y^{k+1} - y^k\|(1 - \eta\lambda)^{\bar{j}/2} \geq \|y^{k+\bar{j}+1} - y^{k+\bar{j}}\| > \bar{\epsilon}^i,$$

where the second inequality is due to the above chain of relations. In turn,

$$\bar{j} < 2 \log_{\left(\frac{1}{1-\eta\lambda}\right)} \frac{D}{\bar{\epsilon}^i} = \frac{2 \ln \frac{D}{\bar{\epsilon}^i}}{\ln \left(\frac{1}{1-\eta\lambda}\right)}. \tag{9}$$

Noting that

$$\bar{\epsilon} \bar{\tau} = \frac{\delta \eta}{(\bar{\tau})^2} = \frac{\delta \frac{\lambda}{L_\Phi^2}}{\bar{\tau}^2} = \frac{\delta^2}{3DL_\Phi^2 \left[\frac{3H}{\delta} \right]^2}, \frac{1}{1 - \eta\lambda} = \frac{1}{1 - \frac{\delta^2}{9(DL_\Phi)^2}} = \frac{9(DL_\Phi)^2}{9(DL_\Phi)^2 - \delta^2},$$

we have

$$\rho \triangleq \left[\frac{2 \ln \frac{3(DL_\Phi)^2 \left[\frac{3H}{\delta} \right]^2}{\delta^2}}{\ln \frac{9(DL_\Phi)^2}{9(DL_\Phi)^2 - \delta^2}} \right] > \bar{J}.$$

Therefore, in at most ρ iterations, the condition in step (S.3) is satisfied and an “outer” iteration is performed, i.e. i is increased. Finally, the claim follows observing that each of the three addenda in the right-hand side of (7) is less or equal than $\delta/3$ in at most $\sigma = \rho \bar{\tau}$ iterations. □

We remark that all the constants appearing in the expression of σ are problem-related and do not depend on δ : thus, no possibly unknown constants and no hidden dependencies on δ are present. Also, reasoning asymptotically, thus considering $\delta \downarrow 0$, we obtain the following rate.

Corollary 1 *Under the same assumptions as in Proposition 1, the maximum number of iterations σ to make the lower-level merit function V less than δ and the condition in step (S.3) satisfied is $\mathcal{O}(\delta^{-3} \ln \delta^{-1})$.*

Proof Recalling that $\sigma = \rho \bar{\tau}$, $\bar{\tau}$ is easily seen to be of order $\mathcal{O}(\delta^{-1})$. We observe that the numerator in the expression of ρ is $\mathcal{O}(\ln \delta^{-1})$, while as for the reciprocal of the denominator, noticing that $\ln \left(\frac{9(DL_\Phi)^2}{9(DL_\Phi)^2 - \delta^2} \right) / \frac{\delta^2}{9(DL_\Phi)^2} \xrightarrow{\delta \rightarrow 0} 1$, we get $\mathcal{O}(\delta^{-2})$. □

5 The simple bilevel instance

A simpler instance of problem (1) worth considering is the case in which G and F are the gradients of convex functions g and f , respectively (see [6] for some recent results concerning simple bilevel optimization). Therefore, problem (1) reduces to the simple bilevel problem

$$\text{minimize}_y g(y) \quad \text{s.t.} \quad y \in \arg \min_w \{f(w) \mid w \in Y\}. \tag{10}$$

In this particular framework, assumptions (A1) and (A2) read as

(A1') g is a convex function on Y , whose gradient is Lipschitz continuous with constant L_G ;

(A2') f is a convex function on Y , whose gradient is Lipschitz continuous with constant L_F .

Differently from the general nested VI framework where convergence can be exclusively obtained with a strictly positive lower bound for the step size η and a strictly positive proximal parameter λ , in the simple bilevel context we have more room to maneuver. Specifically, in the latter case, even setting η and λ equal to zero, convergence and enhanced complexity bounds are achieved. Moreover, as for the accuracy parameter ϵ , one can take it sufficiently large so that the condition in step (S.3) is always fulfilled: in particular we set $\bar{\epsilon}^i \geq D$ for every i . We observe that with the choices $\eta = 0$, $\lambda = 0$ and $\bar{\epsilon}^i \geq D$, Algorithm 1 boils down to a modified version of Solodov's method whose convergence is proved in [27] under assumptions (A1'), (A2') and (A3), and resorting to a line search strategy for the step size γ^k updating rule. Here, we propose two different step size approaches to tailor properly the value of γ^k : fixed and diminishing step size rules.

Theorem 2 *Assuming conditions (A1'), (A2') and (A3) to hold, set*

$$\lambda = 0, \eta = 0, \frac{1}{\bar{\tau}^i} \downarrow 0 \text{ and } \sum_{i=0}^{\infty} \frac{1}{\bar{\tau}^i} = \infty, \bar{\epsilon}^i \geq D \triangleq \max_{v,y \in Y} \|v - y\|,$$

and choose a step size γ^k such that one the following rules hold:

- (i) $\gamma^k = \bar{\gamma} \in \left(0, \frac{2}{L_\Phi}\right)$, for all k , where L_Φ is defined at the beginning of Sect. 4;
- (ii) $\gamma^k \downarrow 0$ with $\sum_{k=0}^{\infty} \gamma^k = \infty$.

Then, each limit point of the sequence $\{w^k\}$ generated by Algorithm 1 is a solution of problem (10).

Proof In order to prove convergence, one can rely on [27, Theorem 3.2]: it suffices to show that the line search condition (8) in Solodov's algorithm [27] with some $\theta \in (0, 1)$ is eventually satisfied. For this to be proven, recall that the characteristic property of the projection

$$\left(y^k - \gamma^k \left[F(y^k) + \frac{1}{\tau^k} G(y^k)\right] - y^{k+1}\right)^\top (y^k - y^{k+1}) \leq 0$$

implies

$$\|y^k - y^{k+1}\|^2 \leq \gamma^k \left[F(y^k) + \frac{1}{\tau^k} G(y^k)\right]^\top (y^k - y^{k+1}).$$

In turn, by the descent lemma, we have

$$f(y^{k+1}) + \frac{1}{\tau^k}g(y^{k+1}) - f(y^k) - \frac{1}{\tau^k}g(y^k) \leq (F(y^k) + \frac{1}{\tau^k}G(y^k))^T(y^{k+1} - y^k) + \frac{L_\Phi}{2} \|y^{k+1} - y^k\|^2 \leq \left(1 - \frac{\gamma^k L_\Phi}{2}\right)(F(y^k) + \frac{1}{\tau^k}G(y^k))^T(y^{k+1} - y^k).$$

In case (i), having a fixed step size $\gamma^k = \bar{\gamma}$, we obtain $0 < (1 - \bar{\gamma}L_\Phi/2) < 1$. Therefore, observing that $(F(y^k) + \frac{1}{\tau^k}G(y^k))^T(y^{k+1} - y^k) < 0$, by setting $\theta \in (0, 1 - \bar{\gamma}L_\Phi/2)$, the line search condition (8) in Solodov’s algorithm [27] is satisfied for every k .

In case (ii), having a diminishing step size γ^k , there exists \bar{k} such that $\theta < (1 - \gamma^{\bar{k}}L_\Phi/2)$, and, recalling that $(F(y^k) + \frac{1}{\tau^k}G(y^k))^T(y^{k+1} - y^k) < 0$, the line search condition (8) in Solodov’s algorithm [27] is satisfied for every $k \geq \bar{k}$.

The (subsequential) convergence of the sequence generated by Algorithm 1 to a solution of problem (10) follows then by [27, Theorem 3.2]. □

Also in this simpler case, if problem (10) is uniquely solvable, such as when e.g. g is strictly convex, the sequence $\{w^k\}$ generated by the algorithm converges strongly to the unique solution of (10).

In the simple bilevel case one can obtain an improved complexity bound, passing from $\mathcal{O}(\delta^{-3} \ln \delta^{-1})$ of the general case to $\mathcal{O}(\delta^{-2})$.

Proposition 2 *Assume conditions (A1’), (A2’) and (A3) to hold. Given a precision $\delta \in (0, 1)$, set $\lambda = 0$, $\eta = 0$, $\gamma^k = \frac{1}{L_\Phi}$ for every k , $\bar{\tau}^i = \max\{1, i\}$ and $\bar{\epsilon}^i = D \triangleq \max_{v,y \in Y} \|v - y\|$ for every i . Then, the lower-level merit function V (see the definition (6)) is driven under δ in at most*

$$\bar{\sigma} \triangleq \max \left\{ \left\lceil \frac{8L_\Phi A_0}{\delta^2} \right\rceil, \left\lceil \frac{2H}{\delta} \right\rceil \right\} \tag{11}$$

iterations, where $A_0 \triangleq f(y^0) - \min_{y \in Y} f(y) + g(y^0) - \min_{y \in Y} g(y)$ and $H \triangleq \max_{y \in Y} \|G(y)\|$.

Proof Our goal is to show that, in at most $\bar{\sigma}$ iterations, both the second and the third terms in the right-hand side of (7) are smaller than $\frac{1}{2}\delta$. Notice that the first addendum in the right-hand side of (7) is null because of $\lambda = 0$.

First of all, reasoning as in the proof of Proposition 1, we know that if

$$k \geq \left\lceil \frac{2H}{\delta} \right\rceil$$

then the second addendum in the right-hand side of (7) is smaller than $\frac{1}{2}\delta$. Now, we count how many iterations are required in order to obtain $\frac{1}{\gamma^k} \|y^{k+1} - y^k\| \leq \frac{\delta}{2}$. So suppose that $\|y^{k+1} - y^k\| > \frac{\delta}{2L_\Phi}$. Thanks to the descent lemma and the characteristic property of the projection, we have

$$\begin{aligned}
 & f(y^{k+1}) + \frac{1}{\tau^k} [g(y^{k+1}) - \min_{y \in Y} g(y)] - f(y^k) - \frac{1}{\tau^k} [g(y^k) - \min_{y \in Y} g(y)] = f(y^{k+1}) \\
 & + \frac{1}{\tau^k} g(y^{k+1}) - f(y^k) - \frac{1}{\tau^k} g(y^k) \leq -\left(\frac{1}{\tau^k} - \frac{L_\Phi}{2}\right) \|y^{k+1} - y^k\|^2 \\
 & \leq -\frac{L_\Phi}{2} \|y^{k+1} - y^k\|^2,
 \end{aligned}$$

and, in turn,

$$\begin{aligned}
 \bar{k} \frac{\delta^2}{4L_\Phi^2} & < \sum_{k=0}^{\bar{k}-1} \|y^{k+1} - y^k\|^2 \\
 & \leq \frac{2}{L_\Phi} \left[f(y^0) + \frac{1}{\tau^0} (g(y^0) - \min_{y \in Y} g(y)) - f(y^{\bar{k}}) - \frac{1}{\tau^{\bar{k}}} (g(y^{\bar{k}}) - \min_{y \in Y} g(y)) \right] \\
 & \leq \frac{2}{L_\Phi} \left[f(y^0) - f(y^{\bar{k}}) + \frac{1}{\tau^0} (g(y^0) - \min_{y \in Y} g(y)) \right] \leq \frac{2}{L_\Phi} \Delta_0.
 \end{aligned}$$

Therefore, whenever

$$k \geq \left\lceil \frac{8L_\Phi \Delta_0}{\delta^2} \right\rceil$$

the third addendum in the right-hand side of (7) is smaller than $\frac{1}{2} \delta$.

Finally, $k \geq \bar{\sigma}$ gives the desired result: $V(y^k) \leq \delta$. □

The convergence rate in (11) complements the results in [2, 21]. Observing that no possibly unknown constants and no hidden dependencies on δ are present in (11), and also in view of the following straightforward Corollary 2, we remark that our convergence rate is in line with the one established in [2].

Corollary 2 *Under the same assumptions as in Proposition 2, the maximum number of iterations $\bar{\sigma}$ to make the lower-level merit function V less than δ is $\mathcal{O}(\delta^{-2})$.*

6 Conclusions

In this work, we provide an explicit, practical version of general Tikhonov-like schemes for the solution of nested variational inequalities, that is the problem of a variational inequality whose feasible set is given by the solution set of another variational inequality. Apart from global convergence guarantees, we establish, as far as we are aware, the first convergence rate results in the context of general nested VIs. Moreover, our enhanced results for the simple bilevel instance, which are in line with the relevant literature, evidenciate the relevance of our analysis.

Numerical results illustrating the practical applicability of our approach are given in [15], where we formulate a multi-portfolio selection model involving several accounts, whose actions are coupled due to market impact via transaction costs, as a Nash game. The different merit functions that are considered yield a selection over the set of equilibria arising from that Nash game. The resulting hierarchical program with a lower-level Nash equilibrium problem is a special instance of problem (1) satisfying assumptions (A1)–(A3).

Finally, to complete the picture, we observe that a more general structure can be obtained when the lower-level VI depends parametrically on some upper-level variables. As this structure is no longer purely hierarchical, the analysis of the present paper does not apply. We leave the investigation of such more general models to future research.

References

1. Aussel, D., Sagratella, S.: Sufficient conditions to compute any solution of a quasivariational inequality via a variational inequality. *Math. Methods Oper. Res.* **85**(1), 3–18 (2017)
2. Beck, A., Sabach, S.: A first order method for finding minimal norm-like solutions of convex optimization problems. *Math. Program.* **147**(1–2), 25–46 (2014)
3. Bertsekas, D.: *Nonlinear Programming*. Athena Scientific, Belmont (1992)
4. Bigi, G., Lampariello, L., Sagratella, S.: Combining approximation and exact penalty in hierarchical programming. Technical report, Preprint (2020)
5. Dempe, S.: *Foundations of Bilevel Programming*. Springer, Berlin (2002)
6. Dempe, S., Dinh, N., Dutta, J., Pandit, T.: Simple bilevel programming and extensions. *Math. Program.*, To appear (2020)
7. Dreves, A., Facchinei, F., Kanzow, C., Sagratella, S.: On the solution of the KKT conditions of generalized Nash equilibrium problems. *SIAM J. Optim.* **21**(3), 1082–1108 (2011)
8. Facchinei, F., Kanzow, C., Karl, S., Sagratella, S.: The semismooth Newton method for the solution of quasi-variational inequalities. *Comput. Optim. Appl.* **62**(1), 85–109 (2015)
9. Facchinei, F., Kanzow, C., Sagratella, S.: Solving quasi-variational inequalities via their KKT conditions. *Math. Program.* **144**(1–2), 369–412 (2014)
10. Facchinei, F., Lampariello, L.: Partial penalization for the solution of generalized Nash equilibrium problems. *J. Global Optim.* **50**(1), 39–57 (2011)
11. Facchinei, F., Pang, J.S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York (2003)
12. Facchinei, F., Pang, J.S., Scutari, G., Lampariello, L.: VI-constrained hemivariational inequalities: distributed algorithms and power control in ad-hoc networks. *Math. Program.* **145**(1–2), 59–96 (2014)
13. Facchinei, F., Sagratella, S.: On the computation of all solutions of jointly convex generalized Nash equilibrium problems. *Optim. Lett.* **5**(3), 531–547 (2011)
14. Kalashnikov, V., Kalashnikova, N.: Solving two-level variational inequality. *J. Global Optim.* **8**(3), 289–294 (1996)
15. Lampariello, L., Neumann, C., Ricci, J., Sagratella, S., Stein, O.: Equilibrium selection for multi-portfolio optimization. Technical report, Optimization Online, Preprint ID 2020-01-7585 (2020)
16. Lampariello, L., Sagratella, S.: A bridge between bilevel programs and Nash games. *J. Optim. Theory Appl.* **174**(2), 613–635 (2017)
17. Lampariello, L., Sagratella, S.: Numerically tractable optimistic bilevel problems. *Comput. Optim. Appl.* **76**, 277–303 (2020)
18. Lampariello, L., Sagratella, S., Stein, O.: The standard pessimistic bilevel problem. *SIAM J. Optim.* **29**, 1634–1656 (2019)
19. Lu, X., Xu, H.K., Yin, X.: Hybrid methods for a class of monotone variational inequalities. *Nonlinear Anal. Theory Methods Appl.* **71**(3–4), 1032–1041 (2009)
20. Marino, G., Xu, H.K.: Explicit hierarchical fixed point approach to variational inequalities. *J. Optim. Theory Appl.* **149**(1), 61–78 (2011)
21. Sabach, S., Shtern, S.: A first order method for solving convex bilevel optimization problems. *SIAM J. Optim.* **27**(2), 640–660 (2017)
22. Sagratella, S.: Computing all solutions of Nash equilibrium problems with discrete strategy sets. *SIAM J. Optim.* **26**(4), 2190–2218 (2016)
23. Sagratella, S.: Algorithms for generalized potential games with mixed-integer variables. *Comput. Optim. Appl.* **68**(3), 689–717 (2017)

24. Sagratella, S.: Computing equilibria of Cournot oligopoly models with mixed-integer quantities. *Math. Methods Oper. Res.* **86**(3), 549–565 (2017)
25. Sagratella, S.: On generalized Nash equilibrium problems with linear coupling constraints and mixed-integer variables. *Optimization* **68**(1), 197–226 (2019)
26. Shehu, Y., Gibali, A., Sagratella, S.: Inertial projection-type methods for solving quasi-variational inequalities in real hilbert spaces. *J. Optim. Theory Appl.* **184**, 877–894 (2019)
27. Solodov, M.: An explicit descent method for bilevel convex optimization. *J. Convex Anal.* **14**(2), 227 (2007)
28. Yamada, I.: The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings. In: *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, vol. **8**, pp. 473–504 (2001)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.