ADDRESSING HIERARCHICAL JOINTLY-CONVEX GENERALIZED NASH EQUILIBRIUM PROBLEMS WITH NONSMOOTH PAYOFFS*

LORENZO LAMPARIELLO†, SIMONE SAGRATELLA‡, AND VALERIO GIUSEPPE SASSO‡

Abstract. We consider a Generalized Nash Equilibrium Problem whose joint feasible region is implicitly defined as the solution set of another Nash game. This structure arises e.g. in multi-portfolio selection contexts, whenever agents interact at different hierarchical levels. We consider nonsmooth terms in all players' objectives, to promote, for example, sparsity in the solution. Under standard assumptions, we show that the equilibrium problems we deal with have a nonempty solution set and turn out to be jointly convex. To compute variational equilibria, we devise different first-order projection Tikhonov-like methods whose convergence properties are studied. We provide complexity bounds and we equip our analysis with numerical tests using real-world financial datasets.

Key words. Generalized Nash Equilibrium Problems, Hierarchical Programming, Generalized Variational Inequality, Numerical Methods, Complexity Bounds

MSC codes. 90C33, 90C25, 90C30, 49J53, 65K15, 65K10, 91A65

1. Introduction.

We address Generalized Nash Equilibrium Problems (GNEP) [6–8], where the shared feasible set is implicitly defined as the equilibrium set of a different Nash Equilibrium Problem (NEP). The resulting GNEP presents a hierarchical structure where the players of the GNEP are the upper-level agents, while the players of the NEP that defines the feasible set are the lower-level ones: the upper-level agents operate a selection among the equilibria of the NEP played by the lower-level agents. Nonsmooth convex terms in both the upper and the lower-level agents' objective functions are considered, in order to include, e.g., sparsity enhancing or exact penalty-like terms. Such hierarchical GNEP, while stemming from real-world applications such as multi-portfolio selection (see e.g. [16, 18] and Example 2 in section 7), to the best of our knowledge has not been explicitly addressed in its full generality yet.

Relying on standard assumptions for the upper and the lower-level agents' problems, the hierarchical GNEP turns out to be jointly convex [10] and with a nonempty equilibrium set (Proposition 3.5 and Proposition 3.8). Mimicking the smooth context, we identify, in our broader framework, variational solutions that can be computed by addressing a suitable (upper-level) Generalized Variational Inequality (GVI), whose feasible set is implicitly defined as the solution set of another (lower-level) GVI ([21] for the definition of a single-level GVI, and [7] where variational solutions of a single-level GNEP are identified in the smooth case). The resulting hierarchical GVI consists of a lower-level GVI reformulating the lower-level NEP, and of an upper-level GVI whose solution set is the set of variational equilibria of the upper-level GNEP.

Concerning hierarchical programs, two main approaches have been developed in the literature: alternating-like techniques [1, 19, 20, 23, 25] and Tikhonov methods [1, 4, 9, 12, 13, 15, 17, 24]. As far as we are aware, considering the level of generality we take into account, there are no methods in the literature for finding variational solutions of hierarchical GNEPs.

We compute variational equilibria of the hierarchical GNEP through the corre-

*Submitted to the editors DATE.

† Department of Business Studies, Roma Tre University, Rome, Italy (lorenzo.lampariello@uniroma3.it).

‡ University of Rome, Rome, Italy Department of Computer, Control and Management Engineering Antonio Ruberti, Sapienza (sagratella@diag.uniroma1.it, sasso@diag.uniroma1.it).

sponding hierarchical GVI described above via a projected gradient Tikhonov-like approach: we derive convergence properties and obtain complexity guarantees. More in detail, we iteratively address single-level GVI subproblems, where the Tikhonov parameter is used to suitably weight the lower and the upper-level GVI operators. We show that using a projected gradient method with a constant Tikhonov parameter, the sequence produced by the algorithm converges to a fixed distance from every solution of the single-level GVI subproblem (Theorem 4.5). As a consequence, either the sequence admits a single limit vector, which turns out to be a solution of the GVI subproblem, or it orbits around the GVI subproblem's solution set. In the latter case, the projected gradient method fails to converge to solutions of the GVI subproblem, and, in the same spirit of [3], we rely on an averaging step to reach the solution set of the GVI subproblem (Theorem 4.7). Notice that, solving the GVI subproblem for positive fixed values of the Tikhonov parameter only corresponds to solving inexactly the hierarchical GNEP. The inexactness in computing variational solutions of the hierarchical GNEP is directly linked to the value of the Tikhonov parameter (Proposition 6.3). Unfortunately, if the Tikhonov parameter is fixed to zero, the solution set of the GVI subproblem corresponds only to the feasible set of the hierarchical GNEP, completely ignoring the payoffs of the upper-level players. In order to compute variational solutions of the hierarchical GNEP, one cannot rely solely on solving the GVI subproblem for any fixed value of the Tikhonov parameter.

Introducing a suitable updating rule that establishes a link between the Tikhonov parameter and the stepsize sequences, and makes them vanish (Assumptions **D**) we prove convergence to a variational solution of the hierarchical GNEP (Theorem 5.2).

Relying on harmonic sequences for the Tikhonov parameter and the stepsize, we provide complexity bounds in terms of maximum number of iterations that the algorithm needs to meet a target accuracy. Specifically, we evaluate the complexity of computing solutions of the GVI subproblem for fixed values of the Tikhonov parameter, for both the standard projected gradient iterations and for the averaging ones. Moreover, we give complexity bounds, under Assumptions **D**, when computing variational solutions of the hierarchical GNEP. The results of our analysis suggest that solutions of the GVI subproblem for fixed values of the Tikhonov parameter can be computed quite efficiently (Table 1). In view of such theoretical insights, we present the Projected Average Single-loop Tikhonov Algorithm (PASTA) that gradually satisfies the requirements in Assumptions **D**. By means of PASTA, we first aim at efficiently approaching the solution set of the GVI subproblem for fixed values of the Tikhonov parameter and, only at a later stage, we seek to achieve convergence to variational solutions of the hierarchical GNEP. Our numerical experiments confirm that such approach works well in practice and results in a faster convergence compared to satisfying Assumptions **D** from the beginning (section 7).

In section 2 we present the hierarchical GNEP model, as well as the main assumptions of our framework, and, in section 3, we introduce the hierarchical GVI we rely on in order to compute variational solutions of the original problem. In section 4, we introduce the Tikhonov approach, and convergence results concerning the GVI subproblem for fixed values of the Tikhonov parameter, while in section 5 we introduce Assumptions **D** and analyze the resulting convergence properties to variational solutions of the hierarchical GNEP. In section 6, we collect the complexity bounds we achieve when considering harmonic sequences for the Tikhonov parameter and the stepsize. In section 7, we introduce PASTA and test it numerically, first addressing a toy example, and then solving a multi-portfolio selection problem, inspired by [16].

2. The hierarchical jointly-convex Generalized Nash Equilibrium model. We define a Generalized Nash Equilibrium Problem (GNEP) whose shared feasible region E is given implicitly by the equilibrium set of a lower-level Nash Equilibrium Problem (NEP). We first deal with the lower-level NEP, highlighting the conditions for its solution set to be nonempty, convex and compact (see Assumptions **A** and developments in section 3). Next, we provide assumptions concerning the upper-level hierarchical GNEP that ensure that make it a jointly-convex problem with nonempty solution set (see Assumptions **B** and developments in section 3).

2.1. The lower-level NEP. The lower-level NEP consists of the collection of N (parametric) optimization problems, each borne by player ν, with $\nu = 1, \ldots, N$, managing n_ν decision variables. We denote by y the vector formed by all the decision variables, and by $y^{-\nu}$ the vector composed by all the players' decision variables except those of player ν: $y \triangleq (y^1 \cdots y^N)^T \in \mathbb{R}^p$, $y^{-\nu} \triangleq (y^1 \cdots y^{\nu-1}, y^{\nu+1} \cdots y^N) \in \mathbb{R}^{p-n_\nu}$, where $p = \sum_{\nu=1}^N n_\nu$. To emphasize player ν's decision variables within y, we sometimes write $(y^\nu, y^{-\nu})$ instead of y. Note that this still stands for the vector y and that, in particular, the notation $(y^\nu, y^{-\nu})$ does not mean that the block components of y are reordered in such a way that y^ν becomes the first block. For each player at the lower level, the objective function is given by the sum of a smooth term $\theta^l_\nu : \mathbb{R}^p \to \mathbb{R}$ depending on variables y^ν as well as on the variables $y^{-\nu}$, and a nonsmooth term $\varphi^l_\nu : \mathbb{R}^{n_\nu} \to \mathbb{R}$ depending on variables y^ν only. Summarizing, the NEP we consider consists of the collection of player ν's parametric optimization problems

$$(\mathrm{P}^l_\nu) \qquad \text{minimize}_{y^\nu}\, \theta^l_\nu(y^\nu, y^{-\nu}) + \varphi^l_\nu(y^\nu) \quad \text{s.t.} \quad y^\nu \in Y_\nu,$$

where $Y_\nu \subseteq \mathbb{R}^{n_\nu}$.

Denoting $Y \triangleq Y_1 \times \cdots \times Y_N \subseteq \mathbb{R}^p$, the lower-level NEP is the following problem

$$(\mathrm{NEP}^l) \quad \text{find } y \in Y\colon \theta^l_\nu(y^\nu, y^{-\nu}) + \varphi^l_\nu(y^\nu) \le \theta^l_\nu(v^\nu, y^{-\nu}) + \varphi^l_\nu(v^\nu),\ \forall v^\nu \in Y_\nu,\ \nu = 1, \ldots, N.$$

Any $y \in Y$ satisfying (NEP^l) is an equilibrium, or a solution of the NEP. A point is therefore an equilibrium if for no player, given the other players' choices, the objective function can be decreased by unilaterally changing their decision variables to any other feasible point. Accordingly, we indicate with $E \triangleq \{y \in Y : \theta^l_\nu(y^\nu, y^{-\nu}) + \varphi^l_\nu(y^\nu) \le \theta^l_\nu(v^\nu, y^{-\nu}) + \varphi^l_\nu(v^\nu), \forall v^\nu \in Y_\nu,\ \nu = 1, \ldots, N\} \subseteq \mathbb{R}^p$ the (non-parametric) set of equilibria of the NEP.

Assumptions A

A1 Y_ν is nonempty, convex and compact, for every $\nu = 1, \ldots, N$;

A2 θ^l_ν is convex with respect to y^ν, for every $\nu = 1, \ldots, N$;

A3 $\left[\nabla_{y^\nu} \theta^l_\nu\right]_{\nu=1}^N$ is monotone on Y;

A4 φ^l_ν is convex and locally Lipchitz, for every $\nu = 1, \ldots, N$.

From assumption **A4**, one can immediately deduce that $\partial_{y^\nu}\varphi^l_\nu$ is locally bounded and outer-semicontinuous for every $\nu = 1, \ldots, N$, where the operator ∂_{y^ν} indicates the set of subgradients with respect to player ν's variables. Furthermore, $\partial_{y^\nu}\varphi^l_\nu$ is a compact and convex nonempty set. Such results can be traced back in [5, Proposition 2.1.2 a] and [5, Proposition 2.1.5 d]. We will show that E is nonempty convex and compact (see section 3).

2.2. The upper-level GNEP. Considering the upper-level hierarchical GNEP, overall, player μ, with $\mu = 1, \ldots, M$, controls the decision variables $x^\mu \in \mathbb{R}^{m_\mu}$, with $\sum_{\mu=1}^M m_\mu = p$, so as to solve the following optimization problem:

$$(\mathrm{P}^u_\mu) \qquad \text{minimize}_{x^\mu}\, \theta^u_\mu(x^\mu, x^{-\mu}) + \varphi^u_\mu(x^\mu) \quad \text{s.t.} \quad (x^\mu, x^{-\mu}) \in E,$$

where $\theta^u_\mu : \mathbb{R}^p \to \mathbb{R}$ is a smooth function depending on variables x^μ as well as on the variables $x^{-\mu}$, and $\varphi^u_\mu : \mathbb{R}^{m_\mu} \to \mathbb{R}$ is a nonsmooth term depending on variables x^μ only. Notice that this is not a simple NEP, but a GNEP, because each player's feasible region depends parametrically on the other players' variables. The variables x^μ belong therefore to the solution set of the lower-level NEP, we denote $x \triangleq (x^1 \cdots x^M) \in \mathbb{R}^p$, $x^{-\mu} \triangleq (x^1 \cdots x^{\mu-1}, x^{\mu+1} \cdots x^M) \in \mathbb{R}^{p-m_\mu}$. The way the lower-level variables are partitioned among the players $(y^1, \dots, y^N)$ is completely independent from the partition of the same variables among the players that happens at upper level $(x^1, \dots, x^M)$. For the sake of notational simplicity, and without loss of generality, we assume that $x = y$, meaning that the variables are ordered (but not partitioned) in the same way at both the levels. The upper-level GNEP is the following problem:

$$\text{(GNEP}^u\text{)} \quad \text{find } x \in E\text{:} \quad \theta^u_\mu(x^\mu, x^{-\mu}) + \varphi^u_\mu(x^\mu) \le \theta^u_\mu(w^\mu, x^{-\mu}) + \varphi^u_\mu(w^\mu), \qquad \forall w^\mu : (w^\mu, x^{-\mu}) \in E, \quad \mu = 1, \dots, M.$$

Assumptions B

B1 θ^u_μ is convex with respect to x^μ, for every $\mu = 1, \dots, M$;

B2 $\left[\nabla_{x^\mu} \theta^u_\mu\right]_{\mu=1}^M$ is monotone on Y;

B3 φ^u_μ is convex and locally Lipchitz, for every $\mu = 1, \dots, M$.

Similarly to the lower level, from assumption **B3** we can deduce that $\partial_{x^\mu} \varphi^u_\mu$ is locally bounded and outer-semicontinuous for every $\mu = 1, \dots, M$. Furthermore $\partial_{x^\mu} \varphi^u_\mu$ is a compact convex nonempty set. We will show that the set of equilibria of the hierarchical GNEP is nonempty (see section 3).

3. The Generalized Variational Inequality Formulation.

The finite-dimensional Generalized Variational Inequality (GVI) provides an analytical tool to address the described hierarchical GNEP. First we focus on reformulating the lower-level NEP as a GVI in order to prove that, under Assumptions **A**, its solution set E is nonempty, convex and compact. We also deal with the solution set of the (upper-level) hierarchical GNEP by showing that the GVI provides a tool to compute its variational equilibria, and we show this subset of equilibria to be nonempty, convex and compact.

3.1. Lower-level GVI formulation.

The lower-level NEP (NEP^l) turns out to be equivalent to the following GVI:

$$\text{(GVI}^l\text{)} \qquad \text{find } y \in Y : \quad \exists f_y \in F(y) : \quad f_y^T(v - y) \ge 0, \quad \forall v \in Y;$$

where $F(y) \triangleq \left[\partial_{y^\nu}\left(\theta^l_\nu(y) + \varphi^l_\nu(y^\nu)\right)\right]_{\nu=1}^N : \quad \mathbb{R}^p \rightrightarrows \mathbb{R}^p$.

Remark 3.1. In view of Assumptions **A**, $\theta^l_\nu(y^\nu, y^{-\nu}) + \varphi^l_\nu(y^\nu)$, for all ν, turns out to be also regular (see [22, Proposition 7.27]). This implies that we can write (see [22, Proposition 10.9])

$$F(y) = \left[\nabla_{y^\nu} \theta^l_\nu(y)\right]_{\nu=1}^N + \left[\partial_{y^\nu} \varphi^l_\nu(y^\nu)\right]_{\nu=1}^N \quad \text{for all } y \in Y. \tag{3.1}$$

Additionally, the operator F turns out to be outer-semicontinuous on Y, since it is the sum of a continuous term $\left[\nabla_{y^\nu} \theta^l_\nu\right]_{\nu=1}^N$ and an outer-semicontinuous one $\left[\partial_{y^\nu} \varphi^l_\nu\right]_{\nu=1}^N$.

In the next proposition, whose proof is given in Appendix A.1, we show that, under Assumptions **A**, (NEP^l) can be recast as (GVI^l), whose solution set is denoted by SOL(F, Y).

PROPOSITION 3.2. *Under assumptions **A1, A2, A4**, $E =$ SOL(F,Y).*

With the following results we list some properties of F and E.

PROPOSITION 3.3. *Under assumptions **A1, A2, A4**, SOL(F,Y), and then E, are nonempty and compact.*

Proof. To prove the nonemptiness of E, we rely on [11, Theorem 3.1], where nonemptiness, compactness and convexity of Y, outer-semicontinuity, convex valuedness (on Y) of F are required for E to be nonempty. These conditions are satisfied under **A1, A2, A4**. E is bounded, since Y is compact.

Regarding closedness of E, the proof is obtained by contradiction. Thanks to Proposition 3.2, if E is not closed, there exists a sequence $\{y_k\} \subset E$ such that

$$\exists f_{y_k} \in F(y_k) : f_{y_k}^T(v - y_k) \geq 0, \quad \forall v \in Y, \tag{3.2}$$

and such that $y_k \to \overline{y} \notin E$, i.e.

$$\forall f_{\overline{y}} \in F(\overline{y}), \quad \exists \overline{v} \in Y : \quad f_{\overline{y}}^T(\overline{v} - \overline{y}) < 0. \tag{3.3}$$

Since F is locally bounded over the bounded set Y, an infinite subset of indices $\mathcal{K}$ exists such that $\lim_{k\in\mathcal{K}} f_{y_k} = \overline{f}$. Moreover, since F is outer-semicontinuous, $\overline{f} \in F(\overline{y})$, taking the subsequential limit on both sides of (3.2), we get $0 \leq \lim_{k\in\mathcal{K}} f_{y_k}^T(v - y_k) = \overline{f}^T(v - \overline{y})$, for all $v \in Y$, which contradicts (3.3). ◻

PROPOSITION 3.4. *Under Assumptions **A**, F is maximal monotone (see Definition* A.1 *in Appendix* A.2*) and SOL(F,Y), and then E, are convex sets.*

Proof. First note that, since under **A3** the operator $\left[\nabla_{y^\nu}\theta_\nu^l\right]_{\nu=1}^N$ is continuous and monotone, it turns out to be also maximal monotone (see [22, Proposition 12.7]). On the other hand, under assumption **A4**, the operator φ_ν^l is continuous and convex, which implies that the point to set map defined by $\partial_{y^\nu}\varphi_\nu^l$ is maximal monotone (see [22, Proposition 12.17]). By Lemma A.2 in Appendix A.2 we therefore have that $\left[\partial\varphi_\nu^l\right]_{\nu=1}^N$ is maximal monotone. Since the sum of maximal monotone operators is maximal monotone under mild conditions (as long as rint (dom$\nabla_{y^\nu}\theta_\nu^l$) $\cap$ rint(dom $\partial_{y^\nu}\varphi_\nu^l$) $\neq \emptyset$) (see [22, Proposition 12.44]), we can deduce that the mapping F is maximal monotone. Recalling [11, Theorem 4.4], the convexity of SOL(F,Y) and E follows, since Y is nonempty and convex, and F is maximally monotone. ◻

PROPOSITION 3.5. *Under Assumptions **A** and **B**,* (GNEP^u) *is jointly-convex.*

Proof. By Proposition 3.4, E is convex, and the thesis holds by Assumptions **B** because the upper-level agents' objectives are convex with respect to their private variables. ◻

3.2. Upper-level GVI formulation. The following GVI can be used to compute solutions of (GNEP^u):

$$(\text{GVI}^u) \quad \text{find } x \in \text{SOL}(F,Y) : \quad \exists g_x \in G(x) : \quad g_x^T(w - x) \geq 0, \quad \forall w \in \text{SOL}(F,Y),$$

where $G(x) \triangleq \left[\partial_{x^\mu}\left(\theta_\mu^u(x) + \varphi_\mu^u(x^\mu)\right)\right]_{\mu=1}^M : \quad \mathbb{R}^p \rightrightarrows \mathbb{R}^p$.

Remark 3.6. Similarly to the lower level, under Assumptions **B**, we have $G(x) = \left[\nabla_{x^\mu}\theta^u_\mu(x)\right]_{\mu=1}^M + \left[\partial_{x^\mu}\varphi^u_\mu(x^\mu)\right]_{\mu=1}^M$, for all $x \in Y$. The operator G is also outer-semicontinuous, by the same reasoning presented in Remark 3.1 for operator F.

With the next result, whose proof is reported in Appendix A.3, under Assumptions **B**, we show that the solution set of (GVI^u), that we denote by $\mathrm{SOL}(G, \mathrm{SOL}(F, Y))$, is included in the solution set of (GNEP^u).

Proposition 3.7. *Under assumptions* ***B1, B3****, every* $x \in \mathrm{SOL}(G, \mathrm{SOL}(F, Y))$ *is a solution of* (GNEP^u)*.*

In particular, we say that the solutions belonging to $\mathrm{SOL}(G, \mathrm{SOL}(F, Y))$ are the variational equilibria of (GNEP^u), mimicking the classical definition in the smooth case. Computing the variational equilibria of a GNEP is relvant for many applications (see e.g. [10], and the references therein). With the following propositions, whose proofs are reported in Appendix A.4 and Appendix A.5 respectively, we establish some properties concerning G and the set of variational equilibria of (GNEP^u).

Proposition 3.8. *Under Assumpions* ***A****,* ***B1****,* ***B3****,* $\mathrm{SOL}(G, \mathrm{SOL}(F, Y))$ *is nonempty and compact and then also the set of equilibria of* (GNEP^u) *is nonempty.*

Proposition 3.9. *Under Assumptions* ***A*** *and* ***B****,* G *is maximal monotone (see Definition* A.1 *in Appendix* A.2*) and* $\mathrm{SOL}(G, \mathrm{SOL}(F, Y))$ *is convex.*

Therefore, we can say that (GNEP^u) is a jointly-convex problem whose solutions can be computed by solving (GVI^u) with a nonempty, convex and compact solution set.

4. On the solution of the Tikhonov single-level GVI subproblem.

By Proposition 3.7 and Proposition 3.8, we can compute variational solutions to (GNEP^u) by addressing (GVI^u). In particular, we employ Tikhonov-like regularization techniques, where the lower-level GVI mapping F is penalized at the same level of the upper-level one G:

$$H_\eta(y) \triangleq F(y) + \eta G(y),$$

where $\eta \geq 0$ is the Tikhonov parameter. The parameter η is used to weight the lower and the upper-level GVI operators F and G. The corresponding single-level GVI subproblem is as follows:

$$\text{(4.1)} \qquad \text{find} \quad y \in Y: \quad \exists h^\eta_y \in H_\eta(y): \quad h^{\eta T}_y (v - y) \geq 0, \quad \forall v \in Y.$$

We denote by $\mathrm{SOL}(H_\eta, Y)$ the solution set of (4.1). We also introduce the Minty counterpart for (4.1), that is instrumental for the forthcoming developments:

$$\text{(4.2)} \qquad \text{find} \quad y \in Y: \quad h^{\eta T}_v (v - y) \geq 0, \quad \forall v \in Y, \quad \forall h^\eta_v \in H_\eta(v).$$

Notice that, as we clarify in the forthcoming developments, solving the GVI subproblem (4.1) and (4.2) corresponds to solving inexactly (GVI^l) and (GVI^u) (see Proposition A.4 and Proposition 6.3).

Proposition 4.1. *Under Assumptions* ***A*** *and* ***B****, for every* $\eta \geq 0$*,* H_η *is maximal monotone, outer-semicontinuous and locally bounded on* Y*. Moreover,* $\mathrm{SOL}(H_\eta, Y)$ *is convex, compact-valued and nonempty.*

Proof. The claim is a consequence of Proposition 3.4 and Proposition 3.9. ∎

The solution sets of (4.1) and the one of the Minty problem (4.2) turn out to coincide, according to the following results whose proofs are given in Appendix A.6 and Appendix A.7, respectively.

THEOREM 4.2. *Under assumptions **A1**, **A3**, **A4**, **B2** and **B3** if a vector $y \in Y$ is a solution of* (4.1)*, then it is a solution of* (4.2)*.*

THEOREM 4.3. *Under assumptions **A1**, **A4**, and **B3**, if a vector $y \in Y$ is a solution of* (4.2)*, it is a solution of* (4.1)*.*

In the rest of the paper, Assumptions **A**, **B** will always be assumed to hold. We define the following finite quantities:

$$\overline{F} \triangleq \max_{y \in Y} \max_{f_y \in F(y)} \|f_y\| \quad \overline{G} \triangleq \max_{y \in Y} \max_{g_y \in G(y)} \|g_y\| \quad D \triangleq \max_{x,v \in Y} \|x - v\|.$$

We remark that the boundedness of Y (see assumption **A1**) is a sufficient condition for $\overline{F}, \overline{G}$ and D to be finite.

To compute a point in $\text{SOL}(H_\eta, Y)$ with $\eta \geq 0$, we investigate different first-order methods. Here we focus only on the solution of the GVI subproblem (4.1), while we provide a convergence analysis for (GNEP^u) in section 5.

We first analyze the properties of the following projected gradient-like procedure when specified to address problem (4.1).

Given $\{\gamma_k\}$, $\{\eta_k\}$, $y_1 \in Y$, for every $k = 1, \ldots$ compute:

$$\begin{array}{l} f_{y_k} \in F(y_k), \quad g_{y_k} \in G(y_k), \quad h_{y_k}^{\eta_k} \leftarrow f_{y_k} + \eta_k g_{y_k} \\ y_{k+1} \leftarrow P_Y(y_k - \gamma_k h_{y_k}^{\eta_k}), \end{array} \tag{4.3}$$

where P_Y denotes the Euclidean projection operator on the convex set Y.

The sequence $\{y_k\}$ produced by Algorithm (4.3) presents strong properties under mild assumptions regarding Tikhonov parameters $\{\eta_k\}$ and stepsizes $\{\gamma_k\}$.

Assumptions C

C1 $\{\gamma_k\}$ is non-increasing, $\gamma_k > 0$ for all k, $\gamma_k \to 0$ and $\{\gamma_k\} \notin \ell_1$, that is, $\sum_{k=1}^\infty \gamma_k = \infty$;

C2 $\{\eta_k\}$ is non-increasing, $\eta_k > 0$ for all k and $\eta_k \to \eta \geq 0$.

The non-summability of $\{\gamma_k\}$ is a condition that, roughly speaking, makes stepsizes vanishing not too fast. Sufficient conditions ensuring **C1** can be readily obtained, see e.g. the example given in (6.1).

When H_η is just maximal monotone, $\{y_k\}$ may not converge to $\text{SOL}(H_\eta, Y)$, see e.g. [15]. However, we show in Theorem 4.5 that the distance of y_k from any $u \in \text{SOL}(H_\eta, Y)$ converges to a constant value, depending on u. In the following theorem, we prove the existence of some bounds which we rely on to prove the claim in Theorem 4.5.

THEOREM 4.4. *Consider the sequences $\{\gamma_k\}$, $\{\eta_k\}$, $\{y_k\}$ and $\{h_{y_k}^{\eta_k}\}$ defined in Algorithm* (4.3) *and assume Assumptions **C** to hold. Let*

$$\Psi_1^k \triangleq \sum_{j=k}^\infty \gamma_j^2, \qquad \Psi_2^k \triangleq \sum_{j=k}^\infty \gamma_j(\eta_j - \eta), \quad \forall k \geq 1.$$

For each $u \in \text{SOL}(H_\eta, Y)$, and for every $k \geq 1$, we have:

$$\limsup_{\Delta \to \infty} \|y_{k+\Delta} - u\|^2 - \|y_k - u\|^2 \leq 2\Lambda_1 \Psi_1^k + 2\Lambda_2 \Psi_2^k, \tag{4.4}$$

with $\Lambda_1 \triangleq (\overline{F}^2 + \eta_1^2 \overline{G}^2)$ and $\Lambda_2 \triangleq \overline{G} D$.

Proof. Due to the non expansiveness of the projection operator, for every $j \geq 1$ we have:

$$\begin{aligned}
\|y_{j+1}-u\|^2 &= \|P_Y(y_j-\gamma_j h_{y_j}^{\eta_j})-P_Y(u)\|^2 \leq \|y_j-\gamma_j h_{y_j}^{\eta_j}-u\|^2 \\
&= \|y_j-u\|^2+\|\gamma_j h_{y_j}^{\eta_j}\|^2+2\gamma_j h_{y_j}^{\eta_j T}(u-y_j)+2\gamma_j\eta g_{y_j}^T(u-y_j) \\
&\quad -2\gamma_j\eta g_{y_j}^T(u-y_j) \\
&= \|y_j-u\|^2+\|\gamma_j h_{y_j}^{\eta_j}\|^2+2\gamma_j h_{y_j}^{\eta T}(u-y_j)+2\gamma_j(\eta_j-\eta)g_{y_j}^T(u-y_j) \\
&\leq \|y_j-u\|^2+2\gamma_j^2\left(\overline{F}^2+\eta_1^2\overline{G}^2\right)+2\gamma_j(\eta_j-\eta)\overline{G}D,
\end{aligned}$$

where the latter inequality holds because $u \in \text{SOL}(H_\eta, Y)$, and due to the following relation, since $\{\eta_j\}$ is non-increasing:

$$\|\gamma_j(f_{y_j}+\eta_j g_{y_j})\|^2 \leq 2\gamma_j^2\left(\|f_{y_j}\|^2+\eta_j^2\|g_{y_j}\|^2\right) \leq 2\gamma_j^2\left(\overline{F}^2+\eta_1^2\overline{G}^2\right). \tag{4.5}$$

Summing j from k to $k+\Delta-1$ we find:

$$\sum_{j=k}^{k+\Delta-1}\|y_{j+1}-u\|^2-\sum_{j=k}^{k+\Delta-1}\|y_j-u\|^2 \leq 2\Lambda_1\sum_{j=k}^{k+\Delta-1}\gamma_j^2+2\Lambda_2\sum_{j=k}^{k+\Delta-1}\gamma_j(\eta_j-\eta)$$

which implies, due to the telescoping series property,

$$\|y_{k+\Delta}-u\|^2 \leq \|y_k-u\|^2+2\Lambda_1\sum_{j=k}^{k+\Delta-1}\gamma_j^2+2\Lambda_2\sum_{j=k}^{k+\Delta-1}\gamma_j(\eta_j-\eta).$$

Relation (4.4) is obtained by letting $\Delta \to \infty$. ∎

In Theorem 4.5 we list the main convergence properties of $\{y_k\}$.

THEOREM 4.5. *Consider the sequences $\{\gamma_k\}$, $\{\eta_k\}$, $\{y_k\}$ and $\{h_{y_k}^{\eta_k}\}$ defined in Algorithm (4.3) and assume Assumptions **C** to hold. The following statements hold:*

- ***a)*** *if $\{\gamma_k\} \in \ell^2$, that is, $\sum_{k=1}^\infty \gamma_k^2 < \infty$ and $\{\gamma_k(\eta_j-\eta)\} \in \ell^1$, given any $u \in \text{SOL}(H_\eta, Y)$, for some l_u depending on u, we have $\lim_{k\to\infty}\|y_k-u\|^2 = l_u$;*
- ***b)*** *if $y_k \to \overline{y}$, then, $\overline{y} \in \text{SOL}(H_\eta, Y)$;*
- ***c)*** $\|y_{k+1}-y_k\| \to 0$.

Proof. The proof of **a)** is obtained from relation (4.4) by observing that $\Psi_1^k \to 0$ and $\Psi_2^k \to 0$. The proof of **b)** is reported in Appendix A.9. As for **c)**: for all $v \in Y$ and $k \geq 1$ we have:

$$\|y_{k+1}-y_k\| = \left\|P_Y(y_k-\gamma_k h_{y_k}^{\eta_k})-P_Y(y_k)\right\| \leq \left\|y_k-\gamma_k h_{y_k}^{\eta_k}-y_k\right\| = \left\|\gamma_k h_{y_k}^{\eta_k}\right\| \to 0,$$

where the inequality is due to the non expansiveness of the projection operator, and the last term goes to zero because H_{η_k} is locally bounded over the compact set Y. ∎

Note that relaxing the assumption on the boundedness of Y, but requiring F and G to be bounded on it, one can still obtain convergence results by slightly modifying the line of reasoning in the results above and in the forthcoming developments.

Under Assumptions **C**, $\{y_k\}$ might orbit around $\text{SOL}(H_\eta, Y)$ thanks to Theorem 4.5 (a), (c), without reaching it eventually. On the other hand, if $\{y_k\}$ converges, then its limit point belongs to the solution set, see Theorem 4.5 (b). This cannot be guaranteed in general, but one might rely on some averaging techniques. Thus, given the sequences $\{\gamma_k\}$ and $\{y_k\}$ defined by Algorithm (4.3), we introduce a further averaging sequence $\{z_k\}$ such that, for $k \geq 1$,

$$z_k \leftarrow \frac{\sum_{j=1}^k \gamma_j y_j}{\sum_{j=1}^k \gamma_j}. \tag{4.6}$$

In Theorem 4.7 we show that $\{z_k\}$ converges to $\text{SOL}(H_\eta, Y)$. With the preliminary Theorem 4.6, we obtain some bounds that are then used to prove Theorem 4.7.

THEOREM 4.6. *Consider the sequences* $\{\gamma_k\}$, $\{\eta_k\}$, $\{y_k\}$, $\{g_{y_k}\}$ *and* $\{h_{y_k}^{\eta_k}\}$ *defined in Algorithm* (4.3) *and* $\{z_k\}$ *defined in* (4.6) *and assume Assumptions* **C** *to hold. Let*

$$\Xi_1^k \triangleq \frac{\sum_{j=1}^k \gamma_j^2}{\sum_{j=1}^k \gamma_j}, \qquad \Xi_2^k \triangleq \frac{\sum_{j=1}^k \gamma_j(\eta_j - \eta)}{\sum_{j=1}^k \gamma_j}, \qquad \Xi_3^k \triangleq \frac{1}{\sum_{j=1}^k \gamma_j}, \qquad k \geq 1.$$

For all $k \geq 1$ *we have:*

$$h_v^{\eta T}(v - z_k) \geq -\Lambda_1 \Xi_1^k - \Lambda_2 \Xi_2^k - \Lambda_3 \Xi_3^k, \quad \forall v \in Y, \quad \forall h_v^\eta \in H_\eta(v), \tag{4.7}$$

with Λ_1 *and* Λ_2 *defined in Theorem* 4.4 *and* $\Lambda_3 \triangleq D^2/2$.

Proof. For all $v \in Y$, $h_v^\eta \in H_\eta(v)$ and for every $j \geq 1$, following the same steps as the ones in the chain of relations at the beginning of the proof of Theorem 4.4,

$$\begin{aligned} \|y_{j+1} - v\|^2 &= \|y_j - v\|^2 + \|\gamma_j h_{y_j}^{\eta_j}\|^2 + 2\gamma_j h_{y_j}^{\eta T}(v - y_j) + 2\gamma_j(\eta_j - \eta) g_{y_j}^T(v - y_j) \\ &\leq \|y_j - v\|^2 + 2\gamma_j^2(\overline{F}^2 + \eta_1^2 \overline{G}^2) + 2\gamma_j h_v^{\eta T}(v - y_j) + 2\gamma_j(\eta_j - \eta)\overline{G}D, \end{aligned}$$

due to the monotonicity of H_η, as well as equation (4.5). Then,

$$-2\gamma_j h_v^{\eta T}(v - y_j) \leq \|y_j - v\|^2 - \|y_{j+1} - v\|^2 + 2\Lambda_1 \gamma_j^2 + 2\Lambda_2 \gamma_j(\eta_j - \eta).$$

Summing j from 1 to k, and dividing by $2\sum_{j=1}^k \gamma_j$, we get

$$\begin{aligned} -h_v^{\eta T}(v - z_k) &\leq \frac{\|y_1 - v\|^2}{2\sum_{j=1}^k \gamma_j} - \frac{\|y_{k+1} - v\|^2}{2\sum_{j=1}^k \gamma_j} + \Lambda_1 \frac{\sum_{j=1}^k \gamma_j^2}{\sum_{j=1}^k \gamma_j} + \Lambda_2 \frac{\sum_{j=1}^k \gamma_j(\eta_j - \eta)}{\sum_{j=1}^k \gamma_j} \\ &\leq \Lambda_1 \frac{\sum_{j=1}^k \gamma_j^2}{\sum_{j=1}^k \gamma_j} + \Lambda_2 \frac{\sum_{j=1}^k \gamma_j(\eta_j - \eta)}{\sum_{j=1}^k \gamma_j} + \frac{D^2}{2} \frac{1}{\sum_{j=1}^k \gamma_j}, \end{aligned}$$

and then (4.7) follows. ∎

THEOREM 4.7. *Consider the sequences* $\{\gamma_k\}$ *and* $\{y_k\}$ *defined in Algorithm* (4.3) *and* $\{z_k\}$ *defined in* (4.6) *and assume Assumptions* **C** *to hold. The limit point of* $\{z_k\}$ *belongs to* $\text{SOL}(H_\eta, Y)$.

Proof. The proof is obtained by observing that $\Xi_1^k, \Xi_2^k \to 0$ in view of Lemma A.3 where we take $b_k = \gamma_k$ and $a_k = \gamma_k$ as far as Ξ_1^k is concerned, while $a_k = \eta_k - \eta$ when considering Ξ_2^k, and $\Xi_3^k \to 0$ due to **C1**. Therefore, Theorem 4.6 yields $\liminf_{k\to\infty} h_v^{\eta T}(v - z_k) \geq 0$, for all $v \in Y$ and for all $h_v^\eta \in H_\eta(v)$. Hence all subsequential limits of $\{z_k\}$ are solutions to the Minty GVI subproblem, and thus, by Theorem 4.3 they belong to $\text{SOL}(H_\eta, Y)$.

In the sequel, we prove that $\{z_k\}$ has actually a single limit point. For every $u_1, u_2 \in \text{SOL}(H_\eta, Y)$, by convexity: $\frac{u_1 + u_2}{2} \in \text{SOL}(H_\eta, Y)$, see Proposition 4.1. Combining point **a)** in Theorem 4.5 and Lemma A.3 in Appendix A.8, we can say that

$\exists\, l_{\left(\frac{u_1+u_2}{2}\right)}, l_{u_1} \in \mathbb{R}$:

$$\frac{\sum_{j=1}^k \gamma_j \left\| y_j - \frac{u_1+u_2}{2} \right\|^2}{\sum_{j=1}^k \gamma_j} \xrightarrow[k\to\infty]{} l_{\left(\frac{u_1+u_2}{2}\right)}, \quad \frac{\sum_{j=1}^k \gamma_j \left\| y_j - u_1 \right\|^2}{\sum_{j=1}^k \gamma_j} \xrightarrow[k\to\infty]{} l_{u_1}.$$

For every $j \geq 1$ we have:

$$\left\| y_j - \frac{u_1+u_2}{2} \right\|^2 = \left\| y_j - u_1 + \frac{u_1-u_2}{2} \right\|^2 = \|y_j - u_1\|^2 + \left\| \frac{u_1-u_2}{2} \right\|^2 + (y_j - u_1)^T(u_1-u_2).$$

Multiplying both sides by γ_j, summing j from 1 to k, and then dividing by $\sum_{j=1}^k \gamma_j$, we get:

$$\frac{\sum_{j=1}^k \gamma_j \left\| y_j - \frac{u_1+u_2}{2} \right\|^2}{\sum_{j=1}^k \gamma_j} - \frac{\sum_{j=1}^k \gamma_j \left\| y_j - u_1 \right\|^2}{\sum_{j=1}^k \gamma_j} - \left\| \frac{u_1-u_2}{2} \right\|^2 = (z_k - u_1)^T(u_1-u_2). \tag{4.8}$$

Taking the limit on both sides, we get

$$l_{\left(\frac{u_1+u_2}{2}\right)} - l_{u_1} - \left\| \frac{u_1-u_2}{2} \right\|^2 = \lim_{k\to\infty} (z_k - u_1)^T(u_1-u_2).$$

Let us assume by contradiction that $\overline{z} \neq \widetilde{z}$ are two limit points of $\{z_k\}$. In the first part of the proof we have shown that $\overline{z}, \widetilde{z} \in \text{SOL}(H_\eta, Y)$. The last equation implies $(\overline{z} - \widetilde{z})^T(u_1-u_2) = (\overline{z} - u_1)^T(u_1-u_2) - (\widetilde{z} - u_1)^T(u_1-u_2) = 0$. Considering $u_1 = \overline{z}$ and $u_2 = \widetilde{z}$, we obtain $\|\overline{z} - \widetilde{z}\|^2 = 0$ that contradicts $\overline{z} \neq \widetilde{z}$. ◻

Under Assumptions **A**, **B** and **C**, the sequence produced by Algorithm (4.3) together with (4.6) converges to $\text{SOL}(H_\eta, Y)$. The points in $\text{SOL}(H_0, Y)$ correspond to the solutions of (GVI^l), therefore they are feasible for (GVI^u), and then they belong to E, but they are not guaranteed to be solutions to (GVI^u). On the other hand, if $\eta > 0$, the sequence produced by Algorithm (4.3) together with (4.6) converges to $\text{SOL}(H_\eta, Y)$, that corresponds to solving, depending on η, (GVI^l) and (GVI^u) inexactly (see Proposition A.4 and Proposition 6.3). Considering relation (6.3), one is not guaranteed to solve (GVI^l) exactly. Therefore, in order to solve the (GVI^u) exactly, and obtain equilibria of (GNEP^u), one cannot focus solely on computing points in $\text{SOL}(H_\eta, Y)$ for any η.

In the following section, we define additional requirements (Assumptions **D**) on $\{\gamma_k\}$ and $\{\eta_k\}$ that let the sequence produced by Algorithm (4.3) together with (4.6) compute points in $\text{SOL}(H_0, Y)$ and in $\text{SOL}(G, \text{SOL}(F, Y))$, and therefore equilibria of (GNEP^u). Note that differently from Assumptions **C**, the conditions in Assumptions **D** require the choices of $\{\gamma_k\}$ and $\{\eta_k\}$ to be related to each other.

5. On the solution of the upper-level GNEP.

We provide assumptions ensuring that the sequence produced by Algorithm (4.3) together with (4.6) converges to a solution of problem (GVI^u), which is also a solution for (GNEP^u) (see Proposition 3.7). We define the following bounds for the Minty versions of (GVI^l) and (GVI^u).

THEOREM 5.1. *Consider the sequences $\{\gamma_k\}$, $\{\eta_k\}$, $\{y_k\}$ and $\{h_{y_k}^{\eta_k}\}$ defined in Algorithm* (4.3) *and $\{z_k\}$ defined in* (4.6) *and assume Assumptions **C** to hold. Let $\eta = 0$ in assumption **C2**, and*

$$\Phi_1^k \triangleq \frac{\sum_{j=1}^k \gamma_j \dfrac{\gamma_j}{\eta_j}}{\sum_{j=1}^k \gamma_j}, \qquad \Phi_2^k \triangleq \frac{1}{\eta_k \sum_{j=1}^k \gamma_j}, \qquad k \geq 1.$$

For all $k \geq 1$ we have:

$$f_v^T(v - z_k) \geq -\Lambda_1 \Xi_1^k - \Lambda_2 \Xi_2^k - \Lambda_3 \Xi_3^k, \qquad \forall v \in Y, \quad \forall f_v \in F(v), \tag{5.1}$$

$$g_v^T(v - z_k) \geq -\Lambda_1 \Phi_1^k - \Lambda_3 \Phi_2^k, \qquad \forall v \in SOL(F, Y), \quad \forall g_v \in G(v), \tag{5.2}$$

with Λ_1, Λ_2, Λ_3, $\{\Xi_1^k\}$, $\{\Xi_2^k\}$ and $\{\Xi_3^k\}$ defined in Theorem 4.4 and Theorem 4.6.

Proof. Relation (5.1) can be obtained by considering Theorem 4.6 with $\eta = 0$.

To prove (5.2), for every $v \in \mathrm{SOL}(F, Y)$, $f_v \in F(v)$, $g_v \in G(v)$, by reasoning similarly to the beginning of the proof of Theorem 4.6, and observing that $\mathrm{SOL}(F, Y) \subseteq Y$ and $f_v + \eta_j g_v \in H_{\eta_j}(v)$, for every $j \geq 1$ we can write $-2\gamma_j(f_v + \eta_j g_v)^T(v - y_j) \leq \|y_j - v\|^2 - \|y_{j+1} - v\|^2 + 2\Lambda_1 \gamma_j^2$. Since $v \in \mathrm{SOL}(F, Y)$, $\overline{f}_v \in F(v)$ exists such that $\overline{f}_v^T(y_j - v) \geq 0$, and then:

$$-2\gamma_j g_v^T(v - y_j) \leq \frac{-2\gamma_j(\overline{f}_v + \eta_j g_v)^T(v - y_j)}{\eta_j} \leq \frac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{\eta_j} + 2\Lambda_1 \frac{\gamma_j^2}{\eta_j}.$$

Summing j from 1 to k and dividing by $\sum_{j=1}^k \gamma_j$ we get:

$$-2g_v^T(v - z_k) \leq \frac{\sum_{j=1}^k \frac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{\eta_j}}{\sum_{j=1}^k \gamma_j} + 2\Lambda_1 \frac{\sum_{j=1}^k \gamma_j \frac{\gamma_j}{\eta_j}}{\sum_{j=1}^k \gamma_j}. \tag{5.3}$$

By observing that

$$\begin{aligned} \sum_{j=1}^k \frac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{\eta_j} &= \frac{\|y_1 - v\|^2}{\eta_1} - \frac{\|y_{k+1} - v\|^2}{\eta_k} + \sum_{j=1}^{k-1} \|y_{j+1} - v\|^2 \left(\frac{1}{\eta_{j+1}} - \frac{1}{\eta_j}\right) \\ &\leq \frac{D^2}{\eta_1} + D^2 \sum_{j=1}^{k-1} \left(\frac{1}{\eta_{j+1}} - \frac{1}{\eta_j}\right) \frac{D^2}{\eta_1} + D^2 \left(\frac{1}{\eta_k} - \frac{1}{\eta_1}\right) = \frac{2\Lambda_3}{\eta_k}, \end{aligned}$$

we obtain $-2g_v^T(v - z_k) \leq 2\Lambda_1 \Phi_1^k + 2\Lambda_3 \frac{1}{\eta_k \sum_{j=1}^k \gamma_j}$, that implies (5.2). ∎

We define the following additional conditions to guarantee the convergence of the sequence produced by Algorithm (4.3) together with (4.6) to solutions of (GVI^u).

Assumptions D

D1 $\eta = 0$;

D2 $\frac{\gamma_k}{\eta_k} \to 0$;

D3 $\eta_k \sum_{j=1}^k \gamma_j \to \infty$.

Differently from the conditions in Assumptions **C**, Assumptions **D** require $\{\gamma_k\}$ and $\{\eta_k\}$ not to be chosen independently of one another . We remark that, in the more restrictive setting of single-valued upper and lower-level operators, as considered in [17], one can control the accuracy in the iterative solution of the Tikhonov subproblems. In this case, an algorithm can be defined to solve the resulting hierarchical Variational Inequality that converges under Assumptions **A**, **B**, **C**, **D1** and $\eta_k \notin \ell^1$, therefore not requiring **D2** and **D3** that relate $\{\gamma_k\}$ and $\{\eta_k\}$. In our general set-valued framework (resulting from nonsmooth payoffs for the players of the Nash problems) it is not practical to control the accuracy in the solution of the Tikhonov subproblems, and therefore Assumptions **D** are required in the following result.

THEOREM 5.2. *Consider the sequences $\{\gamma_k\}$ and $\{\eta_k\}$ defined in Algorithm (4.3) and $\{z_k\}$ defined in (4.6). If Assumptions **C** and **D** hold, then the unique limit point of $\{z_k\}$ is a solution to (GVI^u), and then to (GNEP^u).*

Proof. Sequence $\{z_k\}$ admits a unique limit point by Theorem 4.7. Due to assumptions **C1** and **D3**, $\Xi_3^k, \Phi_2^k \to 0$. Moreover, $\Xi_1^k, \Xi_2^k, \Phi_1^k \to 0$ in view of Lemma A.3, where we take $b_k = \gamma_k$ and $a_k = \gamma_k$ as far as Ξ_1^k is concerned, while $a_k = \eta_k$ when considering Ξ_2^k, and $a_k = \gamma_k/\eta_k$ as for Φ_1^k. The claim then follows from Theorem 4.3.∎

In order to recover solutions of (GVI^u) and then equilibria of (GNEP^u), $\{\eta_k\}$ must be assumed to go to 0. This requirement can be traced back to the lack of standard constraint qualifications for (GVI^u).

THEOREM 5.3. *Consider the sequences $\{\gamma_k\}$, $\{\eta_k\}$ and $\{y_k\}$ defined in Algorithm* (4.3). *If Assumptions **C** and **D** hold, and $y_k \to \overline{y}$, then $\overline{y}$ is a solution to problem* (GVI^u)*, and then to* (GNEP^u).

Proof. The proof is similar to that of Theorem 4.5. ∎

6. Complexity Bounds Considering Harmonic Sequences. In this section we consider the case where $\{\gamma_k\}$ and $\{\eta_k\}$ from Algorithm (4.3) together with (4.6) are defined as harmonic sequences:

$$\gamma_k = \frac{\overline{\gamma}}{k^{\alpha}}, \quad \eta_k = \frac{\overline{\eta}}{k^{\beta}} + \eta, \quad k \geq 1, \tag{6.1}$$

with $\overline{\gamma} > 0$, $\overline{\eta} > 0$ and $\eta \geq 0$. This is done in order to describe a possible practical way to implement the sequences $\{\gamma_k\}$ and $\{\eta_k\}$.

The first theorem deals with the complexity of the distance of $\{y_k\}$ from any solution $u \in \mathrm{SOL}(H_\eta, Y)$, by relying on the bounds defined in Theorem 4.4.

THEOREM 6.1. *Consider $\alpha \in \left(\frac{1}{2}, 1\right)$ and $\beta > 1 - \alpha$ in* (6.1)*, then Assumptions **C** hold. Moreover, given any tolerance $\delta \in (0,1)$ for the bound given in* (4.4)*, it holds that $2\Lambda_1 \Psi_1^k + 2\Lambda_2 \Psi_2^k < \delta$ for every*

$$k > \lambda_1 \left(\frac{1}{\delta}\right)^{\max\left\{\frac{1}{2\alpha-1}, \frac{1}{\alpha+\beta-1}\right\}},$$

with $\lambda_1 \triangleq 1 + \max\left\{ \left(\frac{4\Lambda_1\overline{\gamma}^2}{2\alpha-1}\right)^{\frac{1}{2\alpha-1}}, \left(\frac{4\Lambda_2\overline{\gamma\eta}}{\alpha+\beta-1}\right)^{\frac{1}{\alpha+\beta-1}} \right\}$.

Proof. Assumptions **C** trivially hold under the conditions on α and β.

Let us introduce an upper bound for Ψ_1^k:

$$\Psi_1^k = \overline{\gamma}^2 \sum_{j=k}^{\infty} \frac{1}{j^{2\alpha}} \leq \gamma_0^2 \int_{k-1}^{\infty} x^{-2\alpha} dx = \overline{\gamma}^2 \left[\frac{-1}{(2\alpha-1)x^{2\alpha-1}} \right]_{k-1}^{\infty} = \frac{\overline{\gamma}^2}{(2\alpha-1)(k-1)^{2\alpha-1}}.$$

Therefore, a sufficient condition to have $2\Lambda_1\Psi_1^k < \delta/2$, is $k > 1 + \left(\frac{4\Lambda_1\overline{\gamma}^2}{2\alpha-1}\right)^{\frac{1}{2\alpha-1}} \left(\frac{1}{\delta}\right)^{\frac{1}{2\alpha-1}}$.

Next, we define an upper-bound for Ψ_2^k:

$$\Psi_2^k = \overline{\gamma\eta} \sum_{j=k}^{\infty} \frac{1}{j^{\alpha+\beta}} \leq \overline{\gamma\eta} \int_{k-1}^{\infty} x^{-\alpha-\beta} dx = \overline{\gamma\eta} \left[\frac{-1}{(\alpha+\beta-1)x^{\alpha+\beta-1}} \right]_{k-1}^{\infty} = \frac{\overline{\gamma\eta}}{(\alpha+\beta-1)(k-1)^{\alpha+\beta-1}}.$$

Hence, a sufficient condition to have $2\Lambda_2\Psi_2^k < \delta/2$, is requiring that $k > 1 + \left(\frac{4\Lambda_2\overline{\gamma\eta}}{\alpha+\beta-1}\right)^{\frac{1}{\alpha+\beta-1}} \left(\frac{1}{\delta}\right)^{\frac{1}{\alpha+\beta-1}}$, concluding the proof. ∎

In particular, choosing $\alpha = 1 - \epsilon$ and $\beta = 1 - \epsilon$, with $0 < \epsilon < 1/2$, the maximum number of iterations k to have the distance $\|y_k - u\|^2$ converging with an error lower than δ is $\mathcal{O}(\delta^{-1/(1-2\epsilon)})$, for any $u \in \mathrm{SOL}(H_\eta, Y)$.

In the forthcoming results, we exploit the following bounds for the generic harmonic series with $\alpha > 0$:

$$\frac{k^{(1-\alpha)}}{2(1-\alpha)} \leq \sum_{j=1}^{k} \frac{1}{j^\alpha} \leq \frac{k^{(1-\alpha)}}{1-\alpha} + \frac{-\alpha}{1-\alpha}, \tag{6.2}$$

where the lower bound holds for $k \geq 2^{\frac{2}{1-\alpha}}$. The next result provides complexity bounds for $\{z_k\}$ to converge to $\mathrm{SOL}(H_\eta, Y)$ (see Theorem 4.6).

THEOREM 6.2. *If in* (6.1) $\alpha \in (0,1)$ *and* $\beta > 0$*, then Assumptions* $\boldsymbol{C}$ *hold. Moreover, given any tolerance* $\delta \in (0,1)$ *for the bound given in* (4.7)*, it holds that* $\Lambda_1 \Xi_1^k + \Lambda_2 \Xi_2^k + \Lambda_3 \Xi_3^k < \delta$ *for every*

$$k > \lambda_2 \left(\frac{1}{\delta}\right)^{\max\left\{\frac{1}{\alpha}, \frac{1}{1-\alpha}, \frac{1}{\beta}\right\}},$$

$$\textit{with } \lambda_2 = \max\left\{\left(\frac{12\Lambda_1\overline{\gamma}(1-\alpha)}{1-2\alpha}\right)^{\frac{1}{\alpha}}, \left(\frac{-24\Lambda_1\overline{\gamma}\alpha(1-\alpha)}{1-2\alpha}\right)^{\frac{1}{1-\alpha}},\right.$$
$$\left.\left(\frac{12\Lambda_2\overline{\eta}(1-\alpha)}{1-(\alpha+\beta)}\right)^{\frac{1}{\beta}}, \left(\frac{-12\Lambda_2\overline{\eta}(\alpha+\beta)(1-\alpha)}{1-(\alpha+\beta)}\right)^{\frac{1}{1-\alpha}}, \left(\frac{6\Lambda_3(1-\alpha)}{\overline{\gamma}}\right)^{\frac{1}{1-\alpha}}\right\}.$$

Proof. Assumptions **C** trivially hold under the conditions on α and β.

The bounds defined in (6.2) imply, under our hypotheses on α and β

$$\sum_{j=1}^{k} \gamma_j = \sum_{j=1}^{k} \overline{\gamma}\frac{1}{j^\alpha} \geq \overline{\gamma}\frac{k^{(1-\alpha)}}{2(1-\alpha)},$$
$$\sum_{j=1}^{k} \gamma_j^2 = \sum_{j=1}^{k} \overline{\gamma}^2\frac{1}{j^{2\alpha}} \leq \overline{\gamma}^2\frac{k^{(1-2\alpha)}}{1-2\alpha} + \frac{-\overline{\gamma}^2 2\alpha}{1-2\alpha},$$
$$\sum_{j=1}^{k} \gamma_j(\eta_j - \eta) = \sum_{j=1}^{k} \overline{\gamma\eta}\frac{1}{j^{\alpha+\beta}} \leq \overline{\gamma\eta}\frac{k^{1-(\alpha+\beta)}}{1-(\alpha+\beta)} + \frac{-\overline{\gamma\eta}(\alpha+\beta)}{1-(\alpha+\beta)}.$$

We now define an upper bound for Ξ_1^k:

$$\Xi_1^k = \frac{\sum_{j=1}^k \gamma_j^2}{\sum_{j=1}^k \gamma_j} \leq \frac{2\overline{\gamma}(1-\alpha)}{1-2\alpha}k^{-\alpha} + \frac{-4\overline{\gamma}\alpha(1-\alpha)}{1-2\alpha}k^{\alpha-1},$$

therefore, a sufficient condition to have $\Lambda_1\Xi_1^k < \delta/3$ is to have

$$k > \max\left\{\left(\frac{12\Lambda_1\overline{\gamma}(1-\alpha)}{1-2\alpha}\right)^{\frac{1}{\alpha}}, \left(\frac{-24\Lambda_1\overline{\gamma}\alpha(1-\alpha)}{1-2\alpha}\right)^{\frac{1}{1-\alpha}}\right\}\left(\frac{1}{\delta}\right)^{\max\left\{\frac{1}{\alpha}, \frac{1}{1-\alpha}\right\}}.$$

The upper bound for Ξ_2^k is as follows:

$$\Xi_2^k = \frac{\sum_{j=1}^k \gamma_j(\eta_j - \eta)}{\sum_{j=1}^k \gamma_j} \leq \frac{2\overline{\eta}(1-\alpha)}{1-(\alpha+\beta)}k^{-\beta} + \frac{-2\overline{\eta}(\alpha+\beta)(1-\alpha)}{1-(\alpha+\beta)}k^{\alpha-1},$$

therefore, a sufficient condition to have $\Lambda_2\Xi_2^k < \delta/3$ is to have

$$k > \max\left\{\left(\frac{12\Lambda_2\overline{\eta}(1-\alpha)}{1-(\alpha+\beta)}\right)^{\frac{1}{\beta}}, \left(\frac{-12\Lambda_2\overline{\eta}(\alpha+\beta)(1-\alpha)}{1-(\alpha+\beta)}\right)^{\frac{1}{1-\alpha}}\right\}\left(\frac{1}{\delta}\right)^{\max\left\{\frac{1}{\beta},\frac{1}{1-\alpha}\right\}}.$$

The upper bound for Ξ_3^k is as follows:

$$\Xi_3^k = \frac{1}{\sum_{j=1}^k \gamma_j} \le \frac{2(1-\alpha)}{\overline{\gamma}} k^{\alpha-1},$$

therefore, a sufficient condition to have $\Lambda_3\Xi_3^k < \delta/3$ is to have

$$k > \left(\frac{6\Lambda_3(1-\alpha)}{\overline{\gamma}}\right)^{\frac{1}{1-\alpha}} \left(\frac{1}{\delta}\right)^{\frac{1}{1-\alpha}}.$$

□

Choosing $\alpha = \beta = 1/2$, the maximum number of iterations k to have problem (4.2) solved by z_k with an error of less than δ is $\mathcal{O}(\delta^{-2})$.

We show that solving approximately problem (4.2) yields the approximate fulfillment of optimality conditions for the Minty versions of (GVI^l) and (GVI^u), according to Proposition Proposition 6.3.

PROPOSITION 6.3. *Let $\eta > 0$ and z_k satisfy* (4.7)*, it holds that*

$$f_v^T(v - z_k) \ge -\Lambda_1\Xi_1^k - \Lambda_2\Xi_2^k - \Lambda_3\Xi_3^k - \eta\Lambda_2, \quad \forall v \in Y, \quad \forall f_v \in F(v), \tag{6.3}$$

$$g_v^T(v - z_k) \ge -\frac{\Lambda_1\Xi_1^k + \Lambda_2\Xi_2^k + \Lambda_3\Xi_3^k}{\eta}, \quad \forall v \in SOL(F,Y), \quad \forall g_v \in G(v). \tag{6.4}$$

Proof. See Appendix A.10. □

Notice that Proposition 6.3 works only for $\eta > 0$ and there is no value for η that let the approximation errors given in (6.3) and (6.4) be zero simultaneously.

By considering the bounds obtained in Theorem 5.1, complexity results can be provided as follows.

THEOREM 6.4. *If in* (6.1) *$\alpha \in (0,1)$, $\beta \in (0, \min\{\alpha, 1-\alpha\})$ and $\eta = 0$, then Assumptions **C** and **D** hold. Moreover given any tolerance $\delta \in (0,1)$ for the bounds given in* (5.1) *and* (5.2)*, $\Lambda_1\Xi_1^k + \Lambda_2\Xi_2^k + \Lambda_3\Xi_3^k < \delta$ for every*

$$k > \lambda_2 \left(\frac{1}{\delta}\right)^{\max\left\{\frac{1}{\alpha},\frac{1}{1-\alpha},\frac{1}{\beta}\right\}},$$

with λ_2 defined in Theorem 6.2*, and $\Lambda_1\Phi_1^k + \Lambda_3\Phi_2^k < \delta$ for every*

$$k > \lambda_3 \left(\frac{1}{\delta}\right)^{\max\left\{\frac{1}{\alpha-\beta},\frac{1}{1-\alpha},\frac{1}{1-\alpha-\beta}\right\}},$$

with $\lambda_3 \triangleq \max\left\{\left(\frac{8\Lambda_1\overline{\gamma}(1-\alpha)}{\overline{\eta}(1+\beta-2\alpha)}\right)^{\frac{1}{\alpha-\beta}}, \left(\frac{8\Lambda_1\overline{\gamma}(\beta-2\alpha)(1-\alpha)}{\overline{\eta}(1+\beta-2\alpha)}\right)^{\frac{1}{1-\alpha}}, \left(\frac{4\Lambda_3(1-\alpha)}{\overline{\gamma}\overline{\eta}}\right)^{\frac{1}{(1-\alpha-\beta)}}\right\}.$

Proof. Assumptions **C**, **D1**, **D2** trivially hold under the conditions on α and β. Note that the complexity regarding (5.1) is proved in Theorem 6.2. Using the harmonic series bounds (6.2) we can write:

$$\sum_{j=1}^{k} \gamma_j = \overline{\gamma} \sum_{j=1}^{k} \frac{1}{j^{\alpha}} \geq \overline{\gamma} \frac{k^{1-\alpha}}{2(1-\alpha)} \tag{6.5}$$

$$\sum_{j=1}^{k} \frac{\gamma_j^2}{\eta_j} = \frac{\overline{\gamma}^2}{\overline{\eta}} \sum_{j=1}^{k} \frac{1}{j^{2\alpha-\beta}} \leq \frac{\overline{\gamma}^2}{\overline{\eta}} \frac{k^{1-(2\alpha-\beta)}}{1-(2\alpha-\beta)} + \frac{\overline{\gamma}^2(\beta-2\alpha)}{\overline{\eta}(1+\beta-2\alpha)} \tag{6.6}$$

We can define the following upper bound for Φ_1^k:

$$\Phi_1^k = \frac{\sum_{j=1}^{k} \frac{\gamma_j^2}{\eta_j}}{\sum_{j=1}^{k} \gamma_j} \leq \frac{\overline{\gamma} 2(1-\alpha)}{\overline{\eta}(1+\beta-2\alpha)} k^{\beta-\alpha} + \frac{\overline{\gamma}(\beta-2\alpha)2(1-\alpha)}{\overline{\eta}(1+\beta-2\alpha)} k^{\alpha-1},$$

therefore, a sufficient condition to have $\Lambda_1 \Phi_1^k < \delta/2$ is to have:

$$k > \max\left\{ \left(\tfrac{8\Lambda_1\overline{\gamma}(1-\alpha)}{\overline{\eta}(1+\beta-2\alpha)}\right)^{\frac{1}{\alpha+\beta}}, \left(\tfrac{8\Lambda_1\overline{\gamma}(\beta-2\alpha)(1-\alpha)}{\overline{\eta}(1+\beta-2\alpha)}\right)^{\frac{1}{1-\alpha}} \right\} \left(\tfrac{1}{\delta}\right)^{\max\left\{\frac{1}{\alpha-\beta}, \frac{1}{1-\alpha}\right\}} \Bigg\}.$$

Next, we define an upper bound for Φ_2^k:

$$\Phi_2^k \triangleq \frac{1}{\eta_k \sum_{j=1}^{k} \gamma_j} \leq \frac{2(1-\alpha)}{\overline{\gamma}\overline{\eta}} k^{\alpha+\beta-1}, \tag{6.7}$$

therefore, a sufficient condition to have $\Lambda_3 \Phi_2^k < \delta/2$ is to have

$$k > \left(\frac{4\Lambda_3(1-\alpha)}{\overline{\gamma}\overline{\eta}}\right)^{\frac{1}{(1-\alpha-\beta)}} \left(\frac{1}{\delta}\right)^{\frac{1}{(1-\alpha-\beta)}}.$$

Moreover, assumption **D3** holds due to relation (6.7), since $\alpha + \beta < 1$. ∎

Choosing $\alpha = 1/2$ and $\beta = 1/4$, the maximum number of iterations k to have the Minty versions of (GVI^l) and (GVI^u) solved with an error less than δ is $\mathcal{O}(\delta^{-4})$. Notice that the convergence rate we prove is the same as the one provided, in a more specific case (namely, an optimization problem with variational inequality constraints), in [13].

Summarizing, Algorithm (4.3) together with (4.6), with the harmonic sequences in (6.1), achieves different convergence properties with different complexities for different values of α and β (see Table 1).

7. Numerical Analysis.

We define a practical algorithm to exploit the previous sections' theoretical results. Focusing on Table 1, if α and β are close to 1, one can obtain quite fast convergence of $\{y_k\}$ to an orbit around SOL(H_η, Y). On the other hand, if α and β decrease to 0.5, $\{z_k\}$ converges to SOL(H_η, Y). Finally, if β further decreases to 0.25, the convergence of $\{z_k\}$ is guaranteed to the solutions of (GVI^u), and then the equilibria (GNEP^u), but with worse complexity guarantees. Therefore, a possible way to obtain, at the beginning, fast convergence to partial results, and achieve the convergent setting for α and β once close to the solutions of (GVI^u) (by satisfying Assumptions **C** and **D**), is to consider two decreasing sequences $\{\alpha_k\}$ and $\{\beta_k\}$.

α	β	convergence properties	complexity
$1-\epsilon$	$1-\epsilon$	$\limsup_{\Delta\to\infty} \|y_{k+\Delta}-u\|^2 - \|y_k-u\|^2 \leq \delta$, $u \in \text{SOL}(H_\eta, Y)$	$\mathcal{O}(\delta^{-1/(1-2\epsilon)})$
0.5	0.5	$h_v^{\eta T}(v-z_k) \geq -\delta,\ \forall v \in Y,\ h_v^\eta \in H_\eta(v)$	$\mathcal{O}(\delta^{-2})$
0.5	0.25	$f_v^T(v-y) \geq -\delta,\ \forall v \in Y,\ f_v \in F(v)$ $g_v^T(v-y) \geq -\delta,\ \forall v \in \text{SOL}(F,Y),\ g_v \in G(v)$	$\mathcal{O}(\delta^{-4})$

TABLE 1
Possible settings for α and β and relative convergence properties and complexities

Algorithm 7.1 combines computations (4.3) and (4.6) and employs harmonic sequences for $\{\gamma_k\}$ and $\{\eta_k\}$ with decreasing $\{\alpha_k\}$ and $\{\beta_k\}$, respectively. In particular, $\overline{k}$ is a parameter that indicates the iteration at which the averaging procedure defined in (4.6) starts, and the sequence $\{z_k\}$ is computed. This allows one to start computing $\{z_k\}$ when the sequence $\{y_k\}$ approaches $\text{SOL}(H_{\eta_{\overline{k}}}, Y)$ (see Theorem 4.5). One gets a faster convergence of $\{z_k\}$ as points y_k that are possibly far from the solution set and weight more (since $\{\gamma_k\}$ is monotone non-increasing) are ignored in the average.

In the following result, whose proof is given in Appendix A.11, we provide a practical rule to compute $\{\alpha_k\}$ and $\{\beta_k\}$ in order to satisfy Assumptions **C** and **D**. We focus on the case where $\{\alpha_k\}$ goes from $\overline{\alpha}$ to α and $\{\beta_k\}$ goes from $\overline{\beta}$ to β.

Algorithm 7.1 Projected Average Single-loop Tikhonov Algorithm (PASTA)

Data: $\{\alpha_k\} > 0, \overline{\gamma} > 0, \{\beta_k\} > 0, \overline{\eta} > 0, \overline{k} \in \mathbb{N}, y_1 \in Y$
for $k = 1, 2, \dots$ **do**
  $\gamma_k \leftarrow \overline{\gamma}/k^{\alpha_k}$ and $\eta_k \leftarrow \overline{\eta}/k^{\beta_k}$
  choose $f_{y_k} \in F(y_k)$, $g_{y_k} \in G(y_k)$ and compute $h_{y_k}^{\eta_k} = f_{y_k} + \eta_k g_{y_k}$
  $y_{k+1} = P_Y(y_k - \gamma_k h_{y_k}^{\eta_k})$
end for
for $k = \overline{k}, \overline{k}+1 \dots$ **do**
  $z_k = \dfrac{\sum_{j=\overline{k}}^k \gamma_j y_j}{\sum_{j=\overline{k}}^k \gamma_j}$
end for

PROPOSITION 7.1. *Let $\overline{\alpha} \geq \alpha > 0$, $\overline{\beta} \geq \beta > 0$, $\varepsilon_\alpha, \varepsilon_\alpha > 0$, $I_\alpha, I_\beta \in \mathbb{N}$ and $\gamma_k = \overline{\gamma}/k^{\alpha_k}$, $\eta_k = \overline{\eta}/k^{\beta_k}$, with $\alpha_k = \overline{\alpha} - (\overline{\alpha} - \alpha)(\min\{k, I_\alpha\}/I_\alpha)^{\varepsilon_\alpha}$, $\beta_k = \overline{\beta} - (\overline{\beta} - \beta)(\min\{k, I_\beta\}/I_\beta)^{\varepsilon_\beta}$. Assume $\alpha < 1$, $\beta < \min\{\alpha, 1-\alpha\}$, and $\varepsilon_\alpha \leq \overline{\varepsilon}_\alpha \triangleq log_{I_\alpha}(1 - (1-t_\alpha)\alpha/(t_\alpha(\overline{\alpha}-\alpha)))^{-1}$, $\varepsilon_\beta \leq \overline{\varepsilon}_\beta \triangleq log_{I_\beta}\left(1 - (1-t_\beta)\beta/(t_\beta(\overline{\beta}-\beta))\right)^{-1}$, with $t_\alpha \triangleq log_{I_\alpha}(I_\alpha - 1)$ and $t_\beta \triangleq log_{I_\beta}(I_\beta - 1)$. Assumptions **C** and **D** hold.*

Employing in PASTA $\{\alpha_k\}$ and $\{\beta_k\}$ as defined in Proposition 7.1, with ε_α and ε_β chosen according to Proposition 7.1, Assumptions **C** and **D** hold. Therefore, by Theorem 5.2, Theorem 5.3 and Theorem 6.4, the unique limit point of $\{z_k\}$, that is the limit point of $\{y_k\}$ if it exists, is a solution to (GVI^u) and then it is a variational equilibrium for (GNEP^u) by Theorem 4.2 and Proposition 3.7. Notice that the bounds for ε_α and ε_β provided in Proposition 7.1 are only sufficient to satisfy Assumptions **C** and **D**, and larger values for such parameters can be used in practice. We can employ fixed values by simply setting $\alpha_k = \alpha$ and $\beta_k = \beta$ for all k, and still satisfy Assumptions **C** and **D**, therefore recovering the theoretical convergence properties. In the sequel, we compare these two choices and show, by means of numerical evidences, that PASTA achieves faster convergence than the case of fixed α and β.

We provide numerical experiments to prove the convergence of PASTA in practical settings. In Example 1 we consider a simple hierarchical jointly-convex GNEP, which allows one to evaluate the convergence of the algorithm to the equilibria of (NEP^l) and (GNEP^u), since an analytical description of the lower-level equilibrium set can be readily obtained. In Example 2 we study a more elaborate hierarchical jointly-convex GNEP model in the context of multi-portfolio selection (see [16] for more details regarding multi-portfolio optimization). In this case, one cannot easily evaluate the convergence to equilibria of (GNEP^u), because an analytical description of its feasible set (i.e. the equilibria of (NEP^l)) is not readily available. We focus only on convergence to the equilibria of (NEP^l), but we will also show the influence of the upper level by observing *a posteriori* the computed solutions. All the computations are performed on a Mac mini 8.1, Quad-Core Intel Core i3 3.6 GHz, RAM 8 GB, and took no longer than 10 seconds (Example 1) and 200 seconds (Example 2).

Example 1 We first consider a simple example where it is easy to have an explicit expression for the lower-level equilibrium set E, and to compute the unique variational solution of (GNEP^u). Let us consider $N = 4$ lower-level players and $M = 2$ upper-level players, with $x^1 = (y^2, y^4)$, $x^2 = (y^1, y^3)$,

$$\begin{aligned}
&\theta_1^l(y^1, y^{-1}) = 0.5(y^1)^2 + y^1(y^2 + 2y^3 + y^4 - 100), \quad \varphi_1^l(y^1) = 0, \quad Y_1 = [-100, 50],\\
&\theta_2^l(y^2, y^{-2}) = 0.5(y^2)^2 + y^2(y^1 + y^3 + y^4 - 50), \quad \varphi_2^l(y^2) = \max\{0, -10(y^2 - 15)\}, \quad Y_2 = [0, 50],\\
&\theta_3^l(y^3, y^{-3}) = 0.5(y^3)^2 + y^3(y^2 + y^4 - 100), \quad \varphi_3^l(y^3) = 0, \quad Y_3 = [0, 100],\\
&\theta_4^l(y^4, y^{-4}) = 0.5(y^4)^2 + y^4(y^1 + y^2 + y^3 - 50), \quad \varphi_4^l(y^4) = 0, \quad Y_4 = [0, 50],\\
&\theta_1^u(x^1, x^{-1}) = (y^2 - 20)^2 + (y^4 - 50)^2 + (y^2 + y^4)(y^1 + y^3), \quad \varphi_1^u(x^1) = 0,\\
&\theta_2^u(x^2, x^{-2}) = (y^1)^2 + y^1(y^2 + y^3) + (y^3)^2 + y^3(y^2 + y^4), \quad \varphi_2^u(x^2) = 0.
\end{aligned}$$

For this example Assumptions **A** and **B** are verified. One can obtain an explicit expression for the lower-level equilibrium set $E = \{(-50, y^2, 50, 50 - y^2) : 15 \leq y^2 \leq 50\}$, and thus the unique variational equilibrium of (GNEP^u) is $x^* = (-50, 15, 50, 35)$. Note that at x^*, the second lower-level player's payoff is non differentiable. In this setting, we can test PASTA and monitor the distance from x^*. Concerning the evaluation of the subgradient, in order to deal with the nondifferentiability of the lower-level map, we set $f_y = [\nabla_{y^1}\theta_1^l(y), \nabla_{y^2}\theta_2^l(y) - 5(15 + 10^{-3} - y^2)/(10^{-3}), \nabla_{y^3}\theta_3^l(y), \nabla_{y^4}\theta_4^l(y)]^T$ for every y such that $y^2 \in [15 - 10^{-3}, 15 + 10^{-3}]$. The projection is computed in closed form, since Y_ν are box-sets. We set the maximum number of iterations $\overline{I} = 10^6$, the parameters $\overline{\gamma} = 1$, $\overline{\eta} = 0.1$, and the starting point $y_1 = (0, 0, 0, 0)$. The sequence $\{\alpha_k\}$, used to compute the stepsizes $\{\gamma_k\}$, is defined as in Proposition 7.1 with $\overline{\alpha} = 0.75$, $\alpha = 0.5$, $I_\alpha = \overline{I}/2$ and $\varepsilon_\alpha = 0.05$. On the other hand, $\{\beta_k\}$, used to compute the Tikhonov parameters $\{\eta_k\}$, is defined as in Proposition 7.1, with $\overline{\beta} = 0.75$, $\beta = 0.25$, $I_\beta = \overline{I}$ and $\varepsilon_\beta = 0.03$. These values for ε_α and ε_β are such that the sequences $\{\gamma_k\}$ and $\{\eta_k\}$ are nonincreasing and therefore Assumptions **C** are verified (even though ε_α and ε_β do not verify the sufficient condition given in Proposition 7.1). PASTA, with its variable policies for $\{\alpha_k\}$ and $\{\beta_k\}$, is compared with the fixed case, where $\alpha_k = \alpha$ and $\beta_k = \beta$ for all k. Note that the values for α and β, used for both the variable and fixed settings, ensure Assumptions **D**, and the convergence of the method is guaranteed (see Theorem 6.4). In Table 2 we report $\text{opt}(z_{\overline{I}}) \triangleq \|z_{\overline{I}} - x^*\|_\infty$ for PASTA and for the fixed case, as well as for different choices of the iteration $\overline{k}$ for the averaging procedure $\{z_k\}$ to start. Note that the point $z_{\overline{I}}$ is closer to x^* for higher values of $\overline{k}$,

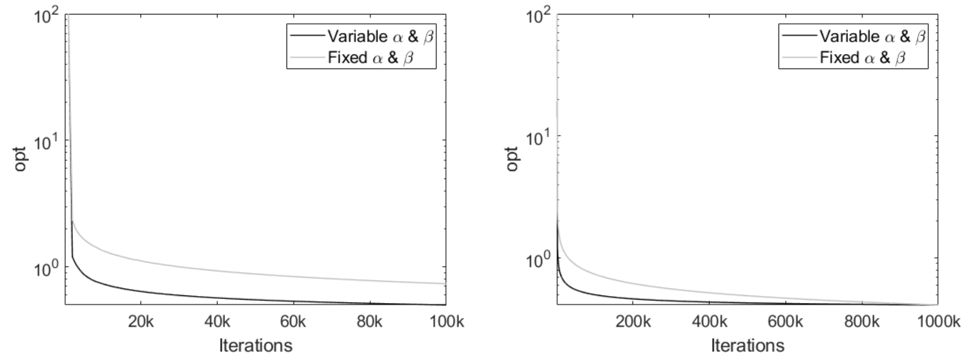

FIG. 1. *Comparison between variable (PASTA) and fixed* $\{\alpha_k\}$ *and* $\{\beta_k\}$ *considering* $\text{opt}(y_k) = \|y_k - x^*\|_\infty$ *for iterations 0- 100k (left-hand side) and for all 1000k iterations (right-hand side)*

		$\text{opt}(z_{\overline{I}})$		$\text{opt}(y_{\overline{I}})$
$\overline{k}$	0	$0.4\overline{I}$	$0.8\overline{I}$	
Variable α & β	0.57434	0.42161	0.41424	0.41219
Fixed α & β	0.84268	0.45928	0.42367	0.41220

TABLE 2

$\text{opt}(w) = \|w - x^*\|_\infty$ *in Example 1, considering variable (PASTA) and fixed* $\{\alpha_k\}$ *and* $\{\beta_k\}$ *for* $z_{\overline{I}}$*, with different starting iterations* $\overline{k}$*, and* $y_{\overline{I}}$

because the early iterations, which are more distant from x^*, are not included in the computation of the average. Moreover, we underline that in all our experiments, $y_{\overline{I}}$ is a better approximation of x^* than every $z_{\overline{I}}$. For this reason, although the averaged sequence $\{z_k\}$ is essential to obtain theoretical convergence guarantees (see section 4 and section 5), in our experiments the sequence $\{y_k\}$ has shown convergent behaviour, and we rely on Theorem 4.5 b) and Theorem 5.3 to justify our choice to focus on $\{y_k\}$ approaching x^*. In Figure 1, we show the comparison between the performances of variable (PASTA) and fixed $\{\alpha_k\}$ and $\{\beta_k\}$ in terms of distance between $\{y_k\}$ and x^*. In Table 3 we report the value of this distance at different iterations. It is evident that using the insights in section 4 concerning the Tikhonov subproblem to develop the algorithm with variable $\{\alpha_k\}$ and $\{\beta_k\}$ (PASTA), one can obtain a faster convergence to the equilibria of (GNEP^u), than using fixed $\{\alpha_k\}$ and $\{\beta_k\}$, see also the explaination at the beginning of section 7, together with Table 1. The output of PASTA is $y_{\overline{I}} = (-49.5878, 15.0010, 50.0124, 34.6699)$.

Example 2 We consider a hierarchical multi-portfolio selection model in the case of financial service providers managing different lower-level clients' portfolios (or accounts) by assigning them to multiple upper-level managers (see [16] for more details about hierarchical multi-portfolio optimization and [18] where the hierarchical GNEP framework is introduced in this context). Following the classical Markowitz approach, as for each lower-level account ν, the weighted sum of linear expected return ($I_\nu(y^\nu)$) and quadratic portfolio volatility ($R_\nu(y^\nu)$) is minimized, by investing the relative budgets in K financial assets. The lower-level variables $y^\nu \in \mathbb{R}^K$ represent the shares of the budget to be invested in each asset. Additionally, each account-related objective depends (parametrically) on the other accounts' problem decision variables via a coupling quadratic transaction cost term ($TC_\nu(y^\nu, y^{-\nu})$). Therefore the accounts-related

Iterations	10k	25k	50k	75k	100k	250k	500k	750k	1000k
Var α & β	0.7342	0.6140	0.5491	0.5186	0.4998	0.4528	0.4283	0.4179	0.4122
Fix α & β	1.3395	1.0513	0.8778	0.7915	0.7359	0.5839	0.4905	0.4431	0.4122

TABLE 3

$\text{opt}(y_k) = \|y_k - x^*\|_\infty$ *at different iterations in Example 1, considering variable (PASTA) and fixed* $\{\alpha_k\}$ *and* $\{\beta_k\}$

lower-level parametric problems form (NEP^l). Upper-level managers $\mu = 1, \dots, M$ are responsible of deciding trades for a subset $\mathcal{S}_\mu$ of lower-level accounts, but selecting only among equilibria of (NEP^l). The objective function of each manager μ measures the performances of the portfolios they manage, and depends not only on each manager's own decision variables, but also on the choices of the other managers, similarly to the lower-level accounts' interplay. The resulting upper-level managers' problems form (GNEP^u), where the shared feasible set is given by the equilibria of the accounts-related (NEP^l). At both the upper and lower level, a sparsity enhancing term is included to reduce monitoring costs and simplify portfolio management.

Consider $N = 25$, $M = 5$ and $x^\mu = [y^\nu]_{\nu \in \mathcal{S}_\mu}$,

$$\theta^l_\nu(y^\nu, y^{-\nu}) = -I_\nu(y^\nu) + \rho_\nu R_\nu(y^\nu) + TC_\nu(y^\nu, y^{-\nu}), \quad \varphi^l_\nu(y^\nu) = \tau_\nu \|y^\nu\|_1,$$

$$Y_\nu \triangleq \left\{ y^\nu \in [l_\nu, u_\nu]^K : \sum_{i=1}^K y^\nu_i \le 1 \right\},$$

$$\theta^u_\mu(x^\mu, x^{-\mu}) = -\sum_{\nu \in \mathcal{S}_\mu} I_\nu(y^\nu) + \rho_\mu \sum_{\nu \in \mathcal{S}_\mu} R_\nu(y^\nu) + TC_\mu(x^\mu, x^{-\mu}), \quad \varphi^u_\mu(x^\mu) = \tau_\mu \sum_{\nu \in \mathcal{S}_\mu} \|y^\nu\|_1,$$

where ρ_ν regulates the risk-aversion of each agent ν, and τ_ν regulates their desire for sparsity. In the following numerical results, $u_\nu = 1$ and $l_\nu = -0.1$ are chosen for each lower-level player ν to allow players to invest at most their whole budget on a single financial asset and to shortsell each asset for at most 10% of their budget. Numerical tests for two data sets are provided, the first one consisting of $K = 10$ assets belonging to Euro Stoxx 50 (SX5E) (from 2/1/2019 to 31/12/2019), resulting in $n_\nu = 10$ variables controlled by each lower-level player, and $p = 250$ total (GNEP^u) variables. The second data set consists of $K = 29$ assets from Dow Jones Industrial Average (DJIA) stock markets (from 2/1/2017 to 31/12/2017), resulting in $n_\nu = 29$ variables controlled by each lower-level player, and $p = 725$ total (GNEP^u) variables. In both cases, the upper-level managers control $N/M = 5$ lower-level accounts each, arranged in such a way that $\mathcal{S}_\mu = \{(\mu - 1)(N/M) + 1, \dots, \mu(N/M)\}$ for all $\mu \in \{1, \dots, M\}$. We have, for the SX5E dataset, $m_\mu = 50$, and for the DIJA dataset $m_\mu = 145$ variables controlled by each upper-level manager. All player-related parameters are computed randomly in order to verify Assumptions **A** and **B** (see [16, Section 3] for further details). We remark that the resulting (NEP^l) and (GNEP^u) are not potential games, and they cannot be reduced to simple optimization problems.

The algorithm's parameters for PASTA are the same as Example 1, except $\overline{\gamma} = 100$ and $\overline{\eta} = 1$, thus satisfying Assumptions **C** and **D**. The equally weighted portfolio $y^\nu = (1/K)\mathbf{1}^K$ for all ν is used as the starting vector. Concerning the subgradients, $f_{y^\nu_i} = \nabla \theta^l_\nu(y)_i + \tau^\nu(y^\nu_i + 10^{-4})/(10^{-4}) - \tau^\nu$ whenever $y^\nu_i \in [-10^{-4}, 10^{-4}]$ for every $\nu \in \{1, \dots, N\}$ and $i \in \{1, \dots, K\}$, and $g_{x^\mu_j} = \nabla \theta^u_\mu(x)_j + \tau^\mu(x^\mu_j + 10^{-4})/(10^{-4}) - \tau^\mu$ whenever $x^\mu_j \in [-10^{-4}, 10^{-4}]$ for every $\mu \in \{1, \dots, M\}$ and $j \in \{1, \dots, (N/M)K\}$. To implement the projection step of PASTA, a finite-steps method, inspired by [16], is implemented, preventing one from having to compute the projection by solving an optimization problem at each iteration.

Portfolios corresponding to clients from 1 to 15 are regularized only at the lower level, while portfolios corresponding to clients from 16 to 25 are regularized only by the upper-level managers: $\tau^l_\nu = \overline{\tau}^l$ for $\nu = 1, \ldots, 15$, $\tau^l_\nu = 0$ for $\nu = 16, \ldots, 25$, $\tau^u_\mu = 0$ for $\mu = 1, \ldots, 3$, $\tau^u_\mu = \overline{\tau}^u$ for $\mu = 4, 5$. This is done in order to observe how the regularization of the two hierarchical levels yields sparsity for the computed portfolios. Depending on $\overline{\tau}^l$ and $\overline{\tau}^u$, we define five different regularization settings: • *No regularization*: $\overline{\tau}^l = \overline{\tau}^u = 0$ • *Lower regularization 1*: $\overline{\tau}^l =$ 2e-04, $\overline{\tau}^u = 0$ • *Lower regularization 2*: $\overline{\tau}^l =$ 3e-04, $\overline{\tau}^u = 0$ • *Full regularization 1*: $\overline{\tau}^l =$ 2e-04, $\overline{\tau}^u =$ 3e-03 • *Full regularization 2*: $\overline{\tau}^l =$ 3e-04, $\overline{\tau}^u =$ 3e-03. It is not reasonable to assume that an analytical expression for E is available, as it is for Example 1, and therefore it is not practical to explicitly compute the distance of $\{y_k\}$ and $\{z_k\}$ from (GNEP^u)'s solution set. A measure of feasibility can still be given as $\text{feas}(y_k, f_{y_k}) \triangleq \|y_k - P_Y(y_k - f_{y_k})\|_2$, with $f_{y_k} \in F(y_k)$. Note that this is an upper bound of the distance from $\{y_k\}$ to E, as $f_{y_k} \in F(y_k)$ was not chosen to minimize this quantity.

Figure 2 and Figure 3 show $\text{feas}(y_k, f_{y_k})$ for the two datasets considered and the five different regularization settings over the iterations. In every picture, we report both the values for the algorithm version with variable $\{\alpha_k\}$ and $\{\beta_k\}$ (PASTA), and the for version with fixed $\{\alpha_k\}$ and $\{\beta_k\}$. Similarly to the results in Example 1, PASTA shows a faster convergence to the feasible set of the hierarchical problem. The erratic behaviour of $\text{feas}(y_k, f_{y_k})$, which happens in the regularized settings, can be explained by the lack of inner semicontinuity of the subgradient point-to-set mappings. In fact, in the *No regularization* setting, the plots turn out to appear quite smooth. Therefore, in the following analysis, we report values obtained by PASTA.

In Table 4 we report $\text{feas}(y_{\overline{I}}, f_{y_{\overline{I}}})$ and $\text{feas}(z_{\overline{I}}, f_{z_{\overline{I}}})$ computed starting from different iterations $\overline{k}$, in all the five regularization settings. Similarly to Example 1, $\{z_k\}$ obtains better feasibility for higher values of $\overline{k}$. Contrarily to Example 1, $\{z_k\}$ can achieve a better feasibility than $\{y_k\}$, because it shows more resilience to the non-continuity of the subgradient and a more stable trend. For this reason, $\{z_k\}$ could be useful to obtain a smoother convergence in the cases where the nonsmoothness of the players' payoffs yields a noisy behaviour of the considered merit function for $\{y_k\}$.

So far, in this numerical example, we only analyzed convergence to the feasible set E of (GNEP^u). To show the influence of the upper-level managers, and consequently of the upper-level objective functions, we measure the sparsity of the portfolio corresponding to $z_{\overline{I}}$ for $\overline{k} = 0.8\overline{I}$ (which is actually the same as the sparsity for $y_{\overline{I}}$) for the five regularization settings considered. Table 5 shows the percentage of zeros (intended as investments of less than 0.1% of the budget) of the final portfolios, regularized by the lower-level agents (accounts 1–15) and upper-level managers (accounts 16–25). Both of the hierarchical levels have an impact on the computed solutions, as witnessed by the different number of zeros depending on the agents' regularization choices. Specifically, in the *No regularization* setting, the computed portfolios require every account to invest in all the assets, resulting in a completely non-sparse solution. In the two *Lower regularization* settings, accounts 1–15 invest in less assets, with a sparser solution for *Lower regularization 2*, as the sparsity enhancing parameter ($\overline{\tau}^l$) is higher. In the two *Full regularization* settings, accounts 1–15 do not modify their behaviour compared to the two *Lower regularization* settings, but for accounts 16–25, controlled by upper-level managers 4 and 5 that enforce sparsity, the number of assets with no investments turns out to be higher. Notice that the regularization operated by the upper-level managers is less effective than the one operated by the lower-level problems, since they can only select porfolios among the lower-level equilibria. None-

		feas$(z_{\overline{I}}, f_{z_{\overline{I}}})$			feas$(y_{\overline{I}}, f_{y_{\overline{I}}})$
		$\overline{k}=0$	$\overline{k}=0.4\overline{I}$	$\overline{k}=0.8\overline{I}$	
SX5E	No reg	3.9860e-03	5.7453e-05	4.7687e-05	4.7442e-05
	Low. reg. 1	3.4648e-03	2.6466e-04	2.6212e-04	4.0399e-04
	Low. reg. 2	3.7396e-03	8.0372e-04	7.8346e-04	1.2101e-03
	Full reg. 1	3.4227e-03	4.1116e-04	4.0964e-04	4.1343e-04
	Full reg. 2	3.5618e-03	5.7343e-04	5.5812e-04	6.8682e-04
DIJA	No reg.	3.0745e-03	2.9030e-05	2.4679e-05	2.4539e-05
	Low. reg. 1	2.3533e-03	1.0974e-04	1.0774e-04	1.8460e-04
	Low. reg. 2	2.5532e-03	4.1920e-04	4.1520e-04	4.7494e-04
	Full reg. 1	6.0450e-03	4.1062e-04	3.6413e-04	4.6802e-04
	Full reg. 2	6.2340e-03	8.3879e-04	8.2669e-04	1.1776e-03

TABLE 4

feas$(w, f_w) = \|w - P_Y(w - f_w)\|_2$, obtained with PASTA for both datasets in Example 2, for $z_{\overline{I}}$, with different starting iterations $\overline{k}$, and $y_{\overline{I}}$, considering the five different regularization settings

	SX5E		DIJA	
#Accounts	1–15	16–25	1–15	16–25
No regularization	0.00%	0.00%	0.00%	0.00%
Lower regularization 1	28.00%	0.00%	38.62%	0.00%
Lower regularization 2	44.67%	0.00%	56.78%	0.00%
Full regularization 1	28.00%	25.00%	38.62%	12.07%
Full regularization 2	44.67%	24.00%	56.55%	11.72%

TABLE 5

Portfolio sparsity (% of assets with an investment lower than 0.1% of the budget), for the first 15 and the last 10 accounts, obtained with PASTA for both datasets in Example 2, considering the five different regularization settings

theless, the sparsity obtained by managers 4 and 5 demonstrates the influence of the upper-level game on the overall solution. This confirms the theoretical properties of PASTA, that ensure theoretical convergence to solutions of (GNEP^u).

8. Conclusions. We list the main contributions of our work below.

1. We focus on the framework of GNEPs with nonsmooth payoffs and having a hierarchical structure, i.e. the shared feasible region is implicitly defined as the set of equilibria of a lower-level NEP with nonsmooth payoffs. These problems naturally arise in real-world applications such as multi-portfolio selection with sparsity enhancing terms. Under standard conditions (see Assumptions **A**), we show that the feasible set of such GNEPs is compact, nonempty and convex (see Proposition 3.3 and Proposition 3.4). Under additional conditions (see Assumptions **B**), the GNEP equilibrium set is nonempty and bounded (see Proposition 3.8). Moreover, there exists a subset of equilibria, that we term variational solutions, which is nonempty, convex and compact. We are not aware of other contributions in this context in the literature.
2. Generalizing a classical result in the smooth context, one can rely on a hierarchical GVI structure to compute variational equilibria of the original hierarchical GNEP. We study conditions that make the hierarchical GVI numerically tractable by exploiting the techniques described below.
3. We combine Tikhonov-like penalization techniques with averaged gradient-like approaches to prove convergence and obtain complexity guarantees under mild conditions (Assumptions **C** and **D**) that, requiring the upper and lower-level mappings to be just maximal monotone, are the most general among the ones relied upon in the literature (see Theorem 5.2 and Theorem 5.3).

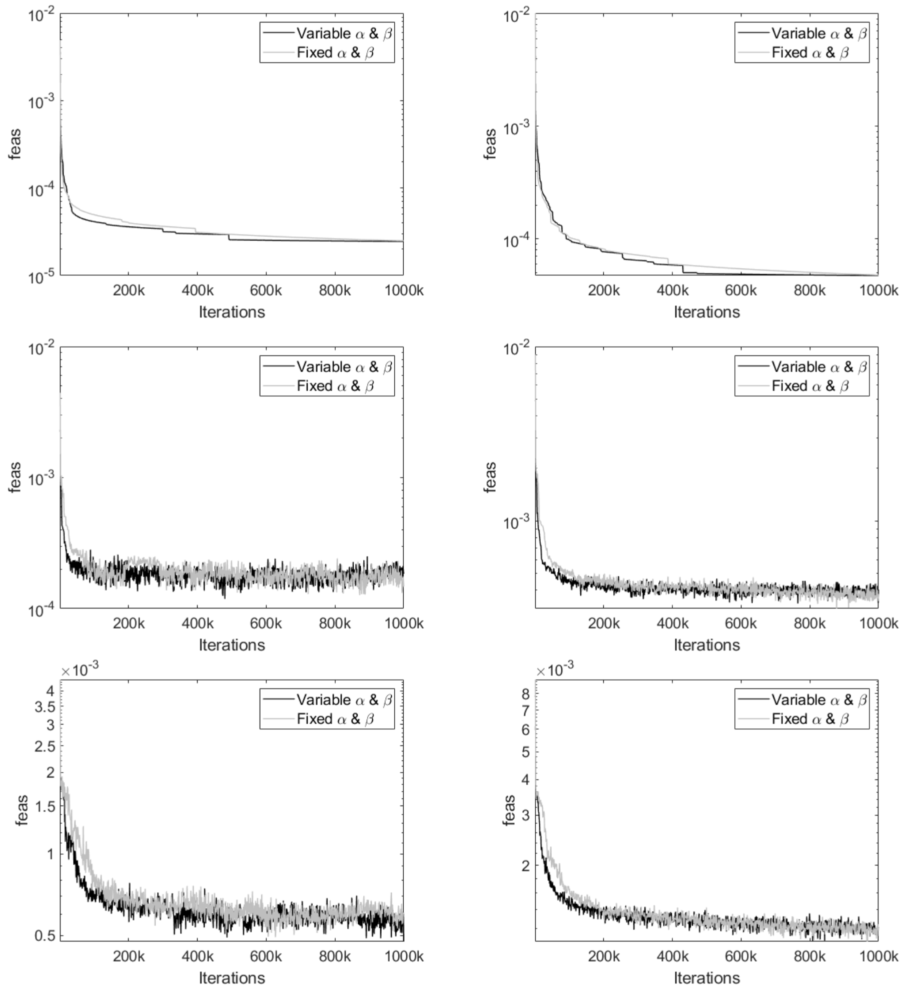

FIG. 2. *Comparison between variable (PASTA) and fixed* $\{\alpha_k\}$ *and* $\{\beta_k\}$ *considering* $\operatorname{feas}(y_k, f_{y_k}) = \|y_k - P_Y(y_k - f_{y_k})\|_2$, *for the SX5E (left-hand side) and the DIJA (right-hand side) datasets, in the cases of No regularization, Lower regularization 1 and 2, respectively*

4. Exploiting the theoretical insights concerning the faster convergence to the subproblem solutions (Theorem 4.5, Theorem 4.7 and Table 1), we propose the Projected Average Single-loop Tikhonov Algorithm that gradually satisfies the requirements in Assumptions **D**. We confirm PASTA's theoretical properties and show that it works well in practice through numerical tests.
5. Focusing on the motivating example of multi-portfolio selection, we apply and test our approach on the novel model presented in [18]. Multi-portfolio selection turns out to be numerically tractable under standard conditions. The numerical results validate the modeling choices: e.g. the computed portfolio turns out to be sparse due to the nonsmooth regularization term.

As future research, we wish to consider Newton-like algorithms to speed up computations and compute non-variational equilibria. We would like to encompass in our analysis enlargements of the set-valued mappings to recover continuity properties.

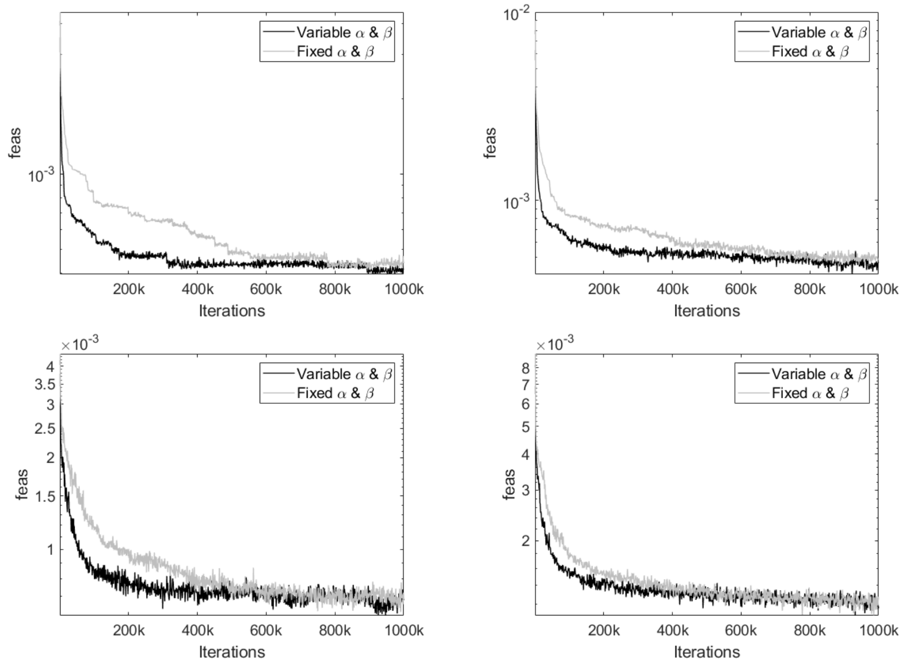

FIG. 3. *Comparison between variable (PASTA) and fixed $\{\alpha_k\}$ and $\{\beta_k\}$ considering* $\text{feas}(y_k, f_{y_k}) = \|y_k - P_Y(y_k - f_{y_k})\|_2$, *for the SX5E (left-hand side) and the DIJA (right-hand side) datasets, in the cases of Full regularization 1 and 2, respectively*

Appendix A. Additional results.

A.1. Proof of Proposition 3.2. If $y \in E$, then $y \in \text{SOL}(F, Y)$. By the convexity of the problems (P^l_ν) and the minimum principle, thanks to (3.1) and the convexity of Y_ν, $y \in E$ if and only if, for all $\nu = 1 \dots N$:

$$\exists \xi_\nu \in \partial_{y^\nu} \varphi^l_\nu(y^\nu) : \quad (\nabla_{y^\nu} \theta^l_\nu(y^\nu, y^{-\nu}) + \xi_\nu)^T (v^\nu - y^\nu) \geq 0 \quad \forall v^\nu \in Y_\nu.$$

Concatenating all these inequalities, (GVI^l) holds with $f_y = \left[\nabla_{y^\nu} \theta^l_\nu(y^\nu, y^{-\nu}) + \xi_\nu\right]_{\nu=1}^N$ and thus $y \in \text{SOL}(F, Y)$. Vice versa, if $y \in \text{SOL}(F, Y)$, for all $\nu = 1 \dots N$ there exists $\exists f_y \in F(y)$ such that $f_y^T((v^\nu, y^{-\nu}) - (y^\nu, y^{-\nu})) \geq 0$, $\forall (v^\nu, y^{-\nu}) \in Y$. By (3.1),

$$\exists f_y^\nu \in \nabla_{y^\nu} \theta^l_\nu + \partial_{y^\nu} \varphi^l_\nu : \quad f_y^{\nu T}(v^\nu - y^\nu) \geq 0, \quad \forall v^\nu \in Y_\nu.$$

By the convexity of player ν's problem, $y \in E$. ◻

A.2. On Maximal Monotonicity.

DEFINITION A.1. *A monotone mapping $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone if for every pair $(\widehat{u}, \widehat{t}) \in (\mathbb{R}^n \times \mathbb{R}^n) \setminus gph(T)$ there exists $(\widetilde{u}, \widetilde{t}) \in gph(T)$, where $gph(T) \triangleq \{(u,t) | u \in \mathbb{R}^n, t \in T(u)\}$, with $(\widehat{u} - \widetilde{u})^T(\widehat{t} - \widetilde{t}) < 0$.*

The following result characterizes the Carthesian product of maximal monotone mappings, and it is used to prove Proposition 3.4 and Proposition 3.9.

LEMMA A.2. *Let $S : X \rightrightarrows \widetilde{X}$ and $T : Y \rightrightarrows \widetilde{Y}$ be maximal monotone mappings. Their Carthesian product is also maximal monotone.*

Proof. If, by contradiction, $S \times T : X \times Y \rightrightarrows \widetilde{X} \times \widetilde{Y}$ is not maximal monotone, then it would mean that there exists an element

$$(\overline{x}, \overline{y}, \overline{s}_x, \overline{t}_y) \notin \text{gph}(S \times T) = \{(x, y, s_x, t_y) | x \in X, y \in Y, s_x \in S(x), t_y \in T(y)\},$$

that does not violate the monotonicity of the operator $S \times T$. That is

$$(s_x - \overline{s}_x)^T(x - \overline{x}) + (t_y - \overline{t}_y)^T(y - \overline{y}) \geq 0, \quad \forall (x, y) \in X \times Y, \quad \forall (s_x, t_y) \in S(x) \times T(y). \tag{A.1}$$

Since $(\overline{x}, \overline{y}, \overline{s}_x, \overline{t}_y) \notin \text{gph}(S \times T)$, we can assume, $(\overline{x}, \overline{s}_x) \notin \text{gph}(S)$. Due to the maximal monotonicity of S, there must exist (x, s_x) with $x \in X$ and $s_x \in S(x)$ such that $(s_x - \overline{s}_x)^T(x - \overline{x}) < 0$. From (A.1), one can deduce $(t_y - \overline{t}_y)^T(y - \overline{y}) > 0$, $\forall y \in Y$ and $\forall t_y \in T(y)$. Due to the maximal monotonicity of mapping T, this would mean $(\overline{y}, \overline{t}_y) \in \text{gph}T$, and it would be possible to choose $(y, t_y) = (\overline{y}, \overline{t}_y)$ and find $(t_y - \overline{t}_y)^T(y - \overline{y}) = (\overline{t}_y - \overline{t}_y)^T(\overline{y} - \overline{y}) = 0$, which is in contradiction with $(t_y - \overline{t}_y)^T(y - \overline{y}) > 0$, $\forall y \in Y$ and $\forall t_y \in T(y)$. ◻

A.3. Proof of Proposition 3.7. For all $\mu = 1 \dots M$, $x \in \text{SOL}(G, \text{SOL}(F, Y))$ means that for every w^μ such that $(w^\mu, x^{-\mu}) \in \text{SOL}(F, Y)$, we have

$$\begin{gathered} \exists g_x \in G(x) : \quad g_x^T((w^\mu, x^{-\mu}) - (x^\mu, x^{-\mu})) \geq 0 \iff \exists g_x^\mu \in G_\mu(x) : \quad g_x^{\mu T}(w^\mu - x^\mu) \geq 0, \\ \theta_\mu^u(x^\mu, x^{-\mu}) + \varphi_\mu^u(x^\mu) \leq \theta_\mu^u(w^\mu, x^{-\mu}) + \varphi_\mu^u(w^\mu), \quad \forall w^\mu : (w^\mu, x^{-\mu}) \in E, \end{gathered}$$

which is due to (Proposition 3.2, Proposition 3.4) convexity of player μ's problem. ◻

A.4. Proof of Proposition 3.8. The proof is obtained similarly to the one for Proposition 3.3, by recalling that, by Assumptions **A**, **B1** and **B3**, the noneptiness, compactness and convexity of $\text{SOL}(F, Y)$, the convex valuedness of G are guaranteed. G is outer-semicontinuous, so that we get the closedness of $\text{SOL}(G, \text{SOL}(F, Y))$. The set of equilibria of problem (GNEP^u) is bounded as its feasible set is compact. ◻

A.5. Proof of Proposition 3.9. Since $\left[\partial \varphi_\mu^u\right]_{\mu=1}^M$ turns out to be maximal monotone, the proof is analogous to the one of Proposition 3.4. ◻

A.6. Proof of Theorem 4.2. We have, for all $v \in Y$, $h_v^\eta \in H_\eta(v)$, $h_y^\eta \in H_\eta(y)$:

$$0 \leq (h_v^\eta - h_y^\eta)^T(v - y) = h_v^{\eta T}(v - y) - h_y^{\eta T}(v - y) \leq h_v^{\eta T}(v - y),$$

which follows from the monotonicity of H_η and since y is a solution of (4.1), and we can select $h_y^\eta \in H_\eta(y)$ such that $h_y^{\eta T}(v - y) \geq 0$, for all $v \in Y$. ◻

A.7. Proof of Theorem 4.3. For any $v \in Y$ we define $u^\tau \triangleq \tau y + (1 - \tau)v$, $\tau \in (0, 1)$. Since $u^\tau \in Y$ by the convexity of Y, if y is a solution of (4.2), for all $h_{u^\tau}^\eta \in H_\eta(u^\tau)$,

$$0 \leq h_{u^\tau}^{\eta T}(u^\tau - y) = h_{u^\tau}^{\eta T}(\tau y + (1 - \tau)v - y) = (1 - \tau)h_{u^\tau}^{\eta T}(v - y) \leq h_{u^\tau}^{\eta T}(v - y)$$

Considering $\tau \to 1$, we have $u^\tau \underset{Y}{\longrightarrow} y$, and because H_η is compact-valued over Y, for an appropriately chosen subsequence of τ, and consequently of u^τ, there exists a sequence of $h_{u^\tau}^\eta$, with $h_{u^\tau}^\eta \in H_\eta(u^\tau)$ such that $h_{u^\tau}^\eta \to \overline{h}_u^\eta$. Since H_η is outer-semicontinuous, $\overline{h}_u^\eta \in H_\eta(y)$. This implies, for all $v \in Y$, $\exists \overline{h}_u^\eta \in H_\eta(y) : \overline{h}_u^{\eta T}(v - y) \geq 0$. ◻

A.8. Averaging Sequences. The proof of the next lemma can be traced back to [14, Point 1 in Section 2.4.2].

LEMMA A.3. *Let* $\{a_k\}$ *and* $\{b_k\}$ *be sequences of positive real numbers such that:* $\lim_{k\to\infty} a_k = \overline{a}$, $\sum_{k=1}^{\infty} b_k = \infty$. *Then,* $\lim_{k\to\infty} \sum_{j=1}^{k} b_j a_j / \sum_{j=1}^{k} b_j = \overline{a}$.

A.9. Proof of point b) in Theorem 4.5. Assume by contradiction $\{y_k\}$ admits a limit vector $\overline{y} \notin \text{SOL}(H_\eta, Y)$. Due to **C1**, together with Lemma A.3, $z_k \to \overline{y}$, and, by Theorem 4.7, we have the contradiction $\overline{y} \in \text{SOL}(H_\eta, Y)$. □

A.10. On Inexactness. First, we give the proof of Proposition 6.3.

Proof of Proposition 6.3. For all $v \in Y$, for all $f_v \in F(v)$, $h_v^\eta = f_v + \eta g_v \in H_\eta(v)$,

$$f_v^T(v - z_k) = h_v^{\eta T}(v - z_k) - \eta g_v^T(v - z_k) \geq -\Lambda_1 \Xi_1^k - \Lambda_2 \Xi_2^k - \Lambda_3 \Xi_3^k - \eta \Lambda_2,$$

where the inequality is due to (4.7), and thus we get (6.3). Moreover, for all $v \in \text{SOL}(F, Y)$, $\overline{f}_v \in F(v)$ exists such that $\overline{f}_v^T(z_k - v) \geq 0$, and for all $g_v \in G(v)$:

$$-\Lambda_1 \Xi_1^k + \Lambda_2 \Xi_2^k + \Lambda_3 \Xi_3^k/\eta \leq \left[\overline{f}_v/\eta + g_v\right]^T (v - z_k) \leq g_v^T(v - z_k),$$

where the first inequality comes from (4.7), and thus we get (6.4). □

We remark that it is difficult to measure how inexactness propagates from Minty-like GVI optimality conditions (like (4.7), (5.1), (5.2), (6.3), (6.4)) to the players' problems' ones. This topic does not seem to have been thoroughly investigated in the literature: some preliminary results can be traced back in [2], where however only the case of single-valued mappings is considered.

We also give the counterpart related to (4.1) of Proposition 6.3.

PROPOSITION A.4. *Given* $\varepsilon \geq 0$, *let* y *be a solution of the inexact version of* (4.1), *i.e.* $y \in Y$, $\exists h_y^\eta \in H_\eta(y)$ *such that* $h_y^{\eta T}(v - y) \geq -\varepsilon$, $\forall v \in Y$. *We have*

$$\exists f_y \in F(y): \quad f_y^T(v - y) \geq -\varepsilon - \eta \Lambda_2, \quad \forall v \in Y,$$

$$\exists g_y \in G(y): \quad g_y^T(v - y) \geq -\varepsilon/\eta, \quad \forall v \in \text{SOL}(F, Y).$$

Proof. Since $h_y^\eta = f_y + \eta g_y$, for some $f_y \in F(y)$ and $g_y \in G(y)$, for all $v \in Y$: $f_y^T(v-y) = h_y^{\eta T}(v-y) - \eta g_y^T(v-y) \geq -\varepsilon - \eta\Lambda_2$, and, as in the proof of Proposition 6.3, $-\varepsilon/\eta \leq [f_y/\eta + g_y]^T (v - y) \leq g_y^T(v - y)$, $\forall v \in \text{SOL}(F, Y)$. □

A.11. Proof of Proposition 7.1. By Theorem 6.4, we only need to prove that sequences $\{\gamma_k\}$ and $\{\eta_k\}$ are nonincreasing. Let us prove this for $\{\gamma_k\}$, therefore focusing on $\{\alpha_k\}$, since the proof for $\{\eta_k\}$ can be obtained following the same reasoning.

Clearly, $\alpha_k = \alpha$, and then $\{\gamma_k\}$ is nonincreasing, for all $k \geq I_\alpha$. For every $k \in (1, I_\alpha)$, and for every $\varepsilon_\alpha \in (0, \overline{\varepsilon}_\alpha]$, we have

$$\begin{aligned} \frac{\overline{\alpha}}{\overline{\alpha} - \alpha} - 1 \quad &= \quad \frac{\alpha}{(\overline{\alpha} - \alpha)} = \frac{t_\alpha}{1 - t_\alpha} \frac{(I_\alpha^{\overline{\varepsilon}_\alpha} - 1)}{I_\alpha^{\overline{\varepsilon}_\alpha}} \geq \frac{t_\alpha}{1 - t_\alpha} \frac{(I_\alpha^{\varepsilon_\alpha} - 1)}{I_\alpha^{\varepsilon_\alpha}} \\ &\geq \quad \frac{t_\alpha}{1 - t_\alpha} \frac{(k^{\varepsilon_\alpha} - (k-1)^{\varepsilon_\alpha})}{I_\alpha^{\varepsilon_\alpha}} \geq \frac{t_\alpha^k}{1 - t_\alpha^k} \frac{(k^{\varepsilon_\alpha} - (k-1)^{\varepsilon_\alpha})}{I_\alpha^{\varepsilon_\alpha}}, \end{aligned}$$

where $t_\alpha^k \triangleq log_k(k-1)$, and the last inequality holds since $t_\alpha^k \leq t_\alpha$, thus $(k/I_\alpha)^{\varepsilon_\alpha} \leq 1 \leq \overline{\alpha}/(\overline{\alpha} - \alpha) - t_\alpha^k/(1 - t_\alpha^k)(k^{\varepsilon_\alpha} - (k-1)^{\varepsilon_\alpha})/(I_\alpha^{\varepsilon_\alpha})$, and by rearranging terms,

$$\alpha_k = \overline{\alpha} - (\overline{\alpha} - \alpha)\,(k/I_\alpha)^{\varepsilon_\alpha} \geq t_\alpha^k \left[\overline{\alpha} - (\overline{\alpha} - \alpha)\,(k - 1/I_\alpha)^{\varepsilon_\alpha}\right] = t_\alpha^k \alpha_{k-1},$$

which implies $k^{\alpha_k} \geq \left[k^{t_\alpha^k}\right]^{\alpha_{k-1}} = (k-1)^{\alpha_{k-1}}$. □

REFERENCES

[1] H. Attouch, M.-O. Czarnecki, and J. Peypouquet, *Prox-penalization and splitting methods for constrained variational problems*, SIAM Journal on Optimization, 21 (2011), pp. 149–173.

[2] G. Bigi, L. Lampariello, S. Sagratella, and V. G. Sasso, *Approximate variational inequalities and equilibria*, to appear, (2023).

[3] R. E. Bruck Jr, *On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in hilbert space*, Journal of Mathematical Analysis and Applications, 61 (1977), pp. 159–164.

[4] A. Cabot, *Proximal point algorithm controlled by a slowly vanishing term: applications to hierarchical minimization*, SIAM Jour. on Opt., (2005), pp. 555–572.

[5] F. H. Clarke, *Optimization and nonsmooth analysis*, SIAM, 1990.

[6] A. Dreves, F. Facchinei, C. Kanzow, and S. Sagratella, *On the solution of the KKT conditions of generalized Nash equilibrium problems*, SIAM Journal on Optimization, 21 (2011), pp. 1082–1108.

[7] F. Facchinei and C. Kanzow, *Generalized nash equilibrium problems*, 4or, 5 (2007), pp. 173–210.

[8] F. Facchinei and L. Lampariello, *Partial penalization for the solution of generalized Nash equilibrium problems*, Jour. of Global Opt., 50 (2011), pp. 39–57.

[9] F. Facchinei, J.-S. Pang, G. Scutari, and L. Lampariello, *Vi-constrained hemivariational inequalities: distributed algorithms and power control in ad-hoc networks*, Mathematical Programming, 145 (2014), pp. 59–96.

[10] F. Facchinei and S. Sagratella, *On the computation of all solutions of jointly convex generalized Nash equilibrium problems*, Optimization Letters, 5 (2011), pp. 531–547.

[11] S. Fang and E. Peterson, *Generalized variational inequalities*, Journal of optimization theory and applications, 38 (1982), pp. 363–383.

[12] V. V. Kalashnikov and N. I. Kalashinikova, *Solving two-level variational inequality*, Journal of Global Optimization, 8 (1996), pp. 289–294.

[13] H. D. Kaushik and F. Yousefian, *A method with convergence rates for optimization problems with variational inequality constraints*, SIAM Journal on Optimization, 31 (2021), pp. 2171–2198.

[14] K. Knopp, *Infinite sequences and series*, Courier Corporation, 1956.

[15] L. Lampariello, C. Neumann, J. M. Ricci, S. Sagratella, and O. Stein, *An explicit Tikhonov algorithm for nested variational inequalities*, Computational Optimization and Applications, 77 (2020), pp. 335–350.

[16] L. Lampariello, C. Neumann, J. M. Ricci, S. Sagratella, and O. Stein, *Equilibrium selection for multi-portfolio optimization*, European Journal of Operational Research, (2021).

[17] L. Lampariello, G. Priori, and S. Sagratella, *On the solution of monotone nested variational inequalities*, Mathematical Methods of Operations Research, (2022), pp. 1–26.

[18] L. Lampariello, S. Sagratella, and V. G. Sasso, *Nonsmooth hierarchical multi portfolio selection*, in Optimization and Decision Science: Operations Research, Inclusion and Equity: ODS, Florence, Italy, August 30 - September 2, 2022, Springer, 2023, p. to appear.

[19] X. Lu, H.-K. Xu, and X. Yin, *Hybrid methods for a class of monotone variational inequalities*, Nonlinear Analysis, (2009), pp. 1032–1041.

[20] G. Marino and H.-K. Xu, *Explicit hierarchical fixed point approach to variational inequalities*, Journal of Optimization Theory and Applications, 149 (2011), pp. 61–78.

[21] J.-S. Pang and D. Chan, *Iterative methods for variational and complementarity problems*, Mathematical programming, 24 (1982), pp. 284–313.

[22] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.

[23] S. Sabach and S. Shtern, *A first order method for solving convex bilevel optimization problems*, SIAM Journal on Optimization, 27 (2017), pp. 640–660.

[24] M. Solodov, *An explicit descent method for bilevel convex optimization*, Journal of Convex Analysis, 14 (2007), p. 227.

[25] I. Yamada, *The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings*, Inherently parallel algorithms in feasibility and optimization and their applications, 8 (2001), pp. 473–504.