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. [17, 19] 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 (see [7] and [11]) 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 ([23] 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, 20, 22, 26, 29] and Tikhonov methods [1, 4, 10, 13, 14, 16, 18, 28]. 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.

*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).

We compute variational equilibria of the hierarchical GNEP through the corresponding 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, 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 5.3).

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 [17].

2. The hierarchical jointly-convex Generalized Nash Equilibrium model. We address Generalized Nash Equilibrium Problems (GNEP) whose shared feasible set 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 study assumptions concerning the upper-level GNEP that ensure 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 faced by player ν, with $\nu = 1, \dots, N$, 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: $y \triangleq (y^1 \cdots y^N)^\top \in \mathbb{R}^p$, $y^{-\nu} \triangleq (y^1 \cdots y^{\nu-1}, y^{\nu+1} \cdots y^N) \in \mathbb{R}^{p-n_\nu}$, 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 $\theta_\nu^l : \mathbb{R}^p \to \mathbb{R}$ depending on variables y^ν as well as on the variables $y^{-\nu}$, and a nonsmooth term $\varphi_\nu^l : \mathbb{R}^{n_\nu} \to \mathbb{R}$ depending on variables y^ν only. Summarizing, the NEP we consider consists of the collection of player ν's parametric optimization problems

$$(\mathrm{P}_\nu^l) \qquad \operatorname{minimize}_{y^\nu} \theta_\nu^l(y^\nu, y^{-\nu}) + \varphi_\nu^l(y^\nu) \quad \text{s.t.} \quad y^\nu \in Y_\nu,$$

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 problem to

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

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 $E \triangleq \{y \in Y : \theta_\nu^l(y^\nu, y^{-\nu}) + \varphi_\nu^l(y^\nu) \le \theta_\nu^l(v^\nu, y^{-\nu}) + \varphi_\nu^l(v^\nu), \forall v^\nu \in Y_\nu,\ \nu = 1, \dots, N\} \subseteq \mathbb{R}^p$ the (non-parametric) set of equilibria of the NEP.

Assumptions A

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

A2 θ_ν^l is convex with respect to y^ν, for every $\nu = 1, \dots, N$;

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

A4 φ_ν^l is convex and locally Lipschitz, for every $\nu = 1, \dots, N$.

From assumption **A4**, one can immediately deduce that $\partial_{y^\nu} \varphi_\nu^l$ is locally bounded and outer-semicontinuous for every $\nu = 1, \dots, N$, where the operator ∂_{y^ν} indicates the set of subgradients with respect to player ν's variables. Furthermore, $\partial_{y^\nu} \varphi_\nu^l$ 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]. In the forthcoming developments, we show that E is nonempty convex and compact (see section 3).

2.2. The upper-level GNEP. Considering the upper-level hierarchical GNEP, player μ, with $\mu = 1, \dots, M$, controls 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:

$$(\mathrm{P}_\mu^u) \qquad \operatorname{minimize}_{x^\mu} \theta_\mu^u(x^\mu, x^{-\mu}) + \varphi_\mu^u(x^\mu) \quad \text{s.t.} \quad (x^\mu, x^{-\mu}) \in E,$$

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 belong therefore to the solution set of the lower-level NEP, where 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:

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

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. In a later stage of the analysis, we 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 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 GVIs are instrumental 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 Stampacchia GVI:

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

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

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 subdifferentially regular (see [25, Proposition 7.27]). So, we can write (see [25, 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 prove that, under Assumptions **A**, (NEP^l) can be recast as (GVI^l), whose solution set is denoted by $\text{SOL}(F, Y)$.

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

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

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

Proof. To prove the nonemptiness of E, we rely on [12, 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**. $E \subseteq Y$ 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

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

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

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

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 $0 \leq \lim_{k\in\mathcal{K}} f_{y_k}^\top (v - y_k) = \overline{f}^\top (v - \overline{y})$, for all $v \in Y$, which contradicts (3.3). ∎

PROPOSITION 3.4. *Under Assumptions* ***A****,* F *is maximal monotone (see Definition* A.2 *in Appendix* A.2*) and* $\mathrm{SOL}(F, Y)$*, 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 [25, Proposition 12.7]). On the other hand, under assumption **A4**, φ_ν^l is continuous and convex so that the point to set map $\partial_{y^\nu} \varphi_\nu^l$ is maximal monotone (see [25, Proposition 12.17]). By Lemma A.4 in Appendix A.2, $\left[\partial \varphi_\nu^l\right]_{\nu=1}^N$ is maximal monotone. Since the sum of maximal monotone operators is maximal monotone under mild conditions (more in detail, as long as $\mathrm{rint}\ (\mathrm{dom} \nabla_{y^\nu} \theta_\nu^l) \cap \mathrm{rint}(\mathrm{dom}\ \partial_{y^\nu} \varphi_\nu^l) \neq \emptyset)$, see [25, Proposition 12.44]), we can deduce that the mapping F is maximal monotone. Recalling [12, Theorem 4.4], the convexity of $\mathrm{SOL}(F, Y)$ and E follows, since Y is nonempty and convex, and F is maximally monotone. ∎

The following result states that (GNEP^u) belongs to the class of problems termed jointly-convex enjoying favorable theoretical properties (see [7] and [11]).

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

Proof. By Proposition 3.4, E is convex, and the claim 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 Stampacchia GVI can be relied on to compute solutions of (GNEP^u):

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

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

Remark 3.6. Similarly to the lower level, under Assumptions **B**, we have $G(x) = \left[\nabla_{x^\mu} \theta_\mu^u(x)\right]_{\mu=1}^M + \left[\partial_{x^\mu} \varphi_\mu^u(x^\mu)\right]_{\mu=1}^M$, for all $x \in Y$. The operator G is also outer-semicontinuous, by the same reasoning in Remark 3.1 for mapping 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), denoted 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. [11], 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.2 *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 Stampacchia GVI subproblem reads as

$$\text{find} \quad y \in Y : \quad \exists h_y^\eta \in H_\eta(y) : \quad h_y^{\eta\top}(v - y) \geq 0, \quad \forall v \in Y. \tag{4.1}$$

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:

$$\text{find} \quad y \in Y : \quad h_v^{\eta\top}(v - y) \geq 0, \quad \forall v \in Y, \quad \forall h_v^\eta \in H_\eta(v). \tag{4.2}$$

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.8 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 result.

THEOREM 4.2. *Under Assumptions **A**, **B** y is a solution of* (4.1) *if and only if it is a solution of* (4.2).

Proof. The claim is a consequence of Proposition 4.1 that ensures the maximal monotonicity of H_η, and follows from Lemma A.5, where $S = H_\eta$ and $K = Y$. ◻

We remark that setting $\eta = 0$, problems (4.1) and (4.2) boil down to (GVI^l) and its Minty counterpart, respectively. Thus, the following straightforward consequence of Theorem 4.2 holds.

THEOREM 4.3. *Under Assumptions A, y is a solution of* (GVI^l) *if and only if it is a solution of its Minty version:*

$$\textit{find} \quad y \in Y : \quad f_v^\top (v - y) \geq 0, \quad \forall v \in Y, \quad \forall f_v \in F(v).$$

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

To compute a point in $\mathrm{SOL}(H_\eta, Y)$ 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:

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

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 $\eta_k \to \eta$, for some $\eta \geq 0$.

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 $\mathrm{SOL}(H_\eta, Y)$, see e.g. [16]. However, we show in Theorem 4.5 that the distance of y_k from any $u \in \mathrm{SOL}(H_\eta, Y)$ 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 $u \in \mathrm{SOL}(H_\eta, Y)$, and for every $k \geq 1$, we have:

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

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:

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

where the latter inequality holds because $u \in \mathrm{SOL}(H_\eta, Y)$, Theorem 4.2, and due to the following relation, since $\{\eta_j\}$ is non-increasing by assumption **C2**:

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

Summing j from k to $k + \Delta - 1$, we get

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

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 let Assumptions* ***C*** *hold. The following statements hold:*

- ***a)*** *if* $\{\gamma_k\} \in \ell^2$, *that is,* $\sum_{k=1}^\infty \gamma_k^2 < \infty$ *and* $\{\gamma_k(\eta_k - \eta)\} \in \ell^1$ *(with* η *defined in Assumption* ***C2****), given any* $u \in \mathrm{SOL}(H_\eta, Y)$, *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,* $\overline{y} \in \mathrm{SOL}(H_\eta, Y)$;
- ***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 $\Psi_2^k \to 0$:

$$\limsup_{\ell\to\infty} \|y_\ell - u\|^2 \leq \liminf_{k\to\infty} \|y_k - u\|^2 + 2\Lambda_1 \lim_{k\to\infty} \Psi_1^k + 2\Lambda_2 \lim_{k\to\infty} \Psi_2^k = \liminf_{k\to\infty} \|y_k - u\|^2.$$

The proof of **b)** is reported in Appendix A.8. 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 $\mathrm{SOL}(H_\eta, Y)$ 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 $k \geq 1$,

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

Note that, since $\gamma_k > 0$ for every k, $z_k \in Y$ due to the convexity of Y.

In Theorem 4.7 we show that $\{z_k\}$ converges to $\mathrm{SOL}(H_\eta, Y)$. 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:

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

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,

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

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

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

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

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

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 $\mathrm{SOL}(H_\eta, Y)$.*

Proof. The proof is obtained by observing that $\Xi_1^k, \Xi_2^k \to 0$ in view of Lemma A.6 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 $\liminf_{k\to\infty} h_v^{\eta\top}(v - z_k) \geq 0$, 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 4.2 they belong to $\mathrm{SOL}(H_\eta, Y)$.

In the sequel, we prove that $\{z_k\}$ has actually a single limit point. For every $u_1, u_2 \in \mathrm{SOL}(H_\eta, Y)$, by convexity, $\frac{u_1+u_2}{2} \in \mathrm{SOL}(H_\eta, Y)$, see Proposition 4.1. Combining point **a)** in Theorem 4.5 and Lemma A.6 in Appendix A.7, we can say that $\exists\, l_{\left(\frac{u_1+u_2}{2}\right)}, l_{u_1} \in \mathbb{R}$ such that

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

For every $j \geq 1$ we have:

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

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

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

Taking the limit on both sides, we get

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

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 $\overline{z}, \widetilde{z} \in \mathrm{SOL}(H_\eta, Y)$. The last equation implies $(\overline{z} - \widetilde{z})^\top (u_1 - u_2) = (\overline{z} - u_1)^\top (u_1 - u_2) - (\widetilde{z} - u_1)^\top (u_1 - u_2) = 0$. 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 $\mathrm{SOL}(H_\eta, Y)$. The points in $\mathrm{SOL}(H_0, Y)$ correspond to the solutions of (GVI^l), and thus are feasible for (GVI^u): 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 $\mathrm{SOL}(H_\eta, Y)$, that corresponds to solving, depending on η, (GVI^l) and (GVI^u) inexactly as will be shown in the forthcoming developments (see Proposition 6.3 and Appendix A.9): namely, considering Proposition 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 $\mathrm{SOL}(H_\eta, Y)$ 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 $\mathrm{SOL}(H_0, Y)$ and in $\mathrm{SOL}(G, \mathrm{SOL}(F, Y))$, 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).

Similarly to the results in Theorem 4.2 and Theorem 4.3, here we state the equivalence between (GVI^u) and its Minty counterpart: this result is crucial to recover solutions to the Stampacchia (GVI^u) from the ones of the Minty version of it.

THEOREM 5.1. *Under Assumptions A and B, y is a solution of* (GVI^u) *if and only if it is a solution of its Minty counterpart:*

$$\textit{find} \quad y \in \mathrm{SOL}(F,Y): \quad g_v^\top(v-y) \geq 0, \quad \forall v \in \mathrm{SOL}(F,Y), \quad \forall g_v \in G(v).$$

Proof. Since G is maximal monotone due to Proposition 3.9, and $\mathrm{SOL}(F,Y)$ is convex due to Proposition 3.4, the claim follows from Lemma A.5 with $S = G$ and $K = \mathrm{SOL}(F,Y)$. ∎

We define the following bounds for the Minty versions of (GVI^l) and (GVI^u).

THEOREM 5.2. *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:

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

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

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 $-2\gamma_j(f_v + \eta_j g_v)^\top(v - y_j) \leq \|y_j - v\|^2 - \|y_{j+1} - v\|^2 + 2\Lambda_1 \gamma_j^2$. Since $v \in \mathrm{SOL}(F,Y)$, $\overline{f}_v \in F(v)$ exists such that $\overline{f}_v^\top(y_j - v) \geq 0$, and then,

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

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

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

By observing that

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

we obtain $-2g_v^\top(v - z_k) \le 2\Lambda_1 \Phi_1^k + 2\Lambda_3 \frac{1}{\eta_k \sum_{j=1}^k \gamma_j}$, 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 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 the ones considered in [18], 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 a general set-valued framework, it is not practical to control the accuracy in the solution of the Tikhonov subproblems, and therefore Assumptions **D** are required to obtain following result.

THEOREM 5.3. *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.6, 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 and Theorem 5.1. ◻

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 5.4. *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:

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

with $\overline{\gamma} > 0$, $\overline{\eta} > 0$, $\alpha, \beta > 0$ and $\eta \ge 0$.

The first result deals with the complexity of satisfying relation (4.4) up to a prescribed accuracy δ: this is related to the distance of $\{y_k\}$ from any solution $u \in \mathrm{SOL}(H_\eta, Y)$.

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:

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

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:

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

A sufficient condition to get $2\Lambda_2\Psi_2^k < \delta/2$, is $k > 1 + \left(\frac{4\Lambda_2\overline{\gamma\eta}}{\alpha+\beta-1}\right)^{\frac{1}{\alpha+\beta-1}} \left(\frac{1}{\delta}\right)^{\frac{1}{\alpha+\beta-1}}$, 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 $\|y_k - u\|^2$ for any $u \in \mathrm{SOL}(H_\eta, Y)$ satisfying

$$\delta \ge l_u - \|y_k - u\|^2, \tag{6.2}$$

where $\lim_{k\to\infty} \|y_k - u\|^2 = l_u$, is $\mathcal{O}(\delta^{-1/(1-2\epsilon)})$. Note that, although, due to Theorem 4.5, the whole bound $|l_u - \|y_k - u\|^2| \le \delta$ is guaranteed to hold for some $k \ge \lambda_1 \left(\frac{1}{\delta}\right)^{\frac{1}{1-2\varepsilon}}$, an estimate for such iteration index is not readily available.

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

$$\frac{k^{(1-\alpha)}}{2(1-\alpha)} \le \sum_{j=1}^{k} \frac{1}{j^\alpha} \le \frac{k^{(1-\alpha)}}{1-\alpha} + \frac{-\alpha}{1-\alpha}, \tag{6.3}$$

where the lower bound holds for $k \ge 2^{\frac{2}{1-\alpha}}$. 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* ***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*

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

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

Proof. Assumptions **C** trivially hold under the conditions on α and β. The bounds defined in (6.3) imply, under our assumptions on α and β,

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

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 for $\Lambda_1\Xi_1^k < \delta/3$ to be satisfied, 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 for $\Lambda_2\Xi_2^k < \delta/3$ to hold, 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: $\Xi_3^k = \dfrac{1}{\sum_{j=1}^{k}\gamma_j} \leq \dfrac{2(1-\alpha)}{\overline{\gamma}}k^{\alpha-1}$, 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 within a tolerance 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 6.3.

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

$$f_v^\top(v-z_k) \geq -\Lambda_1\Xi_1^k - \Lambda_2\Xi_2^k - \Lambda_3\Xi_3^k - \eta\Lambda_2, \quad \forall v \in Y, \quad \forall f_v \in F(v), \tag{6.4}$$

$$g_v^\top(v-z_k) \geq -\frac{\Lambda_1\Xi_1^k + \Lambda_2\Xi_2^k + \Lambda_3\Xi_3^k}{\eta}, \quad \forall v \in SOL(F,Y), \quad \forall g_v \in G(v). \tag{6.5}$$

Proof. See Appendix A.9. □

Notice that Proposition 6.3 works only for $\eta > 0$ and there is no value for η that let the approximation errors given in (6.4) and (6.5) be zero simultaneously. By considering the bounds obtained in Theorem 5.2, 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 $\lambda_3 \triangleq \max\left\{ \left(\frac{8\Lambda_1 \overline{\gamma}(1-\alpha)}{\overline{\eta}(1+\beta-2\alpha)}\right)^{\frac{1}{\alpha-\beta}}, \left(\frac{8\Lambda_1 \overline{\gamma}(\beta-2\alpha)(1-\alpha)}{\overline{\eta}(1+\beta-2\alpha)}\right)^{\frac{1}{1-\alpha}}, \left(\frac{4\Lambda_3(1-\alpha)}{\overline{\gamma\eta}}\right)^{\frac{1}{(1-\alpha-\beta)}} \right\}.$

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 (6.3) we can write:

$$\sum_{j=1}^{k} \gamma_j = \overline{\gamma} \sum_{j=1}^{k} \frac{1}{j^\alpha} \geq \overline{\gamma} \frac{k^{1-\alpha}}{2(1-\alpha)} \tag{6.6}$$

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

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 for $\Lambda_1 \Phi_1^k < \delta/2$ to hold, is to have

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

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

$$\Phi_2^k \triangleq \frac{1}{\eta_k \sum_{j=1}^{k} \gamma_j} \leq \frac{2(1-\alpha)}{\overline{\gamma\eta}} k^{\alpha+\beta-1}, \tag{6.8}$$

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

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

Moreover, assumption **D3** holds due to relation (6.8), since $\alpha + \beta < 1$. $\square$

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 [14].

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

Remark 6.5. We assume F and G to be single-valued, Lipschitz continuous with moduli L_F and L_G, and monotone and strongly monotone with constant μ_G on Y, respectively. Under these conditions, $\text{SOL}(H_\eta, Y) = \{\widehat{y}_\eta\}$. In this case, bounds improving on the ones in Theorem 6.4 can be obtained, as shown next.

We take, for every k, $\eta_k = \eta \in (0,1)$, $\gamma_k = c_1\eta$, $c_1 = \frac{\mu_G}{(L_F+L_G)^2}$. Due to standard reasoning such as, e.g. the one in [9, Theorem 12.1.2], we have $\|y^{k+1} - \widehat{y}_\eta\|^2 \leq \left[1-(c_1\eta)^2\right] \|y^k - \widehat{y}_\eta\|^2$. As a consequence, for every $v \in Y$,

$$
\begin{aligned}
-(\overline{F}+\overline{G})\left(1-c_1^2\eta^2\right)^{\frac{k}{2}}\|y^0-\widehat{y}_\eta\| &\leq -(\overline{F}+\overline{G})\|y^k-\widehat{y}_\eta\| \leq H_\eta(v)^\top(\widehat{y}_\eta - y^k) \\
&\leq H_\eta(v)^\top(v-\widehat{y}_\eta) + H_\eta(v)^\top(\widehat{y}_\eta - y^k) = H_\eta(v)^\top(v-y^k),
\end{aligned}
\tag{6.9}
$$

where the third inequality is due to $\text{SOL}(H_\eta, Y) = \{\widehat{y}_\eta\}$.

Given any tolerance $\delta \in (0,1)$, we count the number of iterations $\overline{k}$ that are needed in order to get, for every $k \geq \overline{k}$,

$$F(v)^\top(v-y_k) \geq -\delta, \qquad \forall v \in Y, \tag{6.10}$$

$$G(v)^\top(v-y_k) \geq -\delta, \qquad \forall v \in \text{SOL}(F,Y). \tag{6.11}$$

Recalling Proposition A.7 and (6.9), the conditions above are satisfied by considering $\delta \geq \varepsilon^k + \eta\Lambda_2 = \frac{\varepsilon^k}{\eta}$, $\varepsilon^k = c_2\left(1-c_1^2\eta^2\right)^{\frac{k}{2}}$, $c_2 = (\overline{F}+\overline{G})\|y^0-\widehat{y}_\eta\|$. Thus, since $\varepsilon^k = \Lambda_2\eta^2/(1-\eta)$, we can take $\eta = \frac{\delta}{c_3}$, $\varepsilon^k \leq \delta\eta = \frac{\delta^2}{c_3}$, where $c_3 = \Lambda_2 + 1$. The last inequality is equivalent to $\left(1-\frac{c_1^2}{c_3^2}\delta^2\right)^{\frac{k}{2}} \leq \frac{\delta^2}{c_2c_3}$. Hence, $k \geq 2\log_{\left(1-\frac{c_1^2}{c_3^2}\delta^2\right)} \frac{\delta^2}{c_2c_3} = \frac{2\ln\frac{\delta^2}{c_2c_3}}{\ln\left(1-\frac{c_1^2}{c_3^2}\delta^2\right)}$ and $\overline{k} = \left\lceil \frac{2\ln\frac{\delta^2}{c_2c_3}}{\ln\left(1-\frac{c_1^2}{c_3^2}\delta^2\right)} \right\rceil = \left\lceil \frac{-\frac{4}{\zeta}\ln\frac{\delta^\zeta}{(c_2c_3)^{\zeta/2}}}{-\ln\left(1-\frac{c_1^2}{c_3^2}\delta^2\right)} \right\rceil$, for any $\zeta > 0$. Since, without loss of generality, we can consider $\frac{\delta^\zeta}{(c_2c_3)^{\zeta/2}} \in (0,1)$, $-\frac{4}{\zeta}\ln\frac{\delta^\zeta}{(c_2c_3)^{\zeta/2}} \leq \frac{4}{\zeta}\frac{(c_2c_3)^{\zeta/2}}{\delta^\zeta}$ and $-\ln\left(1-\frac{c_1^2}{c_3^2}\delta^2\right) \geq \frac{c_1^2}{c_3^2}\delta^2$, we have $\overline{k} \leq \left\lceil \frac{\frac{4}{\zeta}\frac{(c_2c_3)^{\zeta/2}}{\delta^\zeta}}{\frac{c_1^2}{c_3^2}\delta^2} \right\rceil = \left\lceil \frac{4c_2^{\zeta/2}c_3^{2+\zeta/2}}{c_1^2\zeta}\frac{1}{\delta^{2+\zeta}} \right\rceil$. Thus, if for example, we fix $\zeta = 0.01$, 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 $\mathcal{O}(\delta^{-2.01})$, that is a bound improving on the one for the general case (see Theorem 6.4).

All the considerations above are still valid considering a varying $\eta_k \geq \eta \in (0,1)$ such that $\eta_k = \eta$ for every $k \geq \overline{k}$.

Remark 6.6. Relying on conditions (6.1) with $\alpha = \beta = 1/2$ (Assumptions C are verified while Assumptions D are not), one can resort to the results in Theorem 6.2 and Proposition 6.3 in order to present an alternative analysis to obtain similar bounds as the ones in Theorem 6.4, where $\alpha = 1/2$ and $\beta = 1/4$. Following the same line of reasoning as in Remark 6.5, to verify relations (6.10) and (6.11), we consider $\delta \geq \varepsilon^k + \Lambda_2 \eta = \frac{\varepsilon^k}{\eta}$, where $\varepsilon^k = \Lambda_1 \Xi_1^k + \Lambda_2 \Xi_2^k + \Lambda_3 \Xi_3^k$ and, in turn, $\eta = \frac{\delta}{c_3}$, $\varepsilon^k \leq \delta\eta = \frac{\delta^2}{c_3}$. Due to Theorem 6.2, we have $\overline{k} = \left\lceil \lambda_2 \left(\frac{1}{\frac{\delta^2}{c_3}} \right)^2 \right\rceil = \left\lceil \lambda_2 c_3^2 \left(\frac{1}{\delta^4} \right) \right\rceil$, and 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 $\mathcal{O}(\delta^{-4})$.

Remark 6.7. Let us consider the specific case where the lower-level mapping F is single-valued, and the upper level consists in the minimization of a convex, possibly nonsmooth function (i.e. the upper-level mapping G is the subdifferential of that function). The resulting hierarchical optimization problem has been extensively studied in [14], where the authors develop an averaged projected gradient algorithm, achieving complexity bounds that are similar to the ones we obtain in our general framework. The theory we have to exploit to handle multivaluedness in hierarchical settings, in order to establish the main properties of the algorithmic approaches, their convergence and complexity guarantees, presents distinctive challenges.

The link between solutions of Minty and Stampacchia versions of (GVI^l) when multivalued mappings are concerned requires a more elaborate analysis than the one where single-valued operators are involved. In Lemma A.5, the equivalence between these two versions is established, which is instrumental in establishing convergence (see Theorem 5.3 and Theorem 4.3) and to identify bounds (see section 5). For this, we rely on the maximal monotonicity of the multivalued mapping F (S in Lemma A.5). Moreover, to guarantee compactness and nonemptiness of the solution set of (GVI^l), F must be outer-semicontinuous and locally bounded (see Proposition 3.3). These conditions are trivially satisfied if F is continuous, monotone and single-valued (see Appendix A.2), and our assumptions are therefore peculiar to the multivalued case.

As for (GVI^u), since we cannot rely on an upper-level objective function value (as in [14]), we consider Minty GVI optimality-like conditions. Again, to exploit Lemma A.5, one has to rely on the maximal monotonicity of G (S in Lemma A.5). Even more, nonemptiness, closedness, and convexity of $SOL(F, Y)$ (K in Lemma A.5) can be proven by means of Proposition 3.3 and Proposition 3.4, respectively. There, outer-semicontinuity, local boundedness, and maximal monotonicity of F, play again a key role (see also Appendix A.2). To guarantee nonemptiness and compactness of the solution set to (GVI^u), G is required to be outer-semicontinuous and locally bounded (see Proposition 3.8). All these properties are crucial to study the behavior of the solution procedures, as for their convergence and complexity guarantees (see Theorem 5.3 and Theorem 5.1), and to identify bounds (see section 5).

We identify classes of multi-agent optimization problems with nonsmooth payoffs for which the aforementioned theoretical assumptions related to multivalued operators are satisfied (see section 2, section 3 and Appendix A.2).

Remark 6.8. Bounds (5.1) and (5.2) can be stated as

$$\psi^l(z_k) \triangleq \min_{v \in Y} \min_{f_v \in F(v)} f_v^\top (v - z_k) \geq -\delta_k \tag{6.12}$$

$$\psi^u(z_k) \triangleq \min_{v \in \mathrm{SOL}(F,Y)} \min_{g_v \in G(v)} g_v^\top (v - z_k) \geq -\delta_k, \tag{6.13}$$

where $\delta^k \geq \max\left\{\Lambda_1 \Xi_1^k + \Lambda_2 \Xi_2^k + \Lambda_3 \Xi_3^k,\, \Lambda_1 \Phi_1^k + \Lambda_3 \Phi_2^k\right\}$, respectively. Due to $z_k \in Y$, $\psi^l(z_k) \leq 0$, but having just z_k such that $0 \geq \psi^l(z_k) \geq -\delta^k$ does not guarantee $z_k \in \mathrm{SOL}(F, Y)$. At the upper level, $\psi^u(z_k)$ might be positive and an upper bound for it depending on δ_k is not readily available. This does not have an impact on the limiting behaviour of $\{z_k\}$, namely its convergence to solutions of (GVI^u), which is still guaranteed due to $z_k \to \overline{z} \in \mathrm{SOL}(F, Y)$ (see Theorem 5.3). As for the complexity analysis, it is useful to identify classes of problems for which an upper bound depending on δ^k for $\psi^u(z_k)$ can be obtained. Any such bound should be traced back to the complexity results related to the lower-level problem, as hinted by the following chain of relations, taking $w_k = P_{\mathrm{SOL}(F,Y)}(z_k)$:

$$\psi^u(z_k) \leq g_{w_k}^\top (w_k - z_k) \leq \|g_{w_k}\| \|w_k - z_k\| \leq \overline{G}\, \mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k), \tag{6.14}$$

where the first inequality holds for any $g_{w_k} \in G(w_k)$. Establishing an upper bound depending on δ^k for $\psi^u(z^k)$ essentially hinges on identifying, for some classes of lower-level problems, an upper bound depending on δ_k for $\mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k)$, thanks to $\psi^l(z_k) \geq -\delta_k$. This can be done through error bounds theory, as detailed next. In the following developments, we consider matrices $M \in \mathbb{R}^{p\times p}$ and vectors $q \in \mathbb{R}^p$.

- Y polyhedral, $F(y) = M^\top \Gamma(My) + q$, $\Gamma : \mathbb{R}^p \to \mathbb{R}^p$ strongly monotone and Lipschitz continuous with modulus L on the range of M. Relying on [2, (i) in Proposition 3.2 and (ii) in Proposition 3.1], by (6.12), we have $\|z_k - P_Y(z_k - F(z_k))\| \leq (2D)^{1/2} L^{1/4} \delta_k^{1/4}$, and in turn there exists $\overline{\delta} > 0$ such that for every $\delta_k \leq \overline{\delta}$ and for some $A > 0$ (see [9, Proposition 6.2.8]),

  $$\mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k) \leq A\|z_k - P_Y(z_k - F(z_k))\| \leq A(2D)^{1/2} L^{1/4} \delta_k^{1/4}.$$

  In the same setting, if in addition we require $L < 1$, we obtain an improved bound. Recalling [2, Proposition 3.7], due to (6.12), $\|z_k - P_Y(z_k - F(z_k))\| \leq (1 - \|M\|)^{-1/2} \delta_k^{1/2}$, and there exists $\overline{\delta} > 0$ such that for every $\delta_k \leq \overline{\delta}$, we have

  $$\mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k) \leq A\|z_k - P_Y(z_k - F(z_k))\| \leq A(1 - \|L\|)^{-1/2} \delta_k^{1/2}.$$

- Y compact polyhedron, $F(y) = My + q$. Resorting to [2, (i) in Proposition 3.2 and (ii) in Proposition 3.1], (6.12) entails $\|z_k - P_Y(z_k - F(z_k))\| \leq (2D)^{1/2} \|M\|^{1/4} \delta_k^{1/4}$, and in turn,

  $$\mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k) \leq B\|z_k - P_Y(z_k - F(z_k))\| \leq B(2D)^{1/2} \|M\|^{1/4} \delta_k^{1/4},$$

  for some $B > 0$, thanks again to [9, Proposition 6.3.3].
  Whenever $\|M\| < 1$, in view of [2, Proposition 3.7], by (6.12), we have $\|z_k - P_Y(z_k - F(z_k))\| \leq (1 - \|M\|)^{-1/2} \delta_k^{1/2}$, and we improve the bound as follows:

  $$\mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k) \leq B\|z_k - P_Y(z_k - F(z_k))\| \leq B(1 - \|M\|)^{-1/2} \delta_k^{1/2}.$$

- Y polyhedral, $F(y) = My + q$, M positive semidefinite. By [2, (i) in Proposition 3.2], (6.12) yields $\min_{v\in Y} F(z_k)^\top (v - z_k) \geq -2D\|M\|^{1/2} \delta_k^{1/2}$, and in turn, due to [9, Theorem 6.4.1], for some $C > 0$,

  $$\begin{aligned} \mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k) &\leq -C(-\min_{v\in Y} F(v)^\top (v - z_k) + (-\min_{v\in Y} F(v)^\top (v - z_k))^{1/2}) \\ &\leq C(2D\|M\|^{1/2} \delta_k^{1/2} + 2D^{1/2} \|M\|^{1/4} \delta_k^{1/4}). \end{aligned}$$

- Y polyhedral, $N = 1$, θ_1^l quadratic and bounded from below on Y, $\varphi_1^l \equiv 0$. Assume the minimum principle sufficiency to hold, i.e. $\mathrm{SOL}(\nabla\theta_1^l, Y) = \{y \in$

$Y : \nabla\theta_1^l(\overline{z})^\top(y - \overline{z}) \leq 0\}$, for any $\overline{z} \in \mathrm{SOL}(\nabla\theta_1^l, Y)$. By [2, Proposition 3.6], (6.12) yields $\theta_1^l(z_k) \leq \theta_1^l(v) + 2\delta_k$ for every $v \in Y$, and due to [9, Theorem 6.4.4], for some $\Delta > 0$,

$$\mathrm{dist}_{\mathrm{SOL}(\nabla\theta_1^l,Y)}(z_k) \leq \Delta(\theta_1^l(z_k) - \theta_1^l(z)) \leq 2\Delta\delta_k.$$

Another class of problems, for which an upper bound depending on δ_k for ψ^u is available, has been investigated in [27], based on [21]. More in detail, whenever Y is closed and convex, $F : \mathbb{R}^p \to \mathbb{R}^p$ is continuous and the weak-sharpness property holds for (GVI^l), that is, for all $\overline{z} \in \mathrm{SOL}(F, Y)$ and all $v \in Y$, $F(\overline{z})^\top(v - \overline{z}) \geq \Omega\ \mathrm{dist}_{\mathrm{SOL}(F,Y)}(v)$, for some $\Omega > 0$, $\mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k) \leq -\frac{1}{\Omega} \min_{v\in Y} F(v)^\top(v - z_k) \leq \frac{1}{\Omega}\delta_k$. We stress that the discussion above is by no means intended as exhaustive: we wish just to mention some non-trivial instances of the lower-level problem, where single-valued mappings are involved, for which suitable upper bounds for ψ^u are identified. Note that the stronger the assumptions on the problem, the tighter (in terms of δ_k, apart from constants) the upper bound for ψ^u.

Remark 6.9. Assuming F and G single-valued and Lipschitz continuous with moduli L_F and L_G, it is possible to estimate the maximum number of iterations k to verify $H_\eta(z_k)^\top(v - z_k) \geq -\delta$, for all $v \in Y$. Relying on [2, (i) in Proposition 3.2], this inexact Stampacchia VI is guaranteed to hold whenever the Minty counterpart $H_\eta(v)^\top(v - z_k) \geq -(1/(2D\sqrt{L_F + \eta L_G}))\delta^2$, for all $v \in Y$, is verified. The maximum number of iterations k to satisfy this relation (see Theorem 6.2) is $\mathcal{O}\left(\left(\delta^2\right)^{-2}\right)$, that is, $\mathcal{O}\left(\delta^{-4}\right)$.

An analogous result can be immediately derived when (GVI^l) is concerned, considering H_0 in the developments above. On the other hand, focusing on the inexact (GVI^u) $G(z_k)^\top(v - z_k) \geq -\delta$ for all $v \in \mathrm{SOL}(F, Y)$, the picture becomes more complicated because z^k is not guaranteed to belong to $\mathrm{SOL}(F, Y)$ and the result in [2, (i) in Proposition 3.2] cannot be applied directly. Anyway, a link between inexact Minty version of (GVI^u) and (inexact) (GVI^u) can still be established according to the following developments. In the light of Theorem 6.4, the algorithm generates a point $z_k \in Y$ satisfying

$$F(v)^\top(v - z_k) \geq -\widetilde{\delta}, \qquad \forall v \in Y, \tag{6.15}$$

$$G(v)^\top(v - z_k) \geq -\widetilde{\delta}, \qquad \forall v \in \mathrm{SOL}(F, Y), \tag{6.16}$$

in $\mathcal{O}(\widetilde{\delta}^{-4})$ maximum number of iterations.

In view of (6.15) and resorting e.g. to the error bounds results in Remark 6.8, $\mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k)$ can be considered to be bounded by (a function of) the degree of inexactness $\widetilde{\delta}$ in the solution of (GVI^l): let, without loss of generality, $\mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k) \leq b_1\widetilde{\delta}^{b_2}$, where $b_1 > 0$, $b_2 \in (0, 1]$. By introducing the enlargement $\overline{\mathrm{SOL}} \triangleq [\mathrm{SOL}(F, Y) + B_{\mathrm{dist}_{\mathrm{SOL}(F,Y)}(z_k)}(0)] \cap Y$, and taking any $w \in \overline{\mathrm{SOL}}$ and $v = P_{\mathrm{SOL}(F,Y)}(w)$, we have

$$\begin{array}{rcl} G(w)^\top(w - z_k) & = & (G(w) - G(v))^\top(w - v) + (G(w) - G(v))^\top(v - z_k) \\ & & +G(v)^\top(w - v) + G(v)^\top(v - z_k) \\ & \geq & -L_G\|w - v\|\|v - z_k\| - \overline{G}\|w - v\| + G(v)^\top(v - z_k) \\ & \geq & -(L_GD + \overline{G})b_1\widetilde{\delta}^{b_2} - \widetilde{\delta} = -b_3\widetilde{\delta}^{b_2} - \widetilde{\delta}, \end{array}$$

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

TABLE 1

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

where $b_3 \triangleq (L_G D + \overline{G})b_1$ and the first inequality follows from monotonicity and Lipschitz continuity of G. Recalling $z_k \in \overline{\mathrm{SOL}}$, due to [2, (i) in Proposition 3.2] to get

$$G(z_k)^\top(w - z_k) \geq -b_4(b_3\widetilde{\delta}^{b_2} + \widetilde{\delta})^{\frac{1}{2}}, \quad \forall w \in \overline{\mathrm{SOL}}, \tag{6.17}$$

where $b_4 \triangleq 2D\sqrt{L_G}$. Relation (6.17) is *a fortiori* valid for every $w \in \mathrm{SOL}(F,Y) \subseteq \overline{\mathrm{SOL}}$. Following the same steps as in Remark 6.8, also an upper-bound for the upper-level Stampacchia optimality measure can be given: $\min_{w\in\mathrm{SOL}(F,Y)} G(z_k)^\top(w-z_k) \leq \overline{G}b_1\widetilde{\delta}^{b_2}$. Finally, taking $\widetilde{\delta} = \delta^{2/b_2}$, we get $(b_3\widetilde{\delta}^{b_2} + \widetilde{\delta})^{1/2} = (b_3\delta^2 + \delta^{2/b_2})^{1/2} \leq (b_3 + 1)^{1/2}\delta$, and the inexact Stampacchia VI $G(z_k)^\top(w - z_k) \geq -b_4(b_3+1)^{1/2}\delta$, $\forall w \in \mathrm{SOL}(F,Y)$, is verified in $\mathcal{O}((\delta^{2/b_2})^{-4})$, i.e. $\mathcal{O}(\delta^{-8/b_2})$ maximum number of iterations.

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, necessary condition (6.2) for orbiting around $\mathrm{SOL}(H_\eta, Y)$ is, fairly quickly, ensured to be met by $\{y_k\}$. On the other hand, if α and β decrease to 0.5, $\{z_k\}$ converges to $\mathrm{SOL}(H_\eta, Y)$. Finally, if β further decreases to 0.25, the convergence of $\{z_k\}$ is guaranteed to the solutions of (GVI^u), and thus to equilibria of (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 $\{\beta_k\}$.

The Projected Average Single-loop Tikhonov Algorithm (PASTA) 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 $\mathrm{SOL}(H_{\eta_{\overline{k}}}, Y)$ (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.10, 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.

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.*

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

Employing in Algorithm 7.1 $\{\alpha_k\}$ 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 5.3, Theorem 5.4 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 Lemma A.5 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 Algorithm 7.1 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 [17] for more details). 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. However, to analyze the behavior of the method, we are able to measure suitable upper bounds for the feasibility measure $-\psi^l$ and the optimality one $-\psi^u$, and to show the influence of the upper level by studying *a posteriori* the nature of 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 $x^1 = (y^2, y^4)$, $x^2 = (y^1, y^3)$: $\theta_1^l(y^1, y^{-1}) = 0.5(y^1)^2 + y^1(y^2 + 2y^3 + y^4 - 100)$, $\varphi_1^l(y^1) = 0$, $Y_1 = [-100, 50]$; $\theta_2^l(y^2, y^{-2}) = 0.5(y^2)^2 + y^2(y^1 + y^3 + y^4 - 50)$, $\varphi_2^l(y^2) = \max\{0, -10(y^2 - 15)\}$, $Y_2 = [0, 50]$; $\theta_3^l(y^3, y^{-3}) = 0.5(y^3)^2 + y^3(y^2 + y^4 - 100)$, $\varphi_3^l(y^3) = 0$, $Y_3 = [0, 100]$; $\theta_4^l(y^4, y^{-4}) = 0.5(y^4)^2 + y^4(y^1 + y^2 + y^3 - 50)$, $\varphi_4^l(y^4) = 0$, $Y_4 = [0, 50]$; $\theta_1^u(x^1, x^{-1}) = (y^2 - 20)^2 + (y^4 - 50)^2 + (y^2 + y^4)(y^1 + y^3)$, $\varphi_1^u(x^1) = 0$; $\theta_2^u(x^2, x^{-2}) = (y^1)^2 + y^1(y^2 + y^3) + (y^3)^2 + y^3(y^2 + y^4)$, $\varphi_2^u(x^2) = 0$. For this example Assumptions **A** and **B** are verified. One can obtain an explicit expression for the lower-level equilibrium set: $E = \left\{(-50, y^2, 50, 50 - y^2) : 15 \le y^2 \le 50\right\}$. Thus, the

		$\text{opt}(z_{\overline{I}})$		$\text{opt}(y_{\overline{I}})$
$\overline{k}$	0	$0.4\overline{I}$	$0.8\overline{I}$	
Variable $\alpha\,\&\,\beta$	0.57434	0.42161	0.41424	0.41219
Fixed $\alpha\,\&\,\beta$	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* $y_{\overline{I}}$

Iterations	10k	25k	50k	75k	100k	250k	500k	750k	1000k
Var $\alpha\,\&\,\beta$	0.7342	0.6140	0.5491	0.5186	0.4998	0.4528	0.4283	0.4179	0.4122
Fix $\alpha\,\&\,\beta$	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* $\{\beta_k\}$

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^*. Considering the evaluation of the subgradient, in order to deal with the nondifferentiability of the lower-level map, we set $f_y = [\nabla_{y^1}\theta_1^l(y), \nabla_{y^2}\theta_2^l(y) - 5(15+10^{-3}-y^2)/(10^{-3}), \nabla_{y^3}\theta_3^l(y), \nabla_{y^4}\theta_4^l(y)]^\top$ 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, that 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 index $\overline{k}$ we choose for the averaging procedure $\{z_k\}$ to start. Note that the point $z_{\overline{I}}$ is closer to x^* for larger values of $\overline{k}$, because early iterations, which are 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 5.4 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 the 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 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 $y_{\overline{I}} = (-49.5878, 15.0010, 50.0124, 34.6699)$.

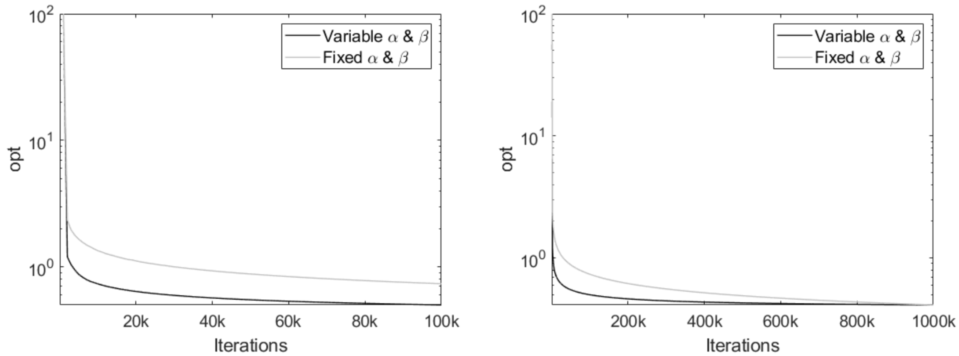

Fig. 1. *Comparison between variable (PASTA) and fixed* $\{\alpha_k\}$ *and* $\{\beta_k\}$ *considering* $\text{opt}(y_k) = \|y_k - x^*\|_\infty$ *for iterations 0- 100k (left-hand side) and for all 1000k iterations (right-hand side)*

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 [17] for more details about hierarchical multi-portfolio optimization and [19] 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 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 $x^\mu = [y^\nu]_{\nu \in \mathcal{S}_\mu}$, $\theta^l_\nu(y^\nu, y^{-\nu}) = -I_\nu(y^\nu) + \rho_\nu R_\nu(y^\nu) + TC_\nu(y^\nu, y^{-\nu})$, $\varphi^l_\nu(y^\nu) = \tau_\nu \|y^\nu\|_1$, $Y_\nu \triangleq \{y^\nu \in [l_\nu, u_\nu]^K : \sum_{i=1}^K y^\nu_i \leq 1\}$, $\theta^u_\mu(x^\mu, x^{-\mu}) = -\sum_{\nu \in \mathcal{S}_\mu} I_\nu(y^\nu) + \rho_\mu \sum_{\nu \in \mathcal{S}_\mu} R_\nu(y^\nu) + TC_\mu(x^\mu, x^{-\mu})$, $\varphi^u_\mu(x^\mu) = \tau_\mu \sum_{\nu \in \mathcal{S}_\mu} \|y^\nu\|_1$, where ρ_ν regulates 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 market (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 [17, 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 the ones in 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 point. Concerning the subgradients, $f_{y_i^\nu} = \nabla\theta_\nu^l(y)_i + \tau^\nu(y_i^\nu + 10^{-4})/(10^{-4}) - \tau^\nu$ whenever $y_i^\nu \in [-10^{-4}, 10^{-4}]$ for every $\nu \in \{1, \dots, N\}$ and $i \in \{1, \dots, K\}$, and $g_{x_j^\mu} = \nabla\theta_\mu^u(x)_j + \tau^\mu(x_j^\mu + 10^{-4})/(10^{-4}) - \tau^\mu$ whenever $x_j^\mu \in [-10^{-4}, 10^{-4}]$ for every $\mu \in \{1, \dots, M\}$ and $j \in \{1, \dots, (N/M)K\}$. To implement the projection step, a finite-steps method, inspired by [17], is relied on, preventing one from having to solve 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: $\tau_\nu^l = \overline{\tau}^l$ for $\nu = 1, \dots, 15$, $\tau_\nu^l = 0$ for $\nu = 16, \dots, 25$, $\tau_\mu^u = 0$ for $\mu = 1, \dots, 3$, $\tau_\mu^u = \overline{\tau}^u$ for $\mu = 4, 5$. This is done in order to observe how the regularization of the two hierarchical levels yields sparsity in 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 variational solution set. However, upper bounds for the feasibility measure $-\psi^l(y)$ and the optimality measure $-\psi^u(y)$ can still be provided. To measure feasibility, we consider $\text{feas}(y, f_y) \triangleq \max_{v\in Y} -f_y^\top(v-y)$. Whenever $f_y \in F(y)$, we have $\text{feas}(y, f_y) \geq -\psi^l(y)$, due to the monotonicity of F. Regarding optimality, we consider $\text{opt}(y, f_y, g_y) \triangleq \max_{v\in Y} -(1/\eta)(f_y + \eta g_y)^\top(v-y)$. Whenever $f_y \in F(y)$ and $g_y \in G(y)$, we have $f_y + \eta g_y \in H_\eta(y)$, and we can resort to Proposition A.8 to show that $\text{opt}(y, f_y, g_y) \geq \max_{v\in Y} -g_y^\top(v-y) \geq -\psi^u(y)$, thanks to the monotonicity of F and G. These measures are easily computed as they only require linear optimization problems to be solved, which is done using the `matlab` built-in function `linprog`. In Figure 2 we show the values of $\text{feas}(y_k, f_{y_k})$ and $\text{opt}(y_k, f_{y_k}, g_{y_k})$ over the iterations, for the SX5E dataset and considering the full regularization 2 setting (we do not report the images relative to the DIJA dataset and to the other regularization settings as they are very similar to Figure 2). In both figures, we report both the values for the algorithm version with variable $\{\alpha_k\}$ and $\{\beta_k\}$ (PASTA), and for the version with fixed $\{\alpha_k\}$ and $\{\beta_k\}$. PASTA achieves a faster decrease of the feasibility measure, as well as a convergence to a better value for the optimality measure we consider for all different regularization settings. In Table 4, we report $\text{feas}(z_{\overline{I}}, f_{z_{\overline{I}}})$ (computed starting from different iterations $\overline{k}$), $\text{feas}(y_{\overline{I}}, f_{y_{\overline{I}}})$ and $\text{opt}(y_{\overline{I}}, f_{y_{\overline{I}}}, g_{y_{\overline{I}}})$, in all the five regularization settings. Similarly to Example 1, $\{z_k\}$ achieves better results in terms of feasibility for larger values of $\overline{k}$. All the result we obtain seem to indicate that the sequence $\{y_k\}$, and *a fortiori* $\{z_k\}$, have a convergent behaviour.

To show the influence of the upper-level managers, and consequently of the upper-

		feas($z_{\overline{I}}, f_{z_{\overline{I}}}$) $\overline{k}=0$	$\overline{k}=0.4\overline{I}$	$\overline{k}=0.8\overline{I}$	feas($y_{\overline{I}}, f_{y_{\overline{I}}}$)	opt($y_{\overline{I}}, f_{y_{\overline{I}}}, g_{y_{\overline{I}}}$)
SX5E	No reg	2.2225e-04	9.8603e-05	9.6364e-05	9.5698e-05	9.1130e-05
	Low. reg. 1	1.0859e-03	2.3361e-04	2.2854e-04	2.2719e-04	8.6980e-05
	Low. reg. 2	2.4301e-03	3.1726e-04	3.0886e-04	3.0703e-04	6.1042e-05
	Full reg. 1	1.5472e-03	4.7882e-04	4.6261e-04	4.5935e-04	1.0150e-04
	Full reg. 2	2.8817e-03	5.5899e-04	5.4486e-04	5.4262e-04	6.9896e-05
DIJA	No reg.	9.2238e-04	3.2620e-04	3.1930e-04	3.1754e-04	1.9082e-04
	Low. reg. 1	3.4512e-03	6.7104e-04	6.7104e-04	6.5528e-04	1.7030e-04
	Low. reg. 2	8.2720e-04	8.2720e-04	8.3136e-04	8.2720e-04	2.0577e-04
	Full reg. 1	3.8326e-03	8.8701e-04	8.6774e-04	8.6284e-04	2.6538e-04
	Full reg. 2	7.6527e-03	1.0600e-03	1.0416e-03	1.0365e-03	2.7482e-04

TABLE 4

Feasibility and optimality performances obtained with PASTA for both datasets in Example 2, considering $z_{\overline{I}}$, with different starting iterations $\overline{k}$, and $y_{\overline{I}}$, focusing on the five regularization settings

#Accounts	SX5E 1–15	SX5E 16–25	DIJA 1–15	DIJA 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 five different regularization settings

level objective functions, we further measure the portfolios sparsity 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 we deal with. In Table 5, we show the percentage of zeros (to be intended as investments of less than 0.1% of the budget) in the final portfolios, regularized by the lower-level agents (accounts 1–15) and upper-level managers (accounts 16–25). Both 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 smaller number of assets, with a sparser solution for *Lower regularization 2*, as the sparsity enhancing parameter ($\overline{\tau}^l$) is larger. 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 larger. 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. Nonetheless, the sparsity obtained by managers 4 and 5 demonstrates the influence of the upper-level game on the overall solution. This is a further confirmation of the theoretical properties of PASTA.

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

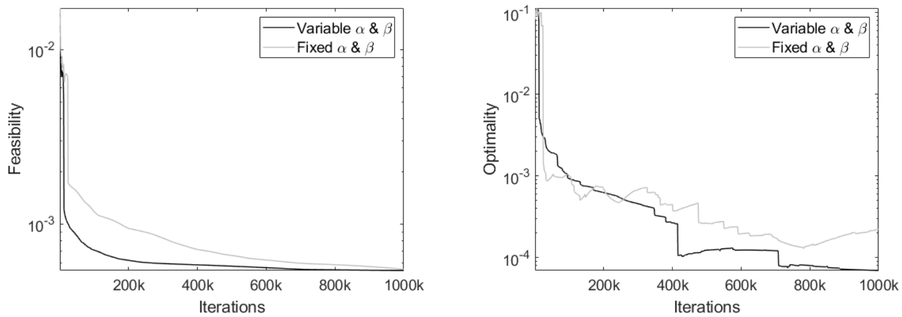

FIG. 2. *Comparison between variable* $\{\alpha_k\}$ *and* $\{\beta_k\}$ *(PASTA) and fixed* $\{\alpha_k\}$ *and* $\{\beta_k\}$ *considering* $\text{feas}(y_k, f_{y_k})$ *and* $\text{opt}(y_k, f_{y_k}, g_{y_k})$*, for the SX5E dataset, in the full regularization 2 setting*

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 5.3 and Theorem 5.4). Focusing on the more restrictive setting of single-valued operators, we obtain improved complexity bounds whenever the upper-level mapping is strongly monotone.
4. Exploiting the theoretical insights concerning the faster convergence behavior to 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. We study the relationship between Minty and Stampacchia versions of (hirarchical) GVIs, both in the exact and in the inexact cases. By making use of error bounds theory, we identify classes of problems for which inexact solutions to the Minty version of both the upper and lower-level GVIs (which is what one obtains as output of the algorithm) are linked to inexact solutions of the Stampacchia GVIs.
6. Focusing on the motivating example of multi-portfolio selection, we apply and test our approach on the novel model presented in [19]. Multi-portfolio se-

lection 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.

Appendix A. Additional results.

A.1. Proof of Proposition 3.2. If $y \in E$, then $y \in \mathrm{SOL}(F, Y)$. 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 $\nu = 1 \dots N$, $\exists \xi_\nu \in \partial_{y^\nu} \varphi^l_\nu(y^\nu) : (\nabla_{y^\nu} \theta^l_\nu(y^\nu, y^{-\nu}) + \xi_\nu)^\top (v^\nu - y^\nu) \geq 0$, $\forall v^\nu \in Y_\nu$. 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 $y \in \mathrm{SOL}(F, Y)$. Vice versa, if $y \in \mathrm{SOL}(F, Y)$, for all $\nu = 1 \dots N$ there exists $\exists f_y \in F(y)$ such that $f_y^\top((v^\nu, y^{-\nu}) - (y^\nu, y^{-\nu})) \geq 0$, $\forall (v^\nu, y^{-\nu}) \in Y$. By (3.1), $\exists f^\nu_y \in \nabla_{y^\nu} \theta^l_\nu + \partial_{y^\nu} \varphi^l_\nu : {f^\nu_y}^\top (v^\nu - y^\nu) \geq 0$, $\forall v^\nu \in Y_\nu$. By the convexity of player ν's problem, $y \in E$. ∎

A.2. On Maximal Monotonicity. We recall some concepts related to the key monotonicity assumptions we rely on.

Definition A.1. *A mapping $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ is monotone if $(\widehat{t} - \widetilde{t})^\top (\widehat{u} - \widetilde{u}) \geq 0$, $\forall\, \widehat{u}, \widetilde{u} \in \mathbb{R}^p$, $\widehat{t} \in T(\widehat{u})$, $\widetilde{t} \in T(\widetilde{u})$.*

Definition A.2. *A mapping $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ is maximal monotone if it is monotone and for every pair $(\widehat{u}, \widehat{t}) \in (\mathbb{R}^p \times \mathbb{R}^p) \setminus \mathrm{gph}(T)$ there exists $(\widetilde{u}, \widetilde{t}) \in \mathrm{gph}(T)$, where $\mathrm{gph}(T) \triangleq \{(u, t) | u \in \mathbb{R}^p, t \in T(u)\}$, with $(\widehat{t} - \widetilde{t})^\top (\widehat{u} - \widetilde{u}) < 0$.*

We remark that every continuous monotone single-valued mapping is maximal monotone (see, e.g., [24, Example 12.7]), while this is not true when dealing with multivalued mappings. The claim in the following lemma allows one to have a better grasp of the rationale behind Definition A.2.

Lemma A.3. *Consider the following conditions on $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$:*

$$(\widehat{u}, \widehat{t}) \in \mathrm{gph}(T), \tag{A.1}$$

$$(\widehat{t} - t)^\top (\widehat{u} - u) \geq 0, \quad \forall (u, t) \in \mathrm{gph}(T). \tag{A.2}$$

If T is monotone, (A.1) yields (A.2). If T is maximal monotone, (A.2) yields (A.1).

Proof. The first implication follows from Definition A.1. As for the second one, in the light of Definition A.2, if (A.1) does not hold, then (A.2) is not verified. ∎

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

Lemma A.4. *Let $S : X \rightrightarrows \widetilde{X}$ and $T : Y \rightrightarrows \widetilde{Y}$ be maximal monotone mappings. Their Cartesian 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

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

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

$$(s_x - \overline{s}_x)^\top (x - \overline{x}) + (t_y - \overline{t}_y)^\top (y - \overline{y}) \geq 0, \quad \forall (x, y) \in X \times Y, \quad \forall (s_x, t_y) \in S(x) \times T(y). \tag{A.3}$$

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 exists (x, s_x) with $x \in X$ and $s_x \in S(x)$ such that $(s_x - \overline{s}_x)^\top (x - \overline{x}) < 0$. From (A.3), one can deduce $(t_y - \overline{t}_y)^\top (y - \overline{y}) > 0$, $\forall y \in Y$ and $\forall t_y \in T(y)$. Due to the maximal monotonicity of 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 $(t_y - \overline{t}_y)^\top (y - \overline{y}) = (\overline{t}_y - \overline{t}_y)^\top (\overline{y} - \overline{y}) = 0$, which contradicts $(t_y - \overline{t}_y)^\top (y - \overline{y}) > 0$, $\forall y \in Y$ 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 $\exists g_x \in G(x) : g_x^\top ((w^\mu, x^{-\mu}) - (x^\mu, x^{-\mu})) \geq 0$ if and only if $\exists g_x^\mu \in G_\mu(x) : g_x^{\mu\top}(w^\mu - x^\mu) \geq 0$ and $\theta_\mu^u(x^\mu, x^{-\mu}) + \varphi_\mu^u(x^\mu) \leq \theta_\mu^u(w^\mu, x^{-\mu}) + \varphi_\mu^u(w^\mu)$, $\forall w^\mu : (w^\mu, x^{-\mu}) \in E$, 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 nonemptiness, 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. On the relations between GVIs and their Minty counterparts. Let $K \subseteq \mathbb{R}^p$ be nonempty, closed and convex, and $S : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$. We define the GVI

$$\text{find} \quad y \in K : \quad \exists s_y \in S(y) : \quad s_y^\top (v - y) \geq 0, \quad \forall v \in K, \tag{A.4}$$

and its Minty counterpart:

$$\text{find} \quad y \in K : \quad s_v^\top (v - y) \geq 0, \quad \forall v \in K, \quad \forall s_v \in S(v). \tag{A.5}$$

We establish the equivalence between problems (A.4) and (A.5), which in turn is instrumental to justify the use of the Minty version of the upper and lower-level GVIs as an optimality-like measure.

LEMMA A.5. *(a) Assuming S to be monotone, if y is a solution of* (A.4)*, then it is a solution of* (A.5)*.*

(b) Assuming S to be maximal monotone, if y is a solution of (A.5)*, then it is a solution of* (A.4)*.*

Proof. We introduce the mapping $T \triangleq S + N_K$, where N_K is the normal (to K) cone mapping from convex analysis: $N_K(v) \triangleq \left\{\zeta_v \in \mathbb{R}^p : \zeta_v^\top (w - v) \leq 0, \forall w \in K\right\}$ whenever $v \in K$. Since N_K is maximal monotone (see, e.g. [24, Corollary 12.18]), the mapping T, whose domain is K, is in turn monotone for point (a) and maximal monotone for point (b) due to [24, Corollary 12.44]. We observe that $y \in K$ is a solution of (A.4) if and only if $(y, 0) \in \text{gph}(T)$. Moreover, recalling that $0 \in N_K(v)$ for every $v \in K$, $y \in K$ is a solution of (A.5) if and only if $(0 - [s_v + \zeta_v])^\top (y - v) \geq 0$, $\forall v \in K$, $\forall s_v \in S(v), \forall \zeta_v \in N_K(v)$, which is equivalent to $(0 - t_v)^\top (y - v) \geq 0$, $\forall (v, t_v) \in \text{gph}(T)$. The claim is true due to Lemma A.3. ∎

The claim in Lemma A.5 suggests the use of $\psi(y) \triangleq \min_{v \in K} \min_{s_v \in S(v)} s_v^\top (v - y)$ as an optimality measure for the problem at hand. The following function $\psi(y) \triangleq \min_{v \in K} \min_{s_v \in S(v)} s_v^\top (v - y) = \min_{(v, s_v) \in \text{gph}(S) \cap (K \times \mathbb{R}^p)} s_v^\top (v - y)$ is shown to be continuous thanks to e.g. [24, proof of Corollary 7.42] due to closedness of $\text{gph}(S)$, local boundedness of S, and compactness of K. This, together with the considerations in Remark 6.9, corroborates the choice of ψ as a suitable optimality measure.

A.7. Averaging Sequences. The proof of the next lemma can be traced back to [15, Point 1 in Section 2.4.2].

LEMMA A.6. *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.8. Proof of point b) in Theorem 4.5. Assume by contradiction $\{y_k\}$ admits a limit $\overline{y} \notin \mathrm{SOL}(H_\eta, Y)$. Due to **C1** and Lemma A.6, $z_k \to \overline{y}$ where z_k is defined in (4.6), and, by Theorem 4.7, we have the contradiction $\overline{y} \in \mathrm{SOL}(H_\eta, Y)$. ∎

A.9. On Inexactness.

PROPOSITION A.7. *Given* $\varepsilon \geq 0$, *let* y *be a solution of the inexact version of* (4.2), *i.e.* $y \in Y$, *such that* $h_v^{\eta\top}(v-y) \geq -\varepsilon$, $\forall v \in Y$, $\forall h_v^\eta \in H_\eta(v)$. *We have*

$$f_v^\top (v-y) \geq -\varepsilon - \eta\Lambda_2, \quad \forall v \in Y, \quad \forall f_v \in F(v);$$

$$g_v^\top (v-y) \geq -\varepsilon/\eta, \quad \forall v \in \mathrm{SOL}(F,Y), \quad \forall g_v \in G(v).$$

Proof. For all $v \in Y$, for all $f_v \in F(v)$, $h_v^\eta = f_v + \eta g_v \in H_\eta(v)$, $f_v^\top(v-z_k) = h_v^{\eta\top}(v-z_k) - \eta g_v^\top(v-z_k) \geq -\varepsilon - \eta\Lambda_2$. For all $v \in \mathrm{SOL}(F,Y)$, $\overline{f}_v \in F(v)$ exists such that $\overline{f}_v^\top(z_k - v) \geq 0$, and for all $g_v \in G(v)$: $-\varepsilon/\eta \leq \left[\overline{f}_v/\eta + g_v\right]^\top (v - z_k) \leq g_v^\top(v-z_k)$. ∎

Proposition 6.3 follows taking $\varepsilon = \Lambda_1 \Xi_1^k + \Lambda_2 \Xi_2^k + \Lambda_3 \Xi_3^k$ in Proposition A.7.

It is difficult to measure how inexactness propagates from Minty-like GVI optimality conditions (like (4.7), (5.1), (5.2), (6.4), (6.5)) 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 only the case of single-valued mappings is considered. We also give the Stampacchia counterpart of Proposition A.7.

PROPOSITION A.8. *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* $h_y^{\eta\top}(v-y) \geq -\varepsilon$, $\forall v \in Y$. *We have* $\exists f_y \in F(y) : f_y^\top(v-y) \geq -\varepsilon - \eta\Lambda_2$, $\forall v \in Y$, *and* $\exists g_y \in G(y) : g_y^\top(v-y) \geq -\varepsilon/\eta$, $\forall v \in \mathrm{SOL}(F,Y)$.

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 $v \in Y$: $f_y^\top(v-y) = h_y^{\eta\top}(v-y) - \eta g_y^\top(v-y) \geq -\varepsilon - \eta\Lambda_2$, and, as in the proof of Proposition 6.3, $-\varepsilon/\eta \leq [f_y/\eta + g_y]^\top (v-y) \leq g_y^\top(v-y)$, $\forall v \in \mathrm{SOL}(F,Y)$. ∎

A.10. Proof of Proposition 7.1. By Theorem 6.4, we only need to prove $\{\gamma_k\}$ and $\{\eta_k\}$ are nonincreasing. Since, $\alpha_k = \alpha$, 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 $\frac{\overline{\alpha}}{\overline{\alpha}-\alpha} - 1 = \frac{\alpha}{(\overline{\alpha}-\alpha)} = \frac{t_\alpha}{1-t_\alpha}\frac{(I_\alpha^{\overline{\varepsilon}_\alpha}-1)}{I_\alpha^{\overline{\varepsilon}_\alpha}} \geq \frac{t_\alpha}{1-t_\alpha}\frac{(I_\alpha^{\varepsilon_\alpha}-1)}{I_\alpha^{\varepsilon_\alpha}} \geq \frac{t_\alpha}{1-t_\alpha}\frac{(k^{\varepsilon_\alpha}-(k-1)^{\varepsilon_\alpha})}{I_\alpha^{\varepsilon_\alpha}} \geq \frac{t_\alpha^k}{1-t_\alpha^k}\frac{(k^{\varepsilon_\alpha}-(k-1)^{\varepsilon_\alpha})}{I_\alpha^{\varepsilon_\alpha}}$, where $t_\alpha^k \triangleq log_k(k-1)$, and the last inequality holds since $t_\alpha^k \leq t_\alpha$, thus $\left(\frac{k}{I_\alpha}\right)^{\varepsilon_\alpha} \leq 1 \leq \frac{\overline{\alpha}}{\overline{\alpha}-\alpha} - \frac{t_\alpha^k}{1-t_\alpha^k}\frac{(k^{\varepsilon_\alpha}-(k-1)^{\varepsilon_\alpha})}{I_\alpha^{\varepsilon_\alpha}}$, and by rearranging terms, $\alpha_k = \overline{\alpha} - (\overline{\alpha}-\alpha)\left(\frac{k}{I_\alpha}\right)^{\varepsilon_\alpha} \geq t_\alpha^k \left[\overline{\alpha} - (\overline{\alpha}-\alpha)\left(\frac{k-1}{I_\alpha}\right)^{\varepsilon_\alpha}\right] = t_\alpha^k \alpha_{k-1}$, which implies $k^{\alpha_k} \geq \left[k^{t_\alpha^k}\right]^{\alpha_{k-1}} = (k-1)^{\alpha_{k-1}}$. The proof for $\{\eta_k\}$ follows by the same reasoning. ∎

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*, Computational Management Science, 20 (2023), p. 43.
[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 and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, 2003.
[10] 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.
[11] 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.
[12] S. Fang and E. Peterson, *Generalized variational inequalities*, Journal of optimization theory and applications, 38 (1982), pp. 363–383.
[13] V. V. Kalashnikov and N. I. Kalashinikova, *Solving two-level variational inequality*, Journal of Global Optimization, 8 (1996), pp. 289–294.
[14] 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.
[15] K. Knopp, *Infinite sequences and series*, Courier Corporation, 1956.
[16] 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.
[17] L. Lampariello, C. Neumann, J. M. Ricci, S. Sagratella, and O. Stein, *Equilibrium selection for multi-portfolio optimization*, European Journal of Operational Research, (2021).
[18] L. Lampariello, G. Priori, and S. Sagratella, *On the solution of monotone nested variational inequalities*, Mathematical Methods of Operations Research, (2022), pp. 1–26.
[19] L. Lampariello, S. Sagratella, and V. G. Sasso, *Nonsmooth hierarchical multi portfolio selection*, in International Conference on Optimization and Decision Science, Springer, 2022, pp. 37–46.
[20] X. Lu, H.-K. Xu, and X. Yin, *Hybrid methods for a class of monotone variational inequalities*, Nonlinear Analysis, (2009), pp. 1032–1041.
[21] P. Marcotte and D. Zhu, *Weak sharp solutions of variational inequalities*, SIAM Journal on Optimization, 9 (1998), pp. 179–189.
[22] 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.
[23] J.-S. Pang and D. Chan, *Iterative methods for variational and complementarity problems*, Mathematical programming, 24 (1982), pp. 284–313.
[24] R. Rockafellar and J. Wets, *Variational Analysis*, Springer, 1998.
[25] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.
[26] S. Sabach and S. Shtern, *A first order method for solving convex bilevel optimization problems*, SIAM Journal on Optimization, 27 (2017), pp. 640–660.
[27] S. Samadi and F. Yousefian, *Improved guarantees for optimal nash equilibrium seeking and bilevel variational inequalities*, arXiv preprint arXiv:2307.12511, (2023).
[28] M. Solodov, *An explicit descent method for bilevel convex optimization*, Journal of Convex Analysis, 14 (2007), p. 227.
[29] 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.