On the Complexity of Subgradient Methods for Trilevel Hierarchical Generalized Variational Inequalities

Lorenzo Lampariello[*], Simone Sagratella[†], Valerio Giuseppe Sasso[*]

March 12, 2026

Abstract

We study generalized variational inequalities with a three-level hierarchical structure. This setting extends nested GVI models beyond the bilevel case, for which $\mathcal{O}(\delta^{-4})$ complexity bounds are known for any prescribed positive tolerance δ, to a fully three-level hierarchical structure. We analyze a projected averaged subgradient method combined with a Tikhonov-like regularization scheme. Under compactness, maximal monotonicity, and a Hölder-type error bound assumption on the bottom-level problem, we derive explicit approximation bounds for the three levels of the hierarchy in terms of corresponding Minty gap functions, prove (subsequential) ergodic convergence, and establish iteration-complexity estimates. In the sharpest regime covered by our analysis, the method achieves complexity of $\mathcal{O}(\delta^{-8})$. We also discuss the underlying error bound condition and provide sufficient conditions for it to hold.

Keywords: Generalized variational inequality; hierarchical variational inequality; maximal monotone mapping; Tikhonov method; complexity analysis.

1 Introduction

Generalized Variational Inequalities (GVIs) provide a broad framework that encompasses several classes of equilibrium, optimization, and complementarity problems. We study a hierarchical setting where the feasible set is itself defined implicitly through lower-level GVI solution sets. More precisely, we consider a problem where the upper-level feasible region is the solution set of a GVI whose own feasible set is, in turn, the solution set of another GVI. We refer to such nested structure as trilevel hierarchical generalized variational inequality. This formulation naturally extends bilevel hierarchical GVIs and allows one to model decision and equilibrium processes with three layers of priority.

[*]Roma Tre University, lorenzo.lampariello, valeriogiuseppe.sasso@uniroma3.it
[†]Universitas Mercatorum, simone.sagratella@unimercatorum.it

For hierarchical problems with two levels, convergence results for subgradient-type methods are already available in the literature; in particular, for bilevel hierarchical GVIs, in [8], complexity bounds of $\mathcal{O}(\delta^{-4})$ are established for any prescribed positive tolerance δ. If one remains within the two-level hierarchical setting but specializes the problem structure, then sharper guarantees can be derived. In particular, the extragradient-type method studied in [12] achieves complexity of $\mathcal{O}(\delta^{-2})$. Under even stronger assumptions, when both levels reduce to optimization problems, accelerated gradient-type techniques yield complexity of $\mathcal{O}(\delta^{-1})$, see [10]. By contrast, the literature on three-level hierarchical problems is still rather limited. To the best of our knowledge, apart from the seminal work [5] that underlies many subsequent developments in this line of research, the main available contribution is [13], where all three hierarchical levels consists in optimization problems under strong convexity assumptions, and a complexity bound of $\mathcal{O}(\delta^{-2})$ is obtained.

Our goal is to address the trilevel problem in a fairly general form. We work in the setting of maximal monotone multivalued mappings and, relying on a Hölder-type error bound assumption on the bottom-level problem, adopt a projected averaged subgradient-type technique combined with a Tikhonov-like regularization scheme. As a result, our analysis covers a class of trilevel hierarchical GVIs that, to the best of our knowledge, has not yet been addressed in the literature. We derive explicit approximation bounds for all three levels of the hierarchy in terms of corresponding Minty gap functions, establish convergence of limit points in an ergodic sense, and obtain iteration-complexity estimates under suitable assumptions on the parameter sequences. In the sharpest regime covered by our analysis, this leads to a complexity bound of $\mathcal{O}(\delta^{-8})$. As expected, this guarantee is worse than the corresponding bilevel subgradient complexity, reflecting the additional difficulty introduced by the third hierarchical level.

Regarding the paper organization, in Section 2, we introduce the trilevel hierarchical GVI, state the standing assumptions, and present the projected averaged subgradient method. In Section 3, we develop approximation bounds for the three nested optimality conditions by means of Minty gap functions, thereby quantifying the degree of inexactness attained by the iterates generated by the algorithm. In Section 4, we build on these bounds to study convergence properties and to derive explicit complexity estimates, including the rate achieved under a harmonic parameter regime. Finally, Section 5 is devoted to the key error-bound condition imposed on the bottom-level GVI: we discuss its meaning, compare it with several related notions already available in the literature, and provide sufficient conditions ensuring that it holds.

2 Problem Definition

Consider the multivalued mappings $F, G, T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and the set $Y \subseteq \mathbb{R}^n$. We focus on the hierarchical trilevel GVI, that is the problem to

$$\text{find } u \in \text{SOL}\big(G, \text{SOL}(F,Y)\big) : \ \exists t_u \in T(u) : \ t_u^\top (v-u) \geq 0, \quad \forall v \in \text{SOL}\big(G, \text{SOL}(F,Y)\big), \tag{GVI$^{\text{Tri}}$}$$

where $\text{SOL}(G, \text{SOL}(F,Y))$ is the solution set of the underlying hierarchical bilevel GVI problem to

$$\text{find } x \in \text{SOL}(F,Y) : \ \exists g_x \in G(x) : \ g_x^\top (v-x) \geq 0, \ \forall v \in \text{SOL}(F,Y), \tag{GVI$^{\text{Bi}}$}$$

and $\text{SOL}(F,Y)$ is in turn the solution set of the (single-level) GVI at the bottom of the hierarchy, that is the problem to

$$\text{find } y \in Y : \ \exists f_y \in F(y) : \ f_y^\top (v-y) \geq 0, \ \forall v \in Y. \tag{GVI}$$

Consistently, we denote by $\text{SOL}(T, \text{SOL}(G, \text{SOL}(F,Y)))$ the solution set of (GVI$^{\text{Tri}}$).

We assume the following blanket conditions for the rest of the developments.

Assumption 1.

- *The set Y is nonempty, compact and convex.*
- *The mappings F, G and T are maximal monotone and locally bounded on Y.*
- *The problem* (GVI) *at the bottom of the hierarchy is such that*

$$\exists \beta > 0, \mathcal{M} \in (0,1] :$$
$$\text{dist}(y, \text{SOL}(F,Y)) \leq \beta \left(f_y^\top (y-v)\right)^{\mathcal{M}}, \forall v \in \text{SOL}(F,Y), y \in Y, f_y \in F(y). \tag{1}$$

The maximal monotonicity of the mappings is a sufficient condition for them to be outer-semicontinuous and closed-convex-valued on Y (see [11, Chapter 12]). Note that condition (1) is, as far as we are aware, novel: it is instrumental for the study of the convergence properties of the method we rely on. In Section 5, we delve into the nature and meaning of (1), relating it to similar assumptions discussed in the literature and providing easy-to-grasp conditions for it to hold. In Appendix A, relying on Assumption 1, we recall and derive some key theoretical properties concerning the GVIs introduced above. Among others, we report the basic existence result, whose proof is omitted since it follows from sequentially applying Proposition 16 to (GVI), (GVI$^{\text{Bi}}$) and (GVI$^{\text{Tri}}$).

Proposition 2. *The sets* $\text{SOL}(F,Y)$*,* $\text{SOL}(G, \text{SOL}(F,Y))$ *and* $\text{SOL}(T, \text{SOL}(G, \text{SOL}(F,Y)))$ *are nonempty, compact and convex.*

To address the trilevel problem, we rely on Tikhonov-like regularization techniques: we consider the mapping $H^{\eta,\zeta} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$:

$$H^{\eta,\zeta}(y) \triangleq F(y) + \eta \left(G(y) + \zeta T(y)\right),$$

which is maximal monotone for any $\eta, \zeta \geq 0$, due to Assumption 1.

The question arises naturally on the convergence properties of first-order-like methods, such as the projected subgradient one, in the more complicated setting of a trilevel GVI problem. We aim at answering this question by focusing on the following algorithm to address our trilevel GVI problem, where P_Y denotes the Euclidean projection operator on Y, which is well-defined since Y is closed and convex.

Algorithm 1: Projected Averaged Gradient

Data: $\{\gamma_k\}, \{\eta_k\}, \{\zeta_k\}$ positive nonincreasing, $y_1 \in Y$

for $k = 1, 2, \dots$ **do**

choose $f_{y_k} \in F(y_k), \quad g_{y_k} \in G(y_k), \quad t_{y_k} \in T(y_k)$;

$h_{y_k}^{\eta_k,\zeta_k} \leftarrow f_{y_k} + \eta_k(g_{y_k} + \zeta_k t_{y_k})$;

$y_{k+1} \leftarrow P_Y(y_k - \gamma_k h_{y_k}^{\eta_k,\zeta_k})$;

$z_{k+1} \leftarrow \dfrac{\sum_{j=1}^{k+1} \gamma_j \, y_j}{\sum_{j=1}^{k+1} \gamma_j}$;

3 Approximation Bounds for Optimality

To measure optimality of the three nested problems, we rely on the Minty gap functions for (GVI), (GVI^{Bi}) and (GVI^{Tri}), respectively: $\Gamma^{F,Y}$, $\Gamma^{G,\text{SOL}(F,Y)}$, $\Gamma^{T,\text{SOL}(G,\text{SOL}(F,Y))}$, see Definition 17. By means of the following proposition, we show that problem (GVI^{Tri}) is equivalent to solving a constrained system of implicitly-defined inequalities.

Proposition 3. *The point u satisfyies*

$$\begin{cases} \Gamma^{T,\text{SOL}(G,\text{SOL}(F,Y))}(u) \leq 0 \\ \Gamma^{G,\text{SOL}(F,Y)}(u) \leq 0 \\ \Gamma^{F,Y}(u) \leq 0 \\ u \in Y \end{cases} \tag{2}$$

if and only if it is a solution to (GVI^{Tri}).

Proof. Let $u \in Y$, $\Gamma^{F,Y}(u) \leq 0$, $\Gamma^{G,\mathrm{SOL}(F,Y)}(u) \leq 0$ and $\Gamma^{T,\mathrm{SOL}(G,\mathrm{SOL}(F,Y))}(u) \leq 0$. Relying on Lemma 18 (b), we have

$$\begin{cases} \Gamma^{T,\mathrm{SOL}(G,\mathrm{SOL}(F,Y))}(u) \leq 0 \\ \Gamma^{G,\mathrm{SOL}(F,Y)}(u) \leq 0 \\ \Gamma^{F,Y}(u) \leq 0 \\ u \in Y \end{cases} \iff \begin{cases} \Gamma^{T,\mathrm{SOL}(G,\mathrm{SOL}(F,Y))}(u) \leq 0 \\ \Gamma^{G,\mathrm{SOL}(F,Y)}(u) \leq 0 \\ u \in \mathrm{SOL}(F,Y) \end{cases}$$

$$\iff \begin{cases} \Gamma^{T,\mathrm{SOL}(G,\mathrm{SOL}(F,Y))}(u) \leq 0 \\ u \in \mathrm{SOL}(G,\mathrm{SOL}(F,Y)) \end{cases} \iff u \in \mathrm{SOL}(T,\mathrm{SOL}(G,\mathrm{SOL}(F,Y)))$$

□

System (2) is just as implicit as the original problem and, as such, inherits the same difficulties as the original formulation. Starting from this system, to quantify how much each point z_k generated by the Algorithm 1 violates the optimality conditions, we define the following system of inexact conditions by introducing the nonnegative tolerances $\varepsilon_k^{\mathrm{Tri}}$, $\varepsilon_k^{\mathrm{Bi}}$, and ε_k:

$$\begin{cases} \Gamma^{T,\mathrm{SOL}(G,\mathrm{SOL}(F,Y))}(z_k) \leq \varepsilon_k^{\mathrm{Tri}} \\ \Gamma^{G,\mathrm{SOL}(F,Y)}(z_k) \leq \varepsilon_k^{\mathrm{Bi}} \\ \Gamma^{F,Y}(z_k) \leq \varepsilon_k \\ y \in Y. \end{cases} \tag{3}$$

In this section, we identify the values for the tolerances, which will be used to establish the convergence results in Section 4, thanks to the continuity of the Minty gap functions (see Lemma (18)).

We define the following finite quantities, which are sometimes called for in the ensuing analysis:

$$\overline{F} \triangleq \max_{(y,f)\in \mathrm{gph}\, F\cap(Y\times\mathbb{R}^n)} \|f\|, \quad \overline{G} \triangleq \max_{(y,g)\in \mathrm{gph}\, G\cap(Y\times\mathbb{R}^n)} \|g\|,$$

$$\overline{T} \triangleq \max_{(y,t)\in \mathrm{gph}\, T\cap(Y\times\mathbb{R}^n)} \|t\|, \quad D \triangleq \max_{x,v\in Y} \|x-v\|.$$

Theorem 4. *Consider the sequences generated by Algorithm 1 and let*

$$\Xi_1^k \triangleq \frac{\sum_{j=1}^k \gamma_j^2}{\sum_{j=1}^k \gamma_j}, \quad \Xi_2^k \triangleq \frac{\sum_{j=1}^k \gamma_j\ \eta_j}{\sum_{j=1}^k \gamma_j}, \quad \Xi_3^k \triangleq \frac{1}{\sum_{j=1}^k \gamma_j}.$$

Then, in system (3),

$$\varepsilon_k = \Lambda_1 \Xi_1^k + \Lambda_2 \Xi_2^k + \Lambda_3 \Xi_3^k, \tag{4}$$

with

$$\Lambda_1 \triangleq \left(\overline{F}^2 + 2\eta_1^2\ \overline{G}^2 +\ 2\eta_1^2\ \zeta_1^2\ \overline{T}^2\right), \quad \Lambda_2 \triangleq (\overline{G} + \zeta_1\ \overline{T})\ D, \quad \Lambda_3 \triangleq \frac{D^2}{2}.$$

Proof. For all $v \in Y$, $h_v^{\eta_j,\zeta_j} \in H_{\eta_j,\zeta_j}(v)$, $j \geq 1$, we have

$$\begin{aligned} \|y_{j+1} - v\|^2 &= \|P_Y(y_j - \gamma_j h_{y_j}^{\eta_j,\zeta_j}) - P_Y(v)\|^2 \leq \|y_j - \gamma_j h_{y_j}^{\eta_j,\zeta_j} - v\|^2 \\ &= \|y_j - v\|^2 + \|\gamma_j h_{y_j}^{\eta_j,\zeta_j}\|^2 + 2\gamma_j h_{y_j}^{\eta_j,\zeta_j \top}(v - y_j) \\ &\leq \|y_j - v\|^2 + \|\gamma_j h_{y_j}^{\eta_j,\zeta_j}\|^2 + 2\gamma_j h_v^{\eta_j,\zeta_j \top}(v - y_j) \end{aligned}$$

where the first inequality is due to the nonexpansiveness of the projection operator and the second one is due to the monotonicity of H_{η_j,ζ_j}.

Observing that $f_v + \eta_j g_v + \eta_j \zeta_j t_v \in H^{\eta_j,\zeta_j}(v)$ for any $f_v \in F(v), g_v \in G(v), t_v \in T(v)$, and leveraging Lemma 21, we get, for all $f_v \in F(v), g_v \in G(v), t_v \in T(v)$.

$$\begin{aligned} -2\gamma_j f_v^\top(v - y_j) &\leq \|y_j - v\|^2 - \|y_{j+1} - v\|^2 + 2\Lambda_1\gamma_j^2 + 2\gamma_j\eta_j g_v^\top(v - y_j) \\ &\quad +2\gamma_j\eta_j\zeta_j t_v^\top(v - y_j) \\ &\leq \|y_j - v\|^2 - \|y_{j+1} - v\|^2 + 2\Lambda_1\gamma_j^2 + 2\gamma_j\ \eta_j(\overline{G} + \zeta_1\ \overline{T})\ D \\ &\leq \|y_j - v\|^2 - \|y_{j+1} - v\|^2 + 2\Lambda_1\gamma_j^2 + 2\Lambda_2\ \gamma_j\ \eta_j. \end{aligned}$$

Summing both sides of the inequality from $j = 1$ to $j = k$, and dividing them by $2\sum_{j=1}^k \gamma_j$, we get

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

because, due to the telescoping series property, $\sum_{j=1}^k \left(\|y_j - v\|^2 - \|y_{j+1} - v\|^2\right) = \|y_1 - v\|^2 - \|y_{k+1} - v\|^2$. The claim follows from Definition 17. □

Theorem 5. *Consider the sequences generated by Algorithm 1, and let*

$$\Phi_1^k \triangleq \frac{\sum_{j=1}^k \frac{\gamma_j}{\eta_j}\gamma_j}{\sum_{j=1}^k \gamma_j}, \quad \Phi_2^k \triangleq \frac{1}{\eta_k \sum_{j=1}^k \gamma_j}, \quad \Phi_3^k \triangleq \frac{\sum_{j=1}^k \zeta_j\gamma_j}{\sum_{j=1}^k \gamma_j}.$$

Then, in system (3)*,*

$$\varepsilon_k^{\mathrm{Bi}} = \Lambda_1\Phi_1^k + \Lambda_3\Phi_2^k + \Lambda_4\Phi_3^k, \tag{5}$$

where Λ_1 and Λ_3 are defined in Theorem 4, and

$$\Lambda_4 \triangleq \overline{T}\ D.$$

Proof. For all $v \in SOL(F,Y)$, $h_v^{\eta_j,\zeta_j} \in H_{\eta_j,\zeta_j}(v)$, $j \geq 1$, and observing that $f_v + \eta_j g_v + \eta_j \zeta_j t_v \in H^{\eta_j,\zeta_j}(v)$ for any $f_v \in F(v), g_v \in G(v), t_v \in T(v)$, we get, for all $f_v \in F(v), g_v \in G(v), t_v \in T(v)$,

$$\begin{aligned}
\|y_{j+1} - v\|^2 &= \|P_Y(y_j - \gamma_j h_{y_j}^{\eta_j,\zeta_j}) - P_Y(v)\|^2 \leq \|y_j - \gamma_j h_{y_j}^{\eta_j,\zeta_j} - v\|^2 \\
&= \|y_j - v\|^2 + \|\gamma_j h_{y_j}^{\eta_j,\zeta_j}\|^2 + 2\gamma_j h_{y_j}^{\eta_j,\zeta_j \top}(v - y_j) \\
&\leq \|y_j - v\|^2 + 2\gamma_j^2 \Lambda_1 + 2\gamma_j h_v^{\eta_j,\zeta_j \top}(v - y_j) \\
&= \|y_j - v\|^2 + 2\gamma_j^2 \Lambda_1 + 2\gamma_j f_v^\top (v - y_j) + 2\gamma_j \ \eta_j \ g_v^\top (v - y_j) \\
&\quad +2\gamma_j \ \eta_j \ \zeta_j \ \overline{T} \ D,
\end{aligned}$$

where the first inequality comes from the non expansiveness of the projection operator and the second one is due to the monotonicity of H_{η_j,ζ_j} and Lemma 21. Observe that since $y_j \in Y$, for every $v \in \text{SOL}(F,Y)$ there exists $\overline{f}_v \in F(v)$ such that $\overline{f}_v^\top (v - y_j) \leq 0$. Rearranging the terms and dividing both sides by $2\eta_j > 0$, and choosing $\overline{f}_v \in F(v)$, we have

$$\begin{aligned}
-\gamma_j \ g_v^\top (v - y_j) &\leq \tfrac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\eta_j} + \Lambda_1 \tfrac{\gamma_j^2}{\eta_j} + \overline{f}_v^\top (v - y_j) + \Lambda_4 \ \gamma_j \ \zeta_j \\
&\leq \tfrac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\eta_j} + \Lambda_1 \tfrac{\gamma_j^2}{\eta_j} + \Lambda_4 \ \gamma_j \ \zeta_j.
\end{aligned}$$

Summing both sides of the inequality from $j = 1$ to k, and dividing by $\sum_{j=1}^k \gamma_j$, we get

$$\begin{aligned}
-g_v^\top (v - z_k) &\leq \frac{1}{\sum_{j=1}^k \gamma_j} \sum_{j=1}^k \frac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\eta_j} + \Lambda_1 \frac{\sum_{j=1}^k \frac{\gamma_j}{\eta_j} \gamma_j}{\sum_{j=1}^k \gamma_j} \\
&\quad +\Lambda_4 \frac{\sum_{j=1}^k \gamma_j \ \zeta_j}{\sum_{j=1}^k \gamma_j} \\
&\leq \Lambda_3 \frac{1}{\eta_k \sum_{j=1}^k \gamma_j} + \Lambda_1 \frac{\sum_{j=1}^k \frac{\gamma_j}{\eta_j} \gamma_j}{\sum_{j=1}^k \gamma_j} + \Lambda_4 \frac{\sum_{j=1}^k \gamma_j \ \zeta_j}{\sum_{j=1}^k \gamma_j},
\end{aligned}$$

where the second inequality holds due to Lemma 22. The claim follows from Definition 17. □

Theorem 6. *Consider the sequences generated by Algorithm 1 and let*

$$\Psi_1^k = \frac{1}{\eta_k \ \zeta_k \sum_{j=1}^k \gamma_j}, \quad \Psi_2^k = \frac{\sum_{j=1}^k \frac{\gamma_j}{\eta_j \zeta_j} \gamma_j}{\sum_{j=1}^k \gamma_j}, \quad \Psi_3^k = \frac{1}{\zeta_k \left(\sum_{j=1}^k \gamma_j \right)^{\mathcal{M}}},$$

$$\Psi_4^k = \left(\frac{\sum_{j=1}^{k} \frac{\gamma_j}{\zeta_j^{\frac{1}{\mathcal{M}}}} \gamma_j}{\sum_{j=1}^{k} \gamma_j} \right)^{\mathcal{M}}, \quad \Psi_5^k = \left(\frac{\sum_{j=1}^{k} \frac{\eta_j}{\zeta_j^{\frac{1}{\mathcal{M}}}} \gamma_j}{\sum_{j=1}^{k} \gamma_j} \right)^{\mathcal{M}}.$$

Then, in system (3),

$$\varepsilon_k^{\mathrm{Tri}} = \Lambda_3 \Psi_1^k + \Lambda_1 \Psi_2^k + \overline{G} \beta \left(\Lambda_3\right)^{\mathcal{M}} \Psi_3^k + \overline{G} \beta \left(\Lambda_1\right)^{\mathcal{M}} \Psi_4^k + \overline{G} \beta \left(\Lambda_2\right)^{\mathcal{M}} \Psi_5^k. \tag{6}$$

Proof. For all $v \in \mathrm{SOL}(G, \mathrm{SOL}(F, Y))$, $h_v^{\eta_j,\zeta_j} \in H_{\eta_j,\zeta_j}(v)$, $j \geq 1$, and observing that $f_v + \eta_j g_v + \eta_j \zeta_j t_v \in H^{\eta_j,\zeta_j}(v)$ for any $f_v \in F(v), g_v \in G(v), t_v \in T(v)$, we get, for all $f_v \in F(v), g_v \in G(v), t_v \in T(v)$, we have

$$\begin{aligned}
\|y_{j+1} - v\|^2 &= \|P_Y(y_j - \gamma_j h_{y_j}^{\eta_j,\zeta_j}) - P_Y(v)\|^2 \leq \|y_j - \gamma_j h_{y_j}^{\eta_j,\zeta_j} - v\|^2 \\
&= \|y_j - v\|^2 + \|\gamma_j h_{y_j}^{\eta_j,\zeta_j}\|^2 + 2\gamma_j h_{y_j}^{\eta_j,\zeta_j \top}(v - y_j) \\
&\leq \|y_j - v\|^2 + 2\gamma_j^2 \Lambda_1 + 2\gamma_j h_v^{\eta_j,\zeta_j \top}(v - y_j) \\
&= \|y_j - v\|^2 + 2\gamma_j^2 \Lambda_1 + 2\gamma_j f_v^\top (v - y_j) + 2\gamma_j \ \eta_j \ g_v^\top (v - y_j) \\
&\quad + 2\gamma_j \ \eta_j \ \zeta_j \ t_v^\top (v - y_j),
\end{aligned}$$

where the first inequality comes from the nonexpansiveness of the projection operator and the second one is due to the monotonicity of H_{η_j,ζ_j} and Lemma 21. Observe that since $y_j \in Y$, for every $v \in \mathrm{SOL}(F, Y)$ there exists $\overline{f}_v \in F(v)$ such that $\overline{f}_v^\top (v - y_j) \leq 0$. Rearranging the terms and dividing both sides by $2\eta_j\zeta_j > 0$, and choosing $\overline{f}_v \in F(v)$, we have

$$\begin{aligned}
-\gamma_j t_v^\top (v - y_j) &\leq \tfrac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\eta_j\zeta_j} + \Lambda_1 \tfrac{\gamma_j^2}{\eta_j\zeta_j} + \tfrac{\gamma_j}{\eta_j\zeta_j} \overline{f}_v^\top (v - y_j) \\
&\quad + \tfrac{\gamma_j}{\zeta_j} \ g_v^\top (v - y_j) \\
&\leq \tfrac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\eta_j\zeta_j} + \Lambda_1 \tfrac{\gamma_j^2}{\eta_j\zeta_j} + \tfrac{\gamma_j}{\zeta_j} \ g_v^\top (v - y_j).
\end{aligned} \tag{7}$$

Consider the following chain, which holds due to the nonexpansiveness of the projection operator

$$\begin{aligned}
\|y_{j+1} - v\|^2 &= \|P_Y(y_j - \gamma_j h_{y_j}^{\eta_j,\zeta_j}) - P_Y(v)\|^2 \leq \|y_j - \gamma_j h_{y_j}^{\eta_j,\zeta_j} - v\|^2 \\
&\leq \|y_j - v\|^2 + 2\Lambda_1 \gamma_j^2 + 2\gamma_j h_{y_j}^{\eta_j,\zeta_j \top}(v - y_j) \\
&\leq \|y_j - v\|^2 + 2\Lambda_1 \gamma_j^2 + 2\gamma_j f_{y_j}^\top (v - y_j) + 2\Lambda_2 \gamma_j \eta_j.
\end{aligned}$$

Rearranging the terms and dividing both sides of the previous inequality by $2\gamma_j$ we get:

$$f_{y_j}^\top (y_j - v) \leq \frac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\gamma_j} + \Lambda_1 \gamma_j + \Lambda_2 \eta_j. \tag{8}$$

Moreover, for any $v \in \mathrm{SOL}(G, \mathrm{SOL}(F,Y))$, $\overline{g}_v \in G(v)$ exists such that $\overline{g}_v^\top(v - u) \leq 0$ for all $u \in \mathrm{SOL}(F,Y)$, and we have the following chain

$$\begin{aligned}
\overline{g}_v^\top(v - y_j) &= \overline{g}_v^\top(v - P_{\mathrm{SOL}(F,Y)}(y_j) + P_{\mathrm{SOL}(F,Y)}(y_j) - y_j) \\
&= \overline{g}_v^\top(v - P_{\mathrm{SOL}(F,Y)}(y_j)) + \overline{g}_v^\top(P_{\mathrm{SOL}(F,Y)}(y_j) - y_j) \\
&\leq \ \overline{g}_v^\top(P_{\mathrm{SOL}(F,Y)}(y_j) - y_j) \leq \ \overline{G}\|P_{\mathrm{SOL}(F,Y)}(y_j) - y_j\| \\
&\leq \overline{G}\beta\left(f_{y_j}^\top(y_j - v)\right)^{\mathcal{M}} \\
&\leq \overline{G}\beta\left(\tfrac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\gamma_j} + \Lambda_1\gamma_j + \Lambda_2\ \eta_j\right)^{\mathcal{M}},
\end{aligned}$$

where the first inequality is due to $v \in \mathrm{SOL}(G, \mathrm{SOL}(F,Y))$ and $P_{\mathrm{SOL}(F,Y)}(y_j) \in \mathrm{SOL}(F,Y)$, and the third one follows from condition (1), and the fourth is true since $\mathcal{M} \in (0,1]$ and therefore $t \to t^{\mathcal{M}}$ is nondecreasing.

Combining the previous chain with (7) (since we can take $g_v = \overline{g}_v$ in (7) and $\zeta_j > 0$ for all j) we get

$$\begin{aligned}
-\gamma_j\ t_v^\top(v - y_j) \quad &\leq \quad \tfrac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\eta_j\ \zeta_j} + \Lambda_1\tfrac{\gamma_j^2}{\eta_j\ \zeta_j} + \tfrac{\gamma_j}{\zeta_j}\overline{G}\beta(\tfrac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\gamma_j} \\
&\qquad +\Lambda_1\gamma_j + \Lambda_2\ \eta_j)^{\mathcal{M}} \\
&= \quad \tfrac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\eta_j\ \zeta_j} + \Lambda_1\tfrac{\gamma_j^2}{\eta_j\ \zeta_j} + \\
&\qquad +\gamma_j\overline{G}\beta\left(\frac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\gamma_j\zeta_j^{\frac{1}{\mathcal{M}}}} + \Lambda_1\frac{\gamma_j}{\zeta_j^{\frac{1}{\mathcal{M}}}} + \Lambda_2\ \frac{\eta_j}{\zeta_j^{\frac{1}{\mathcal{M}}}}\right)^{\mathcal{M}}.
\end{aligned}$$

Summing both sides of the inequality from $j = 1$ to k, and dividing by $\sum_{j=1}^k \gamma_j$, we get

$$\begin{aligned}
-t_v^\top(v - z_k) \quad &\leq \quad \tfrac{1}{\sum_{j=1}^k \gamma_j}\sum_{j=1}^k \tfrac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\eta_j\ \zeta_j} + \Lambda_1\frac{\sum_{j=1}^k \frac{\gamma_j}{\eta_j\ \zeta_j}\gamma_j}{\sum_{j=1}^k \gamma_j} \\
&\qquad +\overline{G}\beta\frac{\sum_{j=1}^k \gamma_j\left(\frac{\|y_j - v\|^2 - \|y_{j+1} - v\|^2}{2\gamma_j\zeta_j^{\frac{1}{\mathcal{M}}}} + \Lambda_1\frac{\gamma_j}{\zeta_j^{\frac{1}{\mathcal{M}}}} + \Lambda_2\ \frac{\eta_j}{\zeta_j^{\frac{1}{\mathcal{M}}}}\right)^{\mathcal{M}}}{\sum_{j=1}^k \gamma_j}.
\end{aligned} \tag{9}$$

We can write the following upper bound for the third term of the right hand side of (9):

$$
\begin{aligned}
&\frac{\sum_{j=1}^{k}\gamma_j\left(\frac{\|y_j-v\|^2-\|y_{j+1}-v\|^2}{2\gamma_j\zeta_j^{\frac{1}{\mathcal{M}}}}+\Lambda_1\frac{\gamma_j}{\zeta_j^{\frac{1}{\mathcal{M}}}}+\Lambda_2\frac{\eta_j}{\zeta_j^{\frac{1}{\mathcal{M}}}}\right)^{\mathcal{M}}}{\sum_{j=1}^{k}\gamma_j} \leq \\
&\left(\frac{\sum_{j=1}^{k}\gamma_j\frac{\|y_j-v\|^2-\|y_{j+1}-v\|^2}{2\gamma_j\zeta_j^{\frac{1}{\mathcal{M}}}}+\Lambda_1\sum_{j=1}^{k}\frac{\gamma_j\gamma_j}{\zeta_j^{\frac{1}{\mathcal{M}}}}+\Lambda_2\sum_{j=1}^{k}\frac{\gamma_j\eta_j}{\zeta_j^{\frac{1}{\mathcal{M}}}}}{\sum_{j=1}^{k}\gamma_j}\right)^{\mathcal{M}} \leq \\
&\left(\Lambda_3\frac{1}{\zeta_k^{\frac{1}{\mathcal{M}}}\sum_{j=1}^{k}\gamma_j}+\Lambda_1\frac{\sum_{j=1}^{k}\frac{\gamma_j^2}{\zeta_j^{\frac{1}{\mathcal{M}}}}}{\sum_{j=1}^{k}\gamma_j}+\Lambda_2\frac{\sum_{j=1}^{k}\frac{\gamma_j\eta_j}{\zeta_j^{\frac{1}{\mathcal{M}}}}}{\sum_{j=1}^{k}\gamma_j}\right)^{\mathcal{M}} \leq \\
&\left(\Lambda_3\frac{1}{\zeta_k^{\frac{1}{\mathcal{M}}}\sum_{j=1}^{k}\gamma_j}\right)^{\mathcal{M}}+\left(\Lambda_1\frac{\sum_{j=1}^{k}\frac{\gamma_j^2}{\zeta_j^{\frac{1}{\mathcal{M}}}}}{\sum_{j=1}^{k}\gamma_j}\right)^{\mathcal{M}}+\left(\Lambda_2\frac{\sum_{j=1}^{k}\frac{\gamma_j\eta_j}{\zeta_j^{\frac{1}{\mathcal{M}}}}}{\sum_{j=1}^{k}\gamma_j}\right)^{\mathcal{M}},
\end{aligned}
\tag{10}
$$

where the first inequality is the Jensen inequality with weights $\gamma_j > 0$ since $\mathcal{M} \in (0, 1]$ implies $t \to t^{\mathcal{M}}$ is concave, the second inequality is due to Lemma 22 and the monotonicity of $t \to t^{\mathcal{M}}$, and the third is due to the subadditivity of $t \to t^{\mathcal{M}}$, since

$$
\Lambda_3\frac{1}{\zeta_k^{\frac{1}{\mathcal{M}}}\sum_{j=1}^{k}\gamma_j} \geq 0, \quad \Lambda_1\frac{\sum_{j=1}^{k}\frac{\gamma_j^2}{\zeta_j^{\frac{1}{\mathcal{M}}}}}{\sum_{j=1}^{k}\gamma_j} \geq 0, \quad \Lambda_2\frac{\sum_{j=1}^{k}\frac{\gamma_j\eta_j}{\zeta_j^{\frac{1}{\mathcal{M}}}}}{\sum_{j=1}^{k}\gamma_j} \geq 0.
$$

Combining (9) and (10) and invoking Lemma 22, we get

$$
\begin{aligned}
-t_v^{\top}(v-z_k) \quad \leq \quad & \Lambda_3\frac{1}{\eta_k\zeta_k\sum_{j=1}^{k}\gamma_j}+\Lambda_1\frac{\sum_{j=1}^{k}\frac{\gamma_j}{\eta_j\zeta_j}\gamma_j}{\sum_{j=1}^{k}\gamma_j}+\overline{G}\beta\Lambda_3^{\mathcal{M}}\left(\frac{1}{\zeta_k^{\frac{1}{\mathcal{M}}}\sum_{j=1}^{k}\gamma_j}\right)^{\mathcal{M}} \\
& +\overline{G}\beta\Lambda_1^{\mathcal{M}}\left(\frac{\sum_{j=1}^{k}\frac{\gamma_j^2}{\zeta_j^{\frac{1}{\mathcal{M}}}}}{\sum_{j=1}^{k}\gamma_j}\right)^{\mathcal{M}}+\overline{G}\beta\Lambda_2^{\mathcal{M}}\left(\frac{\sum_{j=1}^{k}\frac{\gamma_j\eta_j}{\zeta_j^{\frac{1}{\mathcal{M}}}}}{\sum_{j=1}^{k}\gamma_j}\right)^{\mathcal{M}}.
\end{aligned}
$$

The relation (6) holds due to Definition 17. □

4 Convergence and Complexity Results

We study the convergence properties of Algorithm 1 for problem (GVI^{Tri}), in view of the inexact optimality conditions in (3). In Assumption 7, we state the conditions on the parameter sequences in Algorithm 1 that allow us to prove convergence.

Assumption 7. *Consider the sequences $\{\gamma_k\}$, $\{\eta_k\}$ and $\{\zeta_k\}$ in Algorithm 1. They satisfy:*

i) *(a)* $\sum_{j=1}^{\infty} \gamma_j = +\infty$,
(b) $\lim_{k\to\infty} \eta_k = 0$,
(c) $\lim_{k\to\infty} \gamma_k = 0$;

ii) *(a)* $\lim_{k\to\infty} \frac{\gamma_k}{\eta_k} = 0$,
(b) $\lim_{k\to\infty} \eta_k \sum_{j=1}^{k} \gamma_j = +\infty$,
(c) $\lim_{k\to\infty} \zeta_k = 0$;

iii) *(a)* $\lim_{k\to\infty} \eta_k \; \zeta_k \sum_{j=1}^{k} \gamma_j = +\infty$,
(b) $\lim_{k\to\infty} \frac{\gamma_k}{\eta_k \; \zeta_k} = 0$,
(c) $\lim_{k\to\infty} \zeta_k \left(\sum_{j=1}^{k} \gamma_j \right)^{\mathcal{M}} = +\infty$,
(d) $\lim_{k\to\infty} \frac{\gamma_k}{(\zeta_k)^{1/\mathcal{M}}} = 0$,
(e) $\lim_{k\to\infty} \frac{\eta_k}{(\zeta_k)^{1/\mathcal{M}}} = 0$,
where $\mathcal{M} \in (0, 1]$ *is the constant in condition* (1).

Relying on the previous conditions and on Proposition 3, we state the next convergence result.

Theorem 8. *Let Assumption 7 hold and $\overline{z}$ be any limit point of the sequence z_k produced by Algorithm 1. Then* $\overline{z} \in \mathrm{SOL}(T, \mathrm{SOL}(G, \mathrm{SOL}(F, Y)))$.

Proof. Let $\lim z_k = \overline{z}$. Consider the inexact optimality conditions in (3), and the estimates for the tolerances in (4), (5) and (6).

We have $\lim \varepsilon_k = 0$, $\lim \varepsilon_k^{\mathrm{Bi}} = 0$ and $\lim \varepsilon_k^{\mathrm{Tri}} = 0$ thanks to Lemma 23 and to Assumption 7 *i)*, *i)* and *ii)*, and *i)*, *ii)* and *iii)*, respectively. Then $\overline{z}$ solves system (3), due to $z_k \in Y$, the closedness of Y and due to the continuity of the Minty gap functions (see Lemma 18). □

In order to recover convergence rate guarantees for Algorithm 1, we focus on the inexact optimality conditions in (3). Given a fixed positive tolerance δ, we are interested in giving the maximum number of iterates to achieve

$$\max\{\varepsilon_k, \varepsilon_k^{\mathrm{Bi}}, \varepsilon_k^{\mathrm{Tri}}\} \leq \delta. \tag{11}$$

Relying on harmonic sequences for $\{\gamma_k\}$, $\{\eta_k\}$, $\{\zeta_k\}$ and the bounds in Lemma 24, we provide the following upper bounds for $\varepsilon_k, \varepsilon_k^{\mathrm{Bi}}, \varepsilon_k^{\mathrm{Tri}}$.

Theorem 9. *Consider the following choices for the sequences in Algorithm 1:*

$$\gamma_k = \overline{\gamma} k^{-a}, \qquad \eta_k = \overline{\eta} k^{-b}, \qquad \zeta_k = \overline{\zeta} k^{-c},$$

with

$$a \in (0, 0.5], \qquad b \in (0, a), \qquad c \in \Big(0,\ \min\{\, a-b,\ \mathcal{M}a,\ \mathcal{M}b \,\}\Big),$$

where $\mathcal{M} \in (0,1]$ is the constant in condition (1)*. The following hold for all $k \geq 2$:*

i)

$$\varepsilon_k \leq c_1 \frac{1}{k^{1-a}} + c_2 \frac{1}{k^b},$$

where $c_1 \triangleq \Lambda_1 \frac{\overline{\gamma}(1-a)}{1-2^{a-1}} + \Lambda_3 \frac{1-a}{\overline{\gamma}(1-2^{a-1})}$ and $c_2 \triangleq \Lambda_2 \frac{\overline{\eta}(1-a)}{(1-(a+b))(1-2^{a-1})} + \Lambda_1 \frac{\overline{\gamma}(1-a)}{(a-b)(1-2^{a-1})}$;

ii)

$$\varepsilon_k^{\mathrm{Bi}} \leq c_3 \frac{1}{k^{\,a-b}} + c_4 \frac{1}{k^{1-a-b}} + c_5 \frac{1}{k^c},$$

where $c_3 \triangleq \Lambda_1 \frac{\overline{\gamma}\,(1-a)}{\overline{\eta}(1-2^{a-1})[1-(2a-b)]}$, $c_4 \triangleq \Lambda_3 \frac{1-a}{(1-2^{a-1})\overline{\gamma}\overline{\eta}}$ and $c_5 \triangleq \Lambda_4 \frac{\overline{\zeta}(1-a)}{(1-(a+c))(1-2^{a-1})}$;

(iii)

$$\varepsilon_k^{\mathrm{Tri}}(z_k) \leq\ c_6 \frac{1}{k^{1-a-b-c}} + c_7 \frac{1}{k^{a-b-c}} + c_8 \frac{1}{k^{\,\mathcal{M}(1-a)-c}} + c_9 \frac{1}{k^{\mathcal{M}a-c}} + c_{10} \frac{1}{k^{\mathcal{M}b-c}},$$

where $c_6 \triangleq \Lambda_3 \frac{1-a}{(1-2^{a-1})\overline{\gamma}\overline{\eta}\overline{\zeta}}$, $c_7 \triangleq \Lambda_1 \frac{\overline{\gamma}(1-a)}{\overline{\eta}\overline{\zeta}(1-2^{a-1})(1-2a+b+c)}$, $c_8 \triangleq \overline{G}\beta\Lambda_3^{\mathcal{M}} \frac{1}{\overline{\zeta}} \left(\frac{1-a}{(1-2^{a-1})\overline{\gamma}}\right)^{\mathcal{M}}$, $c_9 \triangleq \overline{G}\beta\Lambda_1^{\mathcal{M}} \frac{\overline{\gamma}^{\mathcal{M}}}{\overline{\zeta}} \left(\frac{1-a}{(1-2^{a-1})\,(1-2a+\frac{c}{\mathcal{M}})}\right)^{\mathcal{M}}$, $c_{10} \triangleq \overline{G}\beta\Lambda_2^{\mathcal{M}} \frac{\overline{\eta}^{\mathcal{M}}}{\overline{\zeta}} \left(\frac{1-a}{(1-2^{a-1})\,(1-a-b+\frac{c}{\mathcal{M}})}\right)^{\mathcal{M}}$.

Proof. We rely on the harmonic series bounds of Lemma 24 throughout.

(i) We focus on (4) in Theorem 4 and provide upper bounds for Ξ_1^k, Ξ_2^k and Ξ_3^k. In order to cover all possible $a \in (0, 0.5]$, we rely on points *(a)* and *(b)* of Lemma 24 to derive the following bound:

$$\Xi_1^k = \frac{\sum_{j=1}^k \gamma_j^2}{\sum_{j=1}^k \gamma_j} \leq \frac{\overline{\gamma}(1-a)}{1-2^{a-1}} \frac{1 + k^{1-2a}\log k}{k^{1-a}} = \frac{\overline{\gamma}(1-a)}{1-2^{a-1}} \left(\frac{1}{k^{1-a}} + \frac{\log k}{k^a}\right).$$

For any $k \geq 1$, we have the elementary inequality $k^d = e^{d\log k} \geq d \log k$. Let us apply it with $d = a - b > 0$, and by rearranging the terms, we get $\log k \leq \frac{k^{a-b}}{a-b}$. Therefore the following holds:

$$\frac{\log k}{k^a} = \frac{\log k}{k^b} \frac{1}{k^{a-b}} \leq \frac{k^{a-b}}{k^b(a-b)} \frac{1}{k^{a-b}} \leq \frac{1}{k^b(a-b)},$$

and we get the following upper bound for Ξ_1^k

$$\Xi_1^k \leq \frac{\overline{\gamma}(1-a)}{1-2^{a-1}}\left(\frac{1}{k^{1-a}}+\frac{\log k}{k^a}\right) \leq \frac{\overline{\gamma}(1-a)}{1-2^{a-1}}\frac{1}{k^{1-a}}+\frac{\overline{\gamma}(1-a)}{(a-b)(1-2^{a-1})}\frac{1}{k^b}.$$

Moreover, we have the following upper bounds, since $(a+b) \in (0,1)$ and $a \in (0, 0.5]$:

$$\begin{array}{rcl} \Xi_2^k = \frac{\sum_{j=1}^k \gamma_j \eta_j}{\sum_{j=1}^k \gamma_j} & \leq & \frac{\overline{\eta}(1-a)}{(1-(a+b))(1-2^{a-1})}\frac{1}{k^b}, \\ \Xi_3^k = \dfrac{1}{\sum_{j=1}^k \gamma_j} & \leq & \frac{1-a}{\overline{\gamma}(1-2^{a-1})}\frac{1}{k^{1-a}}. \end{array}$$

(ii) We focus on (5) in Theorem 5 and provide upper bounds for Φ_1^k, Φ_2^k and Φ_3^k. Since $(2a-b) \in (0,1)$ and $(a+c) \in (0,1)$, we have

$$\begin{array}{rcl} \Phi_1^k = \frac{\sum_{j=1}^k \frac{\gamma_j^2}{\eta_j}}{\sum_{j=1}^k \gamma_j} & \leq & \frac{\overline{\gamma}\,(1-a)}{\overline{\eta}(1-2^{a-1})[1-(2a-b)]}\frac{1}{k^{\,a-b}}, \\ \Phi_2^k = \frac{1}{\eta_k \sum_{j=1}^k \gamma_j} & \leq & \frac{1-a}{(1-2^{a-1})\overline{\gamma\eta}}\frac{1}{k^{1-a-b}}, \\ \Phi_3^k = \frac{\sum_{j=1}^k \zeta_j \gamma_j}{\sum_{j=1}^k \gamma_j} & \leq & \frac{\overline{\zeta}(1-a)}{(1-(a+c))(1-2^{a-1})}\frac{1}{k^c}. \end{array}$$

(iii) We focus on (6) in Theorem 6 and provide upper bounds for $\Psi_1^k, \Psi_2^k, \Psi_3^k, \Psi_4^k$ and Ψ_5^k. Since $2a-b-c \in (0,1)$, $2a - \frac{c}{\mathcal{M}} \in (0,1)$ and $b - \frac{c}{\mathcal{M}} \in (0,1)$, we have

$$\begin{array}{rcl} \Psi_1^k = \frac{1}{\eta_k\ \zeta_k \sum_{j=1}^k \gamma_j} & \leq & \frac{1-a}{(1-2^{a-1})\overline{\gamma\eta}\overline{\zeta}}\frac{1}{k^{1-a-b-c}} \\ \Psi_2^k = \frac{\sum_{j=1}^k \frac{\gamma_j}{\eta_j \zeta_j}\gamma_j}{\sum_{j=1}^k \gamma_j} & \leq & \frac{\overline{\gamma}(1-a)}{\overline{\eta}\overline{\zeta}(1-2^{a-1})(1-2a+b+c)}\frac{1}{k^{a-b-c}} \\ \Psi_3^k = \frac{1}{\zeta_k\left(\sum_{j=1}^k \gamma_j\right)^{\mathcal{M}}} & \leq & \frac{1}{\overline{\zeta}}\left(\frac{1-a}{(1-2^{a-1})\overline{\gamma}}\right)^{\mathcal{M}}\frac{1}{k^{\,\mathcal{M}(1-a)-c}} \\ \Psi_4^k = \left(\dfrac{\sum_{j=1}^k \frac{\gamma_j^2}{\zeta_j^{1/\mathcal{M}}}}{\sum_{j=1}^k \gamma_j}\right)^{\mathcal{M}} & \leq & \frac{\overline{\gamma}^{\mathcal{M}}}{\overline{\zeta}}\left(\frac{1-a}{(1-2^{a-1})\,(1-2a+\frac{c}{\mathcal{M}})}\right)^{\mathcal{M}}\ \frac{1}{k^{\mathcal{M}a-c}} \\ \Psi_5^k = \left(\dfrac{\sum_{j=1}^k \frac{\eta_j}{\zeta_j^{1/\mathcal{M}}}\gamma_j}{\sum_{j=1}^k \gamma_j}\right)^{\mathcal{M}} & \leq & \frac{\overline{\eta}^{\mathcal{M}}}{\overline{\zeta}}\left(\frac{1-a}{(1-2^{a-1})\,(1-a-b+\frac{c}{\mathcal{M}})}\right)^{\mathcal{M}}\ \frac{1}{k^{\mathcal{M}b-c}}. \end{array}$$

$\square$

Thanks to the bounds in Theorem 9, we can study the global rate of convergence of Algorithm 1 as a function of $\mathcal{M} \in (0,1]$, which is the constant in condition (1). In fact, we have

$$\max\{\varepsilon_k, \varepsilon_k^{\mathrm{Bi}}, \varepsilon_k^{\mathrm{Tri}}\} \leq \widetilde{C}\frac{1}{k^{\iota}}$$

where $\iota \triangleq \min\{\min\{1-a, b\}, \min\{a-b,\ 1-a-b,\ c\}, \min\{1-a-b-c, a-b-c, \mathcal{M}(1-a)-c, \mathcal{M}a-c, \mathcal{M}b-c\}\}$ and $\widetilde{C} \triangleq 10\max\{c_1, \ldots, c_{10}\}$. The optimal choice of a, b, c to maximize ι turns out to be

$$a^\star = \frac{1}{2}, \qquad b^\star = \frac{1}{2(1+\mathcal{M})}, \qquad c^\star = \frac{\mathcal{M}}{4(1+\mathcal{M})}. \tag{12}$$

Notice that this choice satisfies the constraints required in Theorem 9 since: $a^\star - b^\star = \frac{1}{2} - \frac{1}{2(1+\mathcal{M})} = \frac{\mathcal{M}}{2(1+\mathcal{M})}, \mathcal{M}a^\star = \frac{\mathcal{M}}{2}, \mathcal{M}b^\star = \frac{\mathcal{M}}{2(1+\mathcal{M})}$. Hence $c^\star = \frac{1}{2}\mathcal{M}b^\star = \frac{\mathcal{M}}{4(1+\mathcal{M})}$ satisfies $0 < c^\star < \min\{\, a^\star - b^\star,\ \mathcal{M}a^\star,\ \mathcal{M}b^\star \,\} = \frac{\mathcal{M}}{2(1+\mathcal{M})}$. Thus, the overall convergence rate for Algorithm 1 in the sense of conditions (3) and (11) is

$$\mathcal{O}\left(\frac{1}{\delta^{\frac{4(1+\mathcal{M})}{\mathcal{M}}}}\right).$$

In the best case scenario, i.e. $\mathcal{M} = 1$ in condition (1), we get $a^\star = 1/2$, $b^\star = 1/4$, $c^\star = 1/8$ and the converge rate for Algorithm 1 turns out to be $\mathcal{O}(\delta^{-8})$.

5 On error bound theory and related conditions for the single level GVI at the bottom of the hierarchy

In this section, we examine condition (1) for (GVI) in detail, compare it with error bounds and related conditions available in the literature, and derive easily verifiable sufficient conditions under which it holds.

We consider the following Hölderian bounds of order $\mathcal{M} \in (0, 1]$ on the distance to the solution set of (GVI): there exists $\beta > 0$ such that, for every $y \in Y$,

- (Stampacchia error bound: see e.g. [6] for the single-valued mapping case)

$$\left(\frac{\operatorname{dist}(y, \operatorname{SOL}(F, Y))}{\beta}\right)^{\frac{1}{\mathcal{M}}} \leq \max_{v \in Y} \min_{f_y \in F(y)} f_y^\top (y - v) = \Theta^{F,Y}(y); \tag{13}$$

- (Minty error bound: see e.g. [9] for the single-valued mapping case)

$$\left(\frac{\operatorname{dist}(y, \operatorname{SOL}(F, Y))}{\beta}\right)^{\frac{1}{\mathcal{M}}} \leq \max_{v \in Y} \max_{f_v \in F(v)} f_v^\top (y - v) = \Gamma^{F,Y}(y); \tag{14}$$

- (bound condition in Assumption 1)

$$\left(\frac{\operatorname{dist}(y, \operatorname{SOL}(F, Y))}{\beta}\right)^{\frac{1}{\mathcal{M}}} \leq \min_{v \in \operatorname{SOL}(F,Y)} \min_{f_y \in F(y)} f_y^\top (y - v); \tag{15}$$

- (bound condition from [12])

$$\left(\frac{\operatorname{dist}(y, \operatorname{SOL}(F, Y))}{\beta}\right)^{\frac{1}{\mathcal{M}}} \leq \min_{v \in \operatorname{SOL}(F,Y)} \max_{f_v \in F(v)} f_v^\top (y - v). \tag{16}$$

The following result clarifies the relations between the conditions described above.

Theorem 10. *Under our standing assumptions, the following statements hold:*

- (16) *implies* (15) *and the converse does not hold in general;*
- (15) *implies* (13) *and the converse does not hold in general;*
- (14) *implies* (13)*;*
- (16) *implies* (14) *and the converse does not hold in general.*

Proof.

- ((16) $\Rightarrow$ (15)) By the maximal monotonicity of F, we have $f_v^\top(y - v) \leq f_y^\top(y - v)$ for every $v, y \in Y$, $f_v \in F(v)$ and $f_y \in F(y)$.

  ((16) $\not\Leftarrow$ (15)) Consider $n = 2$ and

$$F(u) = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix} u, \quad Y = \{u \in \mathbb{R}^2 : u \in [0,1]^2\}.$$

  We have $\mathrm{SOL}(F, Y) = \{(0, u) : \ u \in [0,1]\}$ and, therefore, for all $y = (y_1, y_2) \in Y$, $\mathrm{dist}(y, \mathrm{SOL}(F, Y)) = |y_1| = y_1$. Condition (15) holds with $\beta = 1$ and $\mathcal{M} = 1/2$, since, for all $v \in \mathrm{SOL}(F, Y)$,

$$F(y)^\top(y - v) = y_1^2 + y_1 v_2 \geq y_1^2 = (\mathrm{dist}(y, \mathrm{SOL}(S, K)))^2\,,$$

  where the first inequality is true since for all $y, v \in Y$, $y_1 v_2 \geq 0$. On the other hand, (16) does not hold because considering $\overline{y} = (1,1) \in Y$ and $\overline{v} = (0,0) \in \mathrm{SOL}(F, Y)$, we have

$$F(\overline{v})^\top(\overline{y} - \overline{v}) = 0 \not\geq \frac{1}{\beta^{\frac{1}{\mathcal{M}}}} = \left(\frac{\mathrm{dist}(y, \mathrm{SOL}(F, Y))}{\beta}\right)^{\frac{1}{\mathcal{M}}},$$

  for any $\mathcal{M} \in (0,1]$ and any $\beta > 0$.

- ((15) $\Rightarrow$ (13)) Since $\mathrm{SOL}(F, Y) \subseteq Y$, we have

$$\begin{aligned} \min_{v \in \mathrm{SOL}(F,Y)} \min_{f_y \in F(y)} f_y^\top(y - v) &\leq \max_{v \in \mathrm{SOL}(F,Y)} \min_{f_y \in F(y)} f_y^\top(y - v) \\ &\leq \max_{v \in Y} \min_{f_y \in F(y)} f_y^\top(y - v). \end{aligned}$$

  ((15) $\not\Leftarrow$ (13)) Consider, $n = 2$ and

$$F(u) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} u, \quad Y = \{u \in \mathbb{R}^2 : \|u\|_2 \leq 1\}.$$

We have $\mathrm{SOL}(F,Y)=\{(0,0)\}$ and, therefore, for all $y=(y_1,y_2)\in Y$, $\mathrm{dist}(y,\mathrm{SOL}(S,K))=\|y\|_2$. Condition (13) holds with $\beta=1$ and $\mathcal{M}=1$, since, for all $y\in Y$,

$$\begin{aligned}\max_{v\in Y} F(y)^\top(y-v)=\max_{v\in Y}(-y_2 \quad y_1)v=(-y_2 \quad y_1)\frac{(-y_2 \quad y_1)^\top}{\|(-y_2 \quad y_1)\|}\\ =\|y\|_2=\mathrm{dist}(y,\mathrm{SOL}(S,K)).\end{aligned}$$

On the other hand, (15) does not hold because, considering any $\overline{y}\in Y$ such that $\|\overline{y}\|_2>0$ and $v\in\mathrm{SOL}(F,Y)=\{(0,0)\}$, we have

$$F(\overline{y})^\top(\overline{y}-v)=\overline{y}_2\overline{y}_1-\overline{y}_1\overline{y}_2=0\not\geq\left(\frac{\|\overline{y}\|_2}{\beta}\right)^{\frac{1}{\mathcal{M}}}=\left(\frac{\mathrm{dist}(y,\mathrm{SOL}(F,Y))}{\beta}\right)^{\frac{1}{\mathcal{M}}},$$

for any $\mathcal{M}\in(0,1]$ and any $\beta>0$.

- ((14) $\Rightarrow$ (13)) By the maximal monotonicity of F, we get $f_v^\top(y-v)\leq f_y^\top(y-v)$ for every $v,y\in Y$, $f_v\in F(v)$ and $f_y\in F(y)$.
- ((16) $\Rightarrow$ (14)) Since $\mathrm{SOL}(F,Y)\subseteq Y$, we have

$$\begin{aligned}\min_{v\in\mathrm{SOL}(F,Y)}\max_{f_v\in F(v)} f_v^\top(y-v)&\leq\max_{v\in\mathrm{SOL}(F,Y)}\max_{f_v\in F(v)} f_v^\top(y-v)\\ &\leq\max_{v\in Y}\max_{f_v\in F(v)} f_v^\top(y-v).\end{aligned}$$

((16) $\not\Leftarrow$ (14)) consider the same counterexample as the one for the case ((15) $\not\Leftarrow$ (13)). Condition (14) holds with $\beta=1$ and $\mathcal{M}=1$, since, for all $y\in Y$,

$$\begin{aligned}\max_{v\in Y} F(v)^\top(y-v)=\max_{v\in Y}(-v_2 \quad v_1)y=\frac{y^\top}{\|y\|}y\\ =\|y\|_2=\mathrm{dist}(y,\mathrm{SOL}(S,K)).\end{aligned}$$

On the other hand, (16) does not hold because, considering any $\overline{y}\in Y$ such that $\|\overline{y}\|_2>0$ and $v\in\mathrm{SOL}(F,Y)=\{(0,0)\}$, we have

$$F(v)^\top(\overline{y}-v)=0\not\geq\left(\frac{\|\overline{y}\|_2}{\beta}\right)^{\frac{1}{\mathcal{M}}}=\left(\frac{\mathrm{dist}(y,\mathrm{SOL}(F,Y))}{\beta}\right)^{\frac{1}{\mathcal{M}}},$$

for any $\mathcal{M}\in(0,1]$ and any $\beta>0$.

□

We recall that if Y is a compact polyhedron and $F(x)=Ax+b$ is monotone, then Θ and Γ are continuous semialgebraic on Y, by the Kurdyka-Łojasiewicz theory for subanalytic functions, $\mathcal{M}\in(0,1]$ exists such that conditions (13) and (14) hold (see e.g. [2]).

Assuming that F is the subgradient of a convex function $\varphi:\mathbb{R}^n\to(-\infty,+\infty]$, we say that the following Hölderian bound of order $\mathcal{M}\in(0,1]$ for the distance to the solution set of (GVI) holds whenever there exists $\beta>0$ such that, for every $y\in Y$,

- (optimal-value error bound: see e.g. [10])

$$\left(\frac{\operatorname{dist}(y, \operatorname{SOL}(F,Y))}{\beta}\right)^{\frac{1}{\mathcal{M}}} \leq \varphi(y) - \min_{v \in Y} \varphi(v). \tag{17}$$

The following result makes it possible to relate condition (17) to (14) and to our condition (15).

Theorem 11. *Under our standing assumptions, if F is the subgradient of some convex function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$, the following statements hold:*

- (14) *implies* (17)*;*
- (17) *implies* (15)*.*

Proof.

- ((14) $\Rightarrow$ (17)) Let $(\overline{v}, \overline{f}_{\overline{v}}) \in \underset{(v,f)\in \operatorname{gph} F \cap (Y \times \mathbb{R}^n)}{\arg\max} f_v^\top (y - v)$. Then,

$$\max_{v \in Y} \max_{f_v \in F(v)} f_v^\top (y-v) = \overline{f}_{\overline{v}}^\top (y - \overline{v}) \leq \varphi(y) - \varphi(\overline{v}) \leq \varphi(y) - \min_{v \in Y} \varphi(v),$$

  where the first inequality is due to the convexity of φ.

- ((17) $\Rightarrow$ (15)) Consider any $v \in \operatorname{SOL}(F,Y)$: by the convexity of φ, we have $\varphi(v) - \varphi(y) \geq f_y^\top (v-y)$ for all $y \in Y$ and all $f_y \in F(y)$. Therefore, for all $y \in Y$

$$\varphi(y) - \min_{v \in Y} \varphi(v) \leq \min_{v \in \operatorname{SOL}(F,Y)} \min_{f_y \in F(y)} f_y^\top (y-v).$$

$\square$

Remark 12. *Whenever F is the subgradient of some convex function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ (that is, under the assumption in Theorem 11), the bound $\min_{v \in \operatorname{SOL}(F,Y)} \min_{f_\cdot \in F(\cdot)} f_\cdot^\top (\cdot - v)$ in our condition* (15) *turns out to be a gap function, indeed. This can be shown by applying Theorems 10 and 11, since our bound from* (15) *is actually between two gap functions for any $y \in Y$:*

$$\Gamma^{F,Y}(y) \leq \min_{v \in \operatorname{SOL}(F,Y)} \min_{f_y \in F(y)} f_y^\top (y-v) \leq \Theta^{F,Y}(y).$$

We notice that, e,ven under the restrictive conditions of Theorem 11, condition (16) may not hold. In fact, consider $n = 1$ and $F(u) = 2u$, $Y = \{u \in \mathbb{R} : u \in [-1, 1]\}$. Then $\operatorname{SOL}(F,Y) = \{0\}$ and $\operatorname{dist}(y, \operatorname{SOL}(F,Y)) = |y|$. In turn,

$$\min_{v \in \operatorname{SOL}(F,Y)} F(v)(y-v) = 0 \not\geq \left(\frac{|y|}{\beta}\right)^{\frac{1}{\mathcal{M}}} = \left(\frac{\operatorname{dist}(y, \operatorname{SOL}(F,Y))}{\beta}\right)^{\frac{1}{\mathcal{M}}},$$

for any $\mathcal{M} \in (0,1]$ and any $\beta > 0$. On the contrary, for the example at hand, conditions (13), (14), (15) and (17) hold.

Thanks to Theorem 11, condition (15) holds for some $\mathcal{M} \in (0,1]$ whenever F is the subgradient of a convex semialgebraic on $\mathbb{R}^n$ continuous function φ (see e.g. [3]). Moreover, condition (15) holds with $\mathcal{M} = 1$ (i.e. the best-possible bound) whenever (GVI) is a monotone linear complementarity problem, under nondegeneracy assumptions (see [4, Theorem 13]).

A Generalized Variational Inequalities Properties

We summarize the main properties of GVIs. Let $K \subseteq \mathbb{R}^n$ be a nonempty, compact and convex set, and $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a maximal monotone and locally bounded mapping. The maximal monotonicity of the mapping S is a sufficient condition for it to be outer-semicontinuous and closed-convex-valued on Y (see [11, Chapter 12]).

The GVI is the problem to

$$\text{find} \quad y \in K : \quad \exists s_y \in S(y) : \quad s_y^\top (v - y) \geq 0, \quad \forall v \in K; \qquad (\text{GVI}(S, K))$$

and we denote its solution set $\text{SOL}(S, K)$.

The Minty counterpart of (GVI(S, K)) is the problem to

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

Let us introduce the definitions of monotonicity and maximal monotonicity in the case of multivalued mappings.

Definition 13 (Monotonicity). *A mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is monotone if*

$$(\widehat{s} - \widetilde{s})^\top (\widehat{u} - \widetilde{u}) \geq 0, \quad \forall\, \widehat{u},\, \widetilde{u} \in \mathbb{R}^n,\, \widehat{s} \in S(\widehat{u}),\, \widetilde{s} \in S(\widetilde{u}).$$

Definition 14 (Maximal monotonicity). *A mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone if it is monotone and for every pair $(\widehat{u}, \widehat{s}) \in (\mathbb{R}^n \times \mathbb{R}^n) \setminus \text{gph}(S)$ there exists $(\widetilde{u}, \widetilde{s}) \in \text{gph}(S)$, where $\text{gph}(S) \triangleq \{(u, s) | u \in \mathbb{R}^n, s \in S(u)\}$, with $(\widehat{s} - \widetilde{s})^\top (\widehat{u} - \widetilde{u}) < 0$.*

We remark that every continuous monotone single-valued mapping is maximal monotone (see, e.g., [11, Example 12.7]), while this is not true when dealing with multivalued mappings. In the following proposition, whose proof can be traced back to [8], the equivalence between problems (GVI(S, K)) and (MGVI(S, K)) is established.

Proposition 15. *The following statements hold:*

(a) let S be monotone. If y is a solution of (GVI(S, K))*, then it is a solution of* (MGVI(S, K))*;*

(b) let S be maximal monotone. If y is a solution of (MGVI(S, K))*, then it is a solution of* (GVI(S, K)).

In the next proposition, whose proof can be traced back to [8], we state some properties of $\text{SOL}(S, K)$.

Proposition 16. *The set* $\text{SOL}(S, K)$ *is nonempty and compact. If S is maximal monotone, then* $\text{SOL}(S, K)$ *is also convex.*

Next, we introduce the dual gap function for problem GVI(S, K), which is used throughout the paper to derive convergence and complexity guarantees.

Definition 17 (Minty gap function)**.** *The Minty gap function associated with* (GVI(S, K)) *is the function* $\Gamma^{S,K} : \mathbb{R}^n \to \mathbb{R}$ *defined as*

$$\Gamma^{S,K}(y) \triangleq \max_{v \in K} \max_{s \in S(v)} s^\top (y - v).$$

We remark that in the previous definition we use max instead of sup since we assume K to be compact and S to be locally bounded and outer semicontinuous. In fact, we can introduce the finite quantity $\overline{S} \triangleq \max_{s \in S(K)} \|s\|$, which will be called-for in the next lemma.

Proposition 18. *The following statements hold.*

(a) $\Gamma^{S,K}(y) \geq -\overline{S}\,\mathrm{dist}(y, K)$*, where* $\mathrm{dist}(y, K)$ *is the Euclidean distance of* y *from the set* K*;*

(b) assume S *is maximal monotone and* $K = \{y \in \mathbb{R}^n : \varphi(y) \leq 0\} \cap \overline{K}$*, where* $\overline{K}$ *is a compact convex set and* $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ *is a convex function. Then*

$$\begin{cases} \Gamma^{S,K}(y) \leq 0 \\ \varphi(y) \leq 0 \\ y \in \overline{K} \end{cases} \iff y \in \mathrm{SOL}(S, K);$$

(c) the mapping $\Gamma^{S,K}$ *is Lipschitz continuous and convex.*

Proof. (a) Fix $y \in \mathbb{R}^n$ and let $\overline{v} \in K$ be such that $\|y - \overline{v}\| = \mathrm{dist}(y, K)$, which exists because K is nonempty and compact. Take any $\overline{s} \in S(\bar{v})$. Then, by Cauchy-Schwarz inequality and definition of $\overline{S}$,

$$\bar{s}^\top (y - \bar{v}) \;\geq\; -\|\bar{s}\|\,\|y - \bar{v}\| \;\geq\; -\overline{S}\,\|y - \bar{v}\| = -\overline{S}\ \mathrm{dist}(y, K).$$

Since $\Gamma^{S,K}(y)$ is the maximum over all pairs (v, s_v), we have

$$\Gamma^{S,K}(y) = \max_{s \in S(v)} \max_{v \in K} s^\top (y - v) \;\geq\; \bar{s}^\top (y - \bar{v}) \;\geq\; -\overline{S}\ \mathrm{dist}(y, K),$$

which proves (a).

(b) Let $y \in \mathrm{SOL}(S, K)$. Since $y \in K$, then $\varphi(y) \leq 0$ and $y \in \overline{K}$. Morover, since S is maximal monotone, due to Lemma 15(a), y is also a solution of the Minty problem (MGVI(S, K)), which is equivalent to writing

$$s^\top (y - v) \;\leq\; 0, \quad \forall v \in K,\ \forall s \in S(v),$$

so that

$$\Gamma^{S,K}(y) = \max_{s \in S(v)} \max_{v \in K} s^\top (y - v) \;\leq\; 0.$$

Vice versa, let $\varphi(y) \leq 0$, $y \in \overline{K}$ and $\Gamma^{S,K}(y) \leq 0$. Clearly, $y \in K$. By definition of $\Gamma^{S,K}$, for any $v \in K$ and $s \in S(v)$ we have

$$s^\top(y - v) \ \leq \ \Gamma^{S,K}(y) \leq 0,$$

hence

$$s^\top(v - y) \ \geq \ 0, \quad \forall v \in K, \ \forall s \in S(v),$$

that is, y satisfies the Minty variational inequality (MGVI(S, K)). Since S is maximal monotone, Lemma 15(b) guarantees that y is also a solution of (GVI(S, K)), thus proving (b).

(c) Let $y_1, y_2 \in \mathbb{R}^n$. For any $v \in K$ and $s \in S(v)$ we have

$$\begin{aligned} s^\top(y_1 - v) &= s^\top(y_2 - v) + s^\top(y_1 - y_2) \\ &\leq s^\top(y_2 - v) + \|s\| \, \|y_1 - y_2\| \\ &\leq s^\top(y_2 - v) + \overline{S} \, \|y_1 - y_2\|. \end{aligned}$$

Taking the maximum over $v \in K$ and $s \in S(v)$ on both sides yields

$$\begin{aligned} \Gamma^{S,K}(y_1) &= \max\nolimits_{s \in S(v)} \max\nolimits_{v \in K} s^\top(y_1 - v) \\ &\leq \max\nolimits_{s \in S(v)} \max\nolimits_{v \in K} s^\top(y_2 - v) + \overline{S} \, \|y_1 - y_2\| \\ &= \Gamma^{S,K}(y_2) + \overline{S} \, \|y_1 - y_2\|. \end{aligned}$$

Exchanging the roles of y_1 and y_2 we also obtain

$$\Gamma^{S,K}(y_2) \ \leq \ \Gamma^{S,K}(y_1) + \overline{S} \, \|y_1 - y_2\|.$$

Combining the two inequalities we arrive at

$$\left|\Gamma^{S,K}(y_1) - \Gamma^{S,K}(y_2)\right| \ \leq \ \overline{S} \, \|y_1 - y_2\|, \quad \forall y_1, y_2 \in \mathbb{R}^n,$$

so $\Gamma^{S,K}$ is globally Lipschitz continuous. The mapping $\Gamma^{S,K}$ is convex on $\mathbb{R}^n$ by the same reasoning as the proof of [1, Theorem 4.3.8], concluding the proof. $\square$

We also give the definition of the Stampacchia gap function.

Definition 19 (Stampacchia gap function)**.** *The Stampacchia gap function associated with* (GVI(S, K)) *is the function* $\Theta^{S,K} : \mathbb{R}^n \to \mathbb{R}$ *defined as*

$$\Theta^{S,K}(y) \triangleq \min_{s \in S(y)} \max_{v \in K} s^\top(y - v).$$

B Technical Results

Below we collect some technical results that are useful for the developments of our study.

The proof of the following Lemma follows standard reasonings and is therefore omitted.

Lemma 20. *Consider $u, v \in \mathbb{R}^n$, the following inequality holds:*

$$\|u+v\|^2 \leq 2\|u\|^2 + 2\|v\|^2.$$

The result in the following Lemma readily follows by applying Lemma 20 twice.

Lemma 21. *Consider the sequences $\{\gamma_j\}$, $\{\eta_j\}$ and $\{\zeta_j\}$ in Algorithm 1. The following inequality holds for any j*

$$\|\gamma_j h_{y_j}^{\eta_j,\zeta_j}\|^2 \leq 2\gamma_j^2 \left(\overline{F}^2 + 2\eta_1^2\, \overline{G}^2 + \; 2\eta_1^2\, \zeta_1^2\, \overline{T}^2\right). \tag{18}$$

Proof. Considering $h_{y_j}^{\eta_j,\zeta_j} = f_{y_j} + \gamma_j\eta_j g_{y_j} + \gamma_j\eta_j\zeta_j t_{y_j}$ with $f_{y_j} \in F(y_j)$, $g_{y_j} \in G(y_j)$, $t_{y_j} \in T(y_j)$ according to Algorithm 1, the proof follows from the following chain of inequalities:

$$\begin{aligned}
\|\gamma_j h_{y_j}^{\eta_j,\zeta_j}\|^2 &= \|\gamma_j f_{y_j} + \gamma_j\eta_j g_{y_j} + \gamma_j\eta_j\zeta_j t_{y_j}\|^2 \\
&\leq 2\|\gamma_j f_{y_j}\|^2 + 2\|\gamma_j\eta_j g_{y_j} + \gamma_j\eta_j\zeta_j t_{y_j}\|^2 \\
&\leq 2\|\gamma_j f_{y_j}\|^2 + 2\left(2\|\gamma_j\eta_j g_{y_j}\|^2 + 2\|\gamma_j\eta_j\zeta_j t_{y_j}\|^2\right) \\
&\leq 2\gamma_j^2\left(\overline{F}^2 + 2\eta_1^2\, \overline{G}^2 + \; 2\eta_1^2\, \zeta_1^2\, \overline{T}^2\right),
\end{aligned}$$

where the first inequality is due to Lemma 20 with $u = \gamma_j f_{y_j}$ and $v = \gamma_j\eta_j g_{y_j} + \gamma_j\eta_j\zeta_j t_{y_j}$, the second one is due to Lemma 20 with $u = \gamma_j\eta_j g_{y_j}$ and $v = \gamma_j\eta_j\zeta_j t_{y_j}$, and the third one is due to $\{\eta_j\}$ and $\{\zeta_j\}$ being nonincreasing. □

By means of Lemma 22, we obtain a key bound related to a generalized telescoping sum.

Lemma 22. *Let $\{c_j\} \subseteq \mathbb{R}_+$ be a bounded sequence, and let $\{\alpha_j\} \subseteq \mathbb{R}_{++}$ be nonincreasing. For any $k \in \mathbb{N}$, the following inequality holds, where $\overline{c} = \sup_j c_j$,*

$$\sum_{j=1}^{k} \frac{c_j - c_{j+1}}{\alpha_j} \leq \frac{\overline{c}}{\alpha_k}.$$

Proof. The proof follows from the following chain of inequalities

$$\begin{aligned}
\sum_{j=1}^{k} \frac{c_j - c_{j+1}}{\alpha_j} &= \frac{c_1}{\alpha_1} - \frac{c_{k+1}}{\alpha_k} + \sum_{j=1}^{k-1} c_{j+1}\left(\frac{1}{\alpha_{j+1}} - \frac{1}{\alpha_j}\right) \\
&\leq \frac{\overline{c}}{\alpha_1} + \overline{c}\sum_{j=1}^{k-1}\left(\frac{1}{\alpha_{j+1}} - \frac{1}{\alpha_j}\right) = \frac{\overline{c}}{\alpha_1} + \overline{c}\left(\frac{1}{\alpha_k} - \frac{1}{\alpha_1}\right) = \frac{\overline{c}}{\alpha_k}.
\end{aligned}$$

□

The proof of the next Lemma can be traced back to [7, Point 1 in Section 2.4.2].

Lemma 23. *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, $\lim_{k\to\infty} \sum_{j=1}^{k} b_j a_j / \sum_{j=1}^{k} b_j = \overline{a}$.*

Relying on the integral test for convergence of harmonic sequences (see e.g. [7, Chapter 9]), we provide the following bounds for harmonic series, which we rely on to recover convergence rate guarantees for Algorithm 1 in Section 4.

Lemma 24. *Let p be a positive scalar. For any $k \geq 2$, the following relations hold:*

(a) if $p \in (0,1)$, then $\frac{1-2^{p-1}}{1-p} k^{1-p} \leq \sum_{j=1}^{k} \frac{1}{j^p} \leq \frac{1}{1-p} k^{1-p}$;

(b) If $p \in (0,1]$, then $\sum_{j=1}^{k} \frac{1}{j^p} \leq 1 + k^{1-p} \log k$.

References

[1] B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer. *Non-Linear Parametric Optimization.* Akademie-Verlag Berlin, 1982.

[2] J. Bolte, A. Daniilidis, and A. Lewis. The łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007.

[3] J. Bolte, T.P. Nguyen, J. Peypouquet, and B.W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, 165(2):471–507, 2017.

[4] J.V. Burke and M.C. Ferris. Weak sharp minima in mathematical programming. *SIAM Journal on Control and Optimization*, 31(5):1340–1359, 1993.

[5] A. Cabot. Proximal point algorithm controlled by a slowly vanishing term: applications to hierarchical minimization. *SIAM Journal on Optimization*, 15(2):555–572, 2005.

[6] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer, 2003.

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

[8] L. Lampariello, S. Sagratella, and V.G. Sasso. Addressing hierarchical jointly convex generalized nash equilibrium problems with nonsmooth payoffs. *SIAM Journal on Optimization*, 35(1):445–475, 2025.

[9] P. Marcotte and D. Zhu. Weak sharp solutions of variational inequalities. *SIAM Journal on Optimization*, 9(1):179–189, 1998.

[10] R. Merchav, S. Sabach, and M. Teboulle. Dynamic fista for convex composite bi-level optimization. *arXiv preprint arXiv:2407.21221*, 2024.

[11] R.T. Rockafellar and J.B. Wets. *Variational Analysis*. Springer, 1998.

[12] S. Samadi and F. Yousefian. Improved guarantees for optimal nash equilibrium seeking and bilevel variational inequalities. *SIAM Journal on Optimization*, 35(1):369–399, 2025.

[13] A. Shafiei, V. Kungurtsev, and J. Marecek. Trilevel and multilevel optimization using monotone operator theory. *Mathematical Methods of Operations Research*, 99(1):77–114, 2024.