J. Nonlinear Var. Anal. 10 (2026), No. 4, pp. 829-844
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.10.2026.4.07

APPROXIMATE KKT CONDITIONS FOR GENERALIZED BILEVEL OPTIMIZATION WITH A VARIATIONAL-INEQUALITY LOWER LEVEL

RACHID EL IDRISSI[1], EL MOSTAFA KALMOUN[2,*], LAHOUSSINE LAFHIM[3]

[1]*Multidisciplinary Laboratory of Research and Innovation (LPRI), Moroccan School of Engineering Sciences (EMSI), Casablanca 20250, Morocco*
[2]*School of Science and Engineering, Al Akhawayn University in Ifrane, Ifrane 53000, Morocco*
[3]*L2MASI, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco*

Abstract. We study necessary optimality conditions for nonsmooth generalized bilevel optimization problems in which the lower level is a variational inequality. Using a gap-function reformulation and partial calmness, we derive Karush–Kuhn–Tucker (KKT) type conditions via the basic (limiting) subdifferential. We then introduce an approximate KKT (AKKT) framework tailored to this nonsmooth, hierarchical setting and prove that every KKT point is an AKKT point, while the converse may fail in general, as demonstrated by a counterexample. Under natural regularity assumptions, including local Lipschitz continuity of the gap function and a nondegeneracy condition on the argmax set, we recover the equivalence AKKT $\Rightarrow$ KKT. A model example illustrates the applicability of the approach and the sharpness of the assumptions. Our results position AKKT conditions as a robust tool for bilevel analysis and computation when smoothness or classical constraint qualifications are unavailable.

Keywords. Bilevel optimization; Gap function; KKT conditions; Partial calmness; Variational inequality.

1. INTRODUCTION

A generalized bilevel optimization problem (GBOP) is a hierarchical problem with two interdependent levels, where the lower level is, in our setting, a variational inequality (VI). The upper level is an optimization problem whose decisions affect, and are affected, by the response of the lower-level VI. We consider

$$\begin{cases} \min\limits_{x,y} f(x,y) \\ \text{subject to } x \in X,\ y \in S(x), \end{cases} \tag{1.1}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $X \subset \mathbb{R}^n$, and $S(x)$ denotes the solution set of the VI

$$\text{find } y \in Y(x) \text{ such that } \langle F(x,y), y - z \rangle \leq 0 \quad \forall z \in Y(x), \tag{1.2}$$

with $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ and $Y(x) \subset \mathbb{R}^m$.

---

[*] Corresponding author.

E-mail address: E.Kalmoun@aui.ma (E.M. Kalmoun).

Received 22 September 2025; Accepted 4 March 2026; Published online 5 May 2026.

The feasible sets are

$$X = \{x \in \mathbb{R}^n : \ g_i(x) \leq 0, \ i = 1, \ldots, p\}, \qquad Y(x) = \{y \in \mathbb{R}^m : \ G_j(x,y) \leq 0, \ j = 1, \ldots, q\},$$

and the overall feasible region is $\Omega = \{(x,y) : x \in X, \ y \in Y(x)\}$. Unless otherwise stated, all functions are assumed locally Lipschitz around reference points, which allows us to handle nonsmooth data.

Variational inequalities, a broad class of equilibrium problems, have been extensively studied in the optimization community due to their wide applicability, ranging from economics and continuum mechanics to models of biological interactions [1]. In recent years, they have also been employed as lower-level problems in bilevel optimal control frameworks; see, for instance, [2]. These applications motivate us to investigate and extend the framework of generalized bilevel optimization problems. In [3, Theorem 4.8], the authors studied problem (1.1) and derived Karush–Kuhn–Tucker (KKT) type optimality conditions for local minimizers $(\bar{x}, \bar{y})$. However, their results rely on restrictive assumptions: the functions f, F, and G are required to be continuously differentiable, and $\bar{x}$ must lie in the interior of X. These assumptions fail in many applications involving nonsmooth data. This paper addresses this gap by developing necessary optimality conditions for problem (1.1) in a nonsmooth setting, using tools from variational analysis, in particular the basic subdifferential.

Our contributions. We develop necessary optimality conditions for nonsmooth GBOPs with VI lower levels using basic subdifferential calculus. First, we derive KKT-type conditions without differentiability or interiority assumptions. Then, we introduce an approximate KKT (AKKT) framework, prove that every KKT point is an AKKT point, and show by counterexample that the converse fails in general. Finally, we identify verifiable regularity conditions under which AKKT implies KKT. As a consistency check, we show in Remark 4.2 that, under standard smoothness assumptions (e.g., $f, F, G \in C^1$) and the usual feasibility/interiority conditions, our nonsmooth KKT system $(\mathscr{K}_1)$–$(\mathscr{K}_6)$ reduces to the classical optimality conditions of [3, Theorem 4.8].

It is worth noting that AKKT conditions have gained attention in recent years as a relaxation of standard optimality conditions; see, e.g., [4, 5, 6]. AKKT conditions have also been successfully used to define stopping criteria in algorithms such as augmented Lagrangian methods [7], interior-point methods [8], and sequential quadratic programming [9], as well as in modeling real-world problems [10, 11]. In this paper, we extend the AKKT framework to the non-smooth GBOP (1.1), thereby bridging the gap between theory and applications.

Finally, we note that the penalization strategy used in this work can be interpreted as a scalarization of two criteria: the upper-level objective and a measure of violation of the lower-level variational inequality. This perspective is classical in set and vector optimization and is consistent with our use of set-valued analysis tools.

Paper organization. Section 2 recalls notation and tools from variational analysis. Using the gap function, Section 3 provides an equivalent single-level reformulation of problem (1.1) and extends partial calmness to obtain a locally equivalent penalized problem. Section 4 develops KKT-type optimality conditions via the basic subdifferential. Section 5 derives new necessary optimality conditions of AKKT type and establishes the existence of AKKT sequences at local minimizers. Section 6 analyzes the relationship between the derived KKT and AKKT conditions. Finally, Section 7 concludes the paper.

2. PRELIMINARIES

Let $x \in \mathbb{R}^n$ and $\varepsilon > 0$. We denote by $\mathbb{U}$ and $\mathbb{B}$ the open and closed unit balls of $\mathbb{R}^n$, respectively. The corresponding balls centered at x with radius ε are defined by

$$\mathbb{U}^{\varepsilon}(x) = x + \varepsilon\mathbb{U}, \quad \mathbb{B}^{\varepsilon}(x) = x + \varepsilon\mathbb{B}.$$

For the one-dimensional case ($n = 1$), we define the positive part of a scalar as $x_+ = \max\{x, 0\}$. Given a set $A \subset \mathbb{R}^n$, $\operatorname{conv}(A)$ denotes the convex hull of A.

We also recall basic notions for set-valued mappings, which are used later. Let $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping. Its graph is $\operatorname{gph} M := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \ : \ y \in M(x)\}$. For $\bar{u} \in \mathbb{R}^n$, the Painlevé–Kuratowski *outer limit* of $M(u)$ as $u \to \bar{u}$ is

$$\limsup_{u \to \bar{u}} M(u) := \left\{ y \in \mathbb{R}^m \ : \ \exists u^k \to \bar{u},\ \exists y^k \in M(u^k),\ y^k \to y \right\}.$$

We say that M is *outer semicontinuous (o.s.c.)* at $\bar{u}$ if

$$\limsup_{u \to \bar{u}} M(u) \ \subseteq \ M(\bar{u}).$$

Equivalently (sequential characterization), M is o.s.c. at $\bar{u}$ if, for every sequence $u^k \to \bar{u}$ and every $y^k \in M(u^k)$ with $y^k \to y$, $y \in M(\bar{u})$.

Proposition 2.1 (Argmax mapping and Berge's maximum theorem)**.** *Let $U \subset \mathbb{R}^{n+m}$ and let $M : U \rightrightarrows \mathbb{R}^m$ be nonempty, compact-valued and o.s.c. at $\bar{u} \in U$. If $\psi : U \times \mathbb{R}^m \to \mathbb{R}$ is continuous, then the argmax map* $\operatorname{Argmax}_{z \in M(u)} \psi(u,z)$ *is nonempty, compact-valued, and o.s.c. at $\bar{u}$.*

In our setting, we will apply this with $u = (x,y)$, the feasible-set mapping $M(u) \equiv Y(x)$, and $\psi(u,z) = \langle F(x,y), y - z\rangle$, which yields the outer semicontinuity of the argmax map $Z(\cdot,\cdot)$ used later in the proof of Theorem 5.1.

We next recall some notions of normal cones, which play a fundamental role in variational analysis and in the formulation of subdifferentials.

Definition 2.1 ([12])**.** Let $\mathscr{A} \subset \mathbb{R}^n$ be a nonempty set and $\bar{x} \in \mathscr{A}$.

($\mathscr{D}_1$) The Fréchet normal cone to $\mathscr{A}$ at $\bar{x}$ is defined by

$$\hat{N}(\mathscr{A}, \bar{x}) = \left\{ \alpha \in \mathbb{R}^n : \limsup_{x \xrightarrow{\mathscr{A}} \bar{x}} \frac{\langle \alpha, x - \bar{x}\rangle}{\|x - \bar{x}\|} \leq 0 \right\},$$

where $x \xrightarrow{\mathscr{A}} \bar{x}$ means $x \to \bar{x}$ with $x \in \mathscr{A}$.

($\mathscr{D}_2$) The basic normal cone to $\mathscr{A}$ at $\bar{x}$ is given by

$$N(\mathscr{A}, \bar{x}) = \operatorname*{Limsup}_{x \to \bar{x}} \hat{N}(\mathscr{A}, x).$$

We now turn to subdifferentials. Recall that for an extended real-valued function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, its epigraph is the set of points lying on or above its graph:

$$\operatorname{epi} \psi = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : \ \psi(x) \leq \alpha\}.$$

Definition 2.2 ([12])**.** Let $\bar{x} \in \mathbb{R}^n$ be such that $\psi(\bar{x})$ is finite. The basic subdifferential of ψ at $\bar{x}$ is defined by

$$\partial \psi(\bar{x}) = \left\{ \alpha \in \mathbb{R}^n : \ (\alpha, -1) \in N\big(\operatorname{epi} \psi,\ (\bar{x}, \psi(\bar{x}))\big) \right\}.$$

The subdifferential enjoys several fundamental properties. The following lemma ensures compactness under a mild regularity assumption.

Lemma 2.1 ([12])**.** *Let $\bar{x} \in \mathbb{R}^n$ and suppose that ψ is locally Lipschitz around $\bar{x}$. Then, $\partial\psi(\bar{x})$ is a compact set.*

3. ONE-LEVEL REFORMULATION

In this section, we study how generalized bilevel optimization problem (1.1) can be reformulated as an equivalent single-level optimization problem. Our approach is based on the use of a gap function, which provides a convenient reformulation of the lower-level variational inequality and yields a fully equivalent single-level problem. This technique has been widely investigated in the literature on generalized bilevel optimization; see, e.g., [2, 3, 13].

In addition, we highlight the role of the so-called partial calmness property, originally introduced by Ye and Zhu [14, 15] in the context of optimistic bilevel problems as a type of constraint qualification. We extend this concept to the generalized bilevel optimization problem (1.1) in order to obtain an alternative equivalent formulation.

We begin with the following definition. The gap function $\varphi_{GF} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ associated with the variational inequality (1.2) is given by

$$\varphi_{GF}(x,y) = \sup\{\langle F(x,y), y-z\rangle \ : \ z \in Y(x)\}.$$

It is straightforward to check that

$$\begin{cases} \varphi_{GF}(x,y) \geq 0 \text{ for all } y \in Y(x), \\ [\varphi_{GF}(x,y) = 0 \text{ and } y \in Y(x)] \text{ if and only if } y \in S(x). \end{cases}$$

Hence, the solution set of the variational inequality can be written equivalently as

$$S(x) = \{y \in \mathbb{R}^m \ : \ \varphi_{GF}(x,y) = 0 \text{ and } y \in Y(x)\}. \tag{3.1}$$

Substituting (3.1) into the upper-level problem, the generalized bilevel optimization problem (1.1) can be reformulated as the following single-level problem:

$$\min_{x,y} f(x,y) \quad \text{subject to} \quad (x,y) \in \Theta, \tag{3.2}$$

where

$$\Theta = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \ : \ (x,y) \in \Omega, \ \varphi_{GF}(x,y) = 0\}.$$

We next define the notion of partial calmness, which provides a useful stability property of the reformulated problem.

Definition 3.1. Let $(\bar{x},\bar{y}) \in \Theta$. We say that problem (3.2) is partially calm at $(\bar{x},\bar{y})$ if there exist a neighborhood $\mathscr{N}$ of $(\bar{x},\bar{y},0)$ and a scalar $\mu > 0$ such that $f(x,y) - f(\bar{x},\bar{y}) + \mu a \geq 0$, for all $(x,y,a) \in \mathscr{N} \cap (\Omega \times \mathbb{R})$ with $\varphi_{GF}(x,y) = a$.

Remark 3.1. If problem (3.2) is partially calm at $(\bar{x},\bar{y})$ with a parameter $\mu > 0$, then (3.2) is partially calm at $(\bar{x},\bar{y})$ with a parameter $\mu + 1$.

Theorem 3.1. *Let $(\bar{x},\bar{y})$ be a local optimal solution for problem (1.1). Assume that φ_{GF} is locally Lipschitz at $(\bar{x},\bar{y})$. Then, problem (3.2) is partially calm at $(\bar{x},\bar{y})$ if and only if there* *exists a scalar* $\mu > 0$ *such that* $(\bar{x},\bar{y})$ *is a local optimal solution of problem for the partial penalization problem*

$$\min_{x,y} f(x,y) + \mu \varphi_{GF}(x,y) \quad \textit{subject to} \quad (x,y) \in \Omega. \tag{3.3}$$

Proof. We divide the proof into two parts.

Assume first that problem (3.2) is partially calm at $(\bar{x},\bar{y})$. Then, there exist $\mu, \varepsilon > 0$ such that

$$f(x,y) - f(\bar{x},\bar{y}) + \mu a \geq 0, \tag{3.4}$$

for all $(x,y,a) \in \mathbb{U}^{\varepsilon}(\bar{x},\bar{y},0)$ that satisfies $(x,y) \in \Omega$ and $\varphi_{GF}(x,y) = a$.

Since φ_{GF} is locally Lipschitz around $(\bar{x},\bar{y})$, then there exist two scalars $\bar{\varepsilon}, L_{\varphi_{GF}} > 0$ such that

$$|\varphi_{GF}(x_1,y_1) - \varphi_{GF}(x_2,y_2)| \leq L_{\varphi_{GF}} \|(x_1,y_1) - (x_2,y_2)\| \tag{3.5}$$

for each $(x_1,y_1), (x_2,y_2) \in \mathbb{U}^{\bar{\varepsilon}}(\bar{x},\bar{y})$.

Now, take $(x,y) \in \Omega \cap \mathbb{U}^{\varepsilon^\star}(\bar{x},\bar{y})$ with $\varepsilon^\star = \min\left\{\varepsilon, \bar{\varepsilon}, \dfrac{\varepsilon}{L_{\varphi_{GF}}}\right\}$, and let a such that $\varphi_{GF}(x,y) = a$. Then we have

$$\begin{aligned} a &\overset{(1^*)}{=} \varphi_{GF}(x,y) - \varphi_{GF}(\bar{x},\bar{y}) \\ &\overset{(2^*)}{\leq} L_{\varphi_{GF}} \|(x,y) - (\bar{x},\bar{y})\| \\ &\overset{(3^*)}{<} L_{\varphi_{GF}} \varepsilon^\star \\ &\overset{(4^*)}{\leq} \varepsilon, \end{aligned}$$

where (1^*) follows from $\bar{y} \in S(\bar{x})$, (2^*) from (3.5), (3^*) from $(x,y) \in \mathbb{U}^{\varepsilon^\star}(\bar{x},\bar{y})$, and (4^*) from the choice of $\varepsilon^\star$. Therefore, $(x,y,a) \in \mathbb{U}^{\varepsilon}(\bar{x},\bar{y},0)$. Using $\varphi_{GF}(x,y) = a$, $\bar{y} \in S(\bar{x})$, (3.4), and $\varphi_{GF}(\bar{x},\bar{y}) = 0$, we obtain

$$f(x,y) + \mu \varphi_{GF}(x,y) = f(x,y) + \mu a \geq f(\bar{x},\bar{y}) = f(\bar{x},\bar{y}) + \mu \varphi_{GF}(\bar{x},\bar{y}).$$

Hence, $(\bar{x},\bar{y})$ is a local optimal solution to partial penalization problem (3.3).

Conversely, suppose that there exists a scalar $\mu > 0$ such that $(\bar{x},\bar{y})$ is a local optimal solution to partial penalization problem (3.3). Then, there exists a constant $\nu > 0$ such that, for all $(x,y) \in \Omega \cap \mathbb{U}^{\nu}(\bar{x},\bar{y})$,

$$f(\bar{x},\bar{y}) \leq f(x,y) + \mu \varphi_{GF}(x,y), \tag{3.6}$$

Hence, for each $(x,y) \in \Omega \cap \mathbb{U}^{\nu}(\bar{x},\bar{y})$ such that $\varphi_{GF}(x,y) = a$, we obtain from (3.6) that

$$f(x,y) - f(\bar{x},\bar{y}) + \mu a \geq 0.$$

This shows that problem (3.2) is partially calm at $(\bar{x},\bar{y})$. ☐

Remark 3.2. The penalized objective $f(x,y) + \mu\, \varphi_{GF}(x,y)$ can be viewed as a scalarization of the two-criterion model (f, φ_{GF}). Under partial calmness (Theorem 3.1) and local Lipschitz continuity of φ_{GF}, this scalarization is exact in a neighborhood of a local solution, so that minimizers of the penalized problem coincide locally with feasible minimizers of the original bilevel model. This perspective connects the present framework to classical scalarization principles in vector and set optimization.

4. KKT CONDITIONS

In this section, we present necessary optimality conditions for problem (1.1) in the framework of KKT theory. These conditions provide a characterization of local optimal solutions and are particularly useful in hierarchical optimization problems.

We first collect auxiliary sets used in the subdifferential calculus of the gap function φ_{GF} and in the statement of the KKT-type conditions.

Notation 4.1. Let $(\bar{x},\bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$.

- $Z(\bar{x},\bar{y})$ is the set of vectors $z \in Y(\bar{x})$ at which $\varphi_{GF}(\bar{x},\bar{y})$ attains the maximum.
- $\mathscr{M}_{\bar{x}}(z) = \left\{\lambda \in \mathbb{R}^q \,:\, 0 \in \partial_y G(\bar{x},z)^\top \lambda,\ \lambda \geq 0,\ G(\bar{x},z)^\top \lambda = 0\right\}$.
- $\Lambda_{(\bar{x},\bar{y})}(z) = \left\{\lambda \in \mathbb{R}^q \,:\, 0 \in F(\bar{x},\bar{y}) + \partial_y G(\bar{x},z)^\top \lambda,\ \lambda \geq 0,\ G(\bar{x},z)^\top \lambda = 0\right\}$.

Remark 4.1. For fixed $(\bar{x},\bar{y})$, the map $z \mapsto \langle F(\bar{x},\bar{y}), \bar{y} - z\rangle$ is continuous and affine in z. Hence, if $Y(\bar{x})$ is nonempty and compact, the supremum is attained and $Z(\bar{x},\bar{y}) \neq \varnothing$ by Weierstrass' theorem.

From [16, Corollary 1 of Theorem 6.5.2], we get the following result.

Theorem 4.1. *Suppose that $\mathscr{M}_{\bar{x}}(Z(\bar{x},\bar{y}))$ reduces to $\{0\}$. Then, φ_{GF} is locally Lipschitz at $(\bar{x},\bar{y})$ and*

$$\partial \varphi_{GF}(\bar{x},\bar{y}) \subset \operatorname{conv} \bigcup_{z \in Z(\bar{x},\bar{y})} \bigcup_{\lambda \in \Lambda_{(\bar{x},\bar{y})}(z)} \left\{ \begin{pmatrix} \partial_x F(\bar{x},\bar{y})^\top (\bar{y} - z) - \partial_x G(\bar{x},z)^\top \lambda \\ \partial_y F(\bar{x},\bar{y})^\top (\bar{y} - z) + F(\bar{x},\bar{y}) \end{pmatrix} \right\}.$$

Theorem 4.2. *Let $(\bar{x},\bar{y})$ be a local optimal solution of problem (1.1). Suppose that $\mathscr{M}_{\bar{x}}(Z(\bar{x},\bar{y})) = \{0\}$ and problem (3.2) is partially calm at $(\bar{x},\bar{y})$. Then, there exist $\mu > 0$, $(\alpha,\beta) \in \mathbb{R}_+^{p+q}$, $z_r \in Z(\bar{x},\bar{y})$, $(\lambda_r,\gamma_r) \in \mathbb{R}_+^{q+1}$, $r = 1,\ldots,n+m+1$ such that $\sum_{j=1}^{n+m+1} \gamma_j = 1$ and*

$(\mathscr{K}_1)$ $0 \in \partial_x f(\bar{x},\bar{y}) + \partial g(\bar{x})^\top \alpha + \partial_x G(\bar{x},\bar{y})^\top \beta$
$\qquad + \mu \sum_{r=1}^{n+m+1} \gamma_r \left(\partial_x F(\bar{x},\bar{y})^\top (\bar{y} - z_r) - \partial_x G(\bar{x},z_r)^\top \lambda_r\right)$,

$(\mathscr{K}_2)$ $0 \in \partial_y f(\bar{x},\bar{y}) + \partial_y G(\bar{x},\bar{y})^\top \beta + \mu \left(F(\bar{x},\bar{y}) + \sum_{r=1}^{n+m+1} \gamma_r \partial_y F(\bar{x},\bar{y})^\top (\bar{y} - z_r)\right)$,

$(\mathscr{K}_3)$ $0 \in F(\bar{x},\bar{y}) + \partial_y G(\bar{x},z_r)^\top \lambda_r,\ r = 1,\ldots,n+m+1$,

$(\mathscr{K}_4)$ $g_i(\bar{x}) < 0 \Rightarrow \alpha_i = 0$, *for all* $i = 1,\ldots,p$,

$(\mathscr{K}_5)$ $G_j(\bar{x},\bar{y}) < 0 \Rightarrow \beta_j = 0$, *for all* $j = 1,\ldots,q$,

$(\mathscr{K}_6)$ $G_j(\bar{x},z_r) < 0 \Rightarrow \lambda_{r,j} = 0$, *for all* $j = 1,\ldots,q$ *and* $r = 1,\ldots,n+m+1$.

Proof. Since $(\bar{x},\bar{y})$ is a local optimal solution of problem (1.1), then $(\bar{x},\bar{y})$ is a local optimal solution of problem (3.2). Using the fact that problem (3.2) is partially calm at $(\bar{x},\bar{y})$ we can conclude that there exists $\mu > 0$ such that $(\bar{x},\bar{y})$ is a local optimal solution for the partial penalization problem (3.3). Using Theorem 4.1 and applying [12, Proposition 5.3] to problem (3.3), we have

$$0 \in \partial f(\bar{x},\bar{y}) + \mu \partial \varphi_{GF}(\bar{x},\bar{y}) + N(\Omega,(\bar{x},\bar{y})).$$

Finally, using Theorem 4.1 together with [17, Theorem 6.14], we deduce the existence of multipliers $(\alpha,\beta) \in \mathbb{R}_+^{p+q}$, $z_r \in Z(\bar{x},\bar{y})$, $(\lambda_r,\gamma_r) \in \mathbb{R}_+^{q+1}$, $r = 1,\ldots,n+m+1$ such that $\sum_{r=1}^{n+m+1} \gamma_r = 1$ and the conditions $(\mathscr{K}_1) - (\mathscr{K}_6)$ are satisfied. $\square$

Remark 4.2. A key validation of our framework is its consistency with the smooth case. If we assume that all the functions in Theorem 4.2 are continuously differentiable, our necessary optimality conditions $(\mathscr{K}_1)-(\mathscr{K}_6)$ reduce to those established for smooth generalized bilevel optimization problems [3, Theorem 4.8].

Definition 4.1. Let $(\bar{x},\bar{y})$ be a feasible point of problem (1.1). We say that $(\bar{x},\bar{y})$ is a KKT point of problem (1.1) if there exist $\mu>0$, $(\alpha,\beta)\in\mathbb{R}_+^{p+q}$, $z_r\in Z(\bar{x},\bar{y})$, $(\lambda_r,\gamma_r)\in\mathbb{R}_+^{q+1}$, $r=1,\ldots,n+m+1$ such that $\sum_{r=1}^{n+m+1}\gamma_r=1$ and the conditions $(\mathscr{K}_1)-(\mathscr{K}_6)$ of Theorem 4.2 are satisfied.

Example 4.1. Consider the following generalized bilevel optimization problem

$$\begin{cases}\min\limits_{x,y} f(x,y)=x^2+|y| \\ \text{subject to } g(x)=-x\le 0 \\ \qquad\qquad y\in S(x),\end{cases}$$

where, for each $x\in\mathbb{R}$, (x) denotes the solution set of the variational inequality

$$\text{find } y\in Y(x) \text{ such that } \langle F(x,y),y-z\rangle\le 0\ \forall z\in Y(x),$$

with $F(x,y)=|x|$ and

$$\begin{aligned}Y(x)&=\{y\in\mathbb{R}\ :\ G_1(x,y)=-x+y-3\le 0 \quad\text{and}\quad G_2(x,y)=-y\le 0\}\\ &=\begin{cases}\emptyset & \text{if } x<-3\\ [0,x+3] & \text{if } x\ge -3.\end{cases}\end{aligned}$$

One can verify that the solution set of the above variational inequality is

$$S(x)=\begin{cases}\emptyset & \text{if } x<-3\\ [0,3] & \text{if } x=0\\ \{0\} & \text{otherwise,}\end{cases}$$

and the associated gap function takes the form $\varphi_{GF}(x,y)=|x|y$.

It is easy to check that the point $(\bar{x},\bar{y})=(0,0)$ is a local optimal solution of the considered problem. Hence, $Z(\bar{x},\bar{y})=[0,3]$, and therefore $\mathscr{M}_{\bar{x}}(Z(\bar{x},\bar{y}))=\{0\}$. Moreover, we have

$$\begin{aligned}&\Omega=\{(x,y)\in\mathbb{R}\times\mathbb{R}\ :\ x\ge 0 \quad\text{and}\quad 0\le y\le x+3\},\\ &\Theta=\{0\}\times[0,3]\cup\mathbb{R}_+\times\{0\}.\end{aligned}$$

Since $f(\bar{x},\bar{y})=0$, $f(x,y)\ge 0$, and $\varphi_{GF}(x,y)\ge 0$ for all $y\in Y(x)$, then, for each chosen neighborhood $\mathscr{N}$ of $(\bar{x},\bar{y},0)$ and each scalar $\mu>0$, the calmness property at $(\bar{x},\bar{y})$ of the problem

$$\min_{x,y} f(x,y) \quad\text{subject to}\quad (x,y)\in\Theta,$$

is satisfied. Consequently, all the hypotheses of Theorem 4.2 are verified.

Finally, one can verify that $(\bar{x},\bar{y})$ is indeed a KKT point of this problem with the choice of multipliers: $\mu=\frac{1}{3}$, $\alpha=\frac{1}{3}$, $\beta=\left(0,\frac{1}{3}\right)$, $z_1=1$, $z_2=2$, $z_3=3$, $\lambda_1=(0,0)$, $\lambda_2=(0,0)$, $\lambda_3=(1,1)$, and $\gamma_1=\gamma_2=\gamma_3=\frac{1}{3}$.

5. AKKT CONDITIONS

In this section, we introduce the notion of AKKT points for (1.1). This concept generalizes classical KKT conditions by allowing sequences of approximate multipliers.

Definition 5.1. Let $(\bar{x},\bar{y})$ be a feasible point of problem (1.1). We say that $(\bar{x},\bar{y})$ is an AKKT point of problem (1.1) if there exist sequences $\left\{\left(x^k,y^k\right)\right\} \subset \mathbb{R}^{n+m}$, $\left\{\mu^k\right\} \subset \mathbb{R}_+$, $\left\{\left(\alpha^k,\beta^k\right)\right\} \subset \mathbb{R}_+^{p+q}$, $z_r^k \underset{k\to+\infty}{\longrightarrow} z_r \in Z(\bar{x},\bar{y})$, $\left\{\left(\lambda_r^k,\gamma_r^k\right)\right\} \subset \mathbb{R}_+^{q+1}$, $r=1,\ldots,n+m+1$ such that $\sum_{r=1}^{n+m+1}\gamma_r^k = 1$ such that the following conditions hold:

- $(\mathscr{A}_0)$: $\left(x^k,y^k\right) \underset{k\to+\infty}{\longrightarrow} (\bar{x},\bar{y})$,
- $(\mathscr{A}_1)$: $m_1\left(x^k,y^k,z_1^k,\ldots,z_{n+m+1}^k,\mu^k,\alpha^k,\beta^k,\lambda_1^k,\ldots,\lambda_{n+m+1}^k\right) \underset{k\to+\infty}{\longrightarrow} 0$,
- $(\mathscr{A}_2)$: $m_2\left(x^k,y^k,z_1^k,\ldots,z_{n+m+1}^k,\mu^k,\beta^k\right) \underset{k\to+\infty}{\longrightarrow} 0$,
- $(\mathscr{A}_3)$: $m_3\left(x^k,y^k,z_r^k,\lambda_r^k\right) \underset{k\to+\infty}{\longrightarrow} 0$, for $r=1,\ldots,n+m+1$,
- $(\mathscr{A}_4)$: $g_i(\bar{x}) < 0 \Rightarrow \alpha_i^k = 0$ for sufficiently large k and $i=1,\ldots,p$,
- $(\mathscr{A}_5)$: $G_j(\bar{x},\bar{y}) < 0 \Rightarrow \beta_j^k = 0$ for sufficiently large k and $j=1,\ldots,q$,
- $(\mathscr{A}_6)$: $G_j(\bar{x},z_r) < 0 \Rightarrow \lambda_{r,j}^k = 0$ for sufficiently large k, $j=1,\ldots,q$, and $r=1,\ldots,n+m+1$,

where

$$\begin{aligned}
& m_1\left(x^k,y^k,z_1^k,\ldots,z_{n+m+1}^k,\mu^k,\alpha^k,\beta^k,\lambda_1^k,\ldots,\lambda_{n+m+1}^k\right) \\
&= \inf\Bigg\{\left\|\eta_1^k+\sum_{i=1}^{p}\alpha_i^k\rho_i^k+\sum_{j=1}^{q}\beta_j^k\vartheta_{1,j}^k+\mu^k\sum_{r=1}^{n+m+1}\gamma_r^k\left[\sum_{s=1}^{m}\left(y_s^k-z_{r,s}^k\right)\nu_{1,s}^k-\sum_{j=1}^{q}\lambda_{r,j}^k\delta_{1,j}^{r,k}\right]\right\| : \\
&\qquad \eta_1^k \in \partial_x f\left(x^k,y^k\right);\ \rho_i^k \in \partial g_i\left(x^k\right),\ i=1,\ldots,p; \\
&\qquad \vartheta_{1,j}^k \in \partial_x G_j\left(x^k,y^k\right),\ j=1,\ldots,q;\ \nu_{1,s}^k \in \partial_x F_s\left(x^k,y^k\right),\ s=1,\ldots,m; \\
&\qquad \delta_{1,j}^{r,k} \in \partial_x G_j\left(x^k,z_r^k\right),\ j=1,\ldots,q,\ r=1,\ldots,n+m+1\Bigg\}, \\
& m_2\left(x^k,y^k,z_1^k,\ldots,z_{n+m+1}^k,\mu^k,\beta^k\right) \\
&= \inf\Bigg\{\left\|\eta_2^k+\sum_{j=1}^{q}\beta_j^k\vartheta_{2,j}^k+\mu^k\left[F\left(x^k,y^k\right)+\sum_{r=1}^{n+m+1}\gamma_r^k\sum_{s=1}^{m}\left(y_s^k-z_{r,s}^k\right)\nu_{2,s}^k\right]\right\| : \\
&\qquad \eta_2^k \in \partial_y f\left(x^k,y^k\right);\ \vartheta_{2,j}^k \in \partial_y G_j\left(x^k,y^k\right),\ j=1,\ldots,q; \\
&\qquad \nu_{2,s}^k \in \partial_y F_s\left(x^k,y^k\right),\ s=1,\ldots,m\Bigg\}, \\
& m_3\left(x^k,y^k,z_r^k,\lambda_r^k\right) = \inf\Bigg\{\left\|F\left(x^k,y^k\right)+\sum_{j=1}^{q}\lambda_{r,j}^k\delta_{2,j}^{r,k}\right\| : \\
&\qquad\qquad \delta_{2,j}^{r,k} \in \partial_y G_j\left(x^k,z_r^k\right),\ j=1,\ldots,q,\ r=1,\ldots,n+m+1\Bigg\}.
\end{aligned}$$

The AKKT conditions provide a framework in which approximate multipliers and perturbations allow us to describe stationarity for nonsmooth or hierarchical problems.

The next lemma shows how local optimality implies uniqueness of a regularized problem in a neighborhood of a local solution.

Lemma 5.1. *Let* $(\bar{x},\bar{y})$ *be a local optimal solution for problem* (1.1). *Then, there exists a scalar* $\varepsilon > 0$ *such that* $(\bar{x},\bar{y})$ *is the unique optimal solution to the following optimization problem:*

$$\min_{x,y} f(x,y) - f(\bar{x},\bar{y}) + \frac{1}{2}\|(x,y)-(\bar{x},\bar{y})\|^2 \quad \textit{subject to} \quad (x,y) \in \Theta \cap \mathbb{B}^{\varepsilon}(\bar{x},\bar{y}). \tag{5.1}$$

Proof. Since $(\bar{x},\bar{y})$ is a local optimal solution of problem (1.1), then there exists $\varepsilon > 0$ such that $f(\bar{x},\bar{y}) \le f(x,y)$ for all $(x,y) \in \Theta \cap \mathbb{B}^{\varepsilon}(\bar{x},\bar{y})$. Suppose, by contradiction, that problem (5.1) admits another optimal solution $(x^*,y^*) \in \Theta \cap \mathbb{B}^{\varepsilon}(\bar{x},\bar{y})$ different from $(\bar{x},\bar{y})$. Hence, we have

$$f(\bar{x},\bar{y}) \le f(x^*,y^*). \tag{5.2}$$

Remarking that the optimal value of problem (5.1) is 0, one has

$$f(x^*,y^*) - f(\bar{x},\bar{y}) = -\frac{1}{2}\|(x^*,y^*)-(\bar{x},\bar{y})\|^2.$$

Combining this with (5.2) yields $\|(x^*,y^*)-(\bar{x},\bar{y})\| = 0$, i.e., $(x^*,y^*) = (\bar{x},\bar{y})$. Consequently, $(\bar{x},\bar{y})$ is the unique optimal solution of (5.1). ☐

We are now ready to establish that local optimality implies the existence of AKKT sequences.

Theorem 5.1. *Let* $(\bar{x},\bar{y})$ *be a local optimal solution for problem* (1.1). *Assume there exists* $\varepsilon > 0$ *such that* $\operatorname{gph} Y \cap (\mathbb{B}^{\varepsilon}(\bar{x}) \times \mathbb{R}^m)$ *is compact, and* $\mathscr{M}_x(z) = \{0\}$ *for all* (x,z) *with* $x \in \mathbb{B}^{\varepsilon}(\bar{x})$ *and* $z \in Y(x)$. *Then,* $(\bar{x},\bar{y})$ *is an AKKT point of problem* (1.1).

Proof. Since $(\bar{x},\bar{y})$ is a local optimal solution of (1.1), then, according to Lemma 5.1, $(\bar{x},\bar{y})$ is the unique optimal solution of (5.1). For an integer $k > 0$, we consider the following parametric optimization problem:

$$\min_{x,y} J_k(x,y) \quad \text{subject to} \quad (x,y) \in \mathbb{B}^{\varepsilon}(\bar{x},\bar{y}), \tag{5.3}$$

where

$$\begin{aligned} J_k(x,y) =& f(x,y) - f(\bar{x},\bar{y}) + \frac{1}{2}\|(x,y)-(\bar{x},\bar{y})\|^2 + \sum_{i=1}^{p} k\,(g_i(x))_+^2 \\ &+ \sum_{j=1}^{q} k\left(G_j(x,y)\right)_+^2 + k\,(\varphi_{GF}(x,y))_+^2. \end{aligned}$$

One can check that J_k is continuous on the compact set $\mathbb{B}^{\varepsilon}(\bar{x},\bar{y})$. Hence (5.3) admits an optimal solution, denoted as $\left(x^k,y^k\right)$. Let $(\tilde{x},\tilde{y})$ be an accumulation point of the sequence $\{(x^k,y^k)\}$. Then $(\tilde{x},\tilde{y})$ is feasible for (5.1). Since (x^k,y^k) is feasible for (5.3), we have $\|(\bar{x},\bar{y})-(x^k,y^k)\| \le \varepsilon$, so $\|(\bar{x},\bar{y})-(\tilde{x},\tilde{y})\| \le \varepsilon$. Suppose

$$\sum_{i=1}^{p}(g_i(\tilde{x}))_+^2 + \sum_{j=1}^{q}(G_j(\tilde{x},\tilde{y}))_+^2 + (\varphi_{GF}(\tilde{x},\tilde{y}))_+^2 \neq 0.$$

Then, for sufficiently large k, there exists $\zeta > 0$ such that

$$\sum_{i=1}^{p}(g_i(x^k))_+^2 + \sum_{j=1}^{q}(G_j(x^k,y^k))_+^2 + (\varphi_{GF}(x^k,y^k))_+^2 > \zeta.$$

Hence

$$J_k(x^k,y^k) > f(x^k,y^k) - f(\bar{x},\bar{y}) + \frac{1}{2}\|(x^k,y^k)-(\bar{x},\bar{y})\|^2 + k\zeta.$$

Taking the limit as $k \to \infty$, we get $J_k(x^k,y^k) \to +\infty$. However, since $(\bar{x},\bar{y})$ is feasible for (5.3), we have $J_k(x^k,y^k) \le J_k(\bar{x},\bar{y}) = 0$, so $\lim_{k\to\infty} J_k(x^k,y^k) \le 0$, a contradiction. Therefore

$$\sum_{i=1}^{p}(g_i(\tilde{x}))_+^2 + \sum_{j=1}^{q}(G_j(\tilde{x},\tilde{y}))_+^2 + (\varphi_{GF}(\tilde{x},\tilde{y}))_+^2 = 0$$

and $(\tilde{x},\tilde{y}) \in \Theta \cap \mathbb{B}^{\varepsilon}(\bar{x},\bar{y})$. Since $J_k(x^k,y^k) \le 0$, we have $f(x^k,y^k) - f(\bar{x},\bar{y}) + \frac{1}{2}\|(x^k,y^k)-(\bar{x},\bar{y})\|^2 \le 0$. Taking the limit as $k \to \infty$, one sees that

$$f(\tilde{x},\tilde{y}) - f(\bar{x},\bar{y}) + \frac{1}{2}\|(\tilde{x},\tilde{y})-(\bar{x},\bar{y})\|^2 \le 0.$$

Then $(\tilde{x},\tilde{y})$ is optimal for (5.1), so $(\bar{x},\bar{y}) = (\tilde{x},\tilde{y})$. Therefore $(x^k,y^k) \to (\bar{x},\bar{y})$ as $k \to \infty$, and condition $(\mathscr{A}_0)$ is satisfied. Since (x^k,y^k) is optimal for (5.3), we have $0 \in \partial J_k(x^k,y^k)$. Hence

$$\begin{aligned}
0 \in\, & \partial f(x^k,y^k) + (x^k,y^k) - (\bar{x},\bar{y}) + \sum_{i=1}^{p} 2k(g_i(x^k))_+ \partial g_i(x^k) \times \{0\} \\
& + \sum_{j=1}^{q} 2k(G_j(x^k,y^k))_+ \partial G_j(x^k,y^k) + 2k(\varphi_{GF}(x^k,y^k))_+ \partial \varphi_{GF}(x^k,y^k).
\end{aligned}$$

Using this inclusion and Theorem 4.1, there exist $z_r^k \in Z(x^k,y^k) \subset Y(x^k)$ and multipliers $(\lambda_r^k, \gamma_r^k) \in \mathbb{R}_+^{q+1}$, $r = 1,\ldots,n+m+1$, with $\sum_{r=1}^{n+m+1} \gamma_r^k = 1$ and

$$\begin{aligned}
0 \in\ & \partial f(x^k,y^k) + \big((x^k,y^k) - (\bar{x},\bar{y})\big) + \sum_{i=1}^{p} 2k\,(g_i(x^k))_+\, \partial g_i(x^k) \times \{0\} \\
& + \sum_{j=1}^{q} 2k\,(G_j(x^k,y^k))_+\, \partial G_j(x^k,y^k) \\
& + 2k\,(\varphi_{GF}(x^k,y^k))_+ \sum_{r=1}^{n+m+1} \gamma_r^k \begin{pmatrix} \partial_x F(x^k,y^k)^\top (y^k - z_r^k) - \partial_x G(x^k,z_r^k)^\top \lambda_r^k \\ \partial_y F(x^k,y^k)^\top (y^k - z_r^k) + F(x^k,y^k), \end{pmatrix} \qquad (5.4)
\end{aligned}$$

$$0 \in F(x^k,y^k) + \partial_y G(x^k,z_r^k)^\top \lambda_r^k, \qquad r = 1,\ldots,n+m+1, \qquad (5.5)$$

$$G(x^k,z_r^k)^\top \lambda_r^k = 0, \qquad r = 1,\ldots,n+m+1. \qquad (5.6)$$

By local compactness of $\operatorname{gph} Y$ and continuity of $(x,y,z) \mapsto \langle F(x,y), y-z \rangle$, Berge's maximum theorem implies that the argmax mapping $Z(\cdot,\cdot)$ is o.s.c at $(\bar{x},\bar{y})$. Hence, up to a subsequence (chosen commonly for all r), we have $z_r^k \to z_r \in Z(\bar{x},\bar{y})$ for $r = 1,\ldots,n+m+1$. Let

$$\begin{cases} \alpha_i^k = 2k(g_i(x^k))_+, & i = 1,\ldots,p, \\ \beta_j^k = 2k(G_j(x^k,y^k))_+, & j = 1,\ldots,q, \\ \mu^k = 2k(\varphi_{GF}(x^k,y^k))_+. \end{cases} \qquad (5.7)$$

Using inclusions (5.4)-(5.5) and the notations in (5.7), there exist appropriate subgradient elements such that

$$\begin{aligned} & m_1(x^k, y^k, z_1^k, \dots, z_{n+m+1}^k, \mu^k, \alpha^k, \beta^k, \lambda_1^k, \dots, \lambda_{n+m+1}^k) \leq \|\bar{x} - x^k\|, \\ & m_2(x^k, y^k, z_1^k, \dots, z_{n+m+1}^k, \mu^k, \beta^k) \leq \|\bar{y} - y^k\|, \\ & m_3(x^k, y^k, z_r^k, \lambda_r^k) = 0, \quad r = 1, \dots, n+m+1. \end{aligned} \tag{5.8}$$

Taking the limit in (5.8) as $k \to \infty$, one sees that $(\mathscr{A}_1)$-$(\mathscr{A}_3)$ are satisfied. For, $i = 1, \dots, p$, $g_i(\bar{x}) < 0$, we have $g_i(x^k) < 0$ for sufficiently large k. Hence $\alpha_i^k = 0$ for sufficiently large k, so condition $(\mathscr{A}_4)$ is satisfied. For, $j = 1, \dots, q$, $G_j(\bar{x}, \bar{y}) < 0$, we have $G_j(x^k, y^k) < 0$ for sufficiently large k. Hence $\beta_j^k = 0$ for sufficiently large k, so condition $(\mathscr{A}_5)$ is satisfied. For, $j = 1, \dots, q$ and $r = 1, \dots, n+m+1$, $G_j(\bar{x}, z_r) < 0$, we have $G_j(x^k, z_r^k) < 0$ for sufficiently large k. Hence $\lambda_{r,j}^k = 0$ for sufficiently large k, so condition $(\mathscr{A}_6)$ is satisfied. Consequently, $(\bar{x}, \bar{y})$ is an AKKT point of problem (1.1). ☐

In summary, Theorem 5.1 guarantees that any local optimal solution of (1.1) can be characterized by AKKT sequences, providing a foundational tool for both theoretical analysis and numerical algorithms in nonsmooth generalized bilevel optimization.

Example 5.1. Consider the following generalized bilevel optimization problem

$$\begin{cases} \min\limits_{x,y} f(x,y) = |x| + |y| \\ \text{subject to } g(x) = |x| - 1 \leq 0 \\ \qquad\qquad y \in S(x), \end{cases}$$

where, for each $x \in \mathbb{R}$, $S(x)$ represents the solution set of the variational inequality

$$\text{find } y \in Y(x) \text{ such that } \langle F(x,y), y - z \rangle \leq 0 \;\forall z \in Y(x),$$

with $F(x,y) = \left(x^2 - 1\right)|y|$ and $Y(x) = \left\{y \in \mathbb{R} \,:\, G(x,y) = y^2 - 1 \leq 0\right\} = [-1,1]$. One can see that the solution set of the above variational inequality is given by

$$S(x) = \begin{cases} [-1,1] & \text{if } x \in \{-1,1\} \\ \{0,1\} & \text{if } x \in \,]-1,1[\\ \{-1,0\} & \text{otherwise.} \end{cases}$$

Hence, the point $(\bar{x}, \bar{y}) = (0,0)$ is a local optimal solution of the considered problem. Moreover, for all $(x,y) \in \mathbb{R} \times \mathbb{R}$, we have

$$Z(x,y) = \begin{cases} [-1,1] & \text{if } x \in \{-1,1\} \text{ or } y = 0 \\ \{1\} & \text{if } x \in \,]-1,1[\text{ and } y \neq 0 \\ \{-1\} & \text{otherwise.} \end{cases}$$

Fix $\varepsilon > 0$. It follows that $\operatorname{gph} Y \cap (\mathbb{B}^{\varepsilon}(\bar{x}) \times \mathbb{R}) = (\mathbb{R} \times [-1,1]) \cap (\mathbb{B}^{\varepsilon}(\bar{x}) \times \mathbb{R}) = \mathbb{B}^{\varepsilon}(\bar{x}) \times [-1,1]$ is compact. Furthermore, for any (x,z) with $x \in \mathbb{B}^{\varepsilon}(\bar{x})$ and $z \in Y(x)$, it is clear to see that the only solution $\lambda \in \mathbb{R}$ of the following system

$$2\lambda z = 0, \quad \lambda \geq 0, \quad \left(z^2 - 1\right)\lambda = 0,$$

is $\lambda = 0$. Therefore, $\mathscr{M}_x(z) = \{0\}$ for all (x,z) with $x \in \mathbb{B}^{\varepsilon}(\bar{x})$ and $z \in Y(x)$.

Consequently, all the assumptions of Theorem 5.1 are satisfied. The local optimal solution $(\bar{x},\bar{y})$ can be characterized as an AKKT point. Specifically, consider the sequences defined for an integer $k>0$ by: $\left(x^k,y^k\right)=\left(0,\frac{1}{k}\right)$, $\mu^k=k$, $\alpha^k=\beta^k=0$, $z_1^k=z_2^k=z_3^k=\frac{1}{k}$, $\lambda_1^k=\lambda_2^k=\lambda_3^k=0$, $\gamma_1^k=\gamma_2^k=\frac{1}{2}$, and $\gamma_3^k=0$. With these sequences, all the conditions $(\mathscr{A}_0)-(\mathscr{A}_6)$ of the AKKT definition are satisfied, confirming that $(\bar{x},\bar{y})$ is indeed an AKKT point of the considered problem.

Remark 5.1. Our AKKT definition does not require the approximating points z_r^k to be feasible (i.e., $z_r^k\in Y(x_k)$), nor to be exact maximizers in $Z(x_k,y_k)$ at each k. This choice is deliberate and accommodates algorithmic iterates that may be only approximately feasible or maximizing at the z-level. In contrast, the existence proof in Theorem 5.1 constructs AKKT sequences via the gap subdifferential formula, which naturally provides $z_r^k\in Z(x_k,y_k)\subseteq Y(x_k)$. Both viewpoints are consistent: the theoretical construction yields a stronger property, while the definition remains broad enough for numerical purposes. If desired, one may also consider an intermediate variant enforcing asymptotic feasibility, e.g., $\operatorname{dist}(z_r^k,Y(x_k))\to 0$, or approximate maximization, e.g., $\varphi_{GF}(x_k,y_k)-\langle F(x_k,y_k),y_k-z_r^k\rangle\to 0$.

6. RELATIONSHIPS BETWEEN KKT AND AKKT CONDITIONS

To begin the study of the relationships between KKT and AKKT conditions, we first observe that every KKT point of problem (1.1) is automatically an AKKT point of the same problem. Indeed, suppose that $(\bar{x},\bar{y})$ is a KKT point of problem (1.1). Then, by definition, there exist a scalar $\mu>0$, multipliers $(\alpha,\beta)\in\mathbb{R}_+^{p+q}$, points $z_r\in Z(\bar{x},\bar{y})$, and multipliers $(\lambda_r,\gamma_r)\in\mathbb{R}_+^{q+1}$ for $r=1,\ldots,n+m+1$, satisfying $\sum_{j=1}^{n+m+1}\gamma_j=1$, such that the KKT conditions $(\mathscr{K}_1)-(\mathscr{K}_6)$ hold. By setting $x^k=\bar{x}$, $y^k=\bar{y}$, $\mu^k=\mu$, $\alpha^k=\alpha$, $\beta^k=\beta$, $z_r^k=z_r$, $\lambda_r^k=\lambda_r$, and $\gamma_r^k=\gamma_r$, $r=1,\ldots,n+m+1$, it follows that $(\bar{x},\bar{y})$ satisfies all AKKT conditions. Therefore, every KKT point is indeed an AKKT point. However, the converse does not hold in general, as illustrated by the following example.

Example 6.1. Consider the generalized bilevel optimization problem

$$\begin{cases}\min\limits_{x,y} f(x,y)=x\\ \text{subject to } x\in X=\{x\in\mathbb{R}\ :\ g(x)=-x\le 0\}=\mathbb{R}_+\\ \qquad\qquad y\in S(x),\end{cases}$$

where, for each $x\in\mathbb{R}$, $S(x)$ represents the solution set of the variational inequality

$$\text{find } y\in Y(x)\ \text{ such that } \langle F(x,y),y-z\rangle\le 0\ \forall z\in Y(x),$$

with $F(x,y)=-\left(x^2+x+1\right)$ and

$$Y(x)=\left\{y\in\mathbb{R}\ :\ G(x,y)=(y+1)^3\le 0\right\}=]-\infty,-1].$$

One can see that, for $x\in\mathbb{R}$, the solution set of the above variational inequality is $S(x)=\{-1\}$. Moreover, for $(x,y)\in X\times\mathbb{R}$, the associated gap function is given by $\varphi_{GF}(x,y)=-\left(x^2+x+1\right)(y+1)$. It is easy to check that $(\bar{x},\bar{y})=(0,-1)$ is a local optimal solution of the considered problem. And $Z(\bar{x},\bar{y})=\{-1\}$. One can verify also $(\bar{x},\bar{y})$ is also an AKKT point with the sequences defined for an integer $k>0$ by: $\left(x^k,y^k\right)=\left(\frac{1}{k},\frac{1}{k}-1\right)$, $\mu^k=\frac{1}{k}$, $\alpha^k=1$, $\beta^k=\frac{2}{3}k$, $z_1^k=\frac{1}{k^2}-1$, $z_2^k=\frac{1}{k}-1$, $z_3^k=\frac{1}{k}-1$, $\lambda_1^k=\frac{1}{3}k^4$, $\lambda_2^k=\lambda_3^k=\frac{1}{3}k^2$, $\gamma_1^k=0$, $\gamma_2^k=1$, and $\gamma_3^k=0$.

Now, suppose that $(\bar{x},\bar{y})$ is a KKT point. Then, there exist $\mu>0$, $(\alpha,\beta)\in\mathbb{R}_+^2$, $z_r\in Z(\bar{x},\bar{y})$, $(\lambda_r,\gamma_r)\in\mathbb{R}_+^{1+1}$, $r=1,\ldots,3$ such that $\sum_{r=1}^3\gamma_r=1$ and the conditions $(\mathscr{K}_1)-(\mathscr{K}_6)$ are satisfied. Hence, $z_r=-1$ for all $r=1,2,3$. Since $F(\bar{x},\bar{y})=-1$ and $\partial_y G(\bar{x},z_1)=\{0\}$, condition $(\mathscr{K}_3)$ would require $0\in\{-1\}$, a contradiction. Therefore, $(\bar{x},\bar{y})$ cannot be a KKT point. This example shows that not all AKKT points are KKT points.

To formalize conditions under which an AKKT point is also a KKT point, we introduce the following definitions.

Definition 6.1. Let $(\bar{x},\bar{y})\in\mathbb{R}^n\times\mathbb{R}^m$. We say that the upper-level regularity holds at $\bar{x}$ if

$$\left.\begin{array}{r} 0\in\displaystyle\sum_{i=1}^{p}\alpha_i\partial g_i(\bar{x})\\ \alpha_i\geq 0,\ i=1,\ldots,p\\ \alpha_i g_i(\bar{x})=0,\ i=1,\ldots,p \end{array}\right\}\Longrightarrow\alpha_i=0,\ i=1,\ldots,p. \tag{ULR}$$

Similarly, we say that the variational inequality regularity holds at $(\bar{x},\bar{y})$ if

$$\left.\begin{array}{r} 0\in\displaystyle\sum_{j=1}^{q}\beta_j\partial_y G_j(\bar{x},\bar{y})\\ \beta_j\geq 0,\ j=1,\ldots,q\\ \beta_j G_j(\bar{x},\bar{y})=0,\ j=1,\ldots,q \end{array}\right\}\Longrightarrow\beta_j=0,\ j=1,\ldots,q. \tag{VIR}$$

The following theorem provides sufficient conditions for an AKKT point of problem (1.1) to also be a KKT point of problem (1.1).

Theorem 6.1. *Let $(\bar{x},\bar{y})\in\Theta$ be an AKKT point for (1.1), and assume that the sequence $\left\{\mu^k\right\}$ is bounded and bounded below by a strictly positive constant. Suppose moreover that $M_{\bar{x}}(Z(\bar{x},\bar{y}))=\{0\}$, the (ULR) holds at $\bar{x}$, and the (VIR) holds at $(\bar{x},\bar{y})$. Then $(\bar{x},\bar{y})$ is a KKT point for (1.1).*

Proof. Fix an integer k. By Lemma 2.1, all involved subdifferentials, including $\partial f\left(x^k,y^k\right)$, $\partial g_i\left(x^k\right)$, $i=1,\ldots,p$, $\partial G_j\left(x^k,y^k\right)$, $j=1,\ldots,q$, $\partial F_s\left(x^k,y^k\right)$, $s=1,\ldots,m$, and $\partial G_j\left(x^k,z_r^k\right)$, $j=1,\ldots,q$, $r=1,\ldots,n+m+1$, are compact. Hence, since the norms below depend continuously on the chosen subgradients and the Cartesian products of these compact sets are compact, the infima in the definitions of m_1, m_2, m_3 are attained. Therefore, there exist $\left(\eta_1^k,\eta_2^k\right)\in\partial f\left(x^k,y^k\right)$, $\rho_i^k\in\partial g_i\left(x^k\right)$, $i=1,\ldots,p$, $\left(\vartheta_{1,j}^k,\vartheta_{2,j}^k\right)\in\partial G_j\left(x^k,y^k\right)$, $j=1,\ldots,q$, $\left(\nu_{1,s}^k,\nu_{2,s}^k\right)\in\partial F_s\left(x^k,y^k\right)$,

$s=1,\ldots,m$, and $\left(\delta_{1,j}^{r,k},\delta_{2,j}^{r,k}\right)\in\partial G_j\left(x^k,z_r^k\right)$, $j=1,\ldots,q$, $r=1,\ldots,n+m+1$, such that

$$\begin{aligned} m_1^k &= m_1\left(x^k,y^k,z_1^k,\ldots,z_{n+m+1}^k,\mu^k,\alpha^k,\beta^k,\lambda_1^k,\ldots,\lambda_{n+m+1}^k\right) \\ &= \left\|\eta_1^k+\sum_{i=1}^{p}\alpha_i^k\rho_i^k+\sum_{j=1}^{q}\beta_j^k\vartheta_{1,j}^k+\mu^k\sum_{r=1}^{n+m+1}\gamma_r^k\left[\sum_{s=1}^{m}\left(y_s^k-z_{r,s}^k\right)\nu_{1,s}^k-\sum_{j=1}^{q}\lambda_{r,j}^k\delta_{1,j}^{r,k}\right]\right\|, \end{aligned} \tag{6.1}$$

$$\begin{aligned} m_2^k &= m_2\left(x^k,y^k,z_1^k,\ldots,z_{n+m+1}^k,\mu^k,\beta^k\right) \\ &= \left\|\eta_2^k+\sum_{j=1}^{q}\beta_j^k\vartheta_{2,j}^k+\mu^k\left[F\left(x^k,y^k\right)+\sum_{r=1}^{n+m+1}\gamma_r^k\sum_{s=1}^{m}\left(y_s^k-z_{r,s}^k\right)\nu_{2,s}^k\right]\right\|, \end{aligned} \tag{6.2}$$

$$m_3^k=m_3\left(x^k,y^k,z_r^k,\lambda_r^k\right)=\left\|F\left(x^k,y^k\right)+\sum_{j=1}^{q}\lambda_{r,j}^k\delta_{2,j}^{r,k}\right\|. \tag{6.3}$$

Step 1. We show that $\{\varsigma^k\}$ is bounded, where $\varsigma^k:=(\alpha^k,\beta^k,\lambda_1^k,\ldots,\lambda_{n+m+1}^k)$. Suppose by contradiction that $\|\varsigma^k\|\to+\infty$. Since $\|\frac{\varsigma^k}{\|\varsigma^k\|}\|=1$, there exists a subsequence, without reindexing, converging to a vector $\overline{\varsigma}=(\overline{\alpha},\overline{\beta},\overline{\lambda}_1,\ldots,\overline{\lambda}_{n+m+1})$ with $\|\overline{\varsigma}\|=1$.

From condition $(\mathscr{A}_3)$ and (6.3), dividing by $\|\varsigma^k\|$ and using the boundedness of $F(x^k,y^k)$, we get, for each r, $\sum_{j=1}^{q}\overline{\lambda}_{r,j}\,\delta_{2,j}^{r}=0$, with $\delta_{2,j}^{r}\in\partial_y G_j(\bar{x},\bar{z}_r)$, and by $(\mathscr{A}_6)$ we also have the complementarity $\overline{\lambda}_{r,j}\,G_j(\bar{x},\bar{z}_r)=0$. Since $\bar{z}_r\in Z(\bar{x},\bar{y})$ and $M_{\bar{x}}(Z(\bar{x},\bar{y}))=\{0\}$, it follows that $\overline{\lambda}_{r,j}=0$ for all $j=1,\ldots,q$ and $r=1,\ldots,n+m+1$.

From $(\mathscr{A}_2)$ and (6.2), dividing by $\|\varsigma^k\|$ and using that $\mu^k/\|\varsigma^k\|\to0$ (since $\{\mu^k\}$ is bounded) and that η_2^k is bounded (compactness), we obtain

$$\sum_{j=1}^{q}\overline{\beta}_j\,\vartheta_{2,j}=0,\quad\text{with}\;\;\vartheta_{2,j}\in\partial_y G_j(\bar{x},\bar{y}).$$

By $(\mathscr{A}_5)$, $\overline{\beta}_j\,G_j(\bar{x},\bar{y})=0$; since (VIR) holds at $(\bar{x},\bar{y})$, we conclude $\overline{\beta}_j=0$ for all $j=1,\ldots,q$.

From $(\mathscr{A}_1)$ and (6.1), again dividing by $\|\varsigma^k\|$ and using $\mu^k/\|\varsigma^k\|\to0$ and boundedness of η_1^k, we obtain

$$\sum_{i=1}^{p}\overline{\alpha}_i\,\rho_i=0,\quad\text{with}\;\;\rho_i\in\partial g_i(\bar{x}).$$

By $(\mathscr{A}_4)$, $\overline{\alpha}_i\,g_i(\bar{x})=0$; since (ULR) holds at $\bar{x}$, we conclude that $\overline{\alpha}_i=0$ for all $i=1,\ldots,p$. Therefore $\|\overline{\varsigma}\|=0$, a contradiction. Hence $\{\varsigma^k\}$ is bounded.

Step 2. Since $\{\varsigma^k\}$ is bounded and $\{\mu^k\}$ is bounded and bounded below by a positive constant, one sees that there exist subsequences, without reindexing such that

$$(\alpha^k,\beta^k,\lambda_1^k,\ldots,\lambda_{n+m+1}^k)\to(\alpha,\beta,\lambda_1,\ldots,\lambda_{n+m+1}),\quad\mu^k\to\mu>0,$$

and, by compactness, we may also extract convergent subsequences of the selected subgradients $\eta_\ell^k,\rho_i^k,\vartheta_{\ell,j}^k,\nu_{\ell,s}^k,\delta_{\ell,j}^{r,k}$ and of the convex coefficients γ_r^k (simplex compactness: $\sum_r\gamma_r^k=1$, $\gamma_r^k\geq0$). Passing to the limit in (6.1)–(6.3) and invoking $(\mathscr{A}_4)$–$(\mathscr{A}_6)$ yields the KKT conditions $(\mathscr{K}_1)$–$(\mathscr{K}_6)$. Therefore, $(\bar{x},\bar{y})$ is a KKT point for (1.1). $\square$

Remark 6.1. The assumption in Theorem 6.1 that the sequence of penalty multipliers $\{\mu^k\}$ is bounded and bounded away from zero is *not* automatic under Definition 5.1. For instance, in the construction of AKKT sequences in Theorem 5.1, the choice $\mu^k = 2k\left(\varphi_{GF}\left(x^k, y^k\right)\right)_+$ may diverge if the gap function φ_{GF} decays slowly.

This boundedness can be ensured under additional conditions, such as:

- *Partial calmness at* $(\bar{x}, \bar{y})$: if φ_{GF} is locally Lipschitz at a local solution $(\bar{x}, \bar{y})$ and the problem is partially calm there (cf. Theorem 3.1), then there exists $\mu > 0$ such that $(\bar{x}, \bar{y})$ is a local minimizer of the partially penalized problem

  $$\min\{\, f(x,y) + \mu\, \varphi_{GF}(x,y) \;:\; (x,y) \in \Omega \,\}$$

  in a neighborhood of $(\bar{x}, \bar{y})$. In that case, one can select an AKKT sequence with a *fixed* penalty $\mu^k \equiv \mu$, which meets the hypothesis of Theorem 6.1. If $(\bar{x}, \bar{y})$ is not known to be a local solution, partial calmness may be postulated at that point to reach the same conclusion.
- A local *error bound* for the gap function near $(\bar{x}, \bar{y})$ that provides an *upper control* of $\varphi_{GF}\left(x^k, y^k\right)$ along the AKKT sequence when μ^k is chosen as a function of $\varphi_{GF}\left(x^k, y^k\right)$ (e.g., $\mu^k = 2k\left(\varphi_{GF}\left(x^k, y^k\right)\right)_+$). In practice, one may also enforce a uniform positive lower bound on μ^k (e.g., $\mu^k \in [\mu_-, \mu_+]$ with $\mu_- > 0$) to satisfy the hypothesis of Theorem 6.1.

These conditions ensure that the AKKT $\Rightarrow$ KKT implication remains valid and practically verifiable in Theorem 6.1.

7. CONCLUSION

We developed necessary optimality conditions for generalized bilevel optimization problems with variational-inequality lower levels in a nonsmooth setting. Our contributions are threefold: (i) KKT-type conditions based on the basic subdifferential, which extends classical smooth results without requiring differentiability or interiority; (ii) an approximate KKT (AKKT) framework that relaxes standard stationarity and accommodates nonsmoothness; and (iii) verifiable regularity assumptions under which AKKT points coincide with KKT points. We also established that every KKT point is an AKKT point, while the converse may fail in general, as shown by a counterexample. As a consistency check, our KKT system reduces to the classical conditions of [3] under smoothness.

These results highlight AKKT conditions as both a theoretical relaxation and a practical tool for algorithm design when standard constraint qualifications fail. Future research includes developing efficient algorithms, such as augmented Lagrangian, interior-point, or SQP-type methods, using AKKT residuals as stopping criteria, and investigating weaker or alternative regularity conditions that still guarantee AKKT $\Leftrightarrow$ KKT. Another promising direction is to establish error bounds and convergence rates for algorithms exploiting the gap-function and partial calmness framework in large-scale bilevel models. Finally, we note that our gap-based scalarization and partial calmness results align with classical scalarization principles in vector and set optimization, which suggests further developments for multiobjective upper levels.

REFERENCES

[1] B.S. Mordukhovich, J.V. Outrata, Coderivative analysis of quasi-variational inequalities with applications to stability and optimization, SIAM J. Optim. 18(2) (2007), 389-412.

[2] R. El Idrissi, L. Lafhim, Y. Ouakrim, Pontryagin optimality conditions for generalized bilevel optimal control problems with pure state inequality constraints, J. Appl. Numer. Optim. 6 (2024), 229-248.

[3] J.J. Ye, D.L. Zhu, Q.J. Zhu, Exact penalization and necessary optimality conditions for generalized bilevel programming problems, SIAM J. Optim. 7 (1997), 481-507.

[4] R. Andreani, G. Haeser, J.M. Martínez, On sequential optimality conditions for smooth constrained optimization, Optimization, 60 (2011), 627-641.

[5] R. Andreani, G. Haeser, L.D. Secchin, P.J. Silva, New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences, SIAM J. Optim. 29(4) (2019), 3201-3230.

[6] R. El Idrissi, E.M. Kalmoun, L. Lafhim, New approximate optimality conditions of strong type in multiobjective generalized Nash equilibrium problems, Appl. Set-Valued Anal. Optim. 8 (2026), 85-101.

[7] E.G. Birgin, J.M. Martínez, Practical augmented Lagrangian methods for constrained optimization, Society for Industrial and Applied Mathematics, 2014.

[8] L. Chen, D. Goldfarb, Interior-point ℓ_2-penalty methods for nonlinear programming with strong global convergence properties, Math. Program. 108 (2006), 1-36.

[9] L. Qi, Z. Wei, On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. 10 (2000), 963-981.

[10] R. El Idrissi, E.M. Kalmoun, L. Lafhim, Approximate Karush-Kuhn-Tucker conditions for nonsmooth bilevel optimization problems, J. Ind. Manag. Optim. 21(2) (2025) 988-1006.

[11] N. Zhang, Z. Zhang, Manufacturer or retailer? the optimal remanufacturing channel choice under carbon tax policy, J. Ind. Manag. Optim. 21 (2025), 3362-3392.

[12] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation. I. Basic Theory, II. Applications, 2006.

[13] D.W. Hearn, The gap function of a convex program, Oper. Res. Lett. 1 (1982), 67-71.

[14] J.J. Ye, D.L. Zhu, Optimality conditions for bilevel programming problems, Optimization, 33(1) (1995), 9-27.

[15] J.J. Ye, D.L. Zhu, A note on optimality conditions for bilevel programming problems, Optimization, 39 (1997), 361-366.

[16] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley-Interscience, New York, 1983.

[17] R.T. Rockafellar, R.J.B. Wets, Variational Analysis, Springer, 2009.