Control and Cybernetics
vol. **54** (2025) No. 2
pages: 123-155
DOI: 10.2478/candc-2025-0005

Optimality conditions for bilevel optimization with variational inequality constraints using approximations*

by

Zakarya Dardour†, Rachid El Idrissi†, El Mostafa Kalmoun ‡ and Lahoussine Lafhim†

†Laboratoire L2MASI, Department of Mathematics, Sidi Mohammed Ben Abdellah University, Morocco
zakarya.dardour; rachid.elidrissi1; lahoussine.lafhim@usmba.ac.ma
‡(Corresponding Author) School of Science and Engineering, Al Akhawayn University in Ifrane, PO Box 104, Ifrane 53000, Morocco
E.Kalmoun@aui.ma

Abstract: We use first- and second-order approximations to establish necessary and sufficient optimality conditions for nonsmooth generalized bilevel optimization problems with variational inequality constraints. To transform the hierarchical problem into a single-level optimization problem, we employ two approaches: the gap function reformulation and the KKT reformulation. We compare these approaches and provide examples to illustrate the applicability of the results.

Keywords: bilevel optimization; nonsmooth optimization; variational inequality constraints; first- and second-order approximations; optimality conditions

1. Introduction

An optimization problem with variational inequality constraints is a special class of optimization problems, where some or all constraints are defined by a parametric variational inequality, with y as the primary variable and x as the parameter. In this paper, we focus on the following generalized bilevel optimization problem with variational inequality constraints (GBOP):

$$\min_{x,y} \; f(x,y) \;\text{ subject to }\; g(x) \leq 0, \text{ and } y \in S(x), \tag{GBOP}$$

*Submitted: May 2025; Accepted: September 2025

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the leader objective function, $g : \mathbb{R}^n \to \mathbb{R}^q$ is the constraint function of the upper-level problem, and $S(x)$ is the set of all solutions of the optimization problem with variational inequality constraints:

$$\langle F(x,y), y - \xi \rangle \leq 0, \quad \forall \xi \in Y(x), \tag{VI[x]}$$

with $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ and $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ being the follower objective function and the feasible set-valued map, respectively:

$$Y(x) := \{y \in \mathbb{R}^m : G(x,y) \leq 0\},$$

and $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^s$ being the constraint function of the lower-level problem. Hence:

$$S(x) = \{y \in Y(x) : \langle F(x,y), y - \xi \rangle \leq 0, \quad \forall \xi \in Y(x)\}.$$

Let $\mathcal{F}$ denote the set of feasible points for the problem (GBOP), given by:

$$\mathcal{F} = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : g(x) \leq 0 \text{ and } y \in S(x)\}.$$

Variational inequalities, introduced by Stampacchia (1964), are widely used in economics, transportation, network and structural analysis, elasticity, engineering, mechanics, supply chain management, finance, and game theory. Their complexity has spurred the development of quite some numerical methods for various classes of variational inequalities (see, e.g., Huang, Lin and Xiu, 2014; Outrata, 1994; Wan and Chen, 2013; Ye, 2000; Ye and Ye, 1997, and references therein).

Von Stackelberg's influential work (Stackelberg, 1952) is considered as the origin of (GBOP). Then, Bracken and McGill (1973) developed the first mathematical bilevel model in 1973. Many real-world problems in engineering, medicine, and economics can be formulated as (GBOP). The investigation of (GBOP) offers numerous benefits from both theoretical and practical perspectives (see Ben-Ayed, Boyce and Blair, 1988).

Variational inequalities are valuable tools in modeling equilibrium conditions and optimality criteria in complex systems. These inequalities have also been used in terms of the lower-level problem in the context of bilevel optimal control problems, see El Idrissi, Lafhim and Ouakrim (2024) and El Idrissi et al. (2024). This connection shows the practical need to develop optimality conditions for bilevel problems with variational inequality constraints. Our work extends the theoretical framework, so as to include nonsmooth settings and provides conditions independent of classical constraint qualifications. This is crucial for real-world problems, where smoothness may not hold, thus broadening the applicability of bilevel optimization.

Generalized bilevel optimization problems with variational inequality constraints (GBOP) are often reformulated as standard single-level nonlinear programming problems (NLP). However, standard constraint qualifications typically do not hold for the resulting NLP reformulation (Ye, Zhu and Zhu, 1997, Proposition 1.1). A popular approach is the Karush-Kuhn-Tucker (KKT) reformulation by Chen and Florian (1995), which replaces the lower-level problem with its KKT necessary optimality conditions. This requires the lower-level problem's convexity and a standard constraint qualification, such as Slater's condition, see Dempe and Zemkoho (2013, Theorem 3.3) or the Mangasarian-Fromovitz Constraint Qualification (MFCQ), Zemkoho and Zhou (2021, Theorem 2.2), which ensures that the KKT conditions are valid by requiring positive linear independence of constraint gradients. Another is Auslender's gap function approach (Auslender, 1976). This method replaces the lower-level problem with a marginal function, known as the gap function. The challenge with this approach lies in the generally nonsmooth nature of the gap function. In our analysis, we will use these two reformulations of the bilevel problem as single-level problems and highlight the advantages of each approach.

To address the challenges of non-smoothness and the limitations of classical constraint qualifications, we use the concept of approximations of mappings. Introduced by Thibault (1980), approximations provide a way to generalize derivatives, even for discontinuous, nonconvex, or unbounded mappings. Sweetser (1977) considered approximations using subsets of continuous linear maps between Banach spaces, later revised by Jourani and Thibault (1993). Allali and Amahroq (1997) advanced the theory of first- and second-order approximations for continuous functions by introducing the concept of strict prederivatives in order to address the cases, where classical derivatives or standard generalized derivatives may not exist. Their work provides tools for deriving optimality conditions in nonsmooth settings. These methods have proven to be effective in various optimization problems, including those with set-valued mapping constraints, Amahroq and Gadhi (2001, 2003), nonsmooth vector optimization, Khanh and Tuan (2006), vector equilibrium problems, Khanh and Tung (2013), and fractional multiobjective bilevel problems, Lafhim, (2024). We build on this foundation by using first- and second-order approximations as generalized derivatives.

Numerous studies have investigated hierarchical problems, which can be applied to (GBOP), establishing first-order (Ye, 1999, 2000; Ye and Ye, 1997; Zhang, 2014) and second-order (Dardour, Lafhim and Kalmoun, 2024; Dardour et al., 2025; Dempe and Gadhi, 2010; Dempe and Zemkoho, 2013) optimality conditions. However, these conditions often rely on restrictive assumptions, such as twice-differentiable data, convexity, and the satisfaction of classical constraint qualifications. This limits their applicability to a broader class of problems.

In our work, we step outside this framework by deriving optimality conditions for (GBOP) without imposing any regularity assumptions or constraint qualifications, using, instead, approximation tools. Furthermore, to the best of our knowledge, there are no existing contributions that address second-order optimality conditions for (GBOP). Our work is therefore the first to establish such results under assumptions much weaker than the classical ones in the literature.

This paper makes two main contributions. First, we derive first- and second-order optimality conditions for the generalized bilevel optimization problem with variational inequality constraints (GBOP) without relying on constraint qualifications or smoothness assumptions. We achieve this by employing approximation methods based on derivatives of functions, which enable us to address nonsmooth and nonconvex settings more effectively. Second, we apply our approach to two distinct reformulations of the bilevel problem (GBOP) as single-level problems: the gap function reformulation and the KKT reformulation. For each of these two reformulations, we establish necessary and sufficient optimality conditions and provide a detailed comparison of their advantages and limitations. We also present illustrative examples that demonstrate the applicability of our results by showing how our conditions can be verified in specific cases.

The remainder of this paper is organized as follows. Section 2 presents key definitions and results, concerning first- and second-order approximations, and notational conventions. Section 3 uses the gap function reformulation to derive first- and second-order necessary and sufficient optimality conditions. Section 4 focuses on the KKT reformulation, through which we develop optimality conditions analogous to those in Section 3.

2. Preliminaries

In this section, we establish the mathematical foundation necessary for analyzing optimization problems using approximation theory. We examine first-order and second-order approximations, which provide powerful tools for analyzing the local behavior of functions at points where traditional derivatives may not exist or are difficult to compute. These approximation techniques extend classical differential calculus to non-smooth settings, which allows for the study of optimization problems, involving a broader class of objective functions and constraints.

Let $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^m$, and $Z \subseteq \mathbb{R}^p$ denote given sets. We define:

- $L(X,Y)$: the space of continuous linear mappings from X to Y,
- $B(X,Y)$: the space of continuous bilinear mappings from $X \times X$ into Y,
- $\mathbb{B}_{\mathbb{R}^m}$: the closed unit ball in $\mathbb{R}^m$, centered at the origin,
- $\overline{\mathbb{R}} = [-\infty, +\infty]$: the extended real line.

For any subset $A \subseteq L(X,Y)$ and $x \in X$, we define $A(x) := \{\alpha x : \alpha \in A\}$.

2.1. Background materials

To establish our framework, we introduce key geometric concepts involving cones and directional structures. Let Ω be a nonempty subset of $\mathbb{R}^n$ and $\bar{z} \in \Omega$:

- The cone generated by Ω is: $\text{cone}(\Omega) = \{\lambda z : \lambda \geq 0, z \in \Omega\}$.
- The convex hull of the set Ω is:
$\text{co}(\Omega) = \left\{ \sum_{i=1}^{k} \lambda_i x_i : x_i \in \Omega, \lambda_i \geq 0, \sum_{i=1}^{k} \lambda_i = 1, k \in \mathbb{N} \right\}.$
- cl cone(Ω) and cl co(Ω) denote the closure of cone(Ω) and co(Ω), respectively.
- int(Ω) denotes the interior of the set Ω.
- The contingent cone to Ω at $\bar{z}$ is:

$$\mathcal{T}(\Omega, \bar{z}) = \{d \in \mathbb{R}^n : \exists (t_m) \downarrow 0, (d_m) \to d,\ \bar{z} + t_m d_m \in \Omega\ \forall m \in \mathbb{N}\},$$

  with the set of unit tangent directions defined as:

$$\mathcal{T}_1(\Omega, \bar{z}) = \{d \in \mathcal{T}(\Omega, \bar{z}) : \|d\| = 1\}.$$

- The cone of weak feasible directions to Ω at $\bar{z}$ is:

$$\mathcal{D}(\Omega, \bar{z}) = \{d \in \mathbb{R}^n : \exists (t_m) \downarrow 0,\ \bar{z} + t_m z \in \Omega\ \forall m \in \mathbb{N}\}.$$

The following is based on definitions from Khanh and Tung (2015, Definition 2.1).

PROPOSITION 2.1 *Let either $A \subseteq L(\mathbb{R}^n, \mathbb{R}^m)$ or $A \subseteq B(\mathbb{R}^n, \mathbb{R}^m)$. Then, the following assertions hold:*

1. *Each norm bounded sequence $(M_k)_k \subseteq A$ has a subsequence $(M_{k_s})_s$ such that M_{k_s} converges in norm to $M \in A$.*
2. *For each unbounded norm sequence $(M_k)_k \subseteq A$, the sequence $\dfrac{M_k}{\|M_k\|}$ has a subsequence which converges in norm to some $M \in A \setminus \{0\}$.*

Subsequently, for $A \subseteq L(\mathbb{R}^n, \mathbb{R}^m)$ or $A \subseteq B(\mathbb{R}^n, \mathbb{R}^m)$, we adopt the following notations, used in Khanh and Tuan (2006) and Khanh and Tung (2012). We use the notation $\| \| - \lim$ to denote convergence in norm. Specifically, $M = \| \| - \lim M_k$ means that the sequence (M_k) converges to M in the norm topology.

$$\begin{aligned}
\mathrm{cl}(A) &= \{M \in A : \exists\,(M_k)_k \subseteq A,\ M = \| \| - \lim M_k\}\,,\\
A_\infty &= \left\{M \in A : \exists\,(M_k)_k \subseteq A,\ \exists t_k \to 0^+,\ M = \| \| - \lim t_k M_k\right\},\\
p\text{-}A &= \mathrm{cl}(A) \cup (A_\infty \setminus \{0\})\,.
\end{aligned}$$

Observe that $\mathrm{cl}(A)$ is the norm closure, while A_∞ is the norm recession cone of the given set.

2.2. First-order approximations

With the basic mathematical structures established, we now turn to first-order approximations, which generalize the concept of derivatives to broader settings. These approximations prove especially useful for nonsmooth functions, where classical derivatives may not exist. In this subsection, we present their definition and explore their properties, including applications to locally Lipschitz and continuous functions.

Definition 2.1 *Let $f : X \to Y$ be a mapping, $\bar{x} \in X$, and $A_f(\bar{x}) \subseteq L(X, Y)$. The set $A_f(\bar{x})$ is a first-order approximation of f at $\bar{x}$ if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that:*

$$f(x) - f(\bar{x}) \in A_f(\bar{x})(x - \bar{x}) + \varepsilon\|x - \bar{x}\|\mathbb{B}_{\mathbb{R}^m}$$

for all $x \in \bar{x} + \delta\mathbb{B}_{\mathbb{R}^n}$.

Clarke (1990) defines the subdifferential of f at $\bar{x}$ as:

$$\partial f(\bar{x}) := \mathrm{cl\ co}\ \{\lim Df\,(x_n)\ :\ x_n \in \mathrm{dom}\, Df,\ x_n \to \bar{x}\}\,,$$

where $Df(x)$ is the first-order Fréchet derivative of f at x. Using this definition, the following proposition provides a first-order approximation for locally Lipschitz functions.

Proposition 2.2 (Allali and Amahroq, 1997, Proposition 2.1.2) *If $f : X \to \mathbb{R}$ is locally Lipschitz at $\bar{x}$, then the Clarke subdifferential $\partial f(\bar{x})$ of f at $\bar{x}$ is a first-order approximation of f at $\bar{x}$.*

Allali and Amahroq (1997) demonstrated that the first-order approximation $A_f(\bar{x})$ is a singleton if and only if f is Fréchet differentiable at $\bar{x}$.

To give an example of a non-locally Lipschitz function, we recall the following definition.

Definition 2.2 (Jeyakumar and Luc, 1998) *Let $f : X \to Y$ be a continuous mapping and $\bar{x} \in X$. A closed subset $\bar{\partial} f(\bar{x}) \subseteq L(X,Y)$ is called an approximate Jacobian of f at $\bar{x}$ if, for each $u \in X$ and $v \in Y$ we have:*

$$(vf)^{+}(\bar{x},u) \leq \sup_{M \in \bar{\partial} f(\bar{x})} \langle v, Mu \rangle,$$

where $(vf)^{+}(\bar{x},u)$ denotes the upper Dini directional derivative of (vf) at $\bar{x}$ in the direction u, i.e.,

$$(vf)^{+}(\bar{x},u) := \limsup_{t \downarrow 0} \langle v, f(\bar{x}+tu) - f(\bar{x}) \rangle /t.$$

Proposition 2.3 (Khanh and Tuan, 2006) *Let $f : X \longrightarrow Y$ be a continuous mapping that admits an approximate Jacobian $\bar{\partial} f(\cdot)$ which is upper semi-continuous at $\bar{x}$. Then, $\bar{\partial} f(\bar{x})$ is a first-order approximation of f at $\bar{x}$.*

Definition 2.3 (Mordukhovich and Shao, 1995) *Let $f : X \to \overline{\mathbb{R}}$ and $\bar{x} \in$ $\mathrm{dom}(f)$. The symmetric subdifferential of f at $\bar{x}$ is defined by*

$$\partial^{0} f(\bar{x}) := \partial f(\bar{x}) \sqcup [-\partial(-f)(\bar{x})],$$

where $\partial f(\bar{x}) := \limsup_{x \xrightarrow{f} \bar{x}} \widehat{\partial} f(x)$ and $\widehat{\partial} f(x)$ is the Fréchet subdifferential of f at x.

Note that sufficient conditions for the upper semi-continuity of $\partial^{0} f(\cdot)$ can be found in Joffe (1989) and Loewen (1992).

Proposition 2.4 (Amahroq and Gadhi, 2001, Proposition 1) *Let $f : X \to \overline{\mathbb{R}}$ be continuous and $\bar{x} \in$ dom f. Then, $\partial^{0} f(\bar{x})$ is a first-order approximation of f at $\bar{x}$.*

Khanh and Tuan (2006, Remark 2.1) give an example where a first-order approximation exists for a discontinuous mapping. This emphasizes the advantage of using approximations over other generalized derivatives.

We now present the calculus rules for first-order approximations, including rules for sums, products, quotients, and compositions.

Proposition 2.5 *Let $f, g : X \to Y$ with $A_f(\bar{x}), A_g(\bar{x}) \subseteq L(X,Y)$ as first-order approximations of f and g at $\bar{x}$, respectively.*

1. ***Sum** (Amahroq and Gadhi, 2001): $A_f(\bar{x}) + A_g(\bar{x})$ is a first-order approximation of $f+g$ at $\bar{x}$.*
2. ***Product** (Lafhim, 2024): If $Y = \mathbb{R}$, g is continuous at $\bar{x}$, and $A_f(\bar{x})$ is bounded, then $g(\bar{x})A_f(\bar{x}) + f(\bar{x})A_g(\bar{x})$ is a first-order approximation of the function fg at $\bar{x}$.*

3. ***Quotient*** *(Lafhim, 2024): If $Y = \mathbb{R}$, $g(\bar{x}) \neq 0$ and both $A_f(\bar{x})$ and $A_g(\bar{x})$ are bounded, then*

$$\frac{g(\bar{x})A_f(\bar{x}) - f(\bar{x})A_g(\bar{x})}{g^2(\bar{x})}$$

is a first-order approximation of $\dfrac{f}{g}$ at $\bar{x}$.

4. ***Chain Rule*** *(Amahroq and Gadhi, 2001): If $g : Y \to Z$, then $A_g(f(\bar{x})) \circ A_f(\bar{x})$ is a first-order approximation of $g \circ f$ at $\bar{x}$.*

2.3. Second-order approximations

We now consider second-order approximations, which provide more detailed information about local function behavior than the first-order approximations. These are essential for analyzing optimality conditions and the convergence of optimization algorithms. We present their definition, key properties, and existence for important function classes.

Definition 2.4 *Let $f : X \to Y$, $A_f(\bar{x}) \subseteq L(X,Y)$, and $B_f(\bar{x}) \subseteq B(X,Y)$. The pair $(A_f(\bar{x}), B_f(\bar{x}))$ is a second-order approximation of f at $\bar{x}$ if:*

1. *$A_f(\bar{x})$ is a first-order approximation of f at $\bar{x}$.*
2. *For all $\varepsilon > 0$ there exists $\delta > 0$ such that:*

$$f(x) - f(\bar{x}) \in A_f(\bar{x})(x - \bar{x}) + B_f(\bar{x})(x - \bar{x})(x - \bar{x}) + \varepsilon\|x - \bar{x}\|^2\mathbb{B}_{\mathbb{R}^m}$$

for all $x \in \bar{x} + \delta\mathbb{B}_{\mathbb{R}^n}$.

It is called bounded (or compact) if $A_f(\bar{x})$ and $B_f(\bar{x})$ are norm-bounded (or compact, respectively).

Remark 2.1 1. *Every C^2 mapping f at $\bar{x}$ admits $\big(Df(\bar{x}), D^2f(\bar{x})\big)$ as a second-order approximation, where $Df(\bar{x})$ and $D^2f(\bar{x})$ are the first- and second-order Fréchet derivatives of f at $\bar{x}$, respectively.*

2. *Recall that a mapping $f : X \to Y$ is $C^{1,1}$ at $\bar{x}$ if it is Fréchet differentiable near $\bar{x}$ and $Df(\cdot)$ is Lipschitz at $\bar{x}$.*

We now present the second-order approximation for $C^{1,1}$ functions.

Definition 2.5 *Let $f : X \to Y$ be continuously differentiable. The generalized Hessian $\partial^2 f(\overline{x}) \subseteq B(\mathbb{R}^n, \mathbb{R}^m)$ is an approximate Hessian of f at $\overline{x}$ if $\partial^2 f(\overline{x})$ is an approximate Jacobian of f' at $\overline{x}$, where*

$$\partial^2 f(\bar{x}) := co\ \left\{\lim D^2 f(x_n) : x_n \in \operatorname{dom} D^2 f, x_n \to \bar{x}\right\}.$$

PROPOSITION 2.6 (ALLALI AND AMAHROQ, 1997, PROPOSITION 2.2.3) *If* $f : X \to \mathbb{R}$ *is* $C^{1,1}$ *at* $\bar{x}$*, then* $\left(Df(\bar{x}), \frac{1}{2}\partial^2 f(\bar{x})\right)$ *is a second-order approximation of* f *at* $\bar{x}$*.*

REMARK 2.2 1. *A differentiable function with a bounded first-order approximation of its derivative at* $\bar{x}$ *is necessarily* $C^{1,1}$ *at* $\bar{x}$ *and thus has a second-order approximation (see Allali and Amahroq, 1997).*

2. *Under the assumptions of Proposition 2.6,* $\partial^2 f(\bar{x})$ *is nonempty, compact, and convex (Hiriart-Urruty, Strodiot and Nguyen, 1984, Definition 2.1).*

The following proposition provides a second-order approximation for C^1 functions under an additional condition on the Clarke Hessian.

PROPOSITION 2.7 (KHANH AND TUAN, 2006, PROPOSITION 2.3) *Let* $f : X \to \mathbb{R}$ *be continuously differentiable in a neighborhood of* $\bar{x}$*, with* $\partial^2 f(\cdot)$ *upper semicontinuous at* $\bar{x}$*. Then,* $(Df(\bar{x}), \frac{1}{2}\operatorname{co}\partial^2 f(\bar{x}))$ *is a second-order approximation of* f *at* $\bar{x}$*.*

We now present calculation rules for second-order approximations.

PROPOSITION 2.8 *Let* $f, g : X \to Y$ *with* $(A_f(\bar{x}), B_f(\bar{x}))$ *and* $(A_g(\bar{x}), B_g(\bar{x}))$ *as second-order approximations at* $\bar{x}$*, respectively.*

1. ***Sum*** *(Allali and Amahroq, 1997):* $f + g$ *has the second-order approximation* $(A_f(\bar{x}) + A_g(\bar{x}), B_f(\bar{x}) + B_g(\bar{x}))$ *at* $\bar{x}$*.*
2. ***Product*** *(Khanh and Tung, 2015): If* $Y = \mathbb{R}$*,* $A_f(\bar{x})$ *and* $A_g(\bar{x})$ *are bounded, then, the function* fg *has the second-order approximation*

$$(g(\bar{x}) A_f(\bar{x}) + f(\bar{x}) A_g(\bar{x}), g(\bar{x}) B_f(\bar{x}) + f(\bar{x}) B_g(\bar{x}) + A_f(\bar{x}) A_g(\bar{x}))$$

   at $\bar{x}$*.*
3. ***Chain Rule*** *(Jourani and Thibault, 1993): If* $g : Y \to Z$*, and both* $(A_f(\bar{x}), B_f(\bar{x}))$ *and* $(A_g(\bar{y}), B_g(\bar{y}))$ *are bounded with* $\bar{y} := g(\bar{x})$*, then* $g \circ f$ *has the second-order approximation* $(A_g(\bar{y}) \circ A_f(\bar{x}), B_{g\circ f}(\bar{x}))$ *at* $\bar{x}$*, where* $B_{g\circ f}(\bar{x}) := B_g(\bar{y}) \circ \Lambda \circ A_f(\bar{x}) + A_g(\bar{y}) \circ B_f(\bar{x})$ *and* $\Lambda(y) = (y, y)$*.*

The above proposition shows that $f \circ g$ always admits a second-order approximation whenever g is C^2 and f is $C^{1,1}$.

We finally note that first- and second-order approximations are not unique, as any set that includes an approximation also qualifies as an approximation.

3. First- and second-order optimality conditions via gap function

We start this section by presenting an equivalent reformulation of the bilevel problem (GBOP). Reducing a bilevel optimization problem to a single-level problem offers significant theoretical and computational advantages. Here, we apply the gap function reformulation to transform our bilevel problem.

3.1. Gap function reformulation

To tackle the generalized bilevel problem (GBOP), particularly when $Y(x)$ is not convex, we reformulate it into a single-level problem using the gap function. This approach relies on the function:

$$\Psi_\alpha(x, y) = \sup_{\xi \in Y(x)} \phi(x, y, \xi),$$

called the gap function, where:

$$\phi(x, y, \xi) = \langle F(x, y), y - \xi \rangle - \frac{\alpha}{2} \|y - \xi\|^2,$$

and $\alpha > 0$ is a scalar parameter.

The gap function concept, introduced by Auslender (1976) for variational inequalities, was extended by Hearn (1982) to convex programming. Later, Marcotte and Dussault (1987, 1989) used the linear gap function ($\alpha = 0$) for descent methods, and Fukushima (1992) employed the quadratic, differentiable version ($\alpha > 0$) for variational inequalities. In bilevel optimization, Marcotte and Dussault (1989) and Ye, Zhu and Zhu (1997) applied it to derive single-level equivalents of (GBOP). In our analysis, we adopt the linear gap function ($\alpha = 0$), defined as:

$$\Psi_0(x, y) = \sup_{\xi \in Y(x)} \langle F(x, y), y - \xi \rangle.$$

The gap function Ψ_0 is typically nondifferentiable and it satisfies:

$$\Psi_0(x, y) \geq 0 \text{ for all } y \in Y(x), \quad \text{and} \quad \Psi_0(x, y) = 0 \text{ if and only if } y \in S(x).$$

Thus, the solution set $S(x)$ can be written down as:

$$S(x) = \{y \in Y(x) : \Psi_0(x, y) \leq 0\}.$$

This allows us to reformulate (GBOP) as a single-level problem by replacing the lower-level variational inequality with a gap function constraint. The equivalent problem is:

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ \text{s.t.} \quad & g(x) \leq 0, \\ & G(x,y) \leq 0, \\ & \Psi_0(x,y) \leq 0. \end{aligned} \tag{GFR}$$

The feasible set $\mathcal{F}$ becomes:

$$\mathcal{F} = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \ : \ g(x) \leq 0, \ G(x,y) \leq 0, \ \Psi_0(x,y) \leq 0\}.$$

For certain proofs, we use an equivalent form of (GFR) to simplify notation:

$$\min_{x,y} \ f(x,y) \quad \text{s.t.} \quad \Gamma(x,y) \leq 0, \tag{P}$$

where $\Gamma : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^r$, $\Gamma(x,y) = (g(x), G(x,y), \Psi_0(x,y))$, and $r = q + s + 1$. The feasible set is then:

$$\mathcal{F} = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \ : \ \Gamma(x,y) \leq 0\}.$$

3.2. First- and second-order necessary optimality conditions

In this subsection, we use approximations to derive first- and second-order necessary optimality conditions for the generalized bilevel optimization problem (GBOP) without assuming convexity in the lower-level problem or that $S(x)$ is a singleton. We achieve this through the reformulation (GFR).

Theorem 3.1 *Let $\bar{z} = (\bar{x}, \bar{y}) \in \mathcal{F}$ be a local solution of* (GFR). *Assume that f, g, G, and Ψ_0 admit first-order approximations $A_f(\bar{z})$, $A_g(\bar{z})$, $A_G(\bar{z})$, and $A_{\Psi_0}(\bar{z})$ at $\bar{z}$, respectively. Then, for all $z = (x,y) \in \mathbb{R}^n \times \mathbb{R}^m$, there exist $A^f \in p\text{-}A_f(\bar{z})$, $A^g \in A_g(\bar{z})$, $A^G \in A_G(\bar{z})$, $A^{\Psi_0} \in A_{\Psi_0}(\bar{z})$, and scalars $\alpha, \delta \in \mathbb{R}_+$, $\beta \in \mathbb{R}^q_+$, $\gamma \in \mathbb{R}^s_+$, not all zero, such that:*

$$\begin{cases} \alpha A^f(z) + \langle \beta, A^g(z) \rangle + \langle \gamma, A^G(z) \rangle + \delta A^{\Psi_0}(z) \geq 0, \\ \langle \beta, g(\bar{z}) \rangle = 0, \\ \langle \gamma, G(\bar{z}) \rangle = 0. \end{cases}$$

Proof Suppose $\bar{z}$ is a local minimum of (GFR), hence also of (P). For any $z \in \mathbb{R}^n \times \mathbb{R}^m$, we consider two cases.

Case I: $z \in \mathcal{D}(\mathcal{F}, \bar{z})$. There exists a sequence $t_n \downarrow 0$ with $\bar{z} + t_n z \in \mathcal{F}$ for all $n \in \mathbb{N}$. Fix $\epsilon > 0$. By the definition of first-order approximations,

$$f(\bar{z} + t_n z) - f(\bar{z}) \in A_f(\bar{z})(t_n z) + t_n \epsilon \|z\| \mathbb{B}_{\mathbb{R}}. \tag{3.1}$$

Since $\bar{z}$ is a local minimum of (GFR),

$$f(\bar{z}+t_n z)-f(\bar{z})\geq 0. \tag{3.2}$$

From (3.1) and (3.2), there exist $A_n^f \in A_f(\bar{z})$ and $b_n \in \mathbb{B}_{\mathbb{R}}$ such that:

$$A_n^f(z)+\epsilon\|z\|b_n \geq 0. \tag{3.3}$$

We claim that there exists $A^f \in p\text{-}A_f(\bar{z})$ with:

$$A^f(z)\geq 0. \tag{3.4}$$

Indeed, if $(A_n^f)_n$ is bounded and converges to $A^f \in \operatorname{cl} A_f(\bar{z})$, compactness of $\mathbb{B}_{\mathbb{R}}$ implies $b_n \to b \in \mathbb{B}_{\mathbb{R}}$, so $A^f(z)+\epsilon\|z\|b \geq 0$. As $\epsilon \to 0$, $A^f(z)\geq 0$. If $(A_n^f)_n$ is unbounded, assume $\|A_n^f\| \to \infty$ and $A_n^f/\|A_n^f\| \to A^f \in A_f(\bar{z})_\infty \setminus \{0\}$. From (3.3),

$$\frac{A_n^f(z)}{\|A_n^f\|}+\epsilon\|z\|\frac{b_n}{\|A_n^f\|}\geq 0.$$

With $\mathbb{B}_{\mathbb{R}}$ compact and $\epsilon \to 0$, we still have $A^f(z)\geq 0$.

Case II: $z \notin \mathcal{D}(\mathcal{F},\bar{z})$**.** For all $t_n \downarrow 0$, $\bar{z}+t_n z \notin \mathcal{F}$. Fix $\epsilon > 0$. By the definition of first-order approximations,

$$\Gamma(\bar{z}+t_n z)-\Gamma(\bar{z}) \in A_\Gamma(\bar{z})(t_n z)+t_n\epsilon\|z\|\mathbb{B}_{\mathbb{R}^r}. \tag{3.5}$$

To prove that there exists $A^\Gamma \in A_\Gamma(\bar{z})$ such that:

$$A^\Gamma(z)+\Gamma(\bar{z})\geq 0, \tag{3.6}$$

assume the contrary: $A^\Gamma(z)+\Gamma(\bar{z})<0$ for all $A^\Gamma \in A_\Gamma(\bar{z})$. From (3.5),

$$\Gamma(\bar{z}+t_n z)\in(1-t_n)\Gamma(\bar{z})+t_n\left(\Gamma(\bar{z})+A_\Gamma(\bar{z})(z)+\epsilon\|z\|\mathbb{B}_{\mathbb{R}^r}\right).$$

We know that $\bar{z}$ is a local minimum of (GFR), then $\bar{z}$ is also a local minimum of (P). Since $\Gamma(\bar{z})\leq 0$, for small ϵ,

$$\Gamma(\bar{z})+A_\Gamma(\bar{z})(z)+\epsilon\|z\|\mathbb{B}_{\mathbb{R}^r}\subseteq \operatorname{int}(\mathbb{R}^r_-),$$

so:

$$(1-t_n)\Gamma(\bar{z})+t_n\left(\Gamma(\bar{z})+A_\Gamma(\bar{z})(z)+\epsilon\|z\|\mathbb{B}_{\mathbb{R}^r}\right)\subseteq \operatorname{int}(\mathbb{R}^r_-),$$

implying $\Gamma(\bar{z}+t_n z)\leq 0$, a contradiction. Hence, (3.6) holds.

From (3.4) and (3.6),

$$(A^f(z),A^\Gamma(z)+\Gamma(\bar{z}))\notin \operatorname{int}(\mathbb{R}_-\times\mathbb{R}^r_-).$$

By applying a classic separation theorem from convex analysis (see Rockafellar, 1970, Theorem 11.2), there exist $\alpha, \delta \in \mathbb{R}_+$, $\beta \in \mathbb{R}^q_+$, $\gamma \in \mathbb{R}^s_+$, not all zero, such that:

$$\begin{cases} \alpha A^f(z) + \langle (\beta, \gamma, \delta), A^\Gamma(z) \rangle \geq 0, \\ \langle (\beta, \gamma, \delta), \Gamma(\bar{z}) \rangle = 0. \end{cases} \tag{3.7}$$

Since $\Gamma = (g, G, \Psi_0)$, there exist $A^g \in A_g(\bar{z})$, $A^G \in A_G(\bar{z})$, and $A^{\Psi_0} \in A_{\Psi_0}(\bar{z})$ such that:

$$\alpha A^f(z) + \langle \beta, A^g(z) \rangle + \langle \gamma, A^G(z) \rangle + \delta A^{\Psi_0}(z) \geq 0.$$

With $\Psi_0(\bar{z}) = 0$, the second condition in (3.7) implies:

$$\langle \beta, g(\bar{z}) \rangle = 0, \quad \langle \gamma, G(\bar{z}) \rangle = 0.$$

This completes the proof. ■

EXAMPLE 3.1 *Consider the generalized bilevel optimization problem:*

$$\min_{x,y} \{ f(x,y) = x^2 + y \ : \ g_1(x) = x - 4 \leq 0, \ g_2(x) = -x \leq 0, \ y \in S(x) \},$$

where $S(x)$ is the solution set of:

$$\langle F(x,y), y - \xi \rangle \leq 0, \quad \xi \in Y(x),$$

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

$$Y(x) = \{ y \in \mathbb{R}^m : G_1(x,y) = -y - 3 \leq 0, \ G_2(x,y) = -|x| + y - 4 \leq 0 \} = [-3, |x| + 4].$$

The point $\bar{z} = (\bar{x}, \bar{y}) = (0, -3)$ is the global optimum, with:

$$S(x) = \begin{cases} [-3, 4] & \text{if } x = 0, \\ \{-3, 0\} & \text{if } x \neq 0. \end{cases}$$

The gap function is:

$$\Psi_0(x,y) = \sup_{\xi \in Y(x)} \langle F(x,y), y - \xi \rangle = (y+3)|xy|.$$

The reformulated problem becomes:

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ s.t \quad & g_1(x) = x - 4 \leq 0, \\ & g_2(x) = -x \leq 0, \\ & G_1(x,y) = -y - 3 \leq 0, \\ & G_2(x,y) = -|x| + y - 4 \leq 0, \\ & \Psi_0(x,y) = (y+3)|xy| \leq 0. \end{aligned}$$

To verify Theorem 3.1 at $\bar{z} = (0, -3)$, compute the first-order approximations:

- $A_f(\bar{z}) = \{(0,1)\}$,
- $A_{g_1}(\bar{x}) = \{1\}$,
- $A_{g_2}(\bar{x}) = \{-1\}$,
- $A_{G_1}(\bar{z}) = \{(0,-1)\}$,
- $A_{G_2}(\bar{z}) = [-1,1] \times \{1\}$,
- $A_{\Psi_0}(\bar{z}) = \{(0,0)\}$ *(since* $\Psi_0(0,-3) = 0$ *and is locally constant).*

Upon choosing $A^f = (0,1)$*,* $A^{g_1} = 1$*,* $A^{g_2} = -1$*,* $A^{G_1} = (0,-1)$*,* $A^{G_2} = (1,1)$*,* $A^{\Psi_0} = (0,0)$*, and scalars* $\alpha = 1$*,* $\beta = (0,0)$*,* $\gamma = (1,0)$*,* $\delta = \dfrac{3}{2}$*, we get:*

$$\alpha A^f(z) + \langle \beta, A^g(z) \rangle + \langle \gamma, A^G(z) \rangle + \delta A^{\Psi_0}(z) = y - y = 0 \geq 0,$$

$$\langle \beta, g(\bar{z}) \rangle = 0, \quad \langle \gamma, G(\bar{z}) \rangle = 0,$$

satisfying the theorem's conditions. Thus, $\bar{z} = (0,-3)$ *is a candidate optimum.*

It is noteworthy that in this example, the lower-level constraint function $G_2(x,y) = -|x| + y - 4$ *is non-differentiable with respect to* x *at* $\bar{x} = 0$*. Furthermore, the gap function* $\Psi_0(x,y) = (y+3)|xy|$ *is also non-differentiable. Standard optimality conditions relying on smoothness might not be directly applicable or would require more complex constraint qualifications. The use of first-order approximations, as demonstrated, provides a robust framework for handling such non-smoothness.*

Remark 3.1 *In Theorem 3.1, if there exist* $z \in \mathbb{R}^n \times \mathbb{R}^m$ *and scalars* $\alpha, \delta \in \mathbb{R}_+$*,* $\beta \in \mathbb{R}^q_+$*,* $\gamma \in \mathbb{R}^s_+$ *such that:*

$$\langle \beta, A^g(z) \rangle + \langle \gamma, A^G(z) \rangle + \delta A^{\Psi_0}(z) < 0$$

for all $A^g \in A_g(\bar{z})$*,* $A^G \in A_G(\bar{z})$*, and* $A^{\Psi_0} \in A_{\Psi_0}(\bar{z})$*, then* $\alpha \neq 0$*.*

In the following theorem, we derive the second-order necessary optimality conditions for the problem (GBOP) using the reformulated problem (GFR). To this end, we define the following critical direction set:

$$\Lambda_1(\bar{z}) = \left\{ \begin{array}{c|ll} & A^f(z) \leq 0 & \forall A^f \in A_f(\bar{z}) \\ z = (x,y) & A^g(z) + g(\bar{z}) \leq 0 & \forall A^g \in A_g(\bar{z}) \\ & A^G(z) + G(\bar{z}) \leq 0 & \forall A^G \in A_G(\bar{z}) \\ & A^{\Psi_0}(z) \leq 0 & \forall A^{\Psi_0} \in A_{\Psi_0}(\bar{z}) \end{array} \right\}.$$

Theorem 3.2 *Let* $\bar{z} = (\bar{x}, \bar{y}) \in \mathcal{F}$ *be a local solution of* (GFR)*. Assume that* $(A_f(\bar{z}), B_f(\bar{z}))$*,* $(A_g(\bar{z}), B_g(\bar{z}))$*,* $(A_G(\bar{z}), B_G(\bar{z}))$*, and* $(A_{\Psi_0}(\bar{z}), B_{\Psi_0}(\bar{z}))$ *are second-order approximations of* f*,* g*,* G*, and* Ψ_0 *at* $\bar{z}$*, respectively. Then, for all* $z \in \Lambda_1(\bar{z})$*, there exist* $B^f \in p\text{-}B_f(\bar{z})$*,* $B^g \in B_g(\bar{z})$*,* $B^G \in B_G(\bar{z})$*,* $B^{\Psi_0} \in B_{\Psi_0}(\bar{z})$*,* *and scalars* $\alpha, \delta \in \mathbb{R}_+$, $\beta \in \mathbb{R}^q_+$, $\gamma \in \mathbb{R}^s_+$, *not all zero, such that:*

$$\begin{cases} \alpha B^f(z,z) + \langle \beta, B^g(z,z)\rangle + \langle \gamma, B^G(z,z)\rangle + \delta B^{\Psi_0}(z,z) \geq 0, \\ \langle \beta, g(\bar{z})\rangle = 0, \\ \langle \gamma, G(\bar{z})\rangle = 0. \end{cases}$$

Proof Suppose $\bar{z}$ is a local minimum of (GFR), hence also of (P). For any $z \in \mathbb{R}^n \times \mathbb{R}^m$, we consider two cases and adapt the approach from the proof of Theorem 3.1 to second-order approximations.

Case I: $z \in \mathcal{D}(\mathcal{F}, \bar{z})$**.** There exists a sequence $t_n \downarrow 0$ with $\bar{z} + t_n z \in \mathcal{F}$ for all $n \in \mathbb{N}$. Fix $\epsilon > 0$. By the definition of second-order approximations,

$$f(\bar{z} + t_n z) - f(\bar{z}) \in t_n A_f(\bar{z})(z) + t_n^2 B_f(\bar{z})(z,z) + t_n^2 \epsilon \|z\|^2 \mathbb{B}_{\mathbb{R}}.$$

Since $\bar{z}$ is a local minimum of (GFR),

$$f(\bar{z} + t_n z) - f(\bar{z}) \geq 0.$$

Thus, there exist $A_n^f \in A_f(\bar{z})$, $B_n^f \in B_f(\bar{z})$, and $b_n \in \mathbb{B}_{\mathbb{R}}$ such that:

$$t_n A_n^f(z) + t_n^2 B_n^f(z,z) + t_n^2 \epsilon \|z\|^2 b_n \geq 0. \tag{3.8}$$

For $z \in \Lambda_1(\bar{z})$, $A_n^f(z) \leq 0$. Hence, from (3.8),

$$t_n^2 B_n^f(z,z) + t_n^2 \epsilon \|z\|^2 b_n \geq 0.$$

Divide by $t_n^2 > 0$:

$$B_n^f(z,z) + \epsilon \|z\|^2 b_n \geq 0.$$

As in the proof of Theorem 3.1, whether $(B_n^f)_n$ is bounded or unbounded, there exists $B^f \in p\text{-}B_f(\bar{z})$ such that $B^f(z,z) \geq 0$.

Case II: $z \notin \mathcal{D}(\mathcal{F}, \bar{z})$**.** For all $t_n \downarrow 0$, $\bar{z} + t_n z \notin \mathcal{F}$. Fix $\epsilon > 0$. By the definition of second-order approximations,

$$\Gamma(\bar{z} + t_n z) - \Gamma(\bar{z}) \in t_n A_\Gamma(\bar{z})(z) + t_n^2 B_\Gamma(\bar{z})(z,z) + t_n^2 \epsilon \|z\|^2 \mathbb{B}_{\mathbb{R}^r}.$$

Since $z \in \Lambda_1(\bar{z})$, $A^\Gamma(z) + \Gamma(\bar{z}) \leq 0$ for all $A^\Gamma \in A_\Gamma(\bar{z})$. To prove that there exists $B^\Gamma \in B_\Gamma(\bar{z})$ such that:

$$B^\Gamma(z,z) + \Gamma(\bar{z}) \geq 0,$$

assume the contrary: $B^\Gamma(z,z) + \Gamma(\bar{z}) < 0$ for all $B^\Gamma \in B_\Gamma(\bar{z})$. For large n,

$$\left(\frac{1}{2} - t_n\right) \Gamma(\bar{z}) + t_n \left(\Gamma(\bar{z}) + A^\Gamma(z)\right) \leq 0,$$

and, for small ϵ,

$$\left(\frac{1}{2}-t_n^2\right)\Gamma(\bar{z})+t_n^2\left(\Gamma(\bar{z})+B^{\Gamma}(z,z)+\epsilon\|z\|^2\mathbb{B}_{\mathbb{R}^r}\right)\subseteq\mathbb{R}^r_-,$$

since $\Gamma(\bar{z})\leq 0$. Thus,

$$\Gamma(\bar{z}+t_nz)\in t_nA_{\Gamma}(\bar{z})(z)+t_n^2B_{\Gamma}(\bar{z})(z,z)+t_n^2\epsilon\|z\|^2\mathbb{B}_{\mathbb{R}^r}\subseteq\mathbb{R}^r_-,$$

implying $\Gamma(\bar{z}+t_nz)\leq 0$, a contradiction to $z\notin\mathcal{D}(\mathcal{F},\bar{z})$. Hence, such a B^{Γ} exists.

From Case I and II, for $z\in\Lambda_1(\bar{z})$, whether $z\in\mathcal{D}(\mathcal{F},\bar{z})$ or not,

$$(B^f(z,z),B^{\Gamma}(z,z)+\Gamma(\bar{z}))\notin\operatorname{int}(\mathbb{R}_-\times\mathbb{R}^r_-).$$

As in Theorem 3.1 (see Rockafellar, 1970, Theorem 11.2), there exist $\alpha,\delta\in\mathbb{R}_+$, $\beta\in\mathbb{R}^q_+$, $\gamma\in\mathbb{R}^s_+$, not all zero, such that:

$$\alpha B^f(z,z)+\langle(\beta,\gamma,\delta),B^{\Gamma}(z,z)\rangle\geq 0,\quad\langle(\beta,\gamma,\delta),\Gamma(\bar{z})\rangle=0.$$

Since $\Gamma=(g,G,\Psi_0)$, there exist $B^g\in B_g(\bar{z})$, $B^G\in B_G(\bar{z})$, and $B^{\Psi_0}\in B_{\Psi_0}(\bar{z})$ satisfying:

$$\alpha B^f(z,z)+\langle\beta,B^g(z,z)\rangle+\langle\gamma,B^G(z,z)\rangle+\delta B^{\Psi_0}(z,z)\geq 0,$$

with $\langle\beta,g(\bar{z})\rangle=0$ and $\langle\gamma,G(\bar{z})\rangle=0$, which completes the proof. ■

3.3. First- and second-order sufficient optimality conditions

In this subsection, we derive first- and second-order sufficient optimality conditions for the bilevel optimization problem (GBOP) using the gap function reformulation (GFR). Unlike the necessary conditions in the previous subsection, we seek to establish sufficient conditions that ensure optimality. We use first- and second-order approximations through generalized derivatives. This method sidesteps strict convexity assumptions and broadens applicability.

We begin with the first-order sufficient optimality conditions for the reformulated problem (GFR).

Theorem 3.3 *Let $\bar{z}\in\mathcal{F}$ be a feasible point of* (GFR), *and let $A_f(\bar{z})$, $A_g(\bar{z})$, $A_G(\bar{z})$ and $A_{\Psi_0}(\bar{z})$ be first-order approximations of f, g, G and Ψ_0 at $\bar{z}$, respectively. Suppose that for all $d\in\mathcal{T}_1(\mathcal{F},\bar{z})$ and for all $A^f\in p\text{-}A_f(\bar{z})$, $A^g\in p\text{-}A_g(\bar{z})$, $A^G\in p\text{-}A_G(\bar{z})$, and $A^{\Psi_0}\in p\text{-}A_{\Psi_0}(\bar{z})$, there exist $\alpha\in\mathbb{R}_+,\beta\in\mathbb{R}^q_+,\gamma\in\mathbb{R}^s_+,\delta\in\mathbb{R}_+$, not all zero, such that:*

$$\begin{cases}\alpha A^f(d)+\langle\beta,A^g(d)\rangle+\langle\gamma,A^G(d)\rangle+\delta A^{\Psi_0}(d)>0,\\ \langle\beta,g(\bar{z})\rangle=0,\\ \langle\gamma,G(\bar{z})\rangle=0.\end{cases}\tag{3.9}$$

Then, $\bar{z}$ is a local minimum of (GFR).

PROOF Suppose $\bar{z}$ is not a local minimum of (GFR), hence not of (P). Then, there exist $(z_n)_{n\in\mathbb{N}}$ converging to $\bar{z}$ such that $f(z_n) \to f(\bar{z})$, $\Gamma(z_n) \leq 0$, and:

$$f(z_n) - f(\bar{z}) < 0, \quad \forall n \in \mathbb{N}. \tag{3.10}$$

Define $t_n = \|z_n - \bar{z}\|$ and $d_n = (z_n - \bar{z})/\|z_n - \bar{z}\|$. We have $t_n \downarrow 0$ and $d_n \to d$ for some d with $\|d\| = 1$. As $z_n = \bar{z} + t_n d_n \in \mathcal{F}$, it follows that $d \in \mathcal{T}_1(\mathcal{F}, \bar{z}) \setminus \{0\}$.

By the definition of first-order approximations, for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$:

$$f(z_n) - f(\bar{z}) \in t_n A_f(\bar{z})(d_n) + t_n\epsilon\|d_n\|\mathbb{B}_{\mathbb{R}}.$$

From (3.10), this implies that there exist $A_n^f \in A_f(\bar{z})$ and $b_n \in \mathbb{B}_{\mathbb{R}}$ such that:

$$A_n^f(d_n) + \epsilon\|d_n\|b_n < 0.$$

Taking $n \to \infty$ and $\epsilon \downarrow 0$, we extract a subsequence (if necessary), where $A_n^f \to A^f \in p\text{-}A_f(\bar{z})$ and $d_n \to d$. Thus, $A^f(d) \leq 0$.

At $\Gamma(\bar{z}) = (g(\bar{z}), G(\bar{z}), 0_{\mathbb{R}})$, the first-order approximation of Γ gives:

$$\Gamma(z_n) - \Gamma(\bar{z}) \in t_n A_\Gamma(\bar{z})(d_n) + t_n\epsilon\|d_n\|\mathbb{B}_{\mathbb{R}^r}.$$

Since $\Gamma(z_n) \leq 0$, we have:

$$\Gamma(\bar{z}) + t_n A_n^\Gamma(d_n) + t_n\epsilon\|d_n\|b_n^r \leq 0 \quad \text{for some } A_n^\Gamma \in p\text{-}A_\Gamma(\bar{z}),\ b_n^r \in \mathbb{B}_{\mathbb{R}^r}.$$

Upon dividing by t_n and letting $n \to \infty$, $\epsilon \downarrow 0$, we obtain:

$$A^\Gamma(d) \in -\text{cl cone}\left(\mathbb{R}_+^r + \Gamma(\bar{z})\right),$$

where $A^\Gamma \in p\text{-}A_\Gamma(\bar{z})$. By the chain rule for approximations, this implies that there exist $A^g \in p\text{-}A_g(\bar{z})$, $A^G \in p\text{-}A_G(\bar{z})$, and $A^{\Psi_0} \in p\text{-}A_{\Psi_0}(\bar{z})$ such that:

$$\left(A^f(d), A^g(d), A^G(d), A^{\Psi_0}(d)\right) \in -\left(\mathbb{R}_+ \times \text{cl cone}\left(\mathbb{R}_+^q + g(\bar{z}), \mathbb{R}_+^s + G(\bar{z}), \mathbb{R}_+\right)\right).$$

For any non-trivial multipliers $(\alpha, \beta, \gamma, \delta) \in \mathbb{R}_+ \times \mathbb{R}_+^q \times \mathbb{R}_+^s \times \mathbb{R}_+$ satisfying $\langle \beta, g(\bar{z})\rangle = 0$ and $\langle \gamma, G(\bar{z})\rangle = 0$, this leads to:

$$\alpha A^f(d) + \langle \beta, A^g(d)\rangle + \langle \gamma, A^G(d)\rangle + \delta A^{\Psi_0}(d) \leq 0,$$

which contradicts the strict inequality in (3.9). Hence, $\bar{z}$ must be a local minimum of (GFR). ■

Next, we establish the second-order sufficient conditions. Define:

$$\Lambda_2(\bar{z}) = \left\{ z = (x, y) \;\middle|\; \begin{array}{ll} A^f(z) \geq 0 & \forall A^f \in A_f(\bar{z}), \\ A^g(z) - g(\bar{z}) \geq 0 & \forall A^g \in A_g(\bar{z}), \\ A^G(z) - G(\bar{z}) \geq 0 & \forall A^G \in A_G(\bar{z}), \\ A^{\Psi_0}(z) \geq 0 & \forall A^{\Psi_0} \in A_{\Psi_0}(\bar{z}) \end{array} \right\}.$$

Theorem 3.4 *Let $\bar{z} \in \mathcal{F}$ be feasible for* (GFR), *and* $(A_f(\bar{z}), B_f(\bar{z}))$, $(A_g(\bar{z}), B_g(\bar{z}))$, $(A_G(\bar{z}), B_G(\bar{z}))$, *and* $(A_{\Psi_0}(\bar{z}), B_{\Psi_0}(\bar{z}))$ *be second-order approximations of f, g, G, and Ψ_0 at $\bar{z}$, respectively. Suppose that for all $d \in \mathcal{T}_1(\mathcal{F}, \bar{z})$, the following hold:*

(i) There exists a neighborhood V of d such that $z \in \Lambda_2(\bar{z})$ for all $z \in V$ with $\|z\| = 1$.

(ii) For all $B^f \in p\text{-}B_f(\bar{z})$, $B^g \in p\text{-}B_g(\bar{z})$, $B^G \in p\text{-}B_G(\bar{z})$, and $B^{\Psi_0} \in p\text{-}B_{\Psi_0}(\bar{z})$, there exist $\alpha \in \mathbb{R}_+$, $\beta \in \mathbb{R}^q_+$, $\gamma \in \mathbb{R}^s_+$, $\delta \in \mathbb{R}_+$, not all zero, such that:

$$\begin{cases} \alpha B^f(d,d) + \langle \beta, B^g(d,d)\rangle + \langle \gamma, B^G(d,d)\rangle + \delta B^{\Psi_0}(d,d) > 0, \\ \langle \beta, g(\bar{z})\rangle = 0, \\ \langle \gamma, G(\bar{z})\rangle = 0. \end{cases} \tag{3.11}$$

Then, $\bar{z}$ is a local minimum of (GFR).

Proof Suppose $\bar{z}$ is not a local minimum of (GFR), hence not of (P). Then, there exist $(z_n)_{n\in\mathbb{N}}$ converging to $\bar{z}$, such that $f(z_n) \to f(\bar{z})$, $\Gamma(z_n) \leq 0$, and:

$$f(z_n) - f(\bar{z}) < 0 \quad \text{for all } n \in \mathbb{N}. \tag{3.12}$$

Define $t_n = \|z_n - \bar{z}\|$ and $d_n = (z_n - \bar{z})/\|z_n - \bar{z}\|$. Then, $t_n \downarrow 0$, $d_n \to d$ with $\|d\| = 1$, and $z_n = \bar{z} + t_n d_n \in \mathcal{F}$, implying $d \in \mathcal{T}_1(\mathcal{F}, \bar{z}) \setminus \{0\}$.

By the second-order approximation of f at $\bar{z}$, for any $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$:

$$f(z_n) - f(\bar{z}) \in t_n A_f(\bar{z})(d_n) + t_n^2 B_f(\bar{z})(d_n, d_n) + t_n^2 \epsilon \|d_n\|^2 \mathbb{B}_{\mathbb{R}}. \tag{3.13}$$

Since $d_n \in \Lambda_2(\bar{z})$ for sufficiently large n, by hypothesis (i), we have $A^f(d_n) \geq 0$, $\forall A^f \in A_f(\bar{z})$. From (3.12) and (3.13), there exist $B^f_n \in B_f(\bar{z})$ and $b_n \in \mathbb{B}_{\mathbb{R}}$ such that:

$$0 > f(z_n) - f(\bar{z}) \geq t_n^2 \big(B^f_n(d_n, d_n) + \epsilon \|d_n\|^2 b_n\big).$$

Taking $n \to \infty$ and $\epsilon \downarrow 0$, we obtain:

$$B^f(d,d) \leq 0 \quad \text{for some } B^f \in p\text{-}B_f(\bar{z}).$$

At $\Gamma(\bar{z}) = (g(\bar{z}), G(\bar{z}), 0_{\mathbb{R}})$, the second-order approximation yields:

$$\Gamma(z_n) - \Gamma(\bar{z}) \in t_n A_\Gamma(\bar{z})(d_n) + t_n^2 B_\Gamma(\bar{z})(d_n, d_n) + t_n^2 \epsilon \|d_n\|^2 \mathbb{B}_{\mathbb{R}^r}.$$

Since $\Gamma(z_n) \leq 0$, there exist $A^\Gamma_n \in A_\Gamma(\bar{z})$, $B^\Gamma_n \in B_\Gamma(\bar{z})$ and $b^r_n \in \mathbb{B}_{\mathbb{R}^r}$ such that

$$\Gamma(\bar{z}) + t_n A^\Gamma_n (d_n) + t_n^2 B^\Gamma_n (d_n, d_n) + t_n^2 \epsilon \|d_n\|^2 b^r_n \in \mathbb{R}^r_-.$$

Divide by t_n:

$$A_n^\Gamma(d_n) + t_n B_n^\Gamma(d_n, d_n) + t_n \epsilon \|d_n\|^2 b_n^r \in -\text{cl cone}(\mathbb{R}_+^r + \Gamma(\bar{z})).$$

By hypothesis (i), $A_n^\Gamma(d_n) - \Gamma(\bar{z}) \geq 0$. Subtracting yields:

$$t_n B_n^\Gamma(d_n, d_n) + t_n \epsilon \|d_n\|^2 b_n^r \in -\text{cl cone}(\mathbb{R}_+^r + \Gamma(\bar{z})).$$

Divide again by t_n, let $\epsilon \to 0$ and pass to the limit for n to infinity:

$$B^\Gamma(d, d) \in -\text{cl cone}(\mathbb{R}_+^r + \Gamma(\bar{z})).$$

By the chain rule, there exist $B^g \in p\text{-}B_g(\bar{z}), B^G \in p\text{-}B_G(\bar{z})$ and $B^{\Psi_0} \in p\text{-}B_{\Psi_0}(\bar{z})$ such that:

$$\begin{aligned}&(B^f(d,d), B^g(d,d), B^G(d,d), B^{\Psi_0}(d,d)) \in \\ &-\left(\mathbb{R}_+ \times \text{cl cone}(\mathbb{R}_+^q + g(\bar{z}), \mathbb{R}_+^s + G(\bar{z}), \mathbb{R}_+)\right).\end{aligned}$$

For any non-negative multipliers $(\alpha, \beta, \gamma, \delta)$, satisfying $\langle \beta, g(\bar{z}) \rangle = 0$ and $\langle \gamma, G(\bar{z}) \rangle = 0$, this yields

$$\alpha B^f(d,d) + \langle \beta, B^g(d,d) \rangle + \langle \gamma, B^G(d,d) \rangle + \delta B^{\Psi_0}(d,d) \leq 0,$$

which contradicts the strict positivity in (3.11). Thus, $\bar{z}$ is a local minimum. ■

Example 3.2 Let us consider the following generalized bilevel optimization problem:

$$\min_{x,y} \left\{ f(x,y) = \frac{2}{3}|x|^{\frac{3}{2}} + x^2 + \frac{1}{2}(y-1)^2 \ : \ g(x) = x - 1 \leq 0, \ y \in S(x) \right\},$$

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

$$\langle F(x,y), y - \xi \rangle \leq 0, \qquad \text{for all } \xi \in Y(x),$$

with $F(x,y) = [(xy)^+]^2$, where $(xy)^+ = \max \ \{xy, 0\}$. The set-valued map $Y(x)$ is given by

$$\begin{aligned}Y(x) &= \{y \in \mathbb{R} : G_1(x,y) = -y \leq 0, \ G_2(x,y) = x^2 + y - 2 \leq 0\} \\ &= \begin{cases} [0, 2 - x^2] & \text{if } x \in [-\sqrt{2}, \sqrt{2}] \\ \emptyset & \text{otherwise.} \end{cases}\end{aligned}$$

The solution set is:

$$S(x) = \begin{cases} [0, -x^2 + 2] & \text{if } x \in [-\sqrt{2}, 0] \cup \{\sqrt{2}\} \\ \{0\} & \text{if } x \in]0, \sqrt{2}[\\ \emptyset & \text{otherwise.} \end{cases}$$

The point $\bar{z} = (\bar{x}, \bar{y}) = (0, 1)$ is a candidate for the global optimum. The gap function for the lower-level problem is:

$$\Psi_0(x, y) = \sup_{\xi \in Y(x)} \langle F(x, y), y - \xi \rangle = [(xy)^+]^2 y,$$

as ξ should be chosen as small as possible, i.e., $\xi = 0$ (since $y \geq 0$). This yields the equivalent single-level reformulated problem (GFR):

$$\begin{aligned}
\min_{x,y} \quad & f(x, y) = \frac{2}{3}|x|^{3/2} + x^2 + \frac{1}{2}(y-1)^2 \\
\text{s.t.} \quad & g(x) = x - 1 \leq 0, \\
& G_1(x, y) = -y \leq 0, \\
& G_2(x, y) = x^2 + y - 2 \leq 0, \\
& \Psi_0(x, y) = [(xy)^+]^2 y \leq 0.
\end{aligned}$$

The feasible set $\mathcal{F}$ for this reformulated problem is determined by the simultaneous satisfaction of all constraints. The condition $\Psi_0(x, y) \leq 0$, combined with $G_1(x, y) \leq 0$ (i.e., $y \geq 0$), implies that either $y = 0$ or ($y > 0$ and $xy \leq 0$, which means $x \leq 0$). Considering also $g(x) \leq 0$ and $G_2(x, y) \leq 0$, and that $Y(x)$ is non-empty only for $x \in [-\sqrt{2}, \sqrt{2}]$, we find:

$$\mathcal{F} = \Big\{(x, y) \in \mathbb{R}^2 \ : \ x \in [-\sqrt{2}, 0] \ \text{ and } \ y \in [0, 2 - x^2]\Big\} \cup \ (0, 1] \times \{0\}.$$

Note that the interval for the second part is $(0, 1]$, due to $g(x) = x - 1 \leq 0$.

We now demonstrate that $\bar{z} = (0, 1)$ satisfies the second-order sufficient optimality conditions in Theorem 3.4.

The contingent cone to $\mathcal{F}$ and the set of unit tangent directions at $\bar{z} = (0, 1)$ take the following form:

$$\mathcal{T}(\mathcal{F}, \bar{z}) = \mathbb{R}_- \times \mathbb{R}, \quad \mathcal{T}_1(\mathcal{F}, \bar{z}) = \left\{(d_1, d_2) \in \mathbb{R}^2 \ : \ d_1 \leq 0, \|(d_1, d_2)\| = 1\right\}.$$

At $\bar{z} = (0, 1)$, the feasible directions $d = (d_1, d_2)$ must have $d_1 \leq 0$. If $d_1 < 0$, we move into the region where y can vary. If $d_1 = 0$, we can move along the y-axis (since for $x = 0$, $y \in [0, 2]$ is feasible). Movement with $d_1 > 0$ from $(0, 1)$ is not possible as it would require an immediate jump to $y = 0$.

For condition (i) of Theorem 3.4, we consider the first-order approximations at $\bar{z} = (0, 1)$: $A_f(\bar{z}) = \{(0, 0)\}$, $A_g(\bar{z}) = \{(1, 0)\}$, $A_{G_1}(\bar{z}) = \{(0, -1)\}$, $A_{G_2}(\bar{z}) = \{(0, 1)\}$, and $A_{\Psi_0}(\bar{z}) = \{(0, 0)\}$. Let $d = (d_1, d_2) \in \mathcal{T}_1(\mathcal{F}, \bar{z})$. For any unit vector z in a neighborhood V of d, the conditions defining $\Lambda_2(\bar{z})$ are:

- $A^f(z) = 0 \geq 0$.
- $A^g(z) - g(\bar{z}) = z_1 + 1 \geq 0$, since z is near d and $d_1 \leq 0$ (with $||d|| = 1 \implies d_1 \in [-1, 0]$), and so $z_1 \leq 0$ and near d_1.

- $A^{G_1}(z) - G_1(\bar{z}) = -z_2 + 1 \geq 0$, because $z_2 \in [-1, 1]$ as $||z|| = 1$.
- $A^{G_2}(z) - G_2(\bar{z}) = z_2 + 1 \geq 0$.
- $A^{\Psi_0}(z) - \Psi_0(\bar{z}) = (0, 0) \cdot z - 0 = 0 \geq 0$.

Thus, $z \in \Lambda_2(\bar{z})$ for $z \in V$. Consequently, condition (i) of Theorem 3.4 holds.

For condition (ii), direct calculations of second-order approximations at $\bar{z} = (0, 1)$ provide:

$$\partial^2 f(0,1) = \left\{ \begin{pmatrix} a+2 & 0 \\ 0 & 1 \end{pmatrix} : a \geq 0 \right\}, \quad \partial^2 g(0,1) = \{0\},$$

$$\partial^2 G_1(0,1) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \partial^2 G_2(0,1) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \partial^2 \Psi_0(0,1) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The multipliers must satisfy complementarity: $g(\bar{z}) = -1 \neq 0 \implies \beta = 0$; $G_1(\bar{z}) = -1 \neq 0 \implies \gamma_1 = 0$; $G_2(\bar{z}) = -1 \neq 0 \implies \gamma_2 = 0$; $\Psi_0(\bar{z}) = 0$, so $\delta \geq 0$. The second-order condition from Theorem 3.4 (ii) becomes:

$$\alpha B^f(d,d) + \delta B^{\Psi_0}(d,d) > 0.$$

Since $B^{\Psi_0}(d,d) = \dfrac{1}{2} d^T \partial^2 \Psi_0(\bar{z}) d = 0$, we need $\alpha B^f(d,d) > 0$. For $d = (d_1, d_2) \in \mathcal{T}_1(\mathcal{F}, \bar{z})$, we have $B^f(d,d) = \dfrac{1}{2} d^T \partial^2 f(\bar{z}) d = \dfrac{1}{2}\left((a+2)d_1^2 + d_2^2\right)$, with $a \geq 0$. Since $d \neq 0$ and $a + 2 \geq 2$, $(a+2)d_1^2 + d_2^2 > 0$. Upon choosing $\alpha > 0$ (e.g., $\alpha = 1$) and $\delta \geq 0$ (e.g., $\delta = 0$), with $\beta = 0, \gamma_1 = 0, \gamma_2 = 0$, the condition $\alpha B^f(d,d) > 0$ is satisfied for all $d \in \mathcal{T}_1(\mathcal{F}, \bar{z})$.

Thus, for every $d \in \mathcal{T}_1(\mathcal{F}, \bar{z})$, there exist multipliers $\alpha > 0$, $\beta = 0$, $\gamma = (0, 0)$, and $\delta \geq 0$ (not all zero) satisfying the conditions of Theorem 3.4. This proves that $\bar{z} = (0, 1)$ is a strict local minimum for the reformulated problem, and hence also for the original bilevel problem.

This example highlights several key aspects, with respect to which our approxima-tion-based approach is beneficial. The objective function $f(x, y)$ contains the non-differentiable term $|x|^{3/2}$ at $\bar{x} = 0$. The gap function $\Psi_0(x, y) = [(xy)^+]^2 y$ is non-differentiable at $\bar{z} = (0, 1)$, due to the $(xy)^+$ term. Moreover, the feasible set $\mathcal{F}$ is non-convex and has a non-smooth boundary. Classical optimization techniques might struggle with these features, whereas the framework of first- and second-order approximations allows us to establish rigorous optimality conditions.

3.4. First- and second-order approximations of the gap function

The optimality conditions derived earlier rely on first- and second-order approximations of the gap function Ψ_0. This subsection examines these approximations under various regularity conditions.

For the first-order approximation $A_{\Psi_0}(\bar{z})$ at $\bar{z} = (\bar{x}, \bar{y})$:

- If Ψ_0 is Fréchet differentiable at $\bar{z}$, its derivative serves as $A_{\Psi_0}(\bar{z})$.
- If Ψ_0 is locally Lipschitz around $\bar{z}$, Proposition 2.2 establishes the Clarke subdifferential $\partial\Psi_0(\bar{z})$ as $A_{\Psi_0}(\bar{z})$.
- If Ψ_0 is continuous with an approximate Jacobian $\bar{\partial}\Psi_0(\cdot)$ semi-continuous at $\bar{z}$, Proposition 2.3 identifies $\bar{\partial}\Psi_0(\bar{z})$ as $A_{\Psi_0}(\bar{z})$.
- If Ψ_0 is only continuous at $\bar{z}$, Proposition 2.4 shows that the symmetric subdifferential $\partial^0\Psi_0(\bar{z})$ acts as $A_{\Psi_0}(\bar{z})$.
- To determine the first-order approximation of Ψ_0 when it is not continuous at $\bar{z}$, we first introduce the following proposition:

Proposition 3.1 (Clarke, 1990, Proposition 2.3.12) *Let T be a separable set and $\{f_t\}_{t\in T}$ a family of functions with bounded approximations $A_{f_t}(\bar{z})$ at $\bar{z}$. Define $h(z) = \sup_{t\in T} f_t(z)$ and $J(\bar{z}) = \{t \in T : f_t(\bar{z}) = h(\bar{z})\}$. If $t \mapsto f_t(\bar{z})$ is upper semicontinuous, then co $\{A_{f_t}(\bar{z}) : t \in J(\bar{z})\}$ approximates h at $\bar{z}$.*

By setting $h_\xi(x, y) = \langle F(x, y), y - \xi\rangle$, Proposition 3.1 yields a first-order approximation of Ψ_0 at $\bar{z} = (\bar{x}, \bar{y})$, as shown in:

Corollary 3.1 *Let $\bar{z} = (\bar{x}, \bar{y}) \in \mathcal{F}$ be feasible for* (GFR). *Assume $Y(\bar{x})$ is separable and $\{h_\xi\}_{\xi\in Y(\bar{x})}$ admits bounded approximations $A_{h_\xi}(\bar{z})$ at $\bar{z}$. If $\xi \mapsto h_\xi(\bar{z})$ is upper semicontinuous, then co $\{A_{h_\xi}(\bar{z}) : \xi \in J(\bar{z})\}$ is a first-order approximation of Ψ_0 at $\bar{z}$ where $J(\bar{z}) = \{\xi \in Y(\bar{x}) : h_\xi(\bar{z}) = \Psi_0(\bar{z})\}$.*

For the second-order approximation $B_{\Psi_0}(\bar{z})$ at $\bar{z} = (\bar{x}, \bar{y})$:

- If Ψ_0 is C^2 at $\bar{z}$, then $\big(D\Psi_0(\bar{z}), D^2\Psi_0(\bar{z})\big)$ is a second-order approximation of Ψ_0 at $\bar{z}$, as mentioned in Remark 2.1.
- If Ψ_0 is $C^{1,1}$ at $\bar{z}$, Proposition 2.6 shows that $\left(D\Psi_0(\bar{z}), \frac{1}{2}\partial^2\Psi_0(\bar{z})\right)$ is a second-order approximation of Ψ_0 at $\bar{z}$.
- If Ψ_0 is continuously differentiable and the Clarke generalized Hessian $\partial^2\Psi_0(\cdot)$ is upper semi-continuous in an open neighborhood of $\bar{z}$, then $\left(D\Psi_0(\bar{z}), \frac{1}{2}\mathrm{co}\ \partial^2\Psi_0(\bar{z})\right)$ is a second-order approximation of Ψ_0 at $\bar{z}$, according to Proposition 2.7.

Remark 3.2 *In our analysis above, we have provided a second-order approximation of the gap function Ψ_0 under the assumption that it is at least continuously differentiable. However, Ψ_0 can still admit a second-order approximation even when it is nonsmooth, as demonstrated in Example 3.2.*

On the other hand, the regularity of marginal functions (particularly the gap function) has been extensively studied, with various conditions proposed throughout the literature. In the following, we present some of these results:

- *Danskin (1967) established the foundational result that if $F(x, y)$ is continuously differentiable and the solution set to the lower-level problem is a singleton for each x, then Ψ_0 is differentiable with the gradient expressed through the optimal solution.*
- *Bonnans and Shapiro (2000) refined this framework by adding the requirement of second-order sufficient conditions to ensure C^1 smoothness of Ψ_0.*
- *Gauvin and Dubeau (1982) made a significant contribution by examining the differential properties of Ψ_0 in the context of mathematical programming, introducing the concept of pseudo-regularity at a given point and demonstrating that under this condition, plus the Mangasarian-Fromovitz constraint qualification, Ψ_0 is differentiable and locally Lipschitz.*
- *Rockafellar and Wets (1998) proved that if F satisfies certain equidifferentiability properties, Ψ_0 is differentiable without requiring uniqueness of maximizers.*
- *Fiacco and McCormick (1990) contributed through the parametric programming formulation, demonstrating that when the constraints defining $Y(x)$, are sufficiently regular and a constraint qualification holds at the solution, Ψ_0 is differentiable.*
- *Mordukhovich, Nghia and Rockafellar (2013) established C^1 properties using variational analysis and generalized differentiation, requiring only upper-Lip-schitzian behavior of the solution mapping rather than strict uniqueness of solutions, thus relaxing Danskin's original conditions, while still ensuring differentiability of the marginal function Ψ_0.*

Remark 3.3 *The use of the gap function reformulation of the bilevel problem* (GBOP) *provides a single-level problem that proves easier to handle. This approach employs approximations that require neither convexity nor constraint qualifications for our nonsmooth reformulated problem* (GFR). *A difficulty emerges, however: in some cases, approximations of the gap function, particularly the second-order ones, prove hard to determine. These often demand continuous differentiability of the gap function, a property not always clear. For this reason, the subsequent section adopts the second well-known method, the KKT reformulation. This approach requires convexity and a constraint qualification from the lower-level problem but avoids significant complexity in approximation computations.*

4. First- and second-order optimality conditions using KKT reformulation

In this section, we will use another reformulation of the bilevel problem (GBOP) into a single-level problem, namely, the well-known KKT reformulation. By employing this reformulation, we derive the first and second-order optimality conditions using approximations. To proceed, we first introduce the equivalent reformulated problem.

4.1. KKT-reformulation

The Karush-Kuhn-Tucker (KKT) reformulation transforms the generalized bilevel optimization problem (GBOP) into a mathematical program with equilibrium constraints (MPEC). This approach, first introduced by Chen and Florian (1995), demonstrated its applicability to bilevel problems where the lower-level problem is convex. Since then, the KKT reformulation has become a fundamental tool in the study and numerical solution of bilevel optimization problems.

One can check that y solves (VI[x]) if and only if y is an optimal solution of the optimization problem

$$\min_{\xi \in Y(x)} \langle F(x,y), \xi \rangle.$$

Thus, one can equivalently write

$$S(x) = argmin \left\{ \langle F(x,y), \xi \rangle \ : \ \xi \in Y(x) \right\}.$$

This equivalence allows us to view the (GBOP) as a standard bilevel optimization problem (BOP). Therefore, to study our (GBOP), one can use the KKT reformulation introduced in Chen and Florian (1995) for BOP.

Under the condition that the feasible region $Y(x)$ is convex, the core idea behind this approach is to interpret the variational inequality constraint $y \in S(x)$, where y is a solution to the following optimization problem:

$$\min_{\xi \in Y(x)} \langle F(x,y), \xi \rangle,$$

and replace this minimization problem with its corresponding KKT necessary optimality conditions. To proceed, we assume the following for any x and $y \in S(x)$:

- The feasible set $Y(x) = \{\xi \in \mathbb{R}^m : G(x,\xi) \leq 0\}$ is convex.
- The function $G(x,y)$ is continuously differentiable with respect to y.
- For a given x, and for any $y \in S(x)$, a suitable constraint qualification (CQ), such as the Mangasarian-Fromowitz Constraint Qualification (MFCQ), holds for the system of constraints $G(x,\xi) \leq 0$ (defining $Y(x)$) at $\xi = y$. This ensures that $y \in S(x)$ is equivalent to the existence of a multiplier $w \in \mathbb{R}^s_+$ satisfying the KKT conditions for the variational inequality.

Under these assumptions, $(\bar{x},\bar{y})$ is a solution to the generalized bilevel optimization problem (GBOP) if and only if there exists $\bar{w} \in \mathbb{R}^s_+$ such that $(\bar{x},\bar{y},\bar{w})$

is a solution to the following MPEC:

$$\begin{aligned}
\min_{x,y,w} \quad & f(x,y) \\
\text{s.t.} \quad & F(x,y) + \nabla_y G(x,y)^T w = 0, \\
& g(x) \leq 0, \\
& G(x,y) \leq 0, \\
& -w \leq 0 \\
& \langle -w, G(x,y) \rangle = 0.
\end{aligned} \tag{KKTR}$$

Let us note that the new formulation (KKTR) is also stated in Ye, Zhu and Zhu (1997) but without deriving optimality conditions for this reformulation. The relationship between the derived BOP and the KKT reformulation (KKTR) is a standard result in the literature, see e.g. Dempe and Dutta (2012). Consequently, in order to deal with the (GBOP) we will work with its KKT reformulation.

To treat this problem as a one-level optimization problem with inequality constraints, we will rewrite it in the following equivalent form:

$$\begin{aligned}
\min_{x,y,w} \quad & f(x,y) \\
\text{s.t.} \quad & \left\| F(x,y) + \nabla_y G(x.y)^T w \right\| \leq 0, \\
& g(x) \leq 0, \\
& G(x,y) \leq 0, \\
& -w \leq 0, \\
& \langle -w, G(x,y) \rangle \leq 0.
\end{aligned} \tag{KKTR}$$

Let $\mathcal{F}_{KKT}$ be the feasible set of the the problem (KKTR).

4.2. First- and second-order necessary optimality conditions

We begin this subsection by deriving the first-order necessary optimality conditions for the bilevel optimization problem (GBOP) without assuming that the set $S(x)$ is a singleton, using the reformulation (KKTR). In the following, we use the notation $\overline{r} = n + m + s$ to simplify the writing.

THEOREM 4.1 *Let $\bar{z} = (\bar{x}, \bar{y}, \bar{w}) \in \mathcal{F}_{KKT}$ be a local minimum of* (KKTR). *Assume that G is differentiable, and the functions f, F, $\nabla_y G$ and g admit first-order approximations $A_f(\bar{z})$, $A_F(\bar{z})$, $A_{\nabla_y G}(\bar{z})$, $A_g(\bar{z})$ at $\bar{z}$, respectively. Then, for all $z = (x,y,w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^s$, there exist $A^f \in p\text{-}A_f(\bar{z}), A^F \in A_F(\bar{z}), A^{\nabla_y G} \in A_{\nabla_y G}(\bar{z}), A^g \in A_g(\bar{z})$, $b^m \in \mathbb{B}_{\mathbb{R}^m}$, and $\alpha,\ \mu,\ \nu \in \mathbb{R}_+$, $\beta \in \mathbb{R}^q_+$, $\gamma \in \mathbb{R}^s_+$, $\delta \in \mathbb{R}^s_+$ not all zero, such that*

$$\begin{cases} \alpha\left(A^f(z)\right) + \mu\left(\left\langle b^m, A^F(z) + A^{\nabla_y G}(z)^\top w + \nabla_y G(z)^\top 1_s \right\rangle\right) \\ + \langle \beta, A^g(z)\rangle + \langle \gamma, \nabla G(\bar{z})z\rangle - \langle \delta, 1_s\rangle - \nu\left(\nabla G(\bar{z})z^\top w + G(z)^\top 1_s\right) \geq 0, \\ \langle \beta, g(\bar{z})\rangle = 0, \\ \langle \gamma, G(\bar{z})\rangle = 0, \\ \langle \delta, \overline{w}\rangle = 0. \end{cases},$$

where 1_s represents the vector in $\mathbb{R}^s$ with all components equal to 1.

Proof The proof follows the same arguments as in Theorem 3.1, utilizing the calculation rules for first-order approximations, provided in Proposition 2.5. ■

Remark 4.1 *The norm $\|\cdot\|$ is locally Lipschitz around 0. Hence, according to Proposition 2.2, its first-order approximation at 0 is given by the Clarke subdifferential $\partial\|0\|$, which corresponds to the closed unit ball.*

We now proceed to establish second-order necessary optimality conditions using the reformulation (KKTR). For this purpose, we define the following subset $\Lambda^1_{KKT}(\bar{z}) \subseteq \mathbb{R}^{\overline{r}}$ as:

$$\Lambda^1_{KKT}(\bar{z}) = \left\{ z = (x,y,w) \,\middle|\, \begin{array}{l} A^f(z) \leq 0 \\ A^g(z) + g(\bar{z}) \leq 0 \\ \nabla G(\bar{z})z + G(\bar{z}) \leq 0 \\ \left\langle b^m, A^F(z) + A^{\nabla_y G}(z)^\top w + \nabla_y G(z)^\top 1_s \right\rangle \leq 0 \\ \nabla G(\bar{z})z^\top w + G(z)^\top 1_s \leq 0. \end{array} \right\},$$

for all $A^f \in A_f(\bar{z})$, $A^g \in A_g(\bar{z})$, $A^{\nabla_y G} \in A_{\nabla_y G}(\bar{z})$, $A^F \in A_F(\bar{z})$ and $b^m \in \mathbb{B}_{\mathbb{R}^m}$.

Theorem 4.2 *Let $\bar{z} = (\bar{x}, \bar{y}, \bar{w}) \in \mathcal{F}_{KKT}$ be a local minimum of* (KKTR). *Assume that G is differentiable and that $(A_f(\bar{z}), B_f(\bar{z}))$, $(A_g(\bar{z}), B_g(\bar{z}))$, $(\nabla G(\bar{z}), B_G(\bar{z}))$, $(A_F(\bar{z}), B_F(\bar{z}))$, and $(A_{\nabla_y G}(\bar{z}), B_{\nabla_y G}(\bar{z}))$ are second-order approximations of f, g, G, F and $\nabla_y G$ at $\bar{z}$, respectively.*

Then, for all $z = (x,y,w) \in \Lambda^1_{KKT}(\bar{z})$ there exist $B^f \in p\text{-}B_f(\bar{z}), B^g \in B_g(\bar{z})$, $B^G \in B_G(\bar{z})$, $B^F \in B_F(\bar{z})$, $B^{\nabla_y G} \in B_{\nabla_y G}(\bar{z})$, $b^m \in \mathbb{B}_{\mathbb{R}^m}$, and $\alpha,\ \mu,\ \nu \in \mathbb{R}_+$, $\beta \in \mathbb{R}^q_+$, $\gamma \in \mathbb{R}^s_+$ not all zero, such that:

$$\begin{cases} \alpha\left(B^f(z,z)\right) + \mu\left(\left\langle b^m, B^F(z,z) + B^{\nabla_y G}(z,z)^\top w + A^{\nabla_y G}(z)^\top 1_s \right\rangle\right) \\ + \langle \beta, B^g(z,z)\rangle + \left\langle \gamma, B^G(z,z)\right\rangle - \nu\left(B^G(z,z)^\top w\right) \geq 0 \\ \langle \beta, g(\bar{z})\rangle = 0 \\ \langle \gamma, G(\bar{z})\rangle = 0. \end{cases}$$

PROOF The proof employs the same reasoning as in Theorem 3.2, making use of the second-order approximation rules outlined in Proposition 2.8. ■

EXAMPLE 4.1 Let us consider the following generalized bilevel optimization problem:

$$\min_{x,y} \{f(x,y) = x^2 + |y| \mid g_1(x) = |x| - 1 \leq 0,\ g_2(x) = x^2 - 6 \leq 0,\ y \in S(x)\}.$$

Let

$$F(x,y) = |xy|.$$

For every $x \in \mathbb{R}$, let $S(x)$ denote the set of all optimal solutions of the optimization problem with variational inequality constraints

$$\langle F(x,y), y - \xi \rangle \leq 0, \qquad \xi \in Y(x),$$

and

$$Y(x) = \{y \in \mathbb{R}^m : G_1(x,y) = x^2 - y \leq 0,\ G_2(x,y) = y^2 - 4 \leq 0\}.$$

Hence

$$Y(x) = \begin{cases} [x^2, 2]\ ; & -\sqrt{2} \leq x \leq \sqrt{2} \\ \emptyset\ ; & \text{otherwise.} \end{cases}$$

On the one hand, it can be observed that $\bar{z} = (\bar{x}, \bar{y}) = (0,0)$ is an optimum of our problem. The gap function of our problem is given by

$$\Psi_0(x,y) = \sup_{\xi} \{\langle F(x,y), y - \xi \rangle \mid \xi \in Y(x)\} = |xy|(y - x^2).$$

Observe that the gap function Ψ_0 in not differentiable at $\bar{z}$, and so we cannot apply the results from Theorem 3.2.

On the other hand, the KKT reformulation of the considered problem is given by:

$$\begin{aligned}
\min_{x,y,w} \quad & f(x,y) = x^2 + |y| \\
\text{s.t.} \quad & |xy| - w_1 + 2yw_2 = 0 \\
& g_1(x) = |x| - 1 \leq 0 \\
& g_2(x) = x^2 - 6 \leq 0 \\
& G_1(x,y) = x^2 - y \leq 0 \\
& G_2(x,y) = y^2 - 4 \leq 0 \\
& -w_1 \leq 0,\ -w_2 \leq 0 \\
& \left(x^2 - y\right) w_1 + \left(y^2 - 4\right) w_2 = 0.
\end{aligned}$$

We will now show that $\bar{z}$ verifies the necessary optimality conditions, established in Theorem 4.2.

First, we have that

$$\Lambda^1_{KKT}(\bar{z}) = \{0\} \times \{0\} \times \mathbb{R}_+ \times \left\{\frac{1}{2}\right\}.$$

Direct calculations gives:

$$\partial^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \partial^2 g_1(0,0) = \{0\}, \quad \partial^2 g_2(0,0) = \{2\},$$

$$\partial^2 G_1(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad \partial^2 G_2(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix},$$

$$\partial^2 \nabla_y G_1(0,0) = \{0\} \quad \text{and} \quad \partial^2 \nabla_y G_2(0,0) = \{0\}.$$

Thus, for all $z \in \Lambda^1_{KKT}(\bar{z})$, the multipliers $\beta_1 = \beta_2 = 0$, $\gamma_2 = 0$, and for any $\alpha, \mu, \gamma_1, \nu \in \mathbb{R}_+$ and $b \in [-1,1]$ satisfy the optimality conditions, stated in Theorem 4.2.

In this case, while the gap function approach was complicated by the non-differentia-bility of Ψ_0 at $\bar{z} = (0,0)$, the KKT reformulation offers an alternative path. The KKT approach transforms the problem into an MPEC, whose constraints (like the complementarity condition $(-w, G(x,y)) = 0$) are inherently non-smooth. Our approximation-based conditions for the KKTR formulation (Theorems 4.1 and 4.3, and the following Theorems 4.5, 4.6) are designed to handle such structures, without requiring restrictive MPEC-specific constraint qualifications that often involve linear independence or constant rank conditions, which thereby illustrates the broad applicability of the approximation framework.

4.3. First- and second-order sufficient optimality conditions

In this final subsection, we derive first- and second-order sufficient optimality conditions using the reformulation (KKTR) of the problem (GBOP). The following result addresses the first-order conditions:

Theorem 4.3 *Let $\bar{z} = (\bar{x}, \bar{y}, \bar{w})$ be a feasible point of* (KKTR)*. Assume that G is differentiable, and the functions f, F, $\nabla_y G$ and g admit first-order approximations $A_f(\bar{z})$, $A_F(\bar{z})$, $A_{\nabla_y G}(\bar{z})$, $A_g(\bar{z})$ at $\bar{z}$ respectively. Suppose that for all $d = (d_x, d_y, d_w) \in \mathcal{T}_1(\mathcal{F}_{KKT}, \bar{z})$ and for all $A^f \in p\text{-}A_f(\bar{z})$, $A^F \in p\text{-}A_F(\bar{z})$, $A^{\nabla_y G} \in p\text{-}A_{\nabla_y G}(\bar{z})$, $A^g \in p\text{-}A_g(\bar{z})$, $b^m \in \mathbb{B}_{\mathbb{R}^m}$, there exist α, μ, $\nu \in \mathbb{R}_+$, $\beta \in \mathbb{R}^q_+$,*

$\gamma \in \mathbb{R}^s_+$, $\delta \in \mathbb{R}^s_+$, *not all zero, such that*

$$\begin{cases} \alpha\left(A^f(d)\right) + \mu\left(\left\langle b^m, A^F(d) + A^{\nabla_y G}(d)^\top d_w + \nabla_y G(d)^\top 1_s\right\rangle\right) \\ + \langle \beta, A^g(d)\rangle + \langle \gamma, \nabla G(\bar{z})d\rangle - \langle \delta, 1_s\rangle - \nu\left(\nabla G(\bar{z})d^\top d_w + G(d)^\top 1_s\right) > 0, \\ \langle \beta, g(\bar{z})\rangle = 0, \\ \langle \gamma, G(\bar{z})\rangle = 0, \\ \langle \delta, \overline{w}\rangle = 0. \end{cases}$$

Then, $\bar{z}$ is a local minimum of (KKTR).

Proof The proof follows the same arguments as in Theorem 3.3, utilizing the calculation rules for the first-order approximations, provided in Proposition 2.5. ■

In order to derive second-order sufficient optimality conditions using the reformulation (KKTR), we introduce the subset $\Lambda^2_{KKT}(\bar{z}) \subseteq \mathbb{R}^{\overline{r}}$ as follows:

$$\Lambda^2_{KKT}(\bar{z}) = \left\{ z = (x, y, w) \;\middle|\; \begin{array}{l} A^f(z) \geq 0 \\ A^g(z) - g(\bar{z}) \geq 0 \\ \nabla G(\bar{z})z - G(\bar{z}) \geq 0 \\ \left\langle b^m, A^F(z) + A^{\nabla_y G}(z)^\top w + \nabla_y G(z)^\top 1_s\right\rangle \geq 0 \\ \nabla G(\bar{z})z^\top w + G(z)^\top 1_s \geq 0 \end{array} \right\},$$

for all $A^f \in A_f(\bar{z})$, $A^g \in A_g(\bar{z})$, $A^{\nabla_y G} \in A_{\nabla_y G}(\bar{z})$, $A^F \in A_F(\bar{z})$ and $b^m \in \mathbb{B}_{\mathbb{R}^m}$.

Theorem 4.4 *Let $\bar{z} = (\bar{x}, \bar{y}, \bar{w}) \in \mathcal{F}_{KKT}$ be a feasible point of the problem* (KKTR). *Assume that G is differentiable. Let $(A_f(\bar{z}), B_f(\bar{z}))$, $(A_g(\bar{z}),\ B_g(\bar{z}))$, $(\nabla G(\bar{z}), B_G(\bar{z}))$, $(A_F(\bar{z}), B_F(\bar{z}))$ and $(A_{\nabla_y G}(\bar{z}),\ B_{\nabla_y G}(\bar{z}))$ be the second-order approximations of f, g, G, F and $\nabla_y G$ at $\bar{z}$, respectively. Suppose that for all $d = (d_x, d_y, d_w) \in \mathcal{T}_1(\mathcal{F}_{KKT}, \bar{z})$, the following two conditions hold:*

(a) There exists V, a neighbourhood of d, such that for all $z \in V$ with norm one, $z \in \Lambda^1_{KKT}(\bar{z})$.

(b) For all $B^f \in p\text{-}B_f(\bar{z})$, $B^g \in p\text{-}B_g(\bar{z})$, $B^G \in p\text{-}B_G(\bar{z})$, $B^F \in p\text{-}B_F(\bar{z})$ and $B^{\nabla_y G} \in p\text{-}B_{\nabla_y G}(\bar{z})$, $b^m \in \mathbb{B}_{\mathbb{R}^m}$, there exist α, μ, $\nu \in \mathbb{R}_+$, $\beta \in \mathbb{R}^q_+$ and $\gamma \in \mathbb{R}^s_+$, not all zero, such that:

$$\begin{cases} \alpha\left(B^f(d,d)\right) + \mu\left(\left\langle b^{\overline{r}}, B^F(d,d) + B^{\nabla_y G}(d,d)^\top d_w + A^{\nabla_y G}(d)^\top 1_s\right\rangle\right) \\ + \langle \beta, B^g(d,d)\rangle + \left\langle \gamma, B^G(d,d)\right\rangle - \nu\left(B^G(d,d)^\top d_w\right) > 0 \\ \langle \beta, g(\bar{z})\rangle = 0 \\ \langle \gamma, G(\bar{z})\rangle = 0. \end{cases}$$

Then, $\bar{z}$ is a local minimum of (KKTR).

Proof The proof follows the same reasoning as in Theorem 3.4, leveraging the second-order approximation rules, established in Proposition 2.8. ■

5. Conclusion

In this paper, we have investigated first- and second-order necessary and sufficient optimality conditions for generalized bilevel optimization problems with variational inequality constraints. To achieve this, we have employed two distinct single-level reformulations of the hierarchical model (GBOP): the gap function reformulation (GFR) and the KKT reformulation (KKTR). The first reformulation does not require convexity or the satisfaction of constraint qualifications for the lower-level problem. However, it presents challenges in determining the approximations of the gap function, particularly when dealing with second-order approximations. On the other hand, the second approach, based on the KKT reformulation, necessitates convexity and the satisfaction of constraint qualifications for the lower-level problem, but it avoids significant complexity in the computation of approximations. In addition to the theoretical developments, we have provided examples to illustrate the applicability and effectiveness of our results.

Acknowledgements

The authors would like to thank the two anonymous reviewers for their valuable comments and suggestions, which helped to improve the quality of the manuscript.

References

Allali, K. and Amahroq, T. (1997) Second order approximations and dual necessary optimality conditions. *Optimization* **40**, 3, 229–246.

Amahroq, T. and Gadhi, N. (2001) On the regularity condition for vector programming problems. *J. Global Optim.* **21**, 433–441.

Amahroq, T. and Gadhi, N. (2003) Second order optimality conditions for the extremal problem under inclusion constraints. *J. Math. Anal. Appl.* **285**, 1, 74–85.

Auslender, A. (1976) *Optimisation: Méthodes Numériques.* Masson, Paris.

Ben-Ayed, O., Boyce, D. E. and Blair, C. E. (1988) A general bilevel linear programming formulation of the network design problem. *Transp. Res.* **21**, 37–44.

Bonnans, J. F. and Shapiro, A. (2000) *Perturbation Analysis of Optimization Problems. Springer Series in Operations Research* **117**, Springer, New York.

Bracken, J. and McGill, J. T. (1973) Mathematical programs with optimization problems in the constraints. *Oper. Res.* **21**, 37–44.

Chen, Y. and Florian, M. (1995) The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions. *Optimization* **32**, 3, 193–209.

Clarke, F. H. (1990) *Optimization and Nonsmooth Analysis.* Society for Industrial and Applied Mathematics.

Danskin, J. (1967) The theory of max-min, with applications. *SIAM J. Appl. Math.* **14** , 641-664.

Dardour, Z., Lafhim, L. and Kalmoun, E. M. (2024) Primal and dual second-order necessary optimality conditions in bilevel programming. *J. Appl. Numer. Optim.* **6** , 153-175.

Dardour, Z., El Idrissi, R., Kalmoun, E. M. and Lafhim, L. (2025) Second-order necessary conditions for bilevel programs via KKT reformulation. *Carpa-thian J. Math.* **41**, 3, 597–618.

Dempe, S. and Dutta, J. (2012) Is bilevel programming a special case of a mathematical program with complementarity constraints?. *Mathematical Programming.* **131** 37-48.

Dempe, S. and Gadhi, N. (2010) Second order optimality conditions for bilevel set optimization problems. *J. Global Optim.* **47**, 233-245.

Dempe, S. and Zemkoho, A. B. (2013) The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. *Math. Program.* **138**, 1-2, 447-473.

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

El Idrissi, R., Lafhim, L., Kalmoun, E. M. and Ouakrim, Y. (2024) Optimality conditions for bilevel optimal control problems with non-convex quasi-variational inequalities. *RAIRO Oper. Res.* **58**, 2, 1789–1805.

Falk, J. E. and Liu, J. M. (1995) On bilevel programming, Part I: General nonlinear cases. *Math. Program.* **70** , 47-72.

Fiacco, A.V. and McCormick, G.P. (1990) *Nonlinear Programming: Sequential Unconstrained Minimization Techniques. SIAM Classics in Applied Mathematics* **4**, Philadelphia.

Fukushima, M. (1992) Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. *Math. Program.* **53**, 99–110.

Gauvin, J. and Dubeau, F. (1982) Differential properties of the marginal function in mathematical programming. *Math. Program. Stud.* **19**, 101–119.

Hansen, P., Jaumard, B. and Savard, G. (1992) New branch-and-bound rules for linear bilevel programming. *SIAM J. Sci. Stat. Comput.* **13**, 1194–1217.

Hearn, D. W. (1982) The gap function of a convex program. *Oper. Res. Lett.* **1**, 2, 67–71.

Hiriart-Urruty, J. B., Strodiot, J. J. and Nguyen, V. (1984) Generalized Hessian matrix and second order optimality conditions for problems with C1,1 data. *Appl. Math. Optim.* **11**, 43–56.

Huang, Z., Lin, G. and Xiu, N. (2014) Several developments of variational inequalities and complementarity problems, bilevel programming and MPEC. *Oper. Res.* **18**, 1.

Ioffe, A. D. (1989) Approximate subdifferential and applications. III: The metric theory. *Mathematika* **36**, 1–38.

Jeyakumar, V. and Luc, D. T. (1998) Approximate Jacobian matrices for nonsmooth continuous maps and C1-optimization. *SIAM J. Control Optim.* **36** , 1815–1832.

Jourani, A. and Thibault, L. (1993) Approximations and metric regularity in mathematical programming in Banach spaces. *Math. Oper. Res.* **18**, 1, 199–224.

Khanh, P. Q. and Tuan, N. D. (2006) First and second order optimality conditions using approximations for nonsmooth vector optimization in Banach spaces. *J. Optim. Theory Appl.* **130**, 2, 289–308.

Khanh, P. Q. and Tung, L. T. (2013) First and second-order optimality conditions using approximations for vector equilibrium problems with constraints. *J. Global Optim.* **55**, 901–920.

Khanh, P. Q. and Tung, L. T. (2015) First- and second-order optimality conditions for multiobjective fractional programming. *TOP* **23** , 419–440.

Khanh, P. Q. and Tung, N. M. (2015) First and second-order optimality conditions without differentiability in multivalued vector optimization. *Positivity* **19** , 817–841.

Lafhim, L. (2024) First and Second Order Optimality Conditions using Approximations for Fractional Multiobjective Bilevel Problems under Fractional Constraints. *Iran. J. Math. Sci. Inform.* **19**, 1, 211–232.

Loewen, P. D. (1992) Limits of Fréchet normals in nonsmooth analysis. *Optimization and Nonlinear Analysis*, Pitman Res. Notes Math. Ser. **244**, 178–188.

Marcotte, P. and Dussault, J.-P. (1987) A note on a globally convergent method for solving monotone variational inequalities. *Oper. Res. Lett.* **6**, 35–42.

Marcotte, P. and Dussault, J.-P. (1989) A sequential linear programming algorithm for solving monotone variational inequalities. *SIAM J. Control Optim.* **27**: 1260–1278.

Mordukhovich, B. S., Nghia, T. T. A. and Rockafellar, R. T. (2013) Full stability in finite-dimensional optimization. *Math. Oper. Res.* **38**, 4, 768–783.

Mordukhovich, B. S. and Shao, Y. (1995) On nonconvex subdifferential calculus in Banach spaces. *J. Convex Anal.* **2**, 211–227.

Outrata, J. V. (1994) On optimization problems with variational inequality constraints. *SIAM J. Optim.* **4**, 2, 340–357.

Rockafellar, R. T. (1970) *Convex Analysis.* Princeton University Press, Princeton, NJ.

Rockafellar, R. T. and Wets, R. J.-B. (1998) *Variational Analysis. Grundlehren der Mathematischen Wissenschaften* **317**, Springer, Berlin.

Stackelberg, H. (1952) *The Theory of the Market Economy.* Oxford University Press, Oxford, UK.

Stampacchia, G. (1964) Formes bilinéaires coercitives sur les ensembles convexes. *C. R. Math. Acad. Sci. Paris* **258** , 4413–4416.

Sweetser, T. H. (1977) A minimal set-valued strong derivative for set-valued Lipschitz functions. *J. Optim. Theory Appl.* **23** , 539–562.

Thibault, L. (1980) Subdifferentials of compactly Lipschitzian vector-valued functions. *Ann. Mat. Pura Appl.* **4**, 125, 157–192.

Wan, Z. and Chen, J. W. (2013) On bilevel variational inequalities. *J. Oper. Res. Soc. China* **1**, 4, 483–510.

Ye, J. J. (2000) Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints. *SIAM J. Optim.* **10**, 4, 943–962.

Ye, J. J. (1999) Optimality conditions for optimization problems with complementarity constraints. *SIAM J. Optim.* **9**, 374–387.

Ye, J. J. and Ye, X. Y. (1997) Necessary optimality conditions for optimization problems with variational inequality constraints. *Math. Oper. Res.* **22**, 977–997.

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

Zemkoho, A. B. and Zhou, S. (2021) Theoretical and numerical comparison of the Karush–Kuhn–Tucker and value function reformulations in bilevel optimization. *Comput. Optim. Appl.* **78**(2), 625-674.

Zhang, H. (2014) Clarke directional derivatives of regularized gap functions for nonsmooth quasi-variational inequalities. *Math. Control Relat. Fields* **4**, 3, 365–379.