Second-Order Optimality Conditions for Bilevel Optimization Problems Using Parabolic Directional Derivatives

Bhuvnesh Khatana[1] and Geetanjali Panda[2*]

[1,2*]Department of Mathematics, Indian Institute of Technology Kharagpur, Kharagpur, 721302, West Bengal, India.

*Corresponding author(s). E-mail(s): geetanjali@maths.iitkgp.ac.in; dr.geetanjali.p@gmail.com;
Contributing authors: bhuvnesh323001@gmail.com;

Abstract

This paper studies the second-order properties of a class of inequality-constrained bilevel programming problems. First-order optimality conditions for the existence of solutions to bilevel optimization problems are derived using the first-order directional derivative of the optimal solution function of the lower-level problem in the seminal paper by Dempe [1]. In this work, we prove that the optimal solution function of the lower-level problem is a parabolic second-order directionally differentiable function under certain assumptions. The associated second-order necessary and sufficient optimality conditions for the bilevel problem are derived. In this process, the lower-level problem may admit multiple KKT multiplier vectors.

Keywords: Bilevel programming problems, Second-order directional derivatives, Second-order optimality conditions, Solution mapping, Sensitivity analysis

MSC Classification: 90C31 , 90C33 , 90C46

1 Introduction

A bilevel programming problem has a hierarchical two-level structure in which the optimality of the lower-level problem implicitly determines the optimality of the upper-level problem. The following optimistic formulation of the bilevel optimization problem with coupling constraints is considered in this paper.

$$(BOP): \quad \begin{aligned} &\min_{x,y}\ F(x,y) \\ &\text{s.t. } G(x,y) \le 0, \\ &\quad x \in \Psi(y) := \left\{ x \mid x \in \operatorname*{argmin}_{z} f(z,y) \text{ s.t. } g(z,y) \le 0 \right\}, \end{aligned}$$

where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$ are smooth functions. Ψ is a point-to-set mapping $\Psi : \mathbb{R}^m \to 2^{\mathbb{R}^n}$. The set $\Psi(y)$ may be non-singleton for some y. A feasible point $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$ of BOP is called a local optimal solution if there exists $\varepsilon > 0$ such that $F(x,y) \le F(\hat{x},\hat{y})$ for all $(\hat{x},\hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, $\hat{x} \in \Psi(\hat{y})$, $G(\hat{x},\hat{y}) \le 0$ with $\|(x,y) - (\hat{x},\hat{y})\| \le \varepsilon$. A feasible point is a global optimal solution if these conditions hold for all $\varepsilon > 0$. The lower-level problem of BOP is

$$(LP_y): \quad \min_{x} f(x,y) \text{ s.t. } g(x,y) \le 0,$$

which is a parametric programming problem. The global optimal solution of LP_y is obtained for the given upper-level decision vector y to determine a feasible point of BOP. This is often achieved in the literature by formulating the lower-level problem as a convex problem. The BOP becomes intrinsically challenging when the convexity of the lower-level problem is not assumed. Another difficulty with BOP is the lack of regularity when the problem is reformulated into a single-level problem. A detailed discussion of several important aspects of BOP can be found in Dempe [2] and Bard [3].

The common practice of determining the optimality conditions and proposing an approximation scheme for solving any optimization problem relies on a local approximation of the problem at a suitable reference point, under certain assumptions. The approximation of a bilevel problem at any point depends on sensitivity information from the lower-level parametric problem, which is primarily used in bilevel optimization through two approaches. One approach is to transform a bilevel problem into a single-level optimization problem using the optimal value function (marginal function) of the lower-level problem. This is known as the value function approach. However, the single-level problem arising from the reformulation of the bilevel optimization problem is inherently nonconvex and nonsmooth. Hence, using the generalized gradient of the value function, one can derive some optimality conditions for the bilevel optimization problem. Another approach focuses on the sensitivity of the optimal solution function of the lower-level parametric problem. This function is proven to be piecewise differentiable under certain assumptions, and the Jacobian of the optimal solution function can be estimated by solving a finite set of quadratic programming problems, as demonstrated by Dempe and Vogel [4].

There is a vast amount of research on first-order optimality conditions for bilevel optimization problems, including the difficult classes of nonsmooth and nonconvex bilevel frameworks. In the bilevel optimization problem, the optimality of the upper-level problem depends implicitly on the optimal solution function of the lower-level problem, which is generally not known explicitly. The directional differentiability of the optimal solution function becomes an important tool for deriving optimality conditions. Dempe [1] and Mehlitz [5] employed this approach to establish first-order optimality conditions for bilevel problems with uniquely solved lower-level problems. Another widely explored direction is the value function approach for deriving optimality conditions for bilevel problems. Initially, Ye and Zhu [6] derived first-order optimality conditions via the Clarke subdifferential of the value function by introducing a suitable constraint qualification—the *partial calmness condition* on the single-level problem obtained through the value function reformulation. This work has attracted substantial attention and has been further developed in subsequent works. For instance, as a substitute for the restrictive partial calmness condition, Dempe and Zemkoho [7] employed a weaker form of the generalized Mangasarian–Fromovitz constraint qualification (MFCQ) on the single-level problem, along with an additional semicontinuity assumption on the set-valued mapping Ψ, to derive the first-order optimality conditions. Later, in a different work, Dempe and Zemkoho [8] dropped the semicontinuity assumption on Ψ by assuming convexity of the value function. Using tools from nonsmooth analysis, Babahadda and Gadhi [9] employed convexificator calculus for the value function, while Dempe et al. [10] estimated the Mordukhovich subdifferential of the value function of the lower-level problem for the purpose of deriving optimality conditions. The classical KKT-based approach reformulates a bilevel problem into a mathematical program with equilibrium constraints (MPEC) by replacing the convex lower-level problem with its KKT optimality conditions under suitable regularity assumptions. Within this approach, Dempe and Zemkoho [11] studied first-order optimality conditions for nonsmooth bilevel optimization problems by applying various stationary condition concepts of the corresponding MPEC to the bilevel optimization problem. Using Guignard constraint qualification on the MPEC formulation, Dempe and Zemkoho [8] derived first-order optimality conditions for bilevel problems. The combined MPEC–value function approaches are developed by Ye and Zhu [12] and Ma et al. [13] to establish first-order optimality conditions, and various constraint qualification concepts for KKT-based and combined approaches are systematically discussed in the work by Ye [14].

Although first-order methods are useful, they are inherently limited to characterizing stationary points and do not capture the curvature information of the upper and lower-level functions in a bilevel optimization problem. Thus, studying second-order properties of bilevel problems is of significant importance for deriving refined optimality conditions and designing numerical solution methods. An earlier work in this direction is by Falk and Liu [15], who investigated the second-order differentiability properties of the solution function under the Jacobian-uniqueness condition on the lower-level problem. In addition, they developed a quasi-Newton-type scheme, called *adaptive leader predominate algorithm*, which exploits generalized derivative information of the solution function to accelerate convergence. Recently, using the second-order sensitivity properties of the optimal solution function, Dyro et al. [16] conducted an error-bound analysis to solve two-level machine learning problems with the unconstrained lower-level problems.

This motivates the study of second-order sensitivity properties of the optimal solution function for bilevel optimization under weaker assumptions. In this work, we establish second-order sensitivity properties of the optimal solution function without assuming the linear independence constraint qualification (LICQ) for the lower-level problem. Furthermore, we derive no-gap second-order optimality conditions for a class of bilevel optimization problems satisfying the second-order growth condition near local optimal solutions. Recent work has been done in this direction under some restrictive assumptions.

Liu et al. [17, 18] derived second-order optimality conditions for bilevel optimization problems with nonconvex lower-level problems by introducing a new solution concept, the bi-local optimal solution. In [17], the bilevel problem is reformulated as an MPEC in order to establish second-order optimality conditions. To derive the main results, the authors consider MFCQ for the implicit formulation under the Jacobian-uniqueness condition and generalized MFCQ on the MPEC. These assumptions are considered restrictive. In another work, Liu et al. [18] adopt comparatively weaker assumptions, namely MFCQ and the constant rank constraint qualification (CRCQ) for the lower-level problem, along with the metric subregularity constraint qualification (MSCQ) for the implicit formulation of the bilevel problem. By employing second-order tangent approximations of the constraint set, they derive second-order optimality conditions for the bilevel problem in terms of the directional derivatives of the solution function along parabolic paths. However, no computational procedure for these derivatives is provided under the stated assumptions. Instead, to obtain verifiable optimality conditions, the authors impose LICQ on the lower-level problem.

Prior to the recent work based on the optimal solution function, Mehlitz and Zemkoho [19] derived second-order sufficient optimality conditions using second-order directional differentiability and epi-regularity of the value function. In particular, the authors use the second-order growth result of Rückmann and Shapiro [20], which is based on the parabolic second-order directional derivative of the value function. Dempe et al. [21] derived second-order necessary and sufficient optimality conditions using approximate Hessians for a special class of bilevel optimization problems with equality-constrained lower-level problems. In contrast, they reformulate the bilevel problem as an MPEC and impose the Guignard constraint qualification on the resulting formulation.

The primary contribution of this paper is the second-order analysis of bilevel problems with the lower-level problems that admit non-unique KKT multipliers. We establish the second-order directional differentiability along parabolic paths of the optimal solution function for strongly stable parametric problems, extending the results of Ralph and Dempe [22], Liu [23]. The second-order directional derivatives along parabolic paths are obtained by solving a quadratic programming problem. In view of the existing works Liu et al. [17, 18], our analysis does not rely on the uniqueness of KKT multipliers in deriving the main results and does not require constraint qualifications for the implicit reformulation. The resulting second-order necessary and sufficient optimality conditions are characterized as the solution to a quadratic-constrained optimization problem that incorporates both first- and second-order terms in a unified manner. The study of optimality conditions using parabolic paths has been used across various classes of optimization problems. Note that there is, a priori, no reason that optimality should be verified only along such parabolic paths in general. On the other hand, for a broad class of problems, this approach often leads to verifiable second-order growth conditions in the neighborhood of a local optimal solution. Moreover, the parabolic approach can typically yield no-gap second-order optimality conditions associated with second-order tangent approximations; see, for example Liu et al. [18], Rückmann and Shapiro [20], Bonnans and Shapiro [24], Ben-Tal [25].

This paper is organized as follows: Section 2 presents preliminary concepts and notations. Section 3 investigates the existence of the second-order directional differentiability of the optimal solution function for the lower-level problem. Section 4 explores some results to calculate the second-order directional derivatives of the optimal solution function. Finally, Section 5 derives second-order optimality conditions for the bilevel problem BOP, based on the framework developed in Section 4.

2 Preliminaries

2.1 Basic notations and assumptions

The following notations are used throughout this paper.

- $\nabla h := \begin{pmatrix} \nabla h_1^T & \nabla h_2^T & \cdots & \nabla h_q^T \end{pmatrix}^T \in \mathbb{R}^{q \times p}$ is the Jacobian of a function $h : \mathbb{R}^p \to \mathbb{R}^q$, where each $\nabla h_i \in \mathbb{R}^{1\times p}$ is the Jacobian of a function $h_i : \mathbb{R}^p \to \mathbb{R}$.
- $\nabla^2 h := \begin{pmatrix} \nabla^2 h_1^T & \nabla^2 h_2^T & \cdots & \nabla^2 h_q^T \end{pmatrix}^T \in \mathbb{R}^{q\cdot p} \times \mathbb{R}^p$ represents the Hessian of h, where each $\nabla^2 h_i \in \mathbb{R}^{p\times p}$.
- $d^T \nabla^2 h d := \begin{pmatrix} d^T\nabla^2 h_1 d & d^T \nabla^2 h_2 d & \cdots & d^T\nabla^2 h_q d\end{pmatrix}^T \in \mathbb{R}^q$ for any $d \in \mathbb{R}^p$.
- $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$.
- $\Lambda_\beta := \{1, 2, \cdots, \beta\}$ denotes the index set of size β.
- $\widehat{\nabla} := \begin{pmatrix} \nabla_x & \nabla_y & \nabla_\mu \end{pmatrix}$ and $\widehat{\nabla}^2 := \begin{pmatrix} \nabla^2_{xx} & \nabla^2_{xy} & \nabla^2_{x\mu} \\ \nabla^2_{yx} & \nabla^2_{yy} & \nabla^2_{y\mu} \\ \nabla^2_{\mu x} & \nabla^2_{\mu y} & \nabla^2_{\mu\mu} \end{pmatrix}$ denote Jacobian and Hessian with respect to (x, y, μ) respectively.
- Denote $z := \begin{pmatrix} x \\ y \end{pmatrix}$, $d_z := \begin{pmatrix} d_x \\ d_y \end{pmatrix}$, and $d_z^1 := \begin{pmatrix} d_x^1 \\ d_y^1 \end{pmatrix}$.
- $o(t) : \mathbb{R}^+ \to \mathbb{R}$ is a function such that $\frac{o(t)}{t} \to 0$ as $t \to 0$.
- $L(x, y, \mu) := f(x, y) + \mu^T g(x, y)$ is the Lagrangian function associated with LP_y, where $\mu \in \mathbb{R}^q$ is the KKT multiplier vector.
- $x(y)$ denotes the optimal solution function of LP_y.
- $I_g(y) := \{i \in \Lambda_q \ : \ g_i(x(y), y) = 0\}$ is the active set of the constraint g at the feasible point $x(y)$.
- The set of KKT multiplier vectors of LP_y at y is denoted as

$$M(y) := \{\mu \in \mathbb{R}^q \ : \ \nabla_x L(x(y), y, \mu) = 0, \ \mu^T g(x(y), y) = 0, \ \mu \geq 0\}.$$

- The vertex set of $M(y)$ is denoted by $EM(y)$.
- $J(\mu) := \{i \in \Lambda_q : \mu_i > 0\}$.

To study the directional differentiability of the optimal solution function $x(\cdot)$ at y, the following assumptions are considered at point $(x(y), y) \in \mathbb{R}^n \times \mathbb{R}^m$.

($\mathcal{H}$**1**) The functions $f, g \in C^3$ around $(x(y), y)$.

To ensure that $M(y)$ is a nonempty polyhedral set, Mangasarian–Fromovitz constraint qualification (MFCQ) on the problem LP_y is considered. MFCQ for LP_y is satisfied if there exists vector $d_x \in \mathbb{R}^n$ such that $\nabla_x g_i(x(y), y)d_x < 0, \ \forall i \in I_g(y)$.

($\mathcal{H}$**2**) MFCQ holds at $(x(y), y)$ for LP_y.

The KKT optimality conditions of LP_y are sufficient if the second-order optimality conditions for LP_y are satisfied. We assume the strong sufficient optimality condition.

($\mathcal{H}$**3**) For all $\mu \in M(y)$, the following condition is satisfied at $(x(y), y) \in \mathbb{R}^n \times \mathbb{R}^m$.

$$d_x^T \nabla^2_{xx} L(x(y), y, \mu) d_x > 0,$$

for any nonzero $d_x \in \mathbb{R}^n$ satisfying $\nabla_x g_i(x(y), y)d_x = 0, \ i \in J(\mu)$.

A feasible point $x(y)$ is a strict local minimum of LP_y if there exists a neighborhood $N_{x(y)}$ of $x(y)$ such that $f(x(y), y) < f(x', y)$ for all $x' \in N_{x(y)}$ satisfying $g(x', y) \leq 0$. Under Assumptions

$\mathcal{H}1-\mathcal{H}3$, $x(y)$ is a strict local minimum of LP_y.

($\mathcal{H}$**4**) Constant Rank Constraint Qualification ($CRCQ$) holds for LP_y if there exists a neighborhood $N_{(x(y),y)}$ of the point $(x(y),y)$ such that, for any subset $I \subseteq I_g(y)$, the family of gradient vectors $(\nabla_x g_i(x',y'))_{i\in I}$ has a constant rank for all $(x',y') \in N_{(x(y),y)}$.

Additionally, Assumptions $\mathcal{H}1$–$\mathcal{H}3$, along with convexity assumption on LP_y, ensure that $x(\cdot)$ is the unique strict global minimizer of LP_y.

($\mathcal{H}$**5**) LP_y is a convex programming problem for a fixed parameter y.

Assumption $\mathcal{H}2$ allows the presence of non-unique KKT multiplier vectors for LP_y. This leads to an analysis that differs from classical approaches based on LICQ and strict complementarity; see, for example, Falk and Liu [15], Fiacco [26]. In Section 4, we consider suitable selections of multipliers that enable the computation of second-order directional derivatives.

Remark 2.1 Assumption $\mathcal{H}3$ is stronger than the standard second-order sufficient optimality condition, as it requires the Lagrangian function to be positive definite on the critical subspace of the feasible set for each multiplier vector. The framework ($\mathcal{H}1-\mathcal{H}5$) employs two regularity constraint qualifications for LP_y, which are MFCQ and CRCQ. By providing counterexamples, Janin [27] showed that CRCQ is neither stronger nor weaker than MFCQ; that is, neither does MFCQ imply CRCQ, nor does CRCQ imply MFCQ. Clearly, both constraint qualifications are weaker than the linear independent constraint qualification. Shapiro [28] shows that the solution function $x(\cdot)$ is not necessarily Lipschitz continuous (see the example on page 642). This example satisfies the Assumptions $\mathcal{H}1$–$\mathcal{H}3$. The additional Assumption $\mathcal{H}4$ is needed to ensure the Lipschitz continuity of the solution function $x(\cdot)$. In fact, under this framework, the stronger property that the solution function $x(\cdot)$ is piecewise smooth is established by Ralph and Dempe [22] and Liu [23]. The Assumptions $\mathcal{H}3$ and $\mathcal{H}5$ can be omitted if strong convexity is assumed for LP_y.

2.2 Prerequisites

The following discussions are borrowed from Ralph and Dempe [22], which are used in further developments. A function $\chi : \mathbb{R}^m \to \mathbb{R}^n$ is first-order directionally differentiable along the given direction $d_y \in \mathbb{R}^m$ at y if and only if $\lim\limits_{\alpha\downarrow 0} \frac{\chi(y+\alpha d_y)-\chi(y)}{\alpha}$ exists. The first-order directional derivative of χ along the direction d_y at y is denoted by $\chi'(y;d_y)$ and computed as $\chi'(y;d_y) = \lim\limits_{\alpha\downarrow 0} \frac{\chi(y+\alpha d_y)-\chi(y)}{\alpha}$. Further, if χ is also locally Lipschitz continuous, then χ is said to be Bouligand differentiable.
$x(\cdot)$ is a Bouligand-differentiable function under Assumptions $\mathcal{H}1-\mathcal{H}4$ as indicated by Ralph and Dempe [22]. Consider the following set at y along the direction $d_y \in \mathbb{R}^m$ for a fixed $\mu \in M(y)$ as

$$K^1_\mu(y;d_y) := \left\{ d_x \in \mathbb{R}^n \ : \begin{array}{ll} \nabla_x g_i(x(y),y)d_x + \nabla_y g_i(x(y),y)d_y = 0, & i \in J(\mu), \\ \nabla_x g_i(x(y),y)d_x + \nabla_y g_i(x(y),y)d_y \le 0, & i \in I_g(y) \setminus J(\mu) \end{array} \right\},$$

and a subset of $M(y)$ as

$$S^1(d_y) := \operatorname*{argmax}_{\mu\in M(y)} \sum_{i\in I_g(y)} \mu_i \nabla_y g_i(x(y),y)d_y.$$

Theorem 2.1 *(i) (Shapiro [28], Dempe [29]) Under Assumptions $\mathcal{H}1-\mathcal{H}3$, there exist neighborhoods N_{x^*} of $x^* \in \mathbb{R}^n$ and N_{y^*} of $y^* \in \mathbb{R}^m$ such that $x(\cdot) : N_{y^*} \to N_{x^*}$ is first-order directionally differentiable. For a given vector d_y, there exists $\mu \in M(y)$ such that $x'(y;d_y)$ solves the following convex quadratic program uniquely.*

$$\begin{aligned} QP_\mu(y;d_y): \quad \min_{d_x} \quad & \frac{1}{2} d_x^T \nabla^2_{xx} L(x(y),y,\mu) d_x + d_x^T \nabla^2_{yx} L(x(y),y,\mu) d_y \\ s.t. \quad & d_x \in K^1_\mu(y;d_y). \end{aligned}$$

(ii) (Theorem 2, Ralph and Dempe [22]) Under Assumptions $\mathcal{H}1 - \mathcal{H}4$, the directional derivative $x'(y;\cdot)$ is a piecewise linear function such that, for each $\mu \in S^1(d_y)$, $x'(y;d_y)$ solves $QP_\mu(y;d_y)$ uniquely.

The constraint set $K^1_\mu(y;d_y)$ of $QP_\mu(y;d_y)$ is nonempty if and only if $\mu \in S^1(d_y)$ (see Lemma 2.2 of Dempe [29]). Moreover, $QP_\mu(y;d_y)$ is a convex quadratic programming problem over the feasible space at y. In Theorem 2.1 (i), observe that $QP_\mu(y;d_y)$ has a unique solution $x'(y;d_y)$ at given y for some $\mu \in S^1(d_y)$. Hence, the theorem does not compute $x'(y;d_y)$ in a constructive manner, as it remains unclear which $\mu \in S^1(d_y)$ should be used to compute $x'(y;d_y)$ from the quadratic problem $QP_\mu(y;d_y)$. Under the additional Assumption $\mathcal{H}4$, Ralph and Dempe [22] demonstrated that, for any $\mu \in S^1(d_y)$, the solution of $QP_\mu(y;d_y)$ yields $x'(y;d_y)$.

The following theorem is due to Theorem 5.5 of Dempe [2], which we state here in the context of BOP.

Theorem 2.2 *(First-order sufficient optimality condition) Suppose Assumptions $\mathcal{H}1 - \mathcal{H}5$ are satisfied at $(x(y^*), y^*)$. If the optimal value of the optimization problem*

$$\begin{aligned} \min_{d_y} \quad & \nabla_x F(x(y^*),y^*)x'(y^*;d_y) + \nabla_y F(x(y^*),y^*)d_y \\ s.t. \quad & \nabla_x G_i(x(y^*),y^*)x'(y^*;d_y) + \nabla_y G_i(x(y^*),y^*)d_y \le 0, \ \ i \in I_G(y^*), \end{aligned}$$

where $I_G(y^*) = \{i \in \Lambda_p \ : \ G_i(x(y^*),y^*) = 0\}$, is positive then $(x(y^*),y^*)$ is a strict local minimum point of BOP.

3 Existence of second-order directional derivatives of optimal solution function

This section focuses on the second-order properties of the optimal solution function $x(\cdot)$ of LP_y. The second-order directional derivative of a general nonsmooth function is considered along a parabolic path in Ben-Tal and Zowe [30]. In this section, we prove the existence of the second-order directional derivative of the optimal solution function $x(\cdot)$ of BOP in the parabolic sense under some reasonable assumptions. Let $\chi'(y;d_y)$ be the first-order directional derivative of the function $\chi : \mathbb{R}^m \to \mathbb{R}^n$ at y along the direction d_y. The function χ is said to be second-order directionally differentiable in the sense of Ben-Tal and Zowe [30] if $\lim\limits_{\alpha\downarrow 0} \frac{\chi(y+\alpha d_y+\alpha^2 d^1_y)-\chi(y)-\alpha\chi'(y;d_y)}{\alpha^2}$ exists, where $d_y, d^1_y \in \mathbb{R}^m$ are direction vectors. The parabolic second-order directional derivative of χ at y along a parabolic path $y+\alpha d_y+\alpha^2 d^1_y$ is denoted by the vector $\chi''(y;d_y,d^1_y)$, and computed as

$$\chi''(y;d_y,d^1_y) := \lim_{\alpha\downarrow 0} \frac{\chi(y+\alpha d_y+\alpha^2 d^1_y)-\chi(y)-\alpha\chi'(y;d_y)}{\alpha^2}.$$

If $\chi''(y;d_y,d^1_y)$ exists for all d_y and d^1_y, then we say that χ is second-order directionally differentiable at y. By this, we mean that it is second-order directionally differentiable along parabolic paths in the sense of Ben-Tal and Zowe.
Furthermore, using the definitions of $\chi'(y;d_y)$ and $\chi''(y;d_y,d^1_y)$, it can be deduced that $\chi'(y;\cdot)$ and $\chi''(y;\cdot,\cdot)$ are positively homogeneous of degree one and two, respectively. That is, for given $t>0$, we have

$$\chi'(y;td_y) = \lim_{\alpha\downarrow 0} \frac{\chi(y+\alpha t d_y)-\chi(y)}{\alpha} = t\lim_{\alpha\downarrow 0} \frac{\chi(y+\alpha t d_y)-\chi(y)}{t\alpha} = t\lim_{\beta\downarrow 0} \frac{\chi(y+\beta d_y)-\chi(y)}{\beta} = t\chi'(y;d_y), \tag{1}$$

where $\beta = t\alpha$. Similarly, using the definition $\chi''(y;d_y,d^1_y)$, we obtain, for $t>0$,

$$\chi''(y;td_y,t^2d^1_y) = \lim_{\alpha\downarrow 0} \frac{\chi(y+\alpha t d_y+\alpha^2t^2d^1_y)-\chi(y)-\alpha\chi'(y;td_y)}{\alpha^2}$$

$$= \lim_{\alpha \downarrow 0} \frac{\chi(y + \alpha t d_y + \alpha^2 t^2 d_y^1) - \chi(y) - \alpha t \chi'(y; d_y)}{\alpha^2}.$$

Let $\beta = t\alpha$. Substituting this into the above expression, we obtain

$$\chi''(y; td_y, t^2 d_y^1) = t^2 \lim_{\beta \downarrow 0} \frac{\chi\left(y + \beta d_y + \beta^2 d_y^1\right) - \chi(y) - \beta \chi'(y; d_y)}{\beta^2} = t^2 \chi''(y; d_y, d_y^1). \tag{2}$$

To analyze the local behavior of the parametric problem LP_y, an equality-constrained relaxed problem is considered by retaining the constraints corresponding to a subset of the active index set $I_g(y^*)$ at y^*, while the remaining constraints are omitted. Consequently, a solution of the relaxed problem may not necessarily be a solution of the original problem. However, every solution of the original problem corresponds to the solution set of the relaxed problem.

In the following theorem, we show that the optimal solution function to the relaxed problem RLP_y is second-order continuously differentiable.

Theorem 3.1 *Let $(x^*, y^*)(= (x(y^*), y^*))$ and $\mu^* \in EM(y^*)$. Suppose Assumptions $\mathcal{H}1 - \mathcal{H}3$ hold at (x^*, y^*) and the family $(\nabla_x g_i(x^*, y^*))_{i \in I}$ is linearly independent for some I satisfying $J(\mu^*) \subseteq I \subseteq I_g(y^*)$. Then there exist neighborhoods N_{x^*} of x^* and N_{y^*} of y^* such that the optimal solution function $x_I(\cdot) : N_{y^*} \to N_{x^*}$ of the following relaxed problem of LP_y,*

$$(RLP_y): \quad \min_x f(x, y) \text{ s.t. } g_i(x, y) = 0, \ i \in I$$

is second-order continuously differentiable in a neighborhood of y^*.

Proof Assumption $\mathcal{H}2$ implies that $M(y^*)$ is a nonempty polytope. Therefore, its vertex set $EM(y^*)$ is nonempty. Moreover, Assumption $\mathcal{H}3$ ensures that

$$d_x^\top \nabla_{xx}^2 L(x^*, y^*, \mu^*) d_x > 0 \quad \text{for all } d_x \neq 0 \text{ satisfying } \nabla_x g_i(x^*, y^*) d_x = 0, \ i \in I. \tag{3}$$

Under the linear independence condition and the strong second-order sufficient condition for the relaxed problem (RLP_{y^*}), standard results from sensitivity analysis (see Corollary 3.2.5 of Fiacco [26]) imply that the solution function $x_I(\cdot)$ is second-order continuously differentiable in a neighborhood of y^*. □

The above theorem justifies that the optimal solution function of RLP_y is second-order continuously differentiable. However, $x(\cdot)$ is a piecewise smooth function (see Ralph and Dempe [22]), which is not necessarily a continuously differentiable function. The next theorem proves that the optimal solution function of LP_y is second-order directionally differentiable. This result follows from Theorem 2.3 in Liu [23], which concerns piecewise smooth functions. A variant of proof is also provided in Liu et al. [18].

Theorem 3.2 *Let Assumptions $\mathcal{H}1 - \mathcal{H}5$ hold at $(x^*, y^*)(= (x(y^*), y^*))$. Then, there are neighborhoods N_{x^*} of x^* and N_{y^*} of y^* such that the optimal solution function $x(\cdot) : N_{y^*} \to N_{x^*}$ of LP_y is second-order directionally differentiable at y^*.*

Proof Since Assumptions $\mathcal{H}1$–$\mathcal{H}5$ hold, from Theorem 4.10 of Dempe [2], one can conclude that the global optimal solution function $x(\cdot)$ of LP_y is a piecewise smooth function at y^*. Hence, $x(\cdot)$ is a continuous function, and there exists a finite collection of smooth functions $\{x_1(\cdot), x_2(\cdot), \cdots, x_N(\cdot)\}$ such that $x(y) \in \{x_1(y), x_2(y), \cdots, x_N(y)\}$ around y^*, where $x_j(y)$ is the solution of a particular form of RLP_y, which is given by

$$\min_x \quad f(x, y) \text{ s.t. } g_i(x, y) = 0, \ i \in I_j,$$

corresponding to the index set I_j satisfying $J(\mu) \subseteq I_j \subseteq I_g(y^*)$ for some $\mu \in EM(y^*)$, such that the family of gradients $(\nabla_x g_i(x^*, y^*))_{i \in I_j}$ is linearly independent.

By Theorem 3.1, one can conclude that there are neighborhoods N_{x^*} of x^* and N_{y^*} of y^* such that $x_j(.) : N_{y^*} \to N_{x^*}$ is second-order continuously differentiable for each $j = 1, 2, .., N$ at y^*.
Let d_y and d_y^1 be direction vectors at y^*, and denote

$$\mathscr{I}(d_y, d_y^1) := \left\{ j \ : \begin{array}{l} j \in \{1, 2, \cdots, N\} \text{ such that there exists a sequence of positive real numbers} \\ \{\alpha_k\} \downarrow 0 \text{ such that } x(y^* + \alpha_k d_y + \alpha_k^2 d_y^1) = x_j(y^* + \alpha_k d_y + \alpha_k^2 d_y^1) \end{array} \right\}.$$

Clearly, $\mathscr{I}(d_y, d_y^1) \neq \emptyset$. If $x(\cdot)$ is second-order directionally differentiable at x^* then the sequence $\left\{\frac{x(y^*+\alpha_k d_y+\alpha_k^2 d_y^1)-x(y^*)-\alpha_k x'(y^*;d_y)}{\alpha_k^2}\right\}$ must admit unique accumulation point for every sequence $\{\alpha_k\}$ of positive real numbers converging to 0. Hence, to prove the theorem, it suffices to show that

$$x_r''(y^*; d_y, d_y^1) = x_s''(y^*; d_y, d_y^1) \text{ for any } r, s \in \mathscr{I}(d_y, d_y^1). \tag{4}$$

The above considerations yield the existence of $r' > 0$ such that for each $t \in [0, r']$, there is some $j \in \{1, 2, \cdots, N\}$ such that $x(y^* + td_y + t^2 d_y^1) = x_j(y^* + td_y + t^2 d_y^1)$. Now, consider the following set for some $j \in \mathscr{I}(d_y, d_y^1)$ as

$$R_j := \{t \in [0, r'] : \; x(y^* + td_y + t^2 d_y^1) = x_j(y^* + td_y + t^2 d_y^1)\}.$$

Note that R_j is a closed set since $x(y)$ and $x_j(y)$ are continuous at y^* under $\mathcal{H}1 - \mathcal{H}3$ (see Theorem 7.2, Kojima [31]).

For $j_1, j_2 \in \mathscr{I}(d_y, d_y^1)$, define a relation $j_1 \sim j_2$ if there exists a sequence of positive real numbers $\{\alpha_k\}$ such that $\{\alpha_k\} \downarrow 0$ and $\{\alpha_k\} \subseteq R_{j_1} \cap R_{j_2}$. Let $j_1, j_2 \in \mathscr{I}(d_y, d_y^1)$ and $j_1 \sim j_2$. Then we have,

$$x_{j_1}(y^* + \alpha_k d_y + \alpha_k^2 d_y^1) = x_{j_1}(y^*) + \alpha_k x_{j_1}'(y^*; d_y) + \alpha_k^2 x_{j_1}''(y^*; d_y, d_y^1) + o(\alpha_k^2), \;\forall k$$

as x^{j_1} is second-order directionally differentiable at y^*.

Since $j_1 \in \mathscr{I}(d_y, d_y^1)$,

$$x(y^* + \alpha_k d_y + \alpha_k^2 d_y^1) = x(y^*) + \alpha_k x_{j_1}'(y^*; d_y) + \alpha_k^2 x_{j_1}''(y^*; d_y, d_y^1) + o(\alpha_k^2), \;\forall k.$$

Similarly,

$$x(y^* + \alpha_k d_y + \alpha_k^2 d_y^1) = x(y^*) + \alpha_k x_{j_2}'(y^*; d_y) + \alpha_k^2 x_{j_2}''(y^*; d_y, d_y^1) + o(\alpha_k^2), \;\forall k.$$

From the above two expressions,

$$\lim_{k\to\infty} \frac{x(y^* + \alpha_k d_y + \alpha_k^2 d_y^1) - x(y^*)}{\alpha_k} = x_{j_1}'(y^*; d_y) = x_{j_2}'(y^*; d_y), \tag{5}$$

and

$$\lim_{k\to\infty} \frac{x(y^* + \alpha_k d_y + \alpha_k^2 d_y^1) - x(y^*) - \alpha_k x_{j_1}'(y^*, d_y)}{\alpha_k^2} = x_{j_1}''(y^*; d_y, d_y^1) = x_{j_2}''(y^*; d_y, d_y^1). \tag{6}$$

Hence, from the above discussion, it is clear that for any $j_1, j_2 \in \mathscr{I}(d_y, d_y^1)$ satisfying $j_1 \sim j_2$, the expressions (5)-(6) hold. Since $[0, r'] = \bigcup\limits_{\mathscr{I}(d_y, d_y^1)} R_j$, we claim that for any $r, s \in \mathscr{I}(d_y, d_y^1)$, one can find a chain of relations such that $r \sim j_1 \sim j_2, \cdots, \sim j_t \sim s$. On the contrary, suppose that no such chain exists for given $r, s \in \mathscr{I}(d_y, d_y^1)$. Denote $r = r_0$ and $s = s_0$. Then, without loss of generality, the index set $\mathscr{I}(d_y, d_y^1)$ can be divided into two disjoint subsets such that there exist two separate chains

$$r_0 \sim r_1 \sim \cdots \sim r_k \quad \text{and} \quad s_0 \sim s_1 \sim \cdots \sim s_\ell,$$

which satisfy

$$r_i \not\sim s_j \quad \text{for all } i = 0, \ldots, k, \; j = 0, \ldots, \ell.$$

Denote

$$R^r := \bigcup_{i=0}^{k} R_{r_i}, \qquad R^s := \bigcup_{j=0}^{\ell} R_{s_j},$$

which are closed subsets of $[0, r']$ being finite unions of closed sets. Moreover, since $r_i \not\sim s_j$ for all i, j, there does not exist any sequence $\{\alpha_k\} \downarrow 0$ such that $\{\alpha_k\} \subseteq R^r \cap R^s$. Hence, for some $0 < r'' \leq r'$

$$R^r \cap R^s \cap [0, r''] = \emptyset.$$

On the other hand, since $[0, r'] = \bigcup\limits_{\mathscr{I}(d_y, d_y^1)} R_j = R^r \cup R^s$, we have

$$[0, r''] = \left(R^r \cap [0, r'']\right) \cup \left(R^s \cap [0, r'']\right),$$

where both R^r and R^s are nonempty. Therefore, $[0, r'']$ is expressed as the union of two disjoint, nonempty, closed sets, which contradicts the connectedness of $[0, r'']$. Hence, there exists a chain of relation $r \sim j_1 \sim j_2, \cdots, \sim j_t \sim s$ for any $r, s \in \mathscr{I}(d_y, d_y^1)$. Then, (4) holds. This completes the proof of the theorem. □

The epi-regularity condition on the second-order directional differentiable function along parabolic path is used in Mehlitz and Zemkoho [19], Rückmann and Shapiro [20] to obtain sufficient optimality conditions. In this work, we require a stronger condition, which is termed as *second-order gph-regularity condition* on the solution function, to derive the sufficient optimality condition for BOP.

Definition 3.1 (Liu et al. [18]) A locally Lipschitz continuous and second-order directionally differentiable function χ is said to be *second-order gph-regular* if, for any $d_y \in \mathbb{R}^n$ and any path $d_y^1(\cdot) : \mathbb{R}_+ \to \mathbb{R}^n$ satisfying $\lim_{t\downarrow 0} td_y^1(t) = 0$, the following expansion holds

$$\chi(y^* + td_y + t^2 d_y^1(t)) = \chi(y^*) + t\chi'(y^*; d_y) + t^2\chi''(y^*; d_y, d_y^1(t)) + o(t^2),$$

where $\mathbb{R}_+$ denotes the set of nonnegative real numbers.

Hence, for given path $d_y^1(\cdot)$ satisfying $\lim_{t\downarrow 0} td_y^1(t) = 0$, χ is second-order gph-regular if and only if

$$\lim_{t\downarrow 0} \frac{\chi(y^* + td_y + t^2 d_y^1(t)) - \chi(y^*) - t\chi'(y^*; d_y) - \chi''(y^*; d_y, d_y^1(t))}{t^2} = 0.$$

Clearly, if χ is second-order continuously differentiable, then it is second-order gph-regular. The following theorem demonstrates that the solution function $x(\cdot)$ is second-order gph-regular.

Theorem 3.3 *Let the assumptions of Theorem 3.2 hold. Then, there are neighborhoods N_{x^*} of x^* and N_{y^*} of y^* such that the optimal solution function $x(\cdot) : N_{y^*} \to N_{x^*}$ of LP_y is second-order gph-regular at y^*.*

Proof Since $x(\cdot)$ is a piecewise smooth, it is locally Lipschitz continuous (see Ralph and Dempe [22]). Consider a path $y^* + td_y + t^2 d_y^1(t)$ satisfying $\lim_{t\downarrow 0} td_y^1(t) = 0$. To prove $x(\cdot)$ is second-order gph-regular, we claim that the sequence

$$\left\{ \frac{x(y^* + \alpha_k d_y + \alpha_k^2 d_y^1(\alpha_k)) - x(y^*) - \alpha_k x'(y^*; d_y) - \alpha_k^2 x''(y^*; d_y, d_y^1(\alpha_k))}{\alpha_k^2} \right\} \to 0 \text{ for each } \{\alpha_k\} \downarrow 0.$$

Since $x(\cdot)$ is piecewise smooth, one can find a collection of functions $x_j(\cdot) \in C^2$ as in Theorem 3.2 along the path $y^* + td_y + t^2 d_y^1(t)$ for some index set

$$\mathscr{I} := \left\{ j \ : \begin{array}{l} j \in \{1, 2, \cdots, N\} \text{ such that there exists a sequence of positive real numbers} \\ \{\alpha_k\} \downarrow 0 \text{ such that } x(y^* + \alpha_k d_y + \alpha_k^2 d_y^1(\alpha_k)) = x_j(y^* + \alpha_k d_y + \alpha_k^2 d_y^1(\alpha_k)) \end{array} \right\}.$$

Since $x_j(\cdot) \in C^2$, it is second-order gph-regular at y^* for all $j \in \mathscr{I}$. Then we have

$$x_j(y^* + \alpha_k d_y + \alpha_k^2 d_y^1(\alpha_k)) = x_j(y^*) + \alpha_k x_j'(y^*; d_y) + \alpha_k^2 x_j''(y^*; d_y, d_y^1(\alpha_k)) + o(\alpha_k^2), \quad \forall k, \tag{7}$$

where

$$x_j''(y^*; d_y, d_y^1(\alpha_k)) = \lim_{s\downarrow 0} \frac{x_j(y^* + sd_y + s^2 d_y^1(\alpha_k)) - x_j(y^*) - sx_j'(y^*; d_y)}{s^2}.$$

Let $\{\alpha_k\} \downarrow 0$ be any given sequence such that $\{\alpha_k\} \subseteq R_j := \{t \in [0, r'] : \ x(y^* + td_y + t^2 d_y^1(t)) = x_j(y^* + td_y + t^2 d_y^1(t))\}$ for some suitable $r' > 0$. Then, there exists $j \in \mathscr{I}$,

$$\begin{aligned}
&\lim_{k\to\infty} \frac{x(y^* + \alpha_k d_y + \alpha_k^2 d_y^1(\alpha_k)) - x(y^*) - \alpha_k x'(y^*; d_y) - \alpha_k^2 x''(y^*; d_y, d_y^1(\alpha_k))}{\alpha_k^2} \\
&= \lim_{k\to\infty} \frac{x_j(y^* + \alpha_k d_y + \alpha_k^2 d_y^1(\alpha_k)) - x_j(y^*) - \alpha_k x_j'(y^*; d_y) - \alpha_k^2 x''(y^*; d_y, d_y^1(\alpha_k))}{\alpha_k^2} \\
&= \lim_{k\to\infty} x_j''(y^*; d_y, d_y^1(\alpha_k)) - x''(y^*; d_y, d_y^1(\alpha_k)) + \lim_{k\to\infty} \frac{o(\alpha_k)^2}{\alpha_k^2} \quad (\text{using}(7)) \\
&= \lim_{k\to\infty} x_j''(y^*; d_y, d_y^1(\alpha_k)) - x''(y^*; d_y, d_y^1(\alpha_k)).
\end{aligned} \tag{8}$$

Using the arguments in Theorem 3.2, for any $j_1, j_2 \in \mathscr{I}$ satisfying $j_1 \sim j_2$, there exists a common sequence $\{\bar{\alpha}_k\} \downarrow 0$ such that

$$\left. \begin{array}{l} x(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)) = x_{j_1}(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)) = x_{j_2}(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)), \\ x'(y^*; d_x) = x_{j_1}'(y^*; d_x) = x_{j_2}'(y^*; d_x) = \lim_{k\to\infty} \dfrac{x(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)) - x(y^*)}{\bar{\alpha}_k}. \end{array} \right\} \tag{9}$$

Using the second-order expansion for j_1 and j_2, we obtain

$$x_{j_1}(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)) = x_{j_1}(y^*) + \bar{\alpha}_k x_{j_1}'(y^*; d_y) + \bar{\alpha}_k^2 x_{j_1}''(y^*; d_y, d_y^1(\bar{\alpha}_k)) + o(\bar{\alpha}_k^2),$$

$$x_{j_2}(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)) = x_{j_2}(y^*) + \bar{\alpha}_k x'_{j_2}(y^*; d_y) + \bar{\alpha}_k^2 x''_{j_2}(y^*; d_y, d_y^1(\bar{\alpha}_k)) + o(\bar{\alpha}_k^2),\ \forall k.$$

Subtracting the second expression from the first expression, and using (9),

$$\lim_{k\to\infty} x''_{j_2}(y^*; d_y, d_y^1(\bar{\alpha}_k)) - x''_{j_1}(y^*; d_y, d_y^1(\bar{\alpha}_k)) = 0.$$

Using the second-order expansion for $x_j(\cdot)$, we have

$$x''_j(y^*; d_y, d_y^1(t)) = \nabla_y x_j(y^*) d_y^1(t) + 0.5 d_y^T \nabla^2_{yy} x_j(y^*) d_y.$$

From the above expression, it is clear that $x''_j(y^*; d_y, \cdot)$ is continuous in the last argument. Further, since $d_y^1(t)$ is continuous with respect to t, it follows that $x''_j(y^*; d_y, d_y^1(t))$ is also continuous with respect to t. Hence, we have

$$\lim_{t\downarrow 0} x''_{j_2}(y^*; d_y, d_y^1(t)) - x''_{j_1}(y^*; d_y, d_y^1(t)) = 0. \tag{10}$$

Since (10) holds for all $j_1, j_2 \in \mathscr{I}$, and in view of (8), to prove the theorem, it suffices to establish the existence of an index $j \in \mathscr{I}$ such that

$$\lim_{k\to\infty} x''_j(y^*; d_y, d_y^1(\alpha_k)) - x''(y^*; d_y, d_y^1(\alpha_k)) = 0.$$

For the sequence $\{\alpha_k\}$, consider a path $y^* + s d_y + s^2 d_y^1(\alpha_k))$ for $s \geq 0$ and fixed α_k, and the corresponding index set for sufficiently large k as

$$\mathscr{I}^k := \left\{ j \ : \begin{array}{l} j \in \{1, 2, \cdots, N\} \text{ such that there exists a sequence of positive real numbers} \\ \{s_l\} \downarrow 0 \text{ with } s_l \leq \alpha_k \text{ such that } x(y^* + s_l d_y + s_l^2 d_y^1(\alpha_k)) = x_j(y^* + s_l d_y + s_l^2 d_y^1(\alpha_k)) \end{array} \right\}.$$

Clearly, $\mathscr{I}^k \neq \emptyset$ for all k because $x(\cdot)$ is piecewise smooth function. Since $\mathscr{I}^k$ is finite for each k, there exists $N > 0$ sufficiently large so that $j \in \bigcap_{k \geq N} \mathscr{I}^k$, implying that $j \in \mathscr{I}$.
As $x(\cdot)$ is second-order directionally differentiable,

$$x(y^* + s_l d_y + s_l^2 d_y^1(\alpha_k)) = x(y^*) + s_l x'(y^*; d_y) + s_l^2 x''(y^*; d_y, d_y^1(\alpha_k)) + o(s_l^2),\ \forall l.$$

Then

$$\lim_{k\to\infty} \left(\lim_{l\to\infty} \frac{x(y^* + s_l d_y + s_l^2 d_y^1(\alpha_k)) - x(y^*) - s_l x'(y^*; d_y) - s_l^2 x''(y^*; d_y, d_y^1(\alpha_k))}{s_l^2} \right) = 0.$$

Since $j \in \mathscr{I}^k$ for $k \geq N$, we have

$$\lim_{k\to\infty} \left(\lim_{l\to\infty} \frac{x_j(y^* + s_l d_y + s_l^2 d_y^1(\alpha_k)) - x_j(y^*) - s_l x'_j(y^*; d_y) - s_l^2 x''(y^*; d_y, d_y^1(\alpha_k))}{s_l^2} \right) = 0.$$

This gives

$$\lim_{k\to\infty} x''_j(y^*; d_y, d_y^1(\alpha_k)) - x''(y^*; d_y, d_y^1(\alpha_k)) = 0.$$

Hence, the proof of the theorem follows. □

4 Second-order properties of the optimal solution function

Suppose that Assumptions $\mathcal{H}1$–$\mathcal{H}3$ hold at $(x(y), y)$, and let $\mu \in V(d_y, d_y^1) \subseteq M(y)$, where

$$V(d_y, d_y^1) := \left\{ \mu \in M(y) \ : \begin{array}{l} \exists\ \mu^k \in M(y^k) \text{ such that } \{\mu^k\} \to \mu \text{ as } \{\alpha_k\} \downarrow 0,\ \alpha_k > 0, \\ \text{and } y^k = y + \alpha_k d_y + \alpha_k^2 d_y^1,\ \forall k \end{array} \right\}.$$

Let $\{y^k\}$ be the sequence defined by $y^k = y + \alpha_k d_y + \alpha_k^2 d_y^1$ with $\alpha_k > 0$ such that $\{\alpha_k\} \to 0$. As $x(\cdot)$ is continuous around y, we have $\{x(y^k)\} \to x(y)$ as $k \to \infty$. Since $x(\cdot)$ is second-order directionally differentiable at y, as shown in Theorem 3.2, the following limits exist.

$$\lim_{k\to\infty} \frac{x(y^k) - x(y)}{\alpha_k} = \lim_{k\to\infty} \frac{x(y + \alpha_k d_y + \alpha_k^2 d_y^1) - x(y)}{\alpha_k} = d_x \ (say) \ \text{ and}$$

$$\lim_{k\to\infty} \frac{x(y^k) - x(y) - \alpha_k x'(y; d_y)}{\alpha_k^2} = \lim_{k\to\infty} \frac{x(y + \alpha_k d_y + \alpha_k^2 d_y^1) - x(y) - \alpha_k x'(y; d_y)}{\alpha_k^2} = d_x^1 \ (say).$$

The first-order properties associated with KKT multipliers of parametric problems are proved in the monograph by Fiacco [26]. To facilitate the Taylor expansion of the functions in the KKT conditions, we require some second-order result associated with the multiplier set $V(d_y, d_y^1)$, which is stated in the following lemma.

Lemma 4.1 *Suppose Assumptions $\mathcal{H}1-\mathcal{H}4$ hold at $(x(y),y)$ and $\mu \in V(d_y,d_y^1)$. There exists $\mu^k \in M(y^k)$ such that $\lim_{k\to\infty}\frac{\mu^k-\mu}{\alpha_k}=d_\mu$ for some $d_\mu\in\mathbb{R}^q$, and $\lim_{k\to\infty}\frac{\mu^k-\mu-\alpha_k d_\mu}{\alpha_k^2}$ exists.*

Proof The proof follows from Lemma 4.2 of Khatana and Panda [32] by considering $V(d_y,d_y^1)$ in place of $V(d_x)$. □

From Lemma 4.1, $\lim_{k\to\infty}\frac{\mu^k-\mu}{\alpha_k}=d_\mu$ and $\lim_{k\to\infty}\frac{\mu^k-\mu-\alpha_k d_\mu}{\alpha_k^2}=d_\mu^1$ for some sequence $\{\mu^k\}$ and $d_\mu,d_\mu^1\in\mathbb{R}^q$. Hence,

$$x(y^k)=x(y)+\alpha_k d_x+\alpha_k^2 d_x^1+o(\alpha_k^2) \text{ and } \mu^k=\mu+\alpha_k d_\mu+\alpha_k^2 d_\mu^1+o(\alpha_k^2)$$

for some subsequence of $\{\alpha_k\}$. Without loss of generality, assume that the subsequence is $\{\alpha_k\}$. Since $x(y^k)$ solves the problem LP_{y^k} at y^k and $\mu^k\in M(y^k)$,

$$\nabla_x L(x(y^k),y^k,\mu^k)=0, \tag{11a}$$
$$g_i(x(y^k),y^k)=0,\quad \mu_i^k>0,\quad i\in J(\mu^k), \tag{11b}$$
$$g_i(x(y^k),y^k)\le 0,\quad \mu_i^k=0,\quad i\notin J(\mu^k). \tag{11c}$$

Denote $z:=\begin{pmatrix}x\\y\end{pmatrix}$, $d_z:=\begin{pmatrix}d_x\\d_y\end{pmatrix}$, and $d_z^1:=\begin{pmatrix}d_x^1\\d_y^1\end{pmatrix}$. Expanding $\nabla_x L$ around $(x(y),y,\mu)$, we have

$$\begin{aligned}
&\nabla_x L(x(y^k),y^k,\mu^k)^T\\
&\quad=\nabla_x L(x(y),y,\mu)^T+\alpha_k\Big(\nabla^2_{xx}L(x(y),y,\mu)d_x+\nabla^2_{yx}L(x(y),y,\mu)d_y+\nabla^2_{\mu x}L(x(y),y,\mu)d_\mu\Big)\\
&\qquad+\alpha_k^2\left(\frac{1}{2}\left(d_z^T\ d_\mu^T\right)\widehat{\nabla}^2(\nabla_x L(x(y),y,\mu)^T)\begin{pmatrix}d_z\\d_\mu\end{pmatrix}+\nabla^2_{xx}L(x(y),y,\mu)d_x^1\right.\\
&\qquad\quad\left.+\nabla^2_{yx}L(x(y),y,\mu)d_y^1+\nabla^2_{\mu x}L(x(y),y,\mu)d_\mu^1\right)+e_n o(\alpha_k^2),
\end{aligned}\tag{12}$$

where $e_n=(1,1,\dots,1)^T\in\mathbb{R}^n$. Further, expanding $g_i,\ i\in I_g(y)$ around $(x(y),y)$, we have

$$\begin{aligned}
g_i(x(y^k),y^k)=&g_i(x(y),y)+\alpha_k\Big(\nabla_x g_i(x(y),y)d_x+\nabla_y g_i(x(y),y)d_y\Big)\\
&+\alpha_k^2\left(\frac{1}{2}d_z^T\nabla^2_{zz}g_i(x(y),y)d_z+\nabla_z g_i(x(y),y)d_z^1\right)+o(\alpha_k^2).
\end{aligned}\tag{13}$$

Since $\mu\in M(y)$,

$$\nabla_x L(x(y),y,\mu)=0,\ \mu_i\ge 0,\ g_i(x(y),y)=0,\ i\in I_g(y),\ \mu_i=0,\ i\notin I_g(y).$$

From (11a), we have $\nabla_x L(x(y^k),y^k,\mu^k)=0$. Hence, $\lim\limits_{k\to\infty}\frac{\nabla_x L(x(y^k),y^k,\mu^k)-\nabla_x L(x(y),y,\mu)}{\alpha_k}=0$. Then, from (12),

$$\nabla^2_{xx}L(x(y),y,\mu)d_x+\nabla^2_{yx}L(x(y),y,\mu)d_y+\nabla^2_{\mu x}L(x(y),y,\mu)d_\mu=0. \tag{14a}$$

As $J(\mu)\subseteq J(\mu^k)$ for sufficiently large k, (11b) holds for $i\in J(\mu)$. Further, since $g_i(x(y^k),y^k)=0,\ i\in J(\mu)$ by (11b) and $g_i(x(y),y)=0,\ i\in J(\mu)\subseteq I_g(y)$, therefore $\lim\limits_{k\to\infty}\frac{g_i(x(y^k),y^k)-g_i(x(y),y)}{\alpha_k}=0$ for $i\in J(\mu)$. Hence, using (13),

$$\nabla_x g_i(x(y),y)d_x+\nabla_y g_i(x(y),y)d_y=0,\quad i\in J(\mu). \tag{14b}$$

Furthermore, for $i \in I_g(y) \setminus J(\mu)$, it follows from (11c) that $\lim_{k\to\infty} \frac{g_i(x(y^k),y^k)-g_i(x(y),y)}{\alpha_k} \le 0$. Then, from (13),

$$\nabla_x g_i(x(y),y)d_x + \nabla_y g_i(x(y),y)d_y \le 0, \quad i \in I_g(y) \setminus J(\mu). \tag{14c}$$

Since $\mu_i = 0$ for all $i \in I_g(y)\setminus J(\mu)$ and $\mu_i^k \ge 0$ for $i \in \Lambda_q$, we obtain $d_{\mu_i} = \lim_{k\to\infty} \frac{\mu_i^k-\mu_i}{\alpha_k} \ge 0$. Moreover, since $I_g(y^k) \subseteq I_g(y)$ for sufficiently large k, it holds that $\mu_i = 0$ and $\mu_i^k = 0$ for all $i \notin I_g(y)$, implying $d_{\mu_i} = \lim_{k\to\infty} \frac{\mu_i^k-\mu_i}{\alpha_k} = 0$ for $i \notin I_g(y)$.
If $d_{\mu_i} > 0$ for some $i \in I_g(y) \setminus J(\mu)$, then it follows that $\mu_i^k > 0$ for sufficiently large k. Consequently, $i \in J(\mu^k)$, and from (11b), we have $g_i(x(y^k), y^k) = 0$ for sufficiently large k. Thus, using (13), it holds that

$$\nabla_x g_i(x(y),y)d_x + \nabla_y g_i(x(y),y)d_y = 0 \text{ for } i \in \{i \in I_g(y) \setminus J(\mu) \ : \ d_{\mu_i} > 0\}.$$

On the other hand, if $\nabla_x g_i(x(y),y)d_x + \nabla_y g_i(x(y),y)d_y < 0$ for some $i \in I_g(y) \setminus J(\mu)$, then $g_i(x(y^k), y^k) < 0$ for sufficiently large k. Hence, it follows from (11c) that $\mu_i^k = 0$ for sufficiently large k, and therefore $d_{\mu_i} = 0$. Thus,

$$\left.\begin{array}{l} d_{\mu_i}\left(\nabla_x g_i(x(y),y)d_x + \nabla_y g_i(x(y),y)d_y\right) = 0, \\ d_{\mu_i} \ge 0, \ i \in I_g(y) \setminus J(\mu), \qquad d_{\mu_i} = 0, i \notin I_g(y). \end{array}\right\} \tag{14d}$$

It can be verified that the expressions (14a)-(14d) are the KKT optimality conditions of $QP_\mu(y; d_y)$ with KKT multiplier vector d_μ. Hence, this system has a unique solution d_x.

Remark 4.1 Recall the discussion in Section 2.2. The feasible set of $QP_\mu(y; d_y)$ is nonempty if and only if $\mu \in S^1(d_y)$. Further, Theorem 2.1 states that $x'(y; d_y)$ is the unique solution of $QP_\mu(y; d_y)$ for each of $\mu \in S^1(d_y)$. Since (14a)-(14d) are the KKT optimality conditions of $QP_\mu(y; d_y)$ so the system (14a)-(14d) is consistent if and only if $\mu \in S^1(d_y)$. In that case, for given direction vector d_y, if (d_x, d_μ) solves the system (14a)-(14d) then $d_x = x'(y; d_y)$.

We now proceed with a similar argument as in the formulation of (14a)-(14d) with the new index sets $J(\mu; d_\mu)$ and $I_g(y; d_z)$ as

$$J(\mu; d_\mu) := \{i \in I_g(y) \setminus J(\mu) \ : \ d_{\mu_i} > 0\} \cup J(\mu),$$

$$\text{and } I_g(y; d_z) := \{i \in I_g(y) : \nabla_x g_i(x(y),y)d_x + \nabla_y g_i(x(y),y)d_y = 0\},$$

and with the following limits

$$\lim_{k\to\infty} \frac{\nabla_x L(x(y^k), y^k, \mu^k) - \nabla_x L(x(y),y,\mu) - \alpha_k(\nabla^2_{xx}L(x(y),y,\mu)d_x + \nabla^2_{yx}L(x(y),y,\mu)d_y + \nabla^2_{\mu x}L(x(y),y,\mu)d_\mu)}{\alpha_k^2}$$

and

$$\lim_{k\to\infty} \frac{g_i(x(y^k),y^k) - g_i(x(y),y) - \alpha_k(\nabla_x g_i(x(y),y)d_x + \nabla_y g_i(x(y),y)d_y)}{\alpha_k^2}.$$

The following system (15a)-(15d) is derived using (11a)-(13) by considering the above limits and index sets. This is analogous to the process for the derivation of (14a)-(14d). Hence, we omit the detailed derivation.

$$\frac{1}{2}\left(d_z^T \ d_\mu^T\right) \widehat{\nabla}^2(\nabla_x L(x(y),y,\mu)^T) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} + \widehat{\nabla}(\nabla_x L(x(y),y,\mu)^T) \begin{pmatrix} d_z^1 \\ d_\mu^1 \end{pmatrix} = 0, \tag{15a}$$

$$\frac{1}{2} d_z^T \nabla^2_{zz} g_i(x(y),y) d_z + \nabla_z g_i(x(y),y) d_z^1 = 0, \quad i \in J(\mu; d_\mu), \tag{15b}$$

$$\frac{1}{2} d_z^T \nabla^2_{zz} g_i(x(y),y) d_z + \nabla_z g_i(x(y),y) d_z^1 \le 0, \quad i \in I_g(y; d_z) \setminus J(\mu; d_\mu), \tag{15c}$$

$$\left.\begin{array}{l} d^1_{\mu_i}\left(\frac{1}{2}d_z^T\nabla^2_{zz}g_i(x(y),y)d_z + \nabla_z g_i(x(y),y)d_z^1\right) = 0, \\ d^1_{\mu_i} \geq 0, \quad i \in I_g(y;d_z)\setminus J(\mu;d_\mu),\ d^1_{\mu_i} = 0, \quad i \notin I_g(y;d_z). \end{array}\right\} \tag{15d}$$

Consider the following linear programming problem corresponding to the direction vectors d_y and d_y^1,

$$\max_{\mu\in S^1(d_y)} \sum_{i\in I_g(y)} \mu_i \left(\frac{1}{2}\left(x'(y;d_y)^T \ \ d_y^T\right)\nabla^2_{zz}g_i(x(y),y)\begin{pmatrix} x'(y;d_y) \\ d_y\end{pmatrix} + \nabla_y g_i(x(y),y)d_y^1\right) \tag{16}$$

and

$$S^2(d_y,d_y^1) = \{\mu \in S^1(d_y) \ : \ \mu \text{ solves } (16)\}.$$

Clearly $S^2(d_y,d_y^1) \subseteq S^1(d_y)$. Further, $S^2(d_y,d_y^1)$ is a nonempty polytope as $S^1(d_y)$ is a nonempty polytope.

Theorem 4.1 *Suppose Assumptions $\mathcal{H}1-\mathcal{H}5$ are satisfied at the given point $(x(y),y)$. Then,*

1. *$x''(y;d_y,d_y^1)$ solves the following quadratic optimization problem,*

$$\begin{aligned} \min_{d_x^1}\quad & \frac{1}{2}\left(\left(d_z^T \ \ d_\mu^T\right)\widehat{\nabla}^2\left(\nabla_x L(x(y),y,\mu)^T\right)\begin{pmatrix} d_z \\ d_\mu\end{pmatrix}\right)^T d_x^1 + {d_x^1}^T\nabla^2_{yx}L(x(y),y,\mu)d_y^1 + {d_x^1}^T\nabla^2_{xx}L(x(y),y,\mu)d_x^1 \\ \text{s.t.}\quad & \frac{1}{2}d_z^T\nabla^2_{zz}g_i(x(y),y)d_z + \nabla_z g_i(x,y)\begin{pmatrix} d_x^1 \\ d_y^1\end{pmatrix} = 0, \quad i\in J(\mu;d_\mu), \\ & \frac{1}{2}d_z^T\nabla^2_{zz}g_i(x(y),y)d_z + \nabla_z g_i(x,y)\begin{pmatrix} d_x^1 \\ d_y^1\end{pmatrix} \leq 0, \quad i\in I_g(y;d_y)\setminus J(\mu;d_\mu) \end{aligned}$$

   for $\mu \in S^2(d_y,d_y^1)\cap EM(y)$ and $d_\mu \in \overline{EM}_\mu$, where

$$\overline{EM}_\mu = \left\{ d_\mu \ : \ \begin{array}{c} \left(x'(y;d_y),d_\mu\right) \text{ solves } (14a)-(14d) \text{ for given } \mu\in EM(y), \\ \text{such that } (\nabla_x g_i(x(y),y))_{i\in J(\mu)\cup J(d_\mu)} \text{ is linearly independent}\end{array}\right\},$$

$$I_g(y;d_y) := \{i\in I_g(y) : \nabla_y g_i(x(y),y)d_y + \nabla_x g_i(x(y),y)x'(y;d_y) = 0\ \}, \text{ and } d_z = \begin{pmatrix} x'(y;d_y) \\ d_y\end{pmatrix}.$$

2. *Suppose $\mu \in EM(y)$ and the feasible set of the above quadratic programming problem is nonempty, then $\mu \in S^2(d_y,d_y^1)$.*

Proof (1) The first part of the proof follows from Theorem 5.1 of Khatana and Panda [32] by considering the parabolic path $y^k = y + \alpha_k d_y + \alpha_k^2 d_y^1$ in place of linear path, and changing the role of $S^2(d_x)$ by $S^2(d_y,d_y^1)$, and $\overline{EM}_{(u,v)}$ by $\overline{EM}_\mu$ respectively.

(2) The feasible set $M(y)$ is a nonempty polytope under Assumption $\mathcal{H}2$. Recall that $S^1(d_y)$ denotes the solution set to the following problem:

$$\begin{aligned} (P^1):\quad \max_{\mu}\quad & \sum_{i\in I_g(y)} \mu_i\nabla_y g_i(x(y),y)d_y \\ \text{s.t.}\quad & \nabla_x f(x(y),y) + \sum_{i\in I_g(y)}\mu_i\nabla_x g_i(x(y),y) = 0, \\ & \mu_i \geq 0, \quad i\in I_g(y). \end{aligned}$$

The following conditions are the KKT optimality conditions for (P^1).

$$\begin{aligned} & \nabla_y g_i(x(y),y)d_y + \nabla_x g_i(x(y),y)d_w + s_i = 0, \quad i\in I_g(y), \\ & s_i \geq 0, \quad \mu_i s_i = 0, \quad i\in I_g(y), \end{aligned}$$

where $d_w \in \mathbb{R}^n$ and $s_i,\ i\in I_g(y)$ are the KKT multiplier vectors. Hence, each $\mu\in S^1(d_y)$ satisfies the following system.

$$\begin{aligned} & \nabla_x f(x(y),y) + \sum_{i\in I_g(y)}\mu_i\nabla_x g_i(x(y),y) = 0, \\ & \mu_i\left(\nabla_y g_i(x(y),y)d_y + \nabla_x g_i(x(y),y)d_w\right) = 0, \quad i\in I_g(y), \\ & \mu_i \geq 0, \quad \nabla_y g_i(x(y),y)d_y + \nabla_x g_i(x(y),y)d_w \leq 0, \quad i\in I_g(y). \end{aligned}$$

Since (P^1) is a linear programming problem, each solution $\mu \in S^1(d_y)$ satisfies the above optimality system for each multiplier vector d_w. $K^1_\mu(y; d_y)$ is the set of KKT multiplier vectors for (P^1), which is shown in Lemma 2.2 of Dempe [29]. Further, since $x'(y; d_y) \in K^1_\mu(x; d_y)$ therefore $x'(y; d_y)$ satisfies the above KKT optimality conditions of (P^1). Hence, Problem (16) can be written as

$$
\begin{aligned}
(P^2): \quad & \max_{\mu} \quad \sum_{i \in I_g(y)} \mu_i \left(\frac{1}{2} \, d_z^T \nabla^2_{zz} g_i(x(y), y) d_z + \nabla_y g_i(x(y), y) d^1_y \right) \\
& \text{s.t.} \quad \nabla_x f(x(y), y) + \sum_{i \in I_g(y)} \mu_i \nabla_x g_i(x(y), y) = 0, && (17) \\
& \qquad \mu_i = 0, \quad i \in \{ i \in I_g(y) \ : \ \nabla_x g_i(x(y), y) x'(y; d_y) + \nabla_y g_i(x(y), y) d_y < 0 \}, && (18) \\
& \qquad \mu_i \geq 0, \quad i \in \{ i \in I_g(y) \ : \ \nabla_x g_i(x(y), y) x'(y; d_y) + \nabla_y g_i(x(y), y) d_y = 0 \}. && (19)
\end{aligned}
$$

The KKT optimality conditions for (P^2) are

$$
\begin{aligned}
& \frac{1}{2} \, d_z^T \nabla^2_{zz} g_i(x(y), y) d_z + \nabla_y g_i(x(y), y) d^1_y + \nabla_x g_i(x(y), y) d^1_x + b_i = 0, \quad i \in I_g(y), \\
& b_i \in \mathbb{R}, \quad i \in \{ i \in I_g(y) : \mu_i = 0, \ \nabla_x g_i(x(y), y) x'(y; d_y) + \nabla_y g_i(x(y), y) d_y < 0 \}, \\
& b_i = 0, \quad i \in \{ i \in I_g(y; d_y) : \mu_i > 0 \} = J(\mu), \\
& b_i \geq 0, \quad i \in \{ i \in I_g(y; d_y) : \mu_i = 0 \} = I_g(y; d_y) \setminus J(\mu),
\end{aligned}
$$

where d^1_x is the KKT multiplier vector for the constraint (17), and b_i are the KKT multipliers associated with constraints (18)–(19). Given that the feasible region of the quadratic problem is nonempty for some $\mu \in EM(y)$, let ${d^1_x}^*$ be a feasible point for given d_z and d^1_y.
Define $b^*_i = -\left(\frac{1}{2} \, d_z^T \nabla^2_{zz} g_i(x(y), y) d_z + \nabla_y g_i(x(y), y) d^1_y + \nabla_x g_i(x(y), y) {d^1_x}^* \right)$. Then the pair $({d^1_x}^*, b^*)$ satisfies the above KKT optimality conditions, and thus, $\mu \in S^2(d_y, d^1_y)$. □

Remark 4.2 Note that (15a)-(15d) are the KKT optimality conditions of the quadratic optimization problem in Theorem 4.1, where $d^1_x = x''(y; d_y, d^1_y)$ is the optimal solution to this quadratic problem and $d^1_{\mu_i}, \ i \in I_g(y; d_y)$ is the corresponding KKT multipliers. Suppose $\mu \in EM(y)$, then the feasible set of the quadratic problem is nonempty for given $d_\mu \in \overline{EM}_\mu$ if and only if $\mu \in S^2(d_y, d^1_y)$. In that case, for given vector (d_z, d_μ) such that $d_\mu \in \overline{EM}_\mu$ and $\mu \in EM(y)$, if (d^1_x, d^1_μ) solves the system (15a)-(15d) then $d^1_x = x''(y; d_y, d^1_y)$.

From Remarks 4.1 and 4.2 it is clear that for given direction vectors d_y, d^1_y, the first and second order derivatives $x'(y; d_y)$ and $x''(y; d_y, d^1_y)$ satisfy the system (14a)-(14d) and (15a)-(15d) uniquely at (x, y, μ) respectively under certain assumptions. Next, we apply the active-set strategy to derive a unified first- and second-order system.

Using an active set strategy on the complementarity constraints (14d) and (15d) corresponding to some set I satisfying $J(\mu) \subseteq I \subseteq I_g(y)$ so that $\big(\nabla_x g_i(x(y), y)\big)_{i \in I}$ is linearly independent, we can reformulate (14a)-(15d) as (20a)-(20h), which is a system in $(d_x, d_\mu, d^1_x, d^1_\mu)$ at given point $(x(y), y, \mu)$ and direction vectors d_y, d^1_y at y.

$$
\begin{aligned}
& \nabla^2_{xx} L(x(y), y, \mu) d_x + \nabla^2_{yx} L(x(y), y, \mu) d_y + \nabla^2_{\mu x} L(x(y), y, \mu) d_\mu = 0, && (20a) \\
& \nabla_x g_i(x(y), y) d_x + \nabla_y g_i(x(y), y) d_y = 0, \quad i \in I, && (20b) \\
& \nabla_x g_i(x(y), y) d_x + \nabla_y g_i(x(y), y) d_y \leq 0, \quad i \in I_g(y) \setminus I, && (20c) \\
& d_{\mu_i} \geq 0, \quad i \in I \setminus J(\mu), \qquad d_{\mu_i} = 0, \quad i \notin I && (20d) \\
& \frac{1}{2} \begin{pmatrix} d_z^T & d_\mu^T \end{pmatrix} \widehat{\nabla}^2 (\nabla_x L(x(y), y, \mu)^T) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} + \widehat{\nabla} (\nabla_x L(x(y), y, \mu)^T) \begin{pmatrix} d^1_z \\ d^1_\mu \end{pmatrix} = 0, && (20e) \\
& \nabla_x g_i(x(y), y) \, d^1_x + \nabla_y g_i(x(y), y) d^1_y + \frac{1}{2} d_z^T \nabla^2_{zz} g_i(x(y), y) d_z = 0, \quad i \in I, && (20f) \\
& \nabla_x g_i(x(y), y) \, d^1_x + \nabla_y g_i(x(y), y) d^1_y + \frac{1}{2} d_z^T \nabla^2_{zz} g_i(x(y), y) d_z \leq 0, \quad i \in I_g(y; d_z) \setminus I, && (20g) \\
& d^1_{\mu_i} \geq 0, \quad i \in I \setminus J(\mu; d_\mu), \quad d^1_{\mu_i} = 0, \quad i \notin I. && (20h)
\end{aligned}
$$

Let $(d_x, d_\mu, d_x^1, d_\mu^1)$ be the solution of the system (20a)-(20h) for given d_y and d_y^1. Then, for some appropriate small $t > 0$, $(td_x, td_\mu, t^2 d_x^1, t^2 d_\mu^1)$ satisfies (21a)-(21b) for given $td_y, t^2 d_y^1$.

$$\nabla_x g_i(x(y), y)\,(d_x + d_x^1) + \nabla_y g_i(x(y), y)(\,d_y + d_y^1) + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z \le 0, i \in I_g(y) \setminus I_g(y; d_z), \tag{21a}$$

$$\text{and } d_{\mu_i} + d_{\mu_i}^1 \ge 0, \quad i \in J(\mu; d_\mu) \setminus J(\mu). \tag{21b}$$

That is, (20a)-(21b) hold in a neighborhood of d_y and d_y^1. In that case, (20g)-(21b) can be combined and rewritten using $\nabla_x g_i(x(y), y) d_x + \nabla_y g_i(x(y), y) d_y = 0,\ i \in I_g(y; d_z) \setminus I$ and $d_{\mu_i} = 0,\ i \in I \setminus J(\mu; d_\mu)$ as

$$\nabla_x g_i(x(y), y)\,(d_x + d_x^1) + \nabla_y g_i(x(y), y)(\,d_y + d_y^1) + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z \le 0, \quad i \in I_g(y) \setminus I, \tag{21c}$$

$$d_{\mu_i} + d_{\mu_i}^1 \ge 0, \quad i \in I \setminus J(\mu), \quad d_{\mu_i}^1 = 0, \quad i \notin I. \tag{21d}$$

Thus, the system (14a)-(14d) and (15a)-(15d) are reformulated as the system (20a)-(20f) and (21c)-(21d) in a neighborhood of d_y and d_y^1 for $\|d_y\| \le \delta$ and $\|d_y^1\| \le \delta$.

5 Second-order optimality conditions for BOP

The existing literature on second-order optimality conditions using the value function approach to establish the existence of solutions to bilevel programming problems is summarized in Section 1. In this section, we derive the second-order necessary and sufficient conditions for BOP using the parabolic second-order directional derivative of the optimal solution function of the lower-level problem. First, we establish the following second-order necessary optimality conditions for an optimization problem with continuous, second-order directionally differentiable functions, which will be used later to derive the second-order necessary conditions for BOP.

Lemma 5.1 *Let y be a local optimal solution to the following optimization problem*

$$\min_{y'} \quad h_0(y') \ \text{s.t.} \quad h_i(y') \le 0,\ i \in \Lambda_p,$$

where $h_0 : \mathbb{R}^m \to \mathbb{R}$ and $h_i : \mathbb{R}^m \to \mathbb{R},\ i \in \Lambda_p$ are continuous and second-order directional differentiable functions around y. Then there does not exist $d_y^1 \in \mathbb{R}^m$ such that the following system is satisfied.

$$h_0'(y; d_y) + h_0''(y; d_y, d_y^1) < 0, \tag{22a}$$

$$h_i'(y; d_y) + h_i''(y; d_y, d_y^1) < 0, \quad i \in I_h(y), \tag{22b}$$

$$h_0'(y; d_y) \le 0, \quad h_i'(y; d_y) \le 0, \quad i \in I_h(y), \tag{22c}$$

where $I_h(y) = \{i \in \Lambda_p \ : \ h_i(y) = 0\}$.

Proof Let $\{y^k\}$ be a sequence $y^k = y + \alpha_k d_y + \alpha_k^2 d_y^1$ with $\alpha_k > 0$ and $\{\alpha_k\} \downarrow 0$ for some $(d_y, d_y^1) \in \mathbb{R}^m \times \mathbb{R}^m$. Since h_0 and $h_i,\ i \in I_h(y)$ are second-order directional differentiable at y,

$$h_0(y^k) = h_0(y) + \alpha_k h_0'(y; d_y) + \alpha_k^2 h_0''(y; d_y, d_y^1) + o(\alpha_k^2),\ \forall k, \tag{23}$$

$$\text{and } h_i(y^k) = h_i(y) + \alpha_k h_i'(y; d_y) + \alpha_k^2 h_i''(y; d_y, d_y^1) + o(\alpha_k^2),\ \forall k. \tag{24}$$

Suppose, for the purpose of contradiction, that there exists d_y^1 such that (22a)-(22c) is satisfied. From the conditions (22b)-(22c), it follows that for each $i \in I_h(y)$, either $h_i'(y; d_y) < 0$ or $h_i'(y; d_y) = 0,\ h_i''(y; d_y, d_y^1) < 0$. In either case, there exists N_1 such that $h_i(y^k) < h_i(y) = 0$ for all $i \in I_h(y)$ and $k \ge N_1$ from (24). Further, using the continuity of h_i for $i \in \Lambda_p \setminus I_h(y)$, there exists $N_2 \ge N_1$ such that $h_i(y^k) < 0$ for all $i \in \Lambda_p \setminus I_h(y)$ and $k \ge N_2$. Finally, using (22a), (22c), and (23) we can conclude by a similar manner that there is $N_3 \ge N_2$ such that $h_0(y^k) < h_0(y)$ for all $k \ge N_3$. Thus, $h_0(y^k) < h_0(y)$ for $h_i(y^k) < 0$ for all $i \in \Lambda_p$ and $k \ge N_3$. This yields a contradiction since y is a local optimal solution. Hence, the result follows. □

Let $\{\alpha_k\}$ be a sequence with $\alpha_k > 0$ for each k and $\{\alpha_k\} \to 0$, and the sequence $\{y^k\}$ be defined by $y^k = y + \alpha_k d_y + \alpha_k^2 d_y^1$ for each k. From Theorem 3.2, it follows that

$$x(y^k) = x(y) + \alpha_k x'(y; d_y) + \alpha_k^2 x''(y; d_y, d_y^1) + o(\alpha_k^2)$$

for each k. Consider the expansion of F around $(x(y), y)$ as

$$
\begin{aligned}
F(x(y^k), y^k) = F(x(y), y) &+ \alpha_k \Big(\nabla_x F(x(y), y)\, x'(y; d_y) + \nabla_y F(x(y), y)\, d_y \Big) \\
&+ \alpha_k^2 \Big(\nabla_x F(x(y), y)\, x''(y; d_y, d_y^1) + \nabla_y F(x(y), y)\, d_y^1 \\
&\qquad + \frac{1}{2} \left(x'(y; d_y)^T \;\; d_y^T \right) \nabla_{zz} F(x(y), y) \begin{pmatrix} x'(y; d_y) \\ d_y \end{pmatrix} \Big) \\
&+ o(\alpha_k^2).
\end{aligned}
$$

Let $\tilde{F}(\cdot) := F(x(\cdot), \cdot)$ and $\tilde{G}(\cdot) := G(x(\cdot), \cdot)$. Then, the first-order directional derivative and the second-order directional derivative of the implicit function $\tilde{F}(y)$ are given by

$$
\begin{aligned}
\tilde{F}'(y; d_y) = \lim_{k \to \infty} \frac{\tilde{F}(y^k) - \tilde{F}(y)}{\alpha_k} &= \lim_{k \to \infty} \frac{F(x(y^k), y^k) - F(x(y), y)}{\alpha_k} \\
&= \nabla_x F(x(y), y) x'(y; d_y) + \nabla_y F(x(y), y) d_y
\end{aligned} \tag{25a}
$$

$$
\begin{aligned}
\text{and } \tilde{F}''(y; d_y, d_y^1) &= \lim_{k \to \infty} \frac{\tilde{F}(y^k) - \tilde{F}(y) - \alpha_k \tilde{F}'(y; d_y)}{\alpha_k^2} \\
&= \lim_{k \to \infty} \frac{F(x(y^k), y^k) - F(x(y), y) - \alpha_k \Big(\nabla_x F(x(y), y) x'(y; d_y) + \nabla_y F(x(y), y) d_y \Big)}{\alpha_k^2} \\
&= \nabla_x F(x(y), y) x''(y; d_y, d_y^1) + \nabla_y F(x(y), y)\, d_y^1 \\
&\quad + \frac{1}{2} \left(x'(y; d_y)^T \;\; d_y^T \right) \nabla_{zz}^2 F(x(y), y) \begin{pmatrix} x'(y; d_y) \\ d_y \end{pmatrix},
\end{aligned} \tag{25b}
$$

respectively. Similarly, the first-order directional derivative and second-order directional derivative of $\tilde{G}_i(y)$ for $i \in I_G(y)$ are given as follows

$$
\tilde{G}_i'(y; d_y) = \nabla_x G_i(x(y), y) x'(y; d_y) + \nabla_y G_i(x(y), y) d_y \text{ and} \tag{25c}
$$

$$
\begin{aligned}
\tilde{G}_i''(y; d_y, d_y^1) &= \nabla_x G_i(x(y), y) x''(y; d_y, d_y^1) + \nabla_y G_i(x(y), y)\, d_y^1 \\
&\quad + \frac{1}{2} \left(x'(y; d_y)^T \;\; d_y^T \right) \nabla_{zz}^2 G_i(x(y), y) \begin{pmatrix} x'(y; d_y) \\ d_y \end{pmatrix},
\end{aligned} \tag{25d}
$$

respectively.

Theorem 5.1 *(Second-order necessary condition) Let $(x(y), y)$ be a local optimal solution to BOP, and $F, G \in C^2$ in a neighborhood of $(x(y), y)$. Further, assume that Assumptions $(\mathcal{H}1) - (\mathcal{H}5)$ hold at $(x(y), y)$. Then,*

$$
v := \min \left\{ v(\mu, I) \;:\; \begin{array}{l} \mu \in EM(y), \;\; J(\mu) \subseteq I \subseteq I_g(y), \\ \left(\nabla_x g_i(x(y), y) \right)_{i \in I} \text{ is linearly independent} \end{array} \right\} \geq 0, \tag{26}
$$

where $v(\mu, I)$ is the optimal value of the following optimization problem

$$
(P(\mu, I)) : \min_{d_z, d_z^1, d_\mu, d_\mu^1, \alpha} \quad \alpha
$$

$$
\text{s.t.} \quad \begin{aligned} &\nabla_x F(x(y), y)(d_x + d_x^1) + \nabla_y F(x(y), y)\, (d_y + d_y^1) \\ &\qquad + \tfrac{1}{2} d_z^T \nabla_{zz}^2 F(x(y), y) d_z \leq \alpha, \end{aligned} \tag{27a}
$$

$$
\begin{aligned} &\nabla_x G_i(x(y), y)\, (d_x + d_x^1) + \nabla_y G_i(x(y), y)\, (d_y + d_y^1) \\ &\qquad + \tfrac{1}{2} d_z^T \nabla_{zz}^2 G_i(x(y), y) d_z \leq \alpha, \;\; i \in I_G(y), \end{aligned} \tag{27b}
$$

$$
\nabla_x F(x(y), y)\, d_x + \nabla_y F(x(y), y)\, d_y \leq 0, \tag{27c}
$$

$$
\nabla_x G_i(x(y), y) d_x + \nabla_y G_i(x(y), y)\, d_y \leq 0, \quad i \in I_G(y), \tag{27d}
$$

$$(20a) - (20f) \textit{ and } (21c) - (21d) \textit{ hold}, \tag{27e}$$

$$\|d_y\| \le 1, \quad \|d_y^1\| \le 1. \tag{27f}$$

Proof Under Assumptions $\mathcal{H}1 - \mathcal{H}3$ and $\mathcal{H}5$ there exist neighborhoods $N_{x(y)}$ at $x(y)$ and N_y at y such that $x(\cdot) : N_y \to N_{x(y)}$ is strict global optimal solution function. Hence, BOP can be written in an implicit form in the neighborhood of y as

$$\min_{y'} \quad \tilde{F}(y') \text{ s.t. } \tilde{G}_i(y') \le 0, \quad i \in \Lambda_p,$$

where $\tilde{F}(y') = F(x(y'), y'),\ \tilde{G}_i(y') = G_i(x(y'), y'), i \in \Lambda_p$ for $y' \in N_y$.
Given that $(x(y), y)$ solves BOP locally, it follows that y is a local optimal solution to the above implicit bilevel problem. Using Lemma 5.1 in the implicit bilevel problem, one can conclude that there does not exist d_y^1 such that the following conditions are satisfied.

$$\begin{aligned}
&\tilde{F}'(y; d_y) + \tilde{F}''(y; d_y, d_y^1) < 0,\\
&\tilde{G}_i'(y; d_y) + \tilde{G}_i''(y; d_y, d_y^1) < 0, \quad i \in I_G(y),\\
&\tilde{F}'(y; d_y) \le 0,\ \tilde{G}_i'(y; d_y) \le 0, \quad i \in I_G(y).
\end{aligned}$$

Consequently, the optimal value of the following problem is nonnegative.

$$\begin{aligned}
\min_{\alpha, d_y, d_y^1} \quad & \alpha\\
\text{s.t.} \quad & \tilde{F}'(y; d_y) + \tilde{F}''(y; d_y, d_y^1) \le \alpha,\\
& \tilde{G}_i'(y; d_y) + \tilde{G}_i''(y; d_y, d_y^1) \le \alpha, \quad i \in I_G(y),\\
& \tilde{F}'(y; d_y) \le 0,\ \tilde{G}_i'(y; d_y) \le 0, \quad i \in I_G(y),\\
& \|d_y\| \le 1, \quad \|d_y^1\| \le 1.
\end{aligned}$$

Substituting the values for $\tilde{F}'(y; d_y),\ F''(y; d_y, d_y^1),\ \tilde{G}_i'(y; d_y)$ and $\tilde{G}_i''(y; d_y, d_y^1)$ from (25a)-(25d), we obtain

$$\begin{aligned}
(P): \quad \min_{\alpha, d_y, d_y^1} \quad & \alpha\\
\text{s.t.} \quad & \nabla_x F(x(y), y)\, (x'(y; d_y) + x''(y; d_y, d_y^1)) + \nabla_y F(x(y), y)(d_y + d_y^1)\\
& \qquad + \frac{1}{2} \left(x'(y; d_y)^T \ \ d_y^T\right) \nabla_{zz}^2 F(x(y), y) \begin{pmatrix} x'(y; d_y) \\ d_y \end{pmatrix} \le \alpha,\\
& \nabla_x G_i(x(y), y)\, (x'(y; d_y) + x''(y; d_y, d_y^1)) + \nabla_y G_i(x(y), y)(d_y + d_y^1)\\
& \qquad + \frac{1}{2} \left(x'(y; d_y)^T \ \ d_y^T\right) \nabla_{zz}^2 G_i(x(y), y) \begin{pmatrix} x'(y; d_y) \\ d_y \end{pmatrix} \le \alpha, \quad \forall i \in I_G(y),\\
& \nabla_x F(x(y), y) x'(y; d_y) + \nabla_y F(x(y), y) d_y \le 0,\\
& \nabla_x G_i(x(y), y) x'(y; d_y) + \nabla_y G_i(x(y), y) d_y \le 0, \quad \forall i \in I_G(y),\\
& \|d_y\| \le 1, \quad \|d_y^1\| \le 1.
\end{aligned}$$

Replacing $x'(y; d_y) = d_x$ and $x''(y; d_y, d_y^1) = d_x^1$ and incorporating the system (20a)-(20f) and (21c)-(21d) for some index set I satisfying $J(\mu) \subseteq I \subseteq I_g(y)$ such that $\left(\nabla_x g_i(x(y), y)\right)_{i \in I}$ is linearly independent, Problem (P) can be reformulated into $(P(\mu, I))$. Therefore, the optimal value of the optimization problem (26) is always non-negative for any I satisfying $J(\mu) \subseteq I \subseteq I_g(y)$ such that $\left(\nabla_x g_i(x(y), y)\right)_{i \in I}$ is linearly independent as the optimal value of problem (P) is non-negative. This proves the theorem. □

A second-order necessary condition can be useful in certain cases. The following example is a modification of Example 4.1 of Dempe [1]. Here we see that the first-order necessary condition due to (Corollary 3.4, Dempe [1]) is satisfied at a non-optimal point, whereas the second-order necessary optimality condition (Theorem 5.1) is not satisfied at that point.

Example 5.1

$$\begin{aligned}
&\min_{x,y} \{(x+5)^2 + (y-4)^2 \mid x \in \Psi(y)\},\\
&\text{where } \Psi(y) := \arg\min_x \{(x-3)^2 \mid x^2 - y \le 0\}
\end{aligned}$$

Note that $(x, y) = (1, 1)$ is a non-optimal point. Here, $M(y) = \{2\}$ and $J(\mu) = I_g(y)$, which results in a single choice $I = I_g(y)$.

$$
\begin{aligned}
\min \quad & \alpha \\
\text{s.t.} \quad & 12d_x - 6d_y \le \alpha, \\
& 6d_x + 2d_\mu = 0, \quad -2d_x + d_y = 0, \\
& -1 \le d_y \le 1.
\end{aligned}
$$

The optimal value of the above problem is 0, which means that the first-order necessary conditions (Corollary 3.4, Dempe [1]) are satisfied.

Next, we will show that the second-order necessary conditions do not hold at this point.
The problem $(P(\mu, I))$ at $(x, y, \mu) = (1, 1, 2)$ in Theorem 5.1 is

$$
\begin{aligned}
\min \quad & \alpha \\
\text{s.t.} \quad & 12(d_x + d_x^1) - 6(d_y + d_y^1) + (d_x)^2 + (d_y)^2 \le \alpha, \\
& 6d_x + 2d_\mu = 0, \quad -d_y + 2d_x = 0, \\
& 6d_x^1 + 2d_\mu^1 = -2d_x d_\mu, \quad -d_y^1 + 2d_x^1 = -(d_x)^2, \\
& -1 \le d_y \le 1, -1 \le d_y^1 \le 1.
\end{aligned}
$$

The optimal value of the above problem is -0.25, which is negative. Hence, this does not satisfy the necessary condition of Theorem 5.1. $\square$

Let $P'(\mu, I)$ be the optimization problem obtained from $(P(\mu, I))$ by restricting the feasible set of $(P(\mu, I))$ with $\|d_y\| \neq 0$ and putting $\alpha = 0$ in the constraint (27b). This takes the following form after simplifications.

$(P'(\mu, I))$:

$$
\begin{aligned}
\min_{d_z, d_z^1, d_\mu, d_\mu^1} \quad & \nabla_x F(x(y), y)(d_x + d_x^1) + \nabla_y F(x(y), y)\,(d_y + d_y^1) + \frac{1}{2} d_z^T \nabla_{zz}^2 F(x(y), y) d_z \\
\text{s.t.} \quad & \nabla_x G_i(x(y), y)\,(d_x + d_x^1) + \nabla_y G_i(x(y), y)\,(d_y + d_y^1) + \frac{1}{2} d_z^T \nabla_{zz}^2 G_i(x(y), y) d_z \le 0, \ i \in I_G(y), \\
& \|d_y\| \neq 0, \ \|d_y^1\| \le 1, \text{and (27c)} - \text{(27e) hold.}
\end{aligned}
$$

Theorem 5.2 *(Second-order sufficient condition) Let $(x(y), y)$ be a feasible point of BOP. Suppose that Assumptions $\mathcal{H}1 - \mathcal{H}5$ hold, and $F, G \in C^2$ in a neighborhood of $(x(y), y)$. Let*

$$
v' := \min_{\mu, I} \left\{ v'(\mu, I) \ : \ \begin{array}{l} \mu \in EM(y), \ J(\mu) \subseteq I \subseteq I_g(y), \ \left(\nabla_x g_i(x(y), y)\right)_{i \in I} \text{ is linearly independent,} \\ \text{and } v'(\mu, I) \text{ is the optimal value of } P'(\mu, I) \end{array} \right\}.
$$

If $v' > 0$ then $(x(y), y)$ is a strict local minimum of BOP and there exist $\varepsilon > 0$ and $c > 0$ such that

$$
F(x', y') \ge F(x(y), y) + c\|y' - y\|^2
$$

for all (x', y') satisfying $\|y' - y\| \le \varepsilon$ and $x' \in \Psi(y')$, $G(x', y') \le 0$.

Proof Since $v' > 0$, from $P'(\mu, I)$,

$$
\begin{aligned}
& \nabla_x F(x(y), y) \left(x'(y; d_y) + x''(y; d_y, d_y^1) \right) + \nabla_y F(x(y), y) \left(d_y + d_y^1 \right) \\
& \quad + \frac{1}{2} \left(x'(y; d_y)^T \ \ d_y^T \right) \nabla_{zz}^2 F(x(y), y) \begin{pmatrix} x'(y; d_y) \\ d_y \end{pmatrix} > 0
\end{aligned}
\tag{28}
$$

for all (d_y, d_y^1) satisfying

$$
\left.
\begin{aligned}
& \nabla_x G_i(x(y), y)\,(x'(y; d_y) + x''(y; d_y, d_y^1)) + \nabla_y G_i(x(y), y)\,(d_y + d_y^1) \\
& \quad + \frac{1}{2} \left(x'(y; d_y)^T \ \ d_y^T \right) \nabla_{zz}^2 G_i(x(y), y) \begin{pmatrix} x'(y; d_y) \\ d_y \end{pmatrix} \le 0, \quad i \in I_G(y), \\
& \nabla_x G_i(x(y), y)\, x'(y; d_y) + \nabla_y G_i(x(y), y)\, d_y \le 0, \quad i \in I_G(y), \\
& \nabla_x F(x(y), y)\, x'(y; d_y) + \nabla_y F(x(y), y) d_y \le 0.
\end{aligned}
\right\}
\tag{29}
$$

From (1) and (2), it follows that $x'(y;td_y) = tx'(y;d_y)$ and $x''(y;td_y,t^2d_y^1) = t^2x''(y;d_y,d_y^1)$ for any $t > 0$. Suppose $\nabla_x F(x(y),y)\, x'(y;d_y) + \nabla_y F(x(y),y)d_y < 0$ for some (d_y, d_y^1) satisfying (29). Since $\nabla_x F(x(y),y)\, x'(y;d_y) + \nabla_y F(x(y),y)d_y < 0$, the objective value of $P'(\mu, I)$ at the feasible point $(td_y, t^2d_y^1)$ for some appropriate $t > 0$ is

$$\begin{aligned} &t\Big(\nabla_x F(x(y),y)x'(y;d_y) + \nabla_y F(x(y),y)d_y\Big) \\ &\quad + t^2\left(\nabla_x F(x(y),y)x''(y;d_y,d_y^1) + \nabla_y F(x(y),y)d_y^1 + \frac{1}{2}\begin{pmatrix} x'(y;d_y)^T & d_y^T\end{pmatrix} \nabla^2_{zz}F(x(y),y)\begin{pmatrix} x'(y;d_y) \\ d_y\end{pmatrix}\right) < 0. \end{aligned}$$

This contradicts (28). Hence, in view of (29), we obtain $\nabla_x F(x(y),y)\, x'(y;d_y) + \nabla_y F(x(y),y)d_y = 0$. Thus, from (28), it follows that

$$\nabla_x F(x(y),y)x''(y,d_y;d_y^1) + \nabla_y F(x(y),y)d_y^1 + \frac{1}{2}\begin{pmatrix} x'(y;d_y)^T & d_y^T\end{pmatrix} \nabla^2_{zz}F(x(y),y)\begin{pmatrix} x'(y;d_y) \\ d_y\end{pmatrix} > 0$$

for all (d_y, d_y^1) satisfying (29). Suppose on the contrary that there exist a sequence $\{(x(y^k), y^k)\}$ with $x(y^k) \in \Psi(y^k)$ and $G(x(y^k), y^k) \leq 0$, satisfying

$$F(x(y^k), y^k) < F(x(y),y) + o(\|y^k - y\|^2), \tag{30}$$

such that $\{y^k\} \to y$. Using Assumptions $\mathcal{H}1 - \mathcal{H}3$, this follows from the continuity of $x(\cdot)$ that $\{x(y^k)\} \to x(y)$. Consider $\alpha_k = \|y^k - y\|$ then sequence $\{d^k\} := \{\frac{y^k - y}{\alpha_k}\}$ converges to some accumulation point d_y, possibly over a subsequence. Hence, for the sequence $\{{d^1}^k\} := \{\frac{d^k - d_y}{\alpha_k}\}$, there is a subsequence of $\{y^k\}$ such that $\{\alpha_k {d^1}^k\} \to 0$. Then, it follows that $y^k = y + \alpha_k d^k = y + \alpha_k(d_y + \alpha_k {d^1}^k) = y + \alpha_k d_y + \alpha_k^2 {d^1}^k$. As $v' > 0$, so from (30), we have

$$F(x(y^k), y^k) < F(x(y),y) + o(\|y^k - y\|^2) \leq F(x(y),y) + v'\alpha_k^2 + o(\alpha_k^2). \tag{31}$$

From Theorem 3.3, we can obtain

$$x(y^k) = x(y) + \alpha_k x'(y;d_y) + \alpha_k^2 x''(y;d_y,d_k^1) + o(\alpha_k^2).$$

Since $F \in C^2$, incorporating $x(y^k)$ from the above expression and expanding F about $(x(y),y)$ up to second-order,

$$\begin{aligned} F(x(y^k),y^k) &= F\big(x(y) + \alpha_k x'(y;d_y) + \alpha_k^2 x''(y;d_y,{d^1}^k) + o(\alpha_k^2),\ y + \alpha_k d_y + \alpha_k^2 {d^1}^k\big) \\ &= F(x(y),y) + \nabla_z F(x(y),y)\begin{pmatrix} \alpha_k x'(y;d_y) + \alpha_k^2 x''(y;d_y,{d^1}^k) \\ \alpha_k d_y + \alpha_k^2 {d^1}^k \end{pmatrix} \\ &\quad + \frac{1}{2}\begin{pmatrix} \alpha_k x'(y;d_y) + \alpha_k^2 x''(y;d_y,{d^1}^k) \\ \alpha_k d_y + \alpha_k^2 {d^1}^k \end{pmatrix}^T \nabla^2_{zz}F(x(y),y)\begin{pmatrix} \alpha_k x'(y;d_y) + \alpha_k^2 x''(y;d_y,{d^1}^k) \\ \alpha_k d_y + \alpha_k^2 {d^1}^k \end{pmatrix} \\ &\quad + o(\alpha_k^2) \\ &= F(x(y),y) + \alpha_k\Big(\nabla_x F(x(y),y)\, x'(y;d_y) + \nabla_y F(x(y),y)\, d_y\Big) \\ &\quad + \alpha_k^2\left(\nabla_x F(x(y),y)\, x''(y;d_y,{d^1}^k) + \nabla_y F(x(y),y)\, {d^1}^k \right. \\ &\qquad \left. + \frac{1}{2}\begin{pmatrix} x'(y;d_y) \\ d_y\end{pmatrix}^T \nabla^2_{zz}F(x(y),y)\begin{pmatrix} x'(y;d_y) \\ d_y\end{pmatrix}\right) + o(\alpha_k^2). \end{aligned}$$

As $G(x(y^k),y^k) \leq 0$, applying the expansion of G and proceeding as in the above argument, we conclude that $(d_y, {d^1}^k)$ is feasible for (29) after scalarization (if needed). Since $\nabla_x F(x,y)x'(y;d_y) + \nabla_y F(x,y)d_y = 0$,

$$\begin{aligned} F(x(y^k),y^k) &= F(x(y),y) \\ &\quad + \alpha_k^2\left(\nabla_x F(x(y),y)x''(y;d_y,{d^1}^k) + \nabla_y F(x(y),y){d^1}^k \right. \\ &\qquad \left. + \frac{1}{2}\begin{pmatrix} x'(y;d_y)^T & d_y^T\end{pmatrix} \nabla^2_{zz}F(x(y),y)\begin{pmatrix} x'(y;d_y) \\ d_y\end{pmatrix}\right) + o(\alpha_k^2) \\ &\geq F(x(y),y) + v'\alpha_k^2 + o(\alpha_k^2). \end{aligned}$$

The above relation contradicts (31). Hence, there exists $c > 0$ and $\varepsilon > 0$ such that

$$F(x', y') \geq F(x(y),y) + c\|y - y'\|^2$$

for all (x', y') satisfying $\|x' - x\| \leq \varepsilon$ and $y' \in \Psi(y'),\ G(x',y') \leq 0$. From Assumption $\mathcal{H}5$, Ψ is a single-valued global optimal solution function of the lower-level problem. Therefore, $(x(y), y)$ is a strict local minimum of BOP. $\square$

Notice that the constraint set of $P'(\mu, I)$ is not closed due to the presence of the condition $\|d_y\| \neq 0$. This compromises the practicability of Theorem 5.2. To handle this issue, we solve $P'(\mu, I)$ restricted by $\|d_y\| = \alpha$ for some $\alpha > 0$. If the optimal value of $P'(\mu, I)$ is positive for $\alpha > 0$, then the claim of Theorem 5.2 holds.

The following example demonstrates the case when the first-order sufficient condition (Theorem 4.2 of Dempe [1], Theorem 5.5 of Dempe [2]) is not satisfied. However, the second-order sufficient condition of Theorem 5.2 is satisfied at the optimal solution of the example. In this example, the lower-level problem has non-unique multipliers in the KKT optimality conditions at the optimal solution.

Example 5.2

$$
\begin{aligned}
&\min_{x,y} \{(x-3)^2 + (y-1)^2 \mid x \in \Psi(y)\}, \\
&\text{where } \Psi(y) := \underset{x}{\operatorname{argmin}} \quad (x-5)^2 \\
&\qquad\qquad\qquad \text{s.t.} \quad x - 2y + 1 \leq 0, \qquad (i) \\
&\qquad\qquad\qquad\qquad\ \ -2x + y + 1 \leq 0, \qquad (ii) \\
&\qquad\qquad\qquad\qquad\ \ 2x + y - 7 \leq 0. \qquad (iii)
\end{aligned}
$$

Here $F(x, y) = (x-3)^2 + (y-1)^2, f(x, y) = (x-5)^2$ and $g_1(x, y) = x - 2y + 1, g_2(x, y) = -2x + y + 1, g_3(x, y) = 2x + y - 7$. The upper level objective function has a global minimum at $(1, 3)$ $\forall (x, y) \in \mathbb{R}^2$. At $y = 1$, the lower-level problem is $\min_x\ (x-5)^2\ s.t.\ -2 \leq x \leq 3$, which has the minimum value at $x = 3$. Therefore, $(1, 3)$ is the global optimal solution to this bilevel problem. Next, it remains to verify that the second-order sufficient optimality conditions of Theorem 5.2 holds at $(1, 3)$, whereas the first-order conditions of Theorem 2.2 do not hold at this point.
The first-order sufficient optimality condition of Theorem 2.2 is not satisfied at $(1, 3)$ as $\nabla_y F(1, 3) = \nabla_x F(1, 3) = 0$. Observe that Assumptions $\mathcal{H}1 - \mathcal{H}5$ hold at $(1, 3)$. The KKT optimality conditions for the lower-level problem at $y = 1$ are

$$
\begin{aligned}
&2(x-5) + \mu_1 - 2\mu_2 + 2\mu_3 = 0, \\
&\mu_1(x - 2y + 1) = 0,\ \mu_2(-2x + y + 1) = 0,\ \mu_3(2x + y - 7) = 0, \\
&\mu_1 \geq 0,\ \mu_2 \geq 0,\ \mu_3 \geq 0,
\end{aligned}
$$

where μ_1, μ_2 and μ_3 are KKT multipliers associated with the constraints $(i), (ii)$ and (iii) respectively. The above system becomes

$$\mu_1 + 2u_3 = 4,\ \mu_1 \geq 0,\ \mu_2 = 0, \mu_3 \geq 0$$

at $(1, 3)$. Hence, at $(x, y) = (1, 3)$, $EM(y) = \text{convex hull of } \{(4, 0, 0)^T, (0, 0, 2)^T\}$.
Here, two possibilities arise for the index set I. For $\mu = (4, 0, 0)^T$, $J(\mu) = \{(i)\}$, and for $\mu = (0, 0, 2)^T$, $J(\mu) = \{(iii)\}$. Here, $I_g(y) = \{(i), (iii)\}$ at $(x, y) = (1, 3)$. Hence

$$\left\{ I \ | J(\mu) \subseteq I \subseteq I_g(y),\ \mu \in EM(y) \ \ \left(\nabla_x g_i(x(y), y)\right)_{i \in I} \text{ is linearly independent } \right\} = \{\{i\}, \{iii\}\}$$

(i) For $\mu = (4, 0, 0)^T$ and $I = \{(i)\}$,

$$
\begin{aligned}
P'((4, 0, 0)^T, \{(i)\}) : \quad &\min_{d_z, d_z^1, d_\mu, d_\mu^1} \quad d_x^2 + d_y^2 \\
&\text{s.t.} \quad 2d_x + d_{\mu_1} + 2d_{\mu_3} = 0, \\
&\qquad\ \ 2d_x^1 + d_{\mu_1}^1 + 2d_{\mu_3}^1 = 0, \\
&\qquad\ \ d_x - 2d_y = 0, \quad d_x^1 - 2d_y^1 = 0, \\
&\qquad\ \ 2d_x + d_y \leq 0, \quad 2(d_x + d_x^1) + (d_y + d_y^1) \leq 0,
\end{aligned}
$$

$$\|d_y\| = 1, \quad \|d_y^1\| \leq 1.$$

The solution to $P'((4,0,0)^T, \{(i)\})$ is $(d_x, d_y, d_{\mu_1}, d_{\mu_3}, d_x^1, d_y^1, d_{\mu_1}^1, d_{\mu_3}^1) = (-2, -1, 2, 0, 0, 0, 0, 0)$ with the optimal value $v'((4,0,0)^T, \{(i)\}) = 2$.
(ii) For $\mu = (0,0,2)^T$ and $I = \{(iii)\}$,

$$\begin{aligned} P'((0,0,2)^T, \{(iii)\}) \ : \quad & \min_{d_z, d_z^1, d_\mu, d_\mu^1} && d_x^2 + d_y^2 \\ & \text{s.t.} && 2d_x + d_{\mu_1} + 2d_{\mu_3} = 0 \\ &&& 2d_x^1 + d_{\mu_1}^1 + 2d_{\mu_3}^1 = 0, \\ &&& 2d_x + d_y = 0, \quad 2d_x^1 - d_y^1 = 0, \\ &&& d_x - 2d_y \leq 0, \quad (d_x + d_x^1) - (2d_y + 2d_y^1) \leq 0, \\ &&& \|d_y\| = 1, \quad \|d_y^1\| \leq 1. \end{aligned}$$

The solution to $P'((0,0,2)^T, \{(iii)\})$ is $(d_x, d_y, d_{\mu_1}, d_{\mu_3}, d_x^1, d_y^1, d_{\mu_1}^1, d_{\mu_3}^1) = (-\frac{1}{2}, 1, 0, 1, 0, 0, 0, 0)$ with the optimal value $v'((0,0,2)^T, \{(iii)\}) = 1.25$. Hence,

$$v' = \min \left\{ v'((4,0,0)^T, \{(i)\}), v'((0,0,2)^T, \{(iii)\}) \right\} = 1.25.$$

Since $v' > 0$, the second-order sufficient optimality condition of Theorem 5.2 holds.

6 Conclusion

This work derives second-order necessary and sufficient optimality conditions for BOP with a strongly stable lower-level problem that admits non-unique KKT multiplier vectors. The analysis uses the solution function's parabolic second-order directional deviation for the lower-level problem. Specifically, the second-order directional derivative in the parabolic sense of the implicit bilevel problem is applied to derive these optimality conditions. Third-order differentiability information is required for the lower-level problem, which is the limitation of this work. The derivations can easily be extended for upper and lower-level equality-type constraints. Numerical approximation techniques for BOP can be developed using these second-order optimality conditions, which may constitute a future scope of the present contribution. Liu et al. [17] introduces a generalized concept of local optimal solution called bi-local optimal solution for the bilevel problems with nonconvex lower-level problems. The bi-local optimal solution reduces to the standard local optimal solution when the lower-level problem is convex. In a recent work of Liu et al. [18], it is proved that under Assumptions $\mathcal{H}1-\mathcal{H}3$, a bi-local optimal solution and a standard local optimal solution are equivalent. Hence, under these restrictions, the main results of Section 5 can be extended to the case of a nonconvex lower-level problem by characterizing the optimality conditions corresponding to bi-local optimal solutions.

Declarations

Ethics approval and consent to participate Not Applicable

Consent for publication Not applicable

Availability of data and materials Data sharing was not applicable to this article as no data sets were generated or analyzed during the current study.

Competing interests The authors declare that they have no conflicts of interest to this work.

Funding Not applicable

Author contribution Bhuvnesh Khatana: Conceptualization, Formal analysis, Methodology, Software, Validation, Original draft writing, Writing - Review & Editing. Geetanjali Panda: Formal analysis, Investigation, Methodology, Writing - Review & Editing, Supervision.

Acknowledgment The authors thank the reviewers for their helpful comments and suggestions, which have improved the quality of this work.

References

[1] Dempe, S.: A necessary and a sufficient optimality condition for bilevel programming problems. Optimization **25**(4), 341–354 (1992) https://doi.org/10.1080/02331939208843831

[2] Dempe, S.: Foundations of Bilevel Programming. Kluwer Academic Publishers, Dordrecht (2002)

[3] Bard, J.F.: Practical Bilevel Optimization: Algorithms and Applications vol. 30. Springer, Berlin (2013)

[4] Dempe, S., Vogel, S.: The generalized jacobian of the optimal solution in parametric optimization. Optimization **50**(5-6), 387–405 (2001) https://doi.org/10.1080/02331930108844570

[5] Mehlitz, P.: Necessary optimality conditions for a special class of bilevel programming problems with unique lower level solution. Optimization **66**(10), 1533–1562 (2017) https://doi.org/10.1080/02331934.2017.1349123

[6] Ye, J., Zhu, D.: Optimality conditions for bilevel programming problems. Optimization **33**(1), 9–27 (1995) https://doi.org/10.1080/02331939508844060

[7] Dempe, S., Zemkoho, A.B.: The generalized mangasarian-fromowitz constraint qualification and optimality conditions for bilevel programs. Journal of optimization theory and applications **148**, 46–68 (2011) https://doi.org/10.1007/s10957-010-9744-8

[8] Dempe, S., Zemkoho, A.B.: The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. Mathematical Programming **138**(1), 447–473 (2013) https://doi.org/10.1007/s10107-011-0508-5

[9] Babahadda, H., Gadhi, N.: Necessary optimality conditions for bilevel optimization problems using convexificators. Journal of Global Optimization **34**, 535–549 (2006) https://doi.org/10.1007/s10898-005-1650-5

[10] Dempe, S., Dutta, J., Mordukhovich, B.: New necessary optimality conditions in optimistic bilevel programming. Optimization **56**(5-6), 577–604 (2007) https://doi.org/10.1080/02331930701617551

[11] Dempe, S., Zemkoho, A.B.: KKT reformulation and necessary conditions for optimality in nonsmooth bilevel optimization. SIAM Journal on Optimization **24**(4), 1639–1669 (2014) https://doi.org/10.1137/130917715

[12] Ye, J.J., Zhu, D.: New necessary optimality conditions for bilevel programs by combining the MPEC and value function approaches. SIAM Journal on Optimization **20**(4), 1885–1905 (2010) https://doi.org/10.1137/080725088

[13] Ma, X., Yao, W., Ye, J.J., Zhang, J.: Combined approach with second-order optimality conditions for bilevel programming problems. arXiv preprint arXiv:2108.00179 (2021)

[14] Ye, J.J.: Constraint qualifications and optimality conditions in bilevel optimization. Bilevel optimization: advances and next challenges, 227–251 (2020) https://doi.org/10.1007/978-3-030-52119-6_8

[15] Falk, J.E., Liu, J.: On bilevel programming, part i: general nonlinear cases. Mathematical Programming **70**(1), 47–72 (1995) https://doi.org/10.1007/BF01585928

[16] Dyro, R., Schmerling, E., Arechiga, N., Pavone, M.: Second-order sensitivity analysis for bilevel optimization. In: International Conference on Artificial Intelligence and Statistics, pp. 9166–9181 (2022). PMLR

[17] Liu, X., Xu, M., Zhang, L.: Second-order optimality conditions for bi-local solutions of bilevel programs. Optimization, 1–23 (2025) https://doi.org/10.1080/02331934.2025.2468412

[18] Liu, X., Xu, M., Zhang, L.: Second-order optimality conditions for nonsmooth constrained optimization with applications to bilevel programming. arXiv preprint arXiv:2511.02439 (2025)

[19] Mehlitz, P., Zemkoho, A.B.: Sufficient optimality conditions in bilevel programming. Mathematics of operations research **46**(4), 1573–1598 (2021) https://doi.org/10.1287/moor.2021.1122

[20] Rückmann, J.-J., Shapiro, A.: Second-order optimality conditions in generalized semi-infinite programming. Set-Valued Analysis **9**, 169–186 (2001) https://doi.org/10.1023/A:1011239607220

[21] Dempe, S., Gadhi, N., *et al.*: Second order optimality conditions for a bilevel optimization problem in terms of approximate Hessians. Journal of Mathematical Analysis and Applications **509**(1), 125932 (2022) https://doi.org/10.1016/j.jmaa.2021.125932

[22] Ralph, D., Dempe, S.: Directional derivatives of the solution of a parametric nonlinear program. Mathematical programming **70**(1), 159–172 (1995) https://doi.org/10.1007/BF01585934

[23] Liu, J.: Sensitivity analysis in nonlinear programs and variational inequalities via continuous selections. SIAM Journal on Control and Optimization **33**(4), 1040–1060 (1995) https://doi.org/10.1137/S0363012993248876

[24] Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2013). https://doi.org/10.1007/978-1-4612-1394-9

[25] Ben-Tal, A.: Second-order and related extremality conditions in nonlinear programming. Journal of Optimization Theory and Applications **31**(2), 143–165 (1980) https://doi.org/10.1007/BF00934107

[26] Fiacco, A.V.: Introduction to sensitivity and stability analysis in non linear programming (1983)

[27] Janin, R.: In: Fiacco, A.V. (ed.) Directional derivative of the marginal function in nonlinear programming, pp. 110–126. Springer, Berlin, Heidelberg (1984). https://doi.org/10.1007/BFb0121214 . https://doi.org/10.1007/BFb0121214

[28] Shapiro, A.: Sensitivity analysis of nonlinear programs and differentiability properties of metric projections. SIAM Journal on Control and Optimization **26**(3), 628–645 (1988) https://doi.org/10.1137/0326037

[29] Dempe, S.: Directional differentiability of optimal solutions under slater's condition. Mathematical programming **59**, 49–69 (1993) https://doi.org/10.1007/BF01581237

[30] Ben-Tal, A., Zowe, J.: In: Guignard, M. (ed.) A unified theory of first and second order conditions for extremum problems in topological vector spaces, pp. 39–76. Springer, Berlin, Heidelberg (1982). https://doi.org/10.1007/BFb0120982

[31] Kojima, M.: Strongly stable stationary solutions in nonlinear programs. In: Analysis and Computation of Fixed Points, pp. 93–138. Elsevier, the University of Wisconsin–Madison (1980). https://doi.org/10.1016/B978-0-12-590240-3.50009-4

[32] Khatana, B., Panda, G.: Second order directional derivative of optimal solution function in parametric programming problem. Optimization Online (2025). https://optimization-online.org/?p=31185