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 bilevel optimization problems were derived using the first-order directional derivative of the lower-level optimal solution function 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, accommodating lower-level problem that may admit non-unique Karush–Kuhn–Tucker (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.

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

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, and 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 $F(x,y) \leq F(\hat{x},\hat{y})$ for all $(\hat{x},\hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, $\hat{x} \in \Psi(\hat{y})$, $G(\hat{x},\hat{y}) \leq 0$ with $\|(x,y) - (\hat{x},\hat{y})\| \leq \varepsilon$. A feasible point is a global optimal solution if these conditions hold for all $\varepsilon > 0$. The lower-level problem of BOP is

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

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 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. 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 generalized gradients 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 the first-order optimality conditions for bilevel optimization problems, including the difficult classes of nonsmooth and nonconvex bilevel frameworks. In a 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-predominant 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 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 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. The authors impose LICQ on the lower-level problem to obtain verifiable optimality conditions.

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 derivative along a parabolic path is 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. 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.

- $\nabla h := \left(\nabla h_1^T \ \nabla h_2^T \ \cdots \ \nabla h_q^T\right)^T \in \mathbb{R}^{q\times p}$ 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 $h_i : \mathbb{R}^p \to \mathbb{R}$.
- $\nabla^2 h := \left(\nabla^2 h_1^T \ \nabla^2 h_2^T \ \cdots \ \nabla^2 h_q^T\right)^T \in \mathbb{R}^{qp\times p}$ represents the Hessian of h, where each $\nabla^2 h_i \in \mathbb{R}^{p\times p}$.
- $d^T\nabla^2 h d := \left(d^T\nabla^2 h_1 d \ d^T\nabla^2 h_2 d \ \cdots \ d^T\nabla^2 h_q d\right)^T \in \mathbb{R}^q$ for any $d \in \mathbb{R}^p$.
- $|\mathcal{A}|$ denotes the cardinality of the set $\mathcal{A}$.
- $\Lambda_\beta := \{1, 2, \ldots, \beta\}$ denotes the index set of size β.
- $\widehat{\nabla} := \left(\nabla_x \ \nabla_y \ \nabla_\mu\right)$ and $\widehat{\nabla}^2 := \begin{pmatrix} \nabla^2_{xx} & \nabla^2_{yx} & \nabla^2_{\mu x} \\ \nabla^2_{xy} & \nabla^2_{yy} & \nabla^2_{\mu y} \\ \nabla^2_{x\mu} & \nabla^2_{y\mu} & \nabla^2_{\mu\mu} \end{pmatrix}$ 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 the 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 a 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 for local optimality if the second-order optimality conditions for LP_y are satisfied. We assume the strong sufficient optimality condition on LP_y.

($\mathcal{H}$**3**) For all $\mu \in M(y)$, the following condition holds 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 the 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 independence constraint qualification. Shapiro [28] showed 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]. 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.
The solution function $x(\cdot)$ 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 $\mathcal{H}1$-$\mathcal{H}3$ at $(x(y^*), y^*)$, 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.*

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

(ii) (Theorem 2, Ralph and Dempe [22]) Under Assumptions $\mathcal{H}1$-$\mathcal{H}4$ at $(x(y^*), y^*)$, 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*

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

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 $\lim\limits_{\alpha \downarrow 0} \frac{\chi(y+\alpha d_y+\alpha^2 d_y^1)-\chi(y)-\alpha\chi'(y;d_y)}{\alpha^2}$ exists, where $d_y, d_y^1 \in \mathbb{R}^m$ are direction vectors. The parabolic second-order directional derivative of χ at y along a parabolic path $y + \alpha d_y + \alpha^2 d_y^1$ is denoted by the vector $\chi''(y; d_y, d_y^1)$, and computed as

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

If $\chi''(y; d_y, d_y^1)$ exists for all d_y and d_y^1, 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 $\chi''(y; d_y, d_y^1)$, 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

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

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

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

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

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

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

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

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 $\{x_1(\cdot), x_2(\cdot), \ldots, x_N(\cdot)\}$ such that $x(y) \in \{x_1(y), x_2(y), \ldots, x_N(y)\}$ 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

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

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 a 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

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

The above considerations yield the existence of $r' > 0$ such that for each $t \in [0, r']$, there is some $j \in \{1, 2, \ldots, N\}$ 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,

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

and

$$\lim_{k\to\infty} \frac{x(y^* + \alpha_k d_y + \alpha_k^2 d_y^1) - x(y^*) - \alpha_k x'_{j_1}(y^*; d_y)}{\alpha_k^2} = x''_{j_1}(y^*; d_y, d_y^1) = x''_{j_2}(y^*; d_y, d_y^1). \tag{3}$$

Consequently, 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 (2)-(3) hold. Since $[0, r'] = \bigcup_{\mathscr{I}(d_y, d_y^1)} R_j$, we claim that for any $r, s \in \mathscr{I}(d_y, d_y^1)$, one can find a chain of relations such that $r \sim j_1 \sim j_2, \ldots, \sim j_t \sim s$. 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 $\{\alpha_k\} \subseteq R^r \cap R^s$. Therefore, for some $0 < r'' \leq r'$

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

On the other hand, since $[0, r'] = \bigcup_{\mathscr{I}(d_y, d_y^1)} R_j = R^r \cup R^s$, 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 $[0, r'']$. Therefore, there exists a chain of relations $r \sim j_1 \sim j_2, \ldots, \sim j_t \sim s$ for any $r, s \in \mathscr{I}(d_y, d_y^1)$. Then, (1) holds. This completes the proof of the theorem. □

The epi-regularity condition on the second-order directional differentiable function along a 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 $\lim_{t\downarrow 0} t d_y^1(t) = 0$, the following expansion holds

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

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

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

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

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 exist neighborhoods N_{x^*} of x^* and N_{y^*} of y^* such that the optimal solution mapping $x(\cdot) : N_{y^*} \to N_{x^*}$ of LP_y is second-order gph-regular at y^*.*

Proof Since $x(\cdot)$ is 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 that $x(\cdot)$ is second-order gph-regular, we show that the sequence

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

for every sequence $\{\alpha_k\} \downarrow 0$. Since $x(\cdot)$ is piecewise smooth, there exists a collection of C^2 selection functions $\{x_j(\cdot)\}_{j=1}^N$ as in Theorem 3.2. Along the path $y^* + td_y + t^2 d_y^1(t)$, define the active index set

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

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

$$x_j(y^* + \alpha_k d_y + \alpha_k^2 d_y^1(\alpha_k)) = x_j(y^*) + \alpha_k x_j'(y^*; d_y) + \alpha_k^2 x_j''(y^*; d_y, d_y^1(\alpha_k)) + o(\alpha_k^2), \quad \forall k, \tag{4}$$

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 $j \in \mathscr{I}$ and suitable $r' > 0$. Then,

$$\begin{aligned} &\lim_{k\to\infty} \frac{x(y^* + \alpha_k d_y + \alpha_k^2 d_y^1(\alpha_k)) - x(y^*) - \alpha_k x'(y^*; d_y) - \alpha_k^2 x''(y^*; d_y, d_y^1(\alpha_k))}{\alpha_k^2} \\ &= \lim_{k\to\infty} \frac{x_j(y^* + \alpha_k d_y + \alpha_k^2 d_y^1(\alpha_k)) - x_j(y^*) - \alpha_k x_j'(y^*; d_y) - \alpha_k^2 x''(y^*; d_y, d_y^1(\alpha_k))}{\alpha_k^2} \\ &= \lim_{k\to\infty} \left(x_j''(y^*; d_y, d_y^1(\alpha_k)) - x''(y^*; d_y, d_y^1(\alpha_k)) \right) + \lim_{k\to\infty} \frac{o(\alpha_k^2)}{\alpha_k^2} \quad \text{(using (4))} \\ &= \lim_{k\to\infty} \left(x_j''(y^*; d_y, d_y^1(\alpha_k)) - x''(y^*; d_y, d_y^1(\alpha_k)) \right). \end{aligned} \tag{5}$$

Following the arguments from 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 $\{\bar{\alpha}_k\} \subseteq R_{j_1} \cap R_{j_2}$, and

$$\left. \begin{array}{l} x(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)) = x_{j_1}(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)) = x_{j_2}(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)), \\ x'(y^*; d_y) = x_{j_1}'(y^*; d_y) = x_{j_2}'(y^*; d_y) = \lim_{k\to\infty} \dfrac{x(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)) - x(y^*)}{\bar{\alpha}_k}. \end{array} \right\} \tag{6}$$

Writing the second-order Taylor expansion for x_{j_1} and x_{j_2} along $\{\bar{\alpha}_k\}$, we have

$$\begin{aligned} x_{j_1}(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)) &= x_{j_1}(y^*) + \bar{\alpha}_k x_{j_1}'(y^*; d_y) + \bar{\alpha}_k^2 x_{j_1}''(y^*; d_y, d_y^1(\bar{\alpha}_k)) + o(\bar{\alpha}_k^2), \\ x_{j_2}(y^* + \bar{\alpha}_k d_y + \bar{\alpha}_k^2 d_y^1(\bar{\alpha}_k)) &= x_{j_2}(y^*) + \bar{\alpha}_k x_{j_2}'(y^*; d_y) + \bar{\alpha}_k^2 x_{j_2}''(y^*; d_y, d_y^1(\bar{\alpha}_k)) + o(\bar{\alpha}_k^2). \end{aligned}$$

Subtracting the second equation from the first and using (6), we obtain

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

In view of (5) and the above relation, to complete the proof for a given arbitrary sequence, it suffices to show for the sequence $\{\alpha_k\} \downarrow 0$ with $\{\alpha_k\} \subseteq R_j$, that

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

For this sequence $\{\alpha_k\}$, consider the path $y^* + sd_y + s^2 d_y^1(\alpha_k)$ for $s \geq 0$ with α_k fixed. For sufficiently large k, define the corresponding active index set

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

Since $\|s_{l,k} d_y^1(\alpha_k)\| \leq \|\alpha_k d_y^1(\alpha_k)\|$ and $\{\alpha_k d_y^1(\alpha_k)\} \to 0$, it follows that $\{s_{l,k} d_y^1(\alpha_k)\} \to 0$ as $k \to \infty$. As $x(\cdot)$ is piecewise smooth, $\mathscr{I}^k \neq \emptyset$ for all k. Furthermore, because $\{\mathscr{I}^k\}$ is a sequence of non-empty subsets of the finite

set $\{1,\dots,N\}$, there exist an index $t \in \{1,\dots,N\}$, a subsequence $\{k_n\} \subseteq \mathbb{N}$, and an integer $N_0 \in \mathbb{N}$ such that $t \in \bigcap_{n\geq N_0} \mathscr{I}^{k_n}$.

For any two selection indices $j_1, j_2 \in \{1,\dots,N\}$, we define a relation $j_1 \circ j_2$ along the subsequence $\{k_n\}$ if there exists $s_{l,k_n} \in (0, \alpha_{k_n}]$ such that

$$x(y^* + s_{l,k_n} d_y + s_{l,k_n}^2 d_y^1(\alpha_{k_n})) = x_{j_1}(y^* + s_{l,k_n} d_y + s_{l,k_n}^2 d_y^1(\alpha_{k_n})) = x_{j_2}(y^* + s_{l,k_n} d_y + s_{l,k_n}^2 d_y^1(\alpha_{k_n})), \quad \forall n \geq N_0.$$

Since $x_{j_1}, x_{j_2} \in C^2$, their Taylor expansions along $\{y^* + s_{l,k_n} d_y + s_{l,k_n}^2 d_y^1(\alpha_{k_n})\}$ yield

$$x_{j_1}(y^* + s_{l,k_n} d_y + s_{l,k_n}^2 d_y^1(\alpha_{k_n})) = x_{j_1}(y^*) + s_{l,k_n} x'_{j_1}(y^*; d_y) + s_{l,k_n}^2 x''_{j_1}(y^*; d_y, d_y^1(\alpha_{k_n})) + o(s_{l,k_n}^2), \quad \forall n \geq N_0,$$
$$x_{j_2}(y^* + s_{l,k_n} d_y + s_{l,k_n}^2 d_y^1(\alpha_{k_n})) = x_{j_2}(y^*) + s_{l,k_n} x'_{j_2}(y^*; d_y) + s_{l,k_n}^2 x''_{j_2}(y^*; d_y, d_y^1(\alpha_{k_n})) + o(s_{l,k_n}^2), \quad \forall n \geq N_0.$$

Subtracting these two expansions (using $x_{j_1}(y^*) = x_{j_2}(y^*) = x(y^*)$ and $x'_{j_1}(y^*; d_y) = x'_{j_2}(y^*; d_y) = x'(y^*; d_y)$), dividing by s_{l,k_n}^2, and taking $n \to \infty$, we obtain

$$j_1 \circ j_2 \implies \lim_{n\to\infty} \left(x''_{j_1}(y^*; d_y, d_y^1(\alpha_{k_n})) - x''_{j_2}(y^*; d_y, d_y^1(\alpha_{k_n})) \right) = 0.$$

We now claim that j and t can be connected by a finite chain of relations $j = i_0 \circ i_1 \circ \cdots \circ i_M = t$ along a further subsequence $\{\bar{k}_n\} \subseteq \{k_n\}$, where $j \in \mathscr{I}$ and $t \in \bigcap_{n\geq N_0} \mathscr{I}^{\bar{k}_n}$.

For each $n \geq N_0$, consider the continuous path $\gamma_{k_n} : [0, \alpha_{k_n}] \to \mathbb{R}^m$ defined by $\gamma_{k_n}(s) = y^* + s d_y + s^2 d_y^1(\alpha_{k_n})$. Let $A_i^{k_n} := \{s \in [0, \alpha_{k_n}] \ : \ x_i(\gamma_{k_n}(s)) = x(\gamma_{k_n}(s))\}$ for $i = 1,\dots,N$. Due to the piecewise and continuous nature of $x(\cdot)$, the collection $\{A_1^{k_n}, \dots, A_N^{k_n}\}$ forms a finite closed cover of the connected interval $[0, \alpha_{k_n}]$, i.e., $[0, \alpha_{k_n}] = \bigcup_{i=1}^N A_i^{k_n}$.

Fixing $s_0 := \alpha_{k_n} \in A_j^{k_n}$ (since $j \in \mathscr{I}$) and $i_0 := j$, define recursively

$$s_{p+1} := \inf \left\{ s \in [s_{l,k_n}, s_p] \ : \ s \in A_{i_p}^{k_n} \right\},$$

where $s_{l,k_n} = \sup\{s \in [0, \alpha_{k_n}] \ : \ s \in A_t^{k_n}\}$. If $s_{p+1} > s_{l,k_n}$, the connectedness of $[s_{l,k_n}, \alpha_{k_n}]$ and closedness of the covering sets ensure that $s_{p+1} \in A_{i_p}^{k_n} \cap A_{i_{p+1}}^{k_n}$ for some $i_{p+1} \in \{1,\dots,N\}$, which gives

$$x_{i_p}(\gamma_{k_n}(s_{p+1})) = x_{i_{p+1}}(\gamma_{k_n}(s_{p+1})) = x(\gamma_{k_n}(s_{p+1})).$$

The minimality of s_{p+1} and continuity of $x_{i_p}(\cdot)$ ensure that $x_{i_p}(\gamma_{k_n}(s)) \neq x(\gamma_{k_n}(s))$ for $s_{l,k_n} \leq s < s_{p+1}$. Because $\{1,\dots,N\}$ is finite, this procedure terminates at $s_{M_{k_n}+1} = s_{l,k_n}$ in at most $M_{k_n} \leq N-1$ steps with $i_{M_{k_n}} = t \in \mathscr{I}^{k_n}$, yielding an ordered sequence of indices $\{j, i_1^{k_n}, \dots, i_{M_{k_n}-1}^{k_n}, t\} \subseteq \{1,\dots,N\}$ for each $n \geq N_0$. Since the index set $\{1,\dots,N\}$ is finite, there exists a subsequence $\{\bar{k}_n\} \subseteq \{k_n\}$ along which this sequence of indices is constant: $j = i_0, i_1, \dots, i_{M_{\bar{k}_n}} = t$. Hence, $j = i_0 \circ i_1 \circ \cdots \circ i_{M_{\bar{k}_n}} = t$, and using the transitivity over this chain, we get

$$\lim_{n\to\infty} \left(x''_j(y^*; d_y, d_y^1(\alpha_{\bar{k}_n})) - x''_t(y^*; d_y, d_y^1(\alpha_{\bar{k}_n})) \right) = 0. \tag{7}$$

Since $x(\cdot)$ is second-order directionally differentiable (Theorem 3.2), for fixed $\alpha_{\bar{k}_n}$ and along $\{s_{l,\bar{k}_n}\} \downarrow 0$, we have

$$x(y^* + s_{l,\bar{k}_n} d_y + s_{l,\bar{k}_n}^2 d_y^1(\alpha_{\bar{k}_n})) = x(y^*) + s_{l,\bar{k}_n} x'(y^*; d_y) + s_{l,\bar{k}_n}^2 x''(y^*; d_y, d_y^1(\alpha_{\bar{k}_n})) + o(s_{l,\bar{k}_n}^2), \quad \forall l.$$

Because $x(\gamma_{\bar{k}_n}(s_{l,\bar{k}_n})) = x_t(\gamma_{\bar{k}_n}(s_{l,\bar{k}_n}))$ and $x'_t(y^*; d_y) = x'(y^*; d_y)$, subtracting the C^2 Taylor expansion of $x_t(\cdot)$ yields

$$\lim_{n\to\infty} \left(\lim_{l\to\infty} \frac{x_t(y^* + s_{l,\bar{k}_n} d_y + s_{l,\bar{k}_n}^2 d_y^1(\alpha_{\bar{k}_n})) - x(y^*) - s_{l,\bar{k}_n} x'_t(y^*; d_y) - s_{l,\bar{k}_n}^2 x''(y^*; d_y, d_y^1(\alpha_{\bar{k}_n}))}{s_{l,\bar{k}_n}^2} \right) = 0.$$

Evaluating the inner limit gives

$$\lim_{n\to\infty} \left(x''_t(y^*; d_y, d_y^1(\alpha_{\bar{k}_n})) - x''(y^*; d_y, d_y^1(\alpha_{\bar{k}_n})) \right) = 0.$$

Combining this with (7), we arrive at

$$\lim_{n\to\infty} \left(x''_j(y^*; d_y, d_y^1(\alpha_{\bar{k}_n})) - x''(y^*; d_y, d_y^1(\alpha_{\bar{k}_n})) \right) = 0.$$

By the standard subsequence principle, since every sequence $\{\alpha_k\} \downarrow 0$ contains a subsequence along which the error term in (5) converges to 0, the entire sequence converges to 0 as $k \to \infty$. Thus, $x(\cdot)$ is second-order gph-regular at y^*. □

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 expansions of the functions in the KKT conditions, we require some second-order differentiability property associated with the multiplier set $V(d_y, d_y^1)$.

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 the sequence along the parabolic paths. We omit the details. □

From the above lemma, there is a suitable subsequence of multipliers such that $\mu^k \in M(y^k)$, $\lim_{k\to\infty} \frac{\mu^k - \mu}{\alpha_k} = d_\mu$ 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 $\{\alpha_k\}$ itself. Since $x(y^k)$ solves the problem LP_{y^k} at y^k and $\mu^k \in M(y^k)$,

$$\nabla_x L(x(y^k), y^k, \mu^k) = 0, \tag{8a}$$
$$g_i(x(y^k), y^k) = 0, \quad \mu_i^k > 0, \quad i \in J(\mu^k), \tag{8b}$$
$$g_i(x(y^k), y^k) \le 0, \quad \mu_i^k = 0, \quad i \notin J(\mu^k). \tag{8c}$$

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

$$\begin{aligned} &\nabla_x L(x(y^k), y^k, \mu^k)^T \\ &\quad = \nabla_x L(x(y), y, \mu)^T + \alpha_k \Big(\nabla_{xx}^2 L(x(y), y, \mu) d_x + \nabla_{yx}^2 L(x(y), y, \mu) d_y + \nabla_{\mu x}^2 L(x(y), y, \mu) d_\mu \Big) \\ &\qquad + \alpha_k^2 \left(\frac{1}{2} \begin{pmatrix} d_z^T & d_\mu^T \end{pmatrix} \widehat{\nabla}^2 (\nabla_x L(x(y), y, \mu)^T) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} + \nabla_{xx}^2 L(x(y), y, \mu) d_x^1 \right. \\ &\qquad\qquad \left. + \nabla_{yx}^2 L(x(y), y, \mu) d_y^1 + \nabla_{\mu x}^2 L(x(y), y, \mu) d_\mu^1 \right) + e_n o(\alpha_k^2), \ \forall k, \end{aligned} \tag{9}$$

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

$$\begin{aligned} g_i(x(y^k), y^k) =& g_i(x(y), y) + \alpha_k \Big(\nabla_x g_i(x(y), y) d_x + \nabla_y g_i(x(y), y) d_y \Big) \\ &+ \alpha_k^2 \Big(\frac{1}{2} d_z^T \nabla^2_{zz} g_i(x(y), y) d_z + \nabla_z g_i(x(y), y) d_z^1 \Big) + o(\alpha_k^2),\ \forall k. \end{aligned} \tag{10}$$

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

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

From (8a), 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 (9),

$$\nabla^2_{xx} L(x(y), y, \mu) d_x + \nabla^2_{yx} L(x(y), y, \mu) d_y + \nabla^2_{\mu x} L(x(y), y, \mu) d_\mu = 0. \tag{11a}$$

As $J(\mu) \subseteq J(\mu^k)$ for sufficiently large k, (8b) holds for $i \in J(\mu)$. Moreover, since $g_i(x(y^k), y^k) = 0,\ i \in J(\mu)$ by (8b) 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 (10),

$$\nabla_x g_i(x(y), y) d_x + \nabla_y g_i(x(y), y) d_y = 0, \quad i \in J(\mu). \tag{11b}$$

In addition, for $i \in I_g(y) \setminus J(\mu)$, it follows from (8c) that $\lim\limits_{k\to\infty} \frac{g_i(x(y^k), y^k) - g_i(x(y), y)}{\alpha_k} \leq 0$. Then, from (10),

$$\nabla_x g_i(x(y), y) d_x + \nabla_y g_i(x(y), y) d_y \leq 0, \quad i \in I_g(y) \setminus J(\mu). \tag{11c}$$

Since $\mu_i = 0$ for all $i \in I_g(y) \setminus J(\mu)$ and $\mu_i^k \geq 0$ for $i \in \Lambda_q$, we obtain $d_{\mu_i} = \lim_{k\to\infty} \frac{\mu_i^k - \mu_i}{\alpha_k} \geq 0$. 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 (8b), we have $g_i(x(y^k), y^k) = 0$ for sufficiently large k. Thus, using (10), 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 (8c) that $\mu_i^k = 0$ for sufficiently large k, and therefore $d_{\mu_i} = 0$. Thus,

$$\left.\begin{aligned} & d_{\mu_i} \left(\nabla_x g_i(x(y), y) d_x + \nabla_y g_i(x(y), y) d_y \right) = 0, \\ & d_{\mu_i} \geq 0,\ i \in I_g(y) \setminus J(\mu), \quad d_{\mu_i} = 0, i \notin I_g(y). \end{aligned}\right\} \tag{11d}$$

It can be verified that the expressions (11a)-(11d) 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 (11a)-(11d) are the KKT optimality conditions of $QP_\mu(y; d_y)$ so the system (11a)-(11d) 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 (11a)-(11d) then $d_x = x'(y; d_y)$.

We now proceed with a similar argument as in the formulation of (11a)-(11d) with the new index sets $J(\mu; d_\mu)$ and $I_g(y; d_y)$ as

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

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

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 (12a)-(12d) can be derived using (8a)-(10) by considering the above limits and index sets. This is analogous to the process for the derivation of (11a)-(11d). Therefore, we omit the detailed derivation.

$$\frac{1}{2}\begin{pmatrix} d_z^T & d_\mu^T \end{pmatrix} \widehat{\nabla}^2(\nabla_x L(x(y), y, \mu)^T) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} + \widehat{\nabla}(\nabla_x L(x(y), y, \mu)^T) \begin{pmatrix} d_z^1 \\ d_\mu^1 \end{pmatrix} = 0, \tag{12a}$$

$$\frac{1}{2} d_z^T \nabla^2_{zz} g_i(x(y), y)d_z + \nabla_z g_i(x(y), y)d_z^1 = 0, \quad i \in J(\mu; d_\mu), \tag{12b}$$

$$\frac{1}{2} d_z^T \nabla^2_{zz} g_i(x(y), y)d_z + \nabla_z g_i(x(y), y)d_z^1 \le 0, \quad i \in I_g(y; d_y) \setminus J(\mu; d_\mu), \tag{12c}$$

$$\left.\begin{aligned} & d^1_{\mu_i}\left(\frac{1}{2} d_z^T \nabla^2_{zz} g_i(x(y), y)d_z + \nabla_z g_i(x(y), y)d_z^1\right) = 0, \\ & d^1_{\mu_i} \ge 0, \quad i \in I_g(y; d_y) \setminus J(\mu; d_\mu), \ d^1_{\mu_i} = 0, \quad i \notin I_g(y; d_y). \end{aligned}\right\} \tag{12d}$$

Define the set of extreme multipliers associated with the KKT conditions (11a)–(11d) by

$$EM_\mu = \left\{ d_\mu : \begin{array}{l} (x'(y; d_y), d_\mu) \text{ solves (11a)–(11d) for a given } \mu \in EM(y), \\ (\nabla_x g_i(x(y), y))_{i\in J(\mu)\cup J(d_\mu)} \text{ is linearly independent.} \end{array} \right\}.$$

The set EM_μ is indeed the vertex set of the multipliers of the optimization problem $QP_\mu(y; d_y)$. Consider the following linear programming problem corresponding to the direction vectors d_y and d_y^1,

$$\max_{\mu\in S^1(d_y)} \sum_{i\in I_g(y)} \mu_i \left(\frac{1}{2}\begin{pmatrix} x'(y; d_y)^T & d_y^T \end{pmatrix} \nabla^2_{zz} g_i(x(y), y) \begin{pmatrix} x'(y; d_y) \\ d_y \end{pmatrix} + \nabla_y g_i(x(y), y)d_y^1\right) \tag{13}$$

and

$$S^2(d_y, d_y^1) = \{\mu \in S^1(d_y) \ : \ \mu \text{ solves (13)}\}.$$

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

The systems (11a)-(11d) and (12a)-(12d) are obtained from the second-order expansions (9)-(10) of the functions in KKT conditions of the lower-level problem corresponding to multipliers in $V(d_y, d_y^1)$. The first-order directional derivative $d_x (= x'(y; d_y))$ and the second-order directional derivative $d_x^1 (= x''(y; d_y, d_y^1))$ satisfy the systems (11a)-(11d) and (12a)-(12d). In general, the set $V(d_y, d_y^1)$ cannot be obtained explicitly without imposing restrictive assumptions. To derive the second-order directional derivative of the solution function, we introduce a set of multipliers $S^2(d_y, d_y^1)$, which acts as a tractable surrogate for the set $V(d_y, d_y^1)$ under certain assumptions. $S^2(d_y, d_y^1)$ can be obtained by solving the linear programming problem in (13). In Khatana and Panda [32], the second-order directional derivative along linear paths was characterized through a quadratic subproblem. Here, we extend this characterization to the parabolic paths. In particular, we obtain the second-order directional derivatives using the

multiplier set $S^2(d_y, d_y^1)$, which is proved in Lemma 4.2 and Theorem 4.1. This requires constructing the following perturbed parametric problem, which is important in the proof of Lemma 4.2 and Theorem 4.1. Let $y^k = y + \alpha_k d_y + \alpha_k^2 d_y^1$, where $\{\alpha_k\} \downarrow 0$ is a sequence of positive real numbers, and $G(y^k; d_y) \in \mathbb{R}^q$ be defined as

$$G_i(y^k; d_y) := \begin{cases} g_i(x(y^k), y^k) & \text{if } i \in \{i \in \Lambda_q \ : \ \exists \mu \in S^2(d_y, d_y^1),\ \mu_i > 0\} \\ 0 & \text{otherwise} \end{cases},$$

where $S^2(d_y, d_y^1)$ is computed at y. Consider the following problem $\bar{P}_{y^k}$, which is obtained from P_{y^k} by perturbing $g(x, y^k) \leq 0$ as $g(x, y^k) \leq G(y^k; d_y)$,

$$(\bar{P}_{y^k}) : \ \min_x f(x, y^k) \quad \text{s.t.} \quad g(x, y^k) \leq G(y^k; d_y).$$

Since $\bar{P}_{y^k}$ is a restricted problem than LP_{y^k} for which $x(y^k)$ is a feasible vector, it follows that $x^k = x(y^k)$ is an optimal solution to this perturbed problem. The KKT optimality conditions for $\bar{P}_{y^k}$ are

$$\left.\begin{array}{l} \nabla_x f(x, y^k) + \mu^T \nabla_x g(x, y^k) = 0, \quad \mu^T \left(g(x, y^k) - G(y^k; d_y)\right) = 0, \\ \mu \geq 0, \quad g(x, y^k) - G(y^k; d_y) \leq 0. \end{array}\right\} \tag{14}$$

Under the Assumptions $\mathcal{H}2$ and $\mathcal{H}3$ for the parametric problem LP_{y^k}, these KKT optimality conditions are necessary and sufficient for the perturbed problem. Let $\bar{M}(x(y^k); d_y)$ denote the set of Lagrange multipliers of $\bar{P}_{y^k}$,

$$\bar{M}(x(y^k); d_y) := \{\bar{\mu}^k \ : \ (x(y^k), \bar{\mu}^k) \text{ satisfy } (14)\}.$$

We say $\bar{\mu}^k \in \bar{M}(x(y^k); d_y)$ are the perturbed multipliers of the perturbed KKT system (14). For each multiplier $\mu \in S^2(d_y, d_y^1) \cap EM(y)$, we establish in Lemma 4.2 that there exists a sequence of perturbed multipliers $\{\bar{\mu}^k\}$ associated with the perturbed parametric problem $\bar{P}_{y^k}$ such that $\{\bar{\mu}^k\} \to \mu$. Hence, choosing $\mu \in S^2(d_y, d_y^1) \cap EM(y)$ is justified in the second-order Taylor expansion of the parametric problem. Theorem 4.1 demonstrates that the KKT system obtained for the multiplier set $S^2(d_y, d_y^1) \cap EM(y)$ characterizes the second-order directional derivatives of the solution function.

Lemma 4.2 *Let $y^k = y + \alpha_k d_y + \alpha_k^2 d_y^1$ for a sequence of positive real numbers $\{\alpha_k\}$ converging to 0, and assume Assumptions $(\mathcal{H}1)$–$(\mathcal{H}5)$ hold at $(x(y), y)$. Then,*

1. *$S^2(d_y, d_y^1) \cap EM(y)$ is a non-empty set.*
2. *Suppose $\mu \in S^2(d_y, d_y^1) \cap EM(y)$ and $d_\mu \in EM_\mu$. There exists a sequence of perturbed multipliers $\bar{\mu}^k \in \bar{M}(x(y^k); d_y)$ for each k such that*

$$\bar{\mu}^k = \mu + \alpha_k d_\mu + \alpha_k^2 d_\mu^1 + o(\alpha_k^2),$$

   for some $d_\mu^1 \in \mathbb{R}^q$.
3. *For any direction vectors $d_y, d_y^1 \in \mathbb{R}^m$,*

$$V(d_y, d_y^1) \subseteq S^2(d_y, d_y^1) \subseteq S^1(d_y) \subseteq M(y).$$

   In particular, $V(d_y, d_y^1) \cap S^2(d_y, d_y^1) \neq \emptyset$.

Proof **(1).** Recall that $S^1(d_y)$ is the set of optimal solutions to the linear programming problem,

$$\max_{\mu \in M(y)} \sum_{i \in I_g(y)} \mu_i \nabla_y g_i(x(y), y) d_y.$$

Under Assumption $(\mathcal{H}2)$ (MFCQ), $M(y)$ is a non-empty, compact polyhedral set. Hence, the above linear program over the constraint set $M(y)$ attains a finite maximum at an extreme point of $M(y)$, implying that $S^1(d_y)$ is a non-empty, compact polyhedral set and that every extreme point of $S^1(d_y)$ belongs to $EM(y)$.
The set $S^2(d_y, d_y^1)$ is defined as the solution set of the linear program in (13). Since $S^2(d_y, d_y^1)$ maximizes a

linear objective over the non-empty, compact polyhedral set $S^1(d_y)$, $S^2(d_y, d_y^1)$ is a non-empty polyhedral set with every extreme point also an extreme point of $S^1(d_y)$. Hence, $S^2(d_y, d_y^1) \cap EM(y) \neq \emptyset$ follows.

(2). By Remark 4.1, for each $\mu \in S^1(d_y)$, the solution set of the first-order system (11a)–(11d) is nonempty, which implies that $EM_\mu \neq \emptyset$. Let $\mu \in S^2(d_y, d_y^1) \cap EM(y) \subseteq S^1(d_y)$ and $d_\mu \in EM_\mu$. Since $\mu \in EM(y)$ and $d_\mu \in EM_\mu$, one can select an index set $J^* \supseteq J(\mu) \cup J(d_\mu)$ such that the gradient family $\{\nabla_x g_i(x(y), y)\}_{i \in J^*}$ is linearly independent. Consider the parameterized equality-constrained optimization problem $P^*(\alpha)$

$$P^*(\alpha): \quad \min_x f(x, \gamma(\alpha)) \quad \text{s.t.} \quad g_i(x, \gamma(\alpha)) = G_i(\gamma(\alpha); d_y), \quad i \in J^*,$$

where $\gamma(\alpha) = y + \alpha d_y + \alpha^2 d_y^1$. At $\alpha = 0$, $x(y)$ is the optimal solution and $\mu^* = (\mu_i)_{i \in J^*}$ is the associated Lagrange multiplier vector.
Since Assumption ($\mathcal{H}3$) holds and the gradients $\{\nabla_x g_i(x(y), y)\}_{i \in J^*}$ are linearly independent, standard sensitivity analysis (see Corollary 3.2.5 of Fiacco [26]) applies. Thus, there exist unique C^2 vector functions $x^*(\gamma(\alpha))$ and $\mu^*(\gamma(\alpha))$ for $\alpha \in [0, \alpha_0)$ satisfying the KKT conditions of $P^*(\alpha)$ with $\mu^*(\gamma(0)) = \mu^*$ and $\frac{d}{d\alpha}\mu^*(\gamma(0)) = d_\mu^*$ and $\frac{1}{2}\frac{d^2}{d\alpha^2}(\mu^*(\gamma(0)) = d_\mu^{1,*}$for some $d_\mu^*, d_\mu^{1,*} \in \mathbb{R}^{|J^*|}$.
For $y^k = \gamma(\alpha_k) = y + \alpha_k d_y + \alpha_k^2 d_y^1$ with sufficiently large k, define $\bar{\mu}^k \in \mathbb{R}^q$ by

$$\bar{\mu}_i^k = \begin{cases} (\mu^*(y^k))_i & \text{if } i \in J^*, \\ 0 & \text{if } i \notin J^*. \end{cases}$$

By construction, $(x(y^k), \bar{\mu}^k)$ satisfies the perturbed KKT system (14) for large k, ensuring $\bar{\mu}^k \in \bar{M}(x(y^k); d_y)$. Since $\mu^*(\cdot)$ is of class C^2, expanding $\mu^*(y^k)$ about $\alpha = 0$ along $\gamma(\alpha_k)$ up to second-order gives

$$(\mu^*(y^k))_i = \mu_i + \alpha_k d_{\mu_i}^* + \alpha_k^2 d_{\mu_i}^{1,*} + o(\alpha_k^2), \quad i \in J^*.$$

Hence, we have

$$\lim_{k \to \infty} \frac{\bar{\mu}_i^k - \mu_i}{\alpha_k} = \begin{cases} d_{\mu_i}^*, & \text{if } i \in J^*, \\ 0, & \text{if } i \in \Lambda_q \setminus J^*, \end{cases}$$

and

$$\lim_{k \to \infty} \frac{\bar{\mu}_i^k - \mu_i - \alpha_k d_{\mu_i}^*}{\alpha_k^2} = \begin{cases} d_{\mu_i}^{1*}, & \text{if } i \in J^*, \\ 0, & \text{if } i \in \Lambda_q \setminus J^*. \end{cases}$$

It only remains to show that $(d_\mu)_i = d_{\mu_i}^*$ for all $i \in J^*$.
Expanding $\nabla_x L(x(y^k), y^k, \bar{\mu}^k) = 0$ along the sequence $\{(x(y^k), y^k, \bar{\mu}^k)\}$ about $(x(y), y, \mu)$ and using $\nabla_x L(x(y), y, \mu) = 0$, yields

$$\nabla_{xx}^2 L(x(y), y, \mu) d_x + \nabla_{yx}^2 L(x(y), y, \mu) d_y + \sum_{i \in J^*} \nabla_x g_i(x(y), y) d_{\mu_i}^* = 0. \tag{15}$$

On the other hand, since $d_\mu \in EM_\mu$, the pair (d_x, d_μ) satisfies the first-order stationary condition (11a),

$$\nabla_{xx}^2 L(x(y), y, \mu) d_x + \nabla_{yx}^2 L(x(y), y, \mu) d_y + \sum_{i \in J^*} \nabla_x g_i(x(y), y) d_{\mu_i} = 0, \tag{16}$$

with $d_{\mu_i} = 0$ for $i \notin J^*$. Subtracting (15) from (16) gives

$$\sum_{i \in J^*} \nabla_x g_i(x(y), y)\big(d_{\mu_i} - d_{\mu_i}^*\big) = 0.$$

By the linear independence of the active constraint gradients $\{\nabla_x g_i(x(y), y)\}_{i \in J^*}$, it follows that $d_{\mu_i} = d_{\mu_i}^*$ for all $i \in J^*$, which proves our claim. This completes the proof.

(3). The proof of this part is provided in Appendix. □

Theorem 4.1 *Suppose Assumptions (H1)–(H5) are satisfied at the given point $(x(y), y)$. Then,*

1. *(a) Suppose $\mu \in EM(y)$, and the feasible set of the quadratic programming problem of Part (2) is nonempty, then $\mu \in S^2(d_y, d_y^1)$.*
   (b) $d_x^1 = x''(y; d_y, d_y^1)$ satisfies the following system for each $\mu \in S^2(d_y, d_y^1)$,

$$\frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z + \nabla_z g_i(x(y), y) \begin{pmatrix} d_x^1 \\ d_y^1 \end{pmatrix} = 0, \quad i \in J(\mu),$$

$$\frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z + \nabla_z g_i(x(y), y) \begin{pmatrix} d_x^1 \\ d_y^1 \end{pmatrix} \leq 0, \quad i \in I_g(y; d_y) \setminus J(\mu).$$

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

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

for $\mu \in S^2(d_y, d_y^1) \cap EM(y)$ and $d_\mu \in EM_\mu$.

Proof **(1(a).)** 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).

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

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

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

$K_\mu^1(y; d_y)$ is the set of KKT multiplier vectors associated with a primal optimal solution $\mu \in S^1(d_y)$, as shown in Lemma 2.2 of Dempe [29]. Furthermore, since $x'(y; d_y) \in K_\mu^1(y; d_y)$ satisfies the KKT optimality conditions of (P^1) for every $\mu \in S^1(d_y)$ (Theorem 2.1). Since (P^1) is a linear programming problem, its primal solution set satisfying the KKT optimality conditions is independent of the specific choice of the KKT multiplier vector d_w. Hence, by fixing $d_w = x'(y; d_y)$ in the above KKT system, problem (13) can be written as

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

The KKT optimality conditions for (P^2) are

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

where d_x^1 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 ${d_x^1}^*$ be a feasible point of the quadratic problem for given d_z and d_y^1. Define

$$b_i^* := - \left(\frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z + \nabla_y g_i(x(y), y) d_y^1 + \nabla_x g_i(x(y), y) {d_x^1}^* \right).$$

Then the pair $(d_x^{1,*}, b^*)$ satisfies the above KKT optimality conditions. Since $S^2(d_y, d_y^1)$ is the optimal solution set of Problem (P^2), and μ satisfies the KKT system for the multiplier pair $(d_x^{1,*}, b^*)$, it follows that $\mu \in S^2(d_y, d_y^1)$.

(1(b).) $S^2(d_y, d_y^1)$ is the optimal solution set of the linear programming problem (P^2) and (d_x^1, b) is the corresponding Lagrange multiplier of (P^2). Eliminating the vector b, the multiplier set reduces to

$$\left\{ d_x^1 \in \mathbb{R}^n \ : \begin{array}{ll} \frac{1}{2}\, d_z^T \nabla^2_{zz} g_i(x(y),y) d_z + \nabla_z g_i(x(y),y) \begin{pmatrix} d_x^1 \\ d_y^1 \end{pmatrix} = 0, & i \in J(\mu), \\ \frac{1}{2}\, d_z^T \nabla^2_{zz} g_i(x(y),y) d_z + \nabla_z g_i(x(y),y) \begin{pmatrix} d_x^1 \\ d_y^1 \end{pmatrix} \le 0, & i \in I_g(y; d_y) \setminus J(\mu) \end{array} \right\}.$$

It follows from the duality properties of linear programming that the above multiplier set is invariant for every choice of primal solution $\mu \in S^2(d_y, d_y^1)$. In view of (12b)–(12c), for a given $\mu \in V(d_y, d_y^1)$, the directional derivative $x''(y; d_y, d_y^1)$ satisfies the above multiplier set. Moreover, since $V(d_y, d_y^1) \cap S^2(d_y, d_y^1) \neq \emptyset$ (from Lemma 4.2(3)), $d_x^1 = x''(y; d_y, d_y^1)$ satisfies this system for each $\mu \in S^2(d_y, d_y^1)$.

(2) From Assumption $(\mathcal{H}3)$, it follows that

$$w^T \nabla^2_{xx} L(x(y), y, \mu) w > 0, \quad \forall w \in \mathcal{C}(y; d_y) \setminus \{0\},$$

where the critical cone $\mathcal{C}(y; d_y)$ is given by

$$\mathcal{C}(y; d_y) := \left\{ w \in \mathbb{R}^n \ : \begin{array}{ll} \nabla_x g_i(x(y),y) w = 0, & \forall i \in J(\mu; d_\mu), \\ \nabla_x g_i(x(y),y) w \le 0, & \forall i \in I_g(y; d_y) \setminus J(\mu; d_\mu) \end{array} \right\}.$$

Hence, the Hessian matrix of the objective function of the quadratic subproblem is positive definite on the tangent space to the constraints, which ensures that the quadratic subproblem is strictly convex and admits a unique optimal solution.

To prove that $x''(y; d_y, d_y^1)$ solves this quadratic problem, by convexity it is sufficient to show that the associated variational inequality holds at $x''(y; d_y, d_y^1)$ for every feasible vector w, that is,

$$\left(\frac{1}{2} \begin{pmatrix} d_z^T & d_\mu^T \end{pmatrix} \widehat{\nabla}^2 \left(\nabla_x L(x(y), y, \mu)^T \right) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} + \nabla^2_{yx} L(x(y), y, \mu) d_y^1 + \nabla^2_{xx} L(x(y), y, \mu) x''(y; d_y, d_y^1) \right)^T \left(w - x''(y; d_y, d_y^1) \right) \ge 0. \quad (20)$$

Since $(x'(y; d_y), d_\mu)$ solves the first-order KKT system (11a)–(11d) with $d_\mu \in EM_\mu$ at $(x(y), y, \mu)$, by Lemma 4.2(2) there exists a sequence of perturbed multipliers $\bar{\mu}^k \in \bar{M}(x(y^k); d_y)$ such that $\lim_{k\to\infty} \frac{\bar{\mu}^k - \mu}{\alpha_k} = d_\mu$ along the parabolic path $y^k = y + \alpha_k d_y + \alpha_k^2 d_y^1$.

From the stationary condition for each k and $\mu \in S^2(d_y, d_y^1) \subseteq M(y)$, we have

$$\nabla_x L(x(y^k), y^k, \bar{\mu}^k) = \nabla_x f(x(y^k), y^k) + \sum_{i=1}^q \bar{\mu}_i^k \nabla_x g_i(x(y^k), y^k) = 0, \quad (21)$$

$$\nabla_x L(x(y), y, \mu) = \nabla_x f(x(y), y) + \sum_{i=1}^q \mu_i \nabla_x g_i(x(y), y) = 0. \quad (22)$$

Applying the second-order Taylor expansion of $\nabla_x L$ about $(x(y), y, \mu)$ along the sequence $\{(x(y^k), y^k, \bar{\mu}^k)\}$, with $x(y^k) = x(y) + \alpha_k x'(y; d_y) + \alpha_k^2 x''(y; d_y, d_y^1) + o(\alpha_k^2)$, and using (21) and (22), we obtain

$$\frac{1}{2} \begin{pmatrix} d_z^T & d_\mu^T \end{pmatrix} \widehat{\nabla}^2 \left(\nabla_x L(x(y), y, \mu)^T \right) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} + \nabla^2_{yx} L(x(y), y, \mu) d_y^1 + \nabla^2_{xx} L(x(y), y, \mu) x''(y; d_y, d_y^1) + \lim_{k\to\infty} \sum_{i \in I_g(y)} \frac{\bar{\mu}_i^k - \mu_i - \alpha_k d_{\mu_i}}{\alpha_k^2} \nabla_x g_i(x(y), y)^T = 0.$$

Taking the inner product with $(w - x''(y; d_y, d_y^1))$ yields

$$\left(\frac{1}{2} \begin{pmatrix} d_z^T & d_\mu^T \end{pmatrix} \widehat{\nabla}^2 \left(\nabla_x L(x(y), y, \mu)^T \right) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} + \nabla^2_{yx} L(x(y), y, \mu) d_y^1 + \nabla^2_{xx} L(x(y), y, \mu) x''(y; d_y, d_y^1) \right)^T \left(w - x''(y; d_y, d_y^1) \right)$$
$$= - \lim_{k\to\infty} \sum_{i \in I_g(y)} \frac{\bar{\mu}_i^k - \mu_i - \alpha_k d_{\mu_i}}{\alpha_k^2} \nabla_x g_i(x(y), y) \left(w - x''(y; d_y, d_y^1) \right). \quad (23)$$

Define the index set $\mathcal{I}_{S^2} := \{ i \in \Lambda_q \ : \ \exists \mu \in S^2(d_y, d_y^1) \text{ with } \mu_i > 0 \}$. To evaluate the right-hand side of (23), we partition the active index set $I_g(y)$ into three disjoint subsets:

(i) For $i \in J(\mu; d_\mu)$***:***

From Part (1(b)), $d_x^1 = x''(y; d_y, d_y^1)$ satisfy the second-order equality constraint

$$\nabla_x g_i(x(y), y) d_x^1 + \nabla_y g_i(x(y), y) d_y^1 + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z = 0 \quad \text{for } i \in J(\mu)$$

Furthermore, for $i \in J(\mu; d_\mu) \setminus J(\mu)$, if $i \in \mathcal{I}_{S^2}$, this equality holds for $d_x^1 = x''(y; d_y, d_y^1)$ by Part (1(b)). If $i \notin \mathcal{I}_{S^2}$, we have $\mu_i = 0$ and $d_{\mu_i} > 0$, implying $\bar{\mu}_i^k > 0$ and $G_i(y^k; d_y) = 0$ for all sufficiently large k. The complementarity condition $\bar{\mu}_i^k (g_i(x(y^k), y^k) - G_i(y^k; d_y)) = 0$ in (14) then implies $g_i(x(y^k), y^k) = 0$, ensuring the second-order equality holds for $i \notin I_{S^2}$. Hence, both $d_x^1 = x''(y; d_y, d_y^1)$ and feasible vector $d_x^1 = w$ satisfy

$$\nabla_x g_i(x(y), y) d_x^1 + \nabla_y g_i(x(y), y) d_y^1 + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z = 0, \quad \forall i \in J(\mu; d_\mu).$$

This implies

$$\nabla_x g_i(x(y), y) \left(w - x''(y; d_y, d_y^1) \right) = 0, \quad \forall i \in J(\mu; d_\mu). \tag{24}$$

Since $\mu_i = 0$ and $d_{\mu_i} = 0$ for all $i \in I_g(y) \setminus J(\mu; d_\mu)$, substituting (24) into (23) eliminates the terms in $J(\mu; d_\mu)$, which reduces to

$$\left(\frac{1}{2} \begin{pmatrix} d_z^T & d_\mu^T \end{pmatrix} \widehat{\nabla}^2 \left(\nabla_x L(x(y), y, \mu)^T \right) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} + \nabla_{yx}^2 L(x(y), y, \mu) d_y^1 + \nabla_{xx}^2 L(x(y), y, \mu) x''(y; d_y, d_y^1) \right)^T \left(w - x''(y; d_y, d_y^1) \right)$$
$$= - \lim_{k \to \infty} \sum_{i \in I_g(y) \setminus J(\mu; d_\mu)} \frac{\bar{\mu}_i^k}{\alpha_k^2} \nabla_x g_i(x(y), y) \left(w - x''(y; d_y, d_y^1) \right). \tag{25}$$

(ii) For $i \in \mathcal{I}_{S^2} \setminus J(\mu; d_\mu)$***:***

For $i \in \mathcal{I}_{S^2} \setminus J(\mu; d_\mu)$, using Part (1(b)), $x''(y; d_y, d_y^1)$ satisfies the equality

$$\nabla_x g_i(x(y), y) x''(y; d_y, d_y^1) + \nabla_y g_i(x(y), y) d_y^1 + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z = 0,$$

whereas the feasible vector w satisfies the corresponding inequality

$$\nabla_x g_i(x(y), y) w + \nabla_y g_i(x(y), y) d_y^1 + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z \leq 0.$$

Subtracting the equality from the inequality yields $\nabla_x g_i(x(y), y)(w - x''(y; d_y, d_y^1)) \leq 0$. Since $\bar{\mu}_i^k \geq 0$, we have

$$- \frac{\bar{\mu}_i^k}{\alpha_k^2} \nabla_x g_i(x(y), y) \left(w - x''(y; d_y, d_y^1) \right) \geq 0, \quad \forall i \in \mathcal{I}_{S^2} \setminus J(\mu; d_\mu). \tag{26}$$

(iii) For $i \in I_g(y) \setminus (\mathcal{I}_{S^2} \cup J(\mu; d_\mu))$***:***

We consider two cases.

- If either $\nabla_x g_i(x(y), y) x'(y; d_y) + \nabla_y g_i(x(y), y) d_y < 0$ or

  $$\nabla_x g_i(x(y), y) x''(y; d_y, d_y^1) + \nabla_y g_i(x(y), y) d_y^1 + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z < 0,$$

  then by Taylor expansion $g_i(x(y^k), y^k) < 0$ for all sufficiently large k. Since $i \notin \mathcal{I}_{S^2}$, we have $G_i(y^k; d_y) = 0$. The complementarity condition in (14), $\bar{\mu}_i^k g_i(x(y^k), y^k) = 0$ implies $\bar{\mu}_i^k = 0$. Hence $\frac{\bar{\mu}_i^k}{\alpha_k^2} = 0$.

- Otherwise, $i \in I_g(y; d_y) \setminus (\mathcal{I}_{S^2} \cup J(\mu; d_\mu))$ and $x''(y; d_y, d_y^1)$ satisfies the second-order equality, that is,

  $$\nabla_x g_i(x(y), y) x''(y; d_y, d_y^1) + \nabla_y g_i(x(y), y) d_y^1 + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z = 0.$$

  Subtracting this from the inequality satisfied by w yields $\nabla_x g_i(x(y), y)(w - x''(y; d_y, d_y^1)) \leq 0$. In summary, for $i \in I_g(y) \setminus (\mathcal{I}_{S^2} \cup J(\mu; d_\mu))$, since $\bar{\mu}_i^k \geq 0$, we have

  $$- \lim_{k \to \infty} \frac{\bar{\mu}_i^k}{\alpha_k^2} \nabla_x g_i(x(y), y) \left(w - x''(y; d_y, d_y^1) \right) \geq 0. \tag{27}$$

Using (26) and (27), every term in the summation in (25) is non-negative. Combining all three cases, the variational inequality (20) holds. Consequently, $x''(y; d_y, d_y^1)$ is the unique optimal solution to the quadratic subproblem for any $\mu \in S^2(d_y, d_y^1) \cap EM(y)$ and $d_\mu \in EM_\mu$. □

From part (1) of Theorem 4.1, it follows that whenever the feasible set of the quadratic subproblem is non-empty and $\mu \in EM(y)$, we have $\mu \in S^2(d_y, d_y^1)$. On the other hand, the arguments used in part (2) of Theorem 4.1 imply that $x''(y; d_y, d_y^1)$ is feasible for the quadratic subproblem for any $\mu \in S^2(d_y, d_y^1) \cap EM(y)$. Thus, the following remark is justified.

Remark 4.2 Note that (12a)-(12d) are the KKT optimality conditions of the quadratic optimization problem in Theorem 4.1, where $d_x^1 = x''(y; d_y, d_y^1)$ is the optimal solution to this quadratic problem and $d_{\mu_i}^1,\ i \in I_g(y; d_y)$ is the corresponding KKT multipliers. Suppose $\mu \in EM(y)$, then the feasible set of the quadratic problem is nonempty for given $d_\mu \in EM_\mu$ if and only if $\mu \in S^2(d_y, d_y^1)$ (from the arguments used in Theorem 4.1). In that case, for given vector (d_z, d_μ) such that $d_\mu \in EM_\mu$ and $\mu \in EM(y)$, if (d_x^1, d_μ^1) solves the system (12a)-(12d) then $d_x^1 = x''(y; d_y, d_y^1)$.

From Remarks 4.1 and 4.2 it is clear that for given direction vectors d_y, d_y^1, the first- and second-order derivatives $x'(y; d_y)$ and $x''(y; d_y, d_y^1)$ satisfy the system (11a)-(11d) and (12a)-(12d) 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 (11d) and (12d) corresponding to some set I satisfying $J(\mu) \subseteq I \subseteq I_g(y)$ so that $\left(\nabla_x g_i(x(y), y)\right)_{i \in I}$ is linearly independent, we can reformulate (11a)-(12d) as (28a)-(28h), which is a system in $(d_x, d_\mu, d_x^1, d_\mu^1)$ at given point $(x(y), y, \mu)$ and direction vectors d_y, d_y^1 at y, which is proved in Proposition 4.1. Here, $d_x (= x'(y; d_y))$ and $d_x^1 (= x''(y; d_y, d_y^1))$ are the first- and second-order directional derivatives of the solution function $x(\cdot)$, respectively, and d_μ and d_μ^1 represent the first and second-order terms in the expansion of the multiplier vector μ along the parabolic path $y + \alpha d_y + \alpha^2 d_y^1$, respectively.

$$\nabla_{xx}^2 L(x(y), y, \mu) d_x + \nabla_{yx}^2 L(x(y), y, \mu) d_y + \nabla_{\mu x}^2 L(x(y), y, \mu) d_\mu = 0, \tag{28a}$$

$$\nabla_x g_i(x(y), y) d_x + \nabla_y g_i(x(y), y) d_y = 0, \quad i \in I, \tag{28b}$$

$$\nabla_x g_i(x(y), y) d_x + \nabla_y g_i(x(y), y) d_y \le 0, \quad i \in I_g(y) \setminus I, \tag{28c}$$

$$d_{\mu_i} \ge 0, \quad i \in I \setminus J(\mu), \qquad d_{\mu_i} = 0, \quad i \notin I \tag{28d}$$

$$\frac{1}{2} \begin{pmatrix} d_z^T & d_\mu^T \end{pmatrix} \widehat{\nabla}^2 (\nabla_x L(x(y), y, \mu)^T) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} + \widehat{\nabla} (\nabla_x L(x(y), y, \mu)^T) \begin{pmatrix} d_z^1 \\ d_\mu^1 \end{pmatrix} = 0, \tag{28e}$$

$$\nabla_x g_i(x(y), y)\, d_x^1 + \nabla_y g_i(x(y), y) d_y^1 + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z = 0, \quad i \in I, \tag{28f}$$

$$\nabla_x g_i(x(y), y)\, d_x^1 + \nabla_y g_i(x(y), y) d_y^1 + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z \le 0, \quad i \in I_g(y; d_y) \setminus I, \tag{28g}$$

$$d_{\mu_i}^1 \ge 0, \quad i \in I \setminus J(\mu; d_\mu), \quad d_{\mu_i}^1 = 0, \quad i \notin I. \tag{28h}$$

Proposition 4.1 (Equivalence of Active-Set Reformulation) *Suppose Assumptions* ($\mathcal{H}1$)-($\mathcal{H}5$) *hold at* $(x(y), y)$ *for* $\mu \in S^2(d_y, d_y^1) \cap EM(y)$ *and* $d_\mu \in EM_\mu$. *Then*

1. *For every solution* $(d_x, d_\mu, d_x^1, d_\mu^1)$ *of the first- and second-order systems* (11a)–(11d) *and* (12a)–(12d), *there exists at least one index set* I *satisfying* $J(\mu) \subseteq I \subseteq I_g(y)$ *such that the family of gradients* $\{\nabla_x g_i(x(y), y)\}_{i \in I}$ *is linearly independent and* $(d_x, d_\mu, d_x^1, d_\mu^1)$ *satisfies* (28a)–(28h).
2. *For any admissible index set* I *so that* $J(\mu) \subseteq I \subseteq I_g(y)$ *and* $\left(\nabla_x g_i(x(y), y)\right)_{i \in I}$ *is linearly independent, the system* (28a)–(28h) *has a unique solution* $(d_x, d_\mu, d_x^1, d_\mu^1)$, *where* $d_x = x'(y; d_y)$ *and* $d_x^1 = x''(y; d_y, d_y^1)$.
3. *For* $\|d_y\| \le \delta$ *and* $\|d_y^1\| \le \delta$ *with suitable sufficiently small* $\delta > 0$, *the system* (28a)–(28f) *combined with*

$$\nabla_x g_i(x(y), y)(d_x + d_x^1) + \nabla_y g_i(x(y), y)(d_y + d_y^1) + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z \le 0, \quad i \in I_g(y) \setminus I, \tag{29a}$$

$$d_{\mu_i} + d_{\mu_i}^1 \ge 0, \quad i \in I \setminus J(\mu), \qquad d_{\mu_i}^1 = 0, \quad i \notin I. \tag{29b}$$

is locally equivalent to (11a)–(11d) *and* (12a)–(12d).

Proof **(1).** Let $(d_x, d_\mu, d_x^1, d_\mu^1)$ be a solution satisfying systems (11a)–(11d) and (12a)–(12d) for a given $\mu \in S^2(d_y, d_y^1) \cap EM(y)$ and $d_\mu \in EM_\mu$ and d_μ^1 as extreme point of the multipliers of KKT system (12a)–(12d).

Choose the active index set $\hat{I}$ as

$$\hat{I} = \{i \in I_g(y) : \mu_i > 0 \text{ or } d_{\mu_i} > 0 \text{ or } d^1_{\mu_i} > 0\}.$$

By construction, the chain of inclusions $J(\mu) \subseteq J(\mu; d_\mu) \subseteq \hat{I} \subseteq I_g(y; d_y) \subseteq I_g(y)$ holds. Since $\mu \in EM(y)$, $d_\mu \in EM_\mu$ is a vertex solution of (11a)–(11d), and d^1_μ is a vertex solution of (12a)–(12d); therefore the gradient family $\{\nabla_x g_i(x(y), y)\}_{i \in \hat{I}}$ is linearly independent in $\mathbb{R}^n$. We derive (28a)–(28h) from (11a)–(11d) and (12a)–(12d).

(i) Equation (28a) coincides with the stationary condition (11a). To establish (28b), we consider any index $i \in \hat{I}$. If $i \in J(\mu)$, the equality $\nabla_x g_i(x(y), y)d_x + \nabla_y g_i(x(y), y)d_y = 0$ holds directly from (11b). If $i \in \hat{I} \setminus J(\mu)$ with $d_{\mu_i} > 0$, the complementarity condition (11d) ensures that $\nabla_x g_i(x(y), y)d_x + \nabla_y g_i(x(y), y)d_y = 0$. If $i \in \hat{I} \setminus J(\mu; d_\mu)$, then $d^1_{\mu_i} > 0$, which implies $i \in I_g(y; d_y)$ by (12d), and hence $\nabla_x g_i(x(y), y)d_x + \nabla_y g_i(x(y), y)d_y = 0$. Thus, (28b) holds for all $i \in \hat{I}$. Next, the inequality (28c) follows immediately from (11c). Finally, for $i \in \hat{I} \setminus J(\mu)$, non-negativity $d_{\mu_i} \geq 0$ holds by (11d), while for $i \notin \hat{I}$, we have $d_{\mu_i} = 0$, implying (28d).

(ii) Equation (28e) follows directly from the second-order stationary relation (12a). For equation (28f), let $i \in \hat{I}$. If $i \in J(\mu; d_\mu)$, the equality $\nabla_x g_i(x(y), y)d^1_x + \nabla_y g_i(x(y), y)d^1_y + \frac{1}{2}d_z^T \nabla^2_{zz} g_i(x(y), y)d_z = 0$ holds by (12b). If $i \in \hat{I} \setminus J(\mu; d_\mu)$, then $d^1_{\mu_i} > 0$, and the second-order complementarity condition (12d) implies that $\nabla_x g_i(x(y), y)d^1_x + \nabla_y g_i(x(y), y)d^1_y + \frac{1}{2}d_z^T \nabla^2_{zz} g_i(x(y), y)d_z = 0$. For $i \in I_g(y; d_y) \setminus \hat{I} \subseteq I_g(y; d_y) \setminus J(\mu; d_\mu)$, condition (12c) gives the second-order inequality (28g). Moreover, for $i \in \hat{I} \setminus J(\mu; d_\mu)$, non-negativity $d^1_{\mu_i} \geq 0$ holds by (12d), while for $i \notin \hat{I}$, we have $d^1_{\mu_i} = 0$, ensuring (28h).

(2). Let I be any fixed admissible index set satisfying $J(\mu) \subseteq I \subseteq I_g(y)$ such that $\{\nabla_x g_i(x(y), y)\}_{i \in I}$ is linearly independent. Suppose $(\tilde{d}_x, \tilde{d}_\mu, \tilde{d}^1_x, \tilde{d}^1_\mu)$ is an arbitrary solution to the active-set system (28a)–(28h). It is straightforward to verify that $(\tilde{d}_x, \tilde{d}_\mu, \tilde{d}^1_x, \tilde{d}^1_\mu)$ satisfies (11a)–(11d) and (12a)–(12d). By Theorem 2.1, $QP_\mu(y; d_y)$ is a strictly convex quadratic program under Assumption ($\mathcal{H}3$), so its KKT system (11a)–(11d) admits a unique solution $\tilde{d}_x = x'(y; d_y)$ and $\tilde{d}_\mu \in EM_\mu$ as vertex solution of the multiplier set. Similarly, by Theorem 4.1, the second-order parabolic subproblem is strictly convex under Assumption ($\mathcal{H}3$), ensuring that its KKT system (12a)–(12d) admits a unique primal solution $\tilde{d}^1_x = x''(y; d_y, d^1_y)$ and $\tilde{d}^1_\mu$ as a vertex solution of the corresponding multiplier set. Consequently, using the result of first part,(11a)–(11d) and (12a)–(12d) is equivalent to (28a)–(28h). Hence, any solution to system (28a)–(28h) must be unique.

(3). Let $(d_x, d_\mu, d^1_x, d^1_\mu)$ be the solution of the system (28a)–(28h) for given direction vectors d_y and d^1_y. Then, it can be easily verified that $(td_x, td_\mu, t^2 d^1_x, t^2 d^1_\mu)$ also solves (28a)–(28h) for any scalar $t > 0$ with directions td_y and $t^2 d^1_y$. For $i \in I_g(y) \setminus I_g(y; d_y)$, we have $\nabla_x g_i(x(y), y)d_x + \nabla_y g_i(x(y), y)d_y < 0$. Therefore, for all sufficiently small $t > 0$,

$$\nabla_x g_i(x(y), y)(td_x + t^2 d^1_x) + \nabla_y g_i(x(y), y)(td_y + t^2 d^1_y) + \frac{1}{2}t^2 d_z^T \nabla^2_{zz} g_i(x(y), y)d_z \leq 0, \quad i \in I_g(y) \setminus I_g(y; d_y). \tag{30}$$

Similarly, for $i \in J(\mu; d_\mu) \setminus J(\mu)$, we have $d_{\mu_i} > 0$, which ensures that for all sufficiently small $t > 0$,

$$td_{\mu_i} + t^2 d^1_{\mu_i} \geq 0, \quad i \in J(\mu; d_\mu) \setminus J(\mu). \tag{31}$$

Denote $(\hat{d}_x, \hat{d}_y, \hat{d}_\mu, \hat{d}^1_x, \hat{d}^1_y, \hat{d}^1_\mu) = (td_x, td_y, td_\mu, t^2 d^1_x, t^2 d^1_y, t^2 d^1_\mu)$ and $\hat{d}_z = (\hat{d}_x, \hat{d}_y)$. Set $\delta_1 = \|\hat{d}_y\| = \|td_y\|$ and $\delta_2 = \|\hat{d}^1_y\| = \|t^2 d^1_y\|$. Combining (30)–(31) with (28g)–(28h) (satisfied by $(\hat{d}_x, \hat{d}_y, \hat{d}_\mu, \hat{d}^1_x, \hat{d}^1_y, \hat{d}^1_\mu)$) and using the facts that $\nabla_x g_i(x(y), y)\hat{d}_x + \nabla_y g_i(x(y), y)\hat{d}_y = 0$ for $i \in I_g(y; \hat{d}_y) \setminus I$ and $\hat{d}_{\mu_i} = 0$ for $i \in I \setminus J(\mu; \hat{d}_\mu)$, we obtain (29a)–(29b).

Conversely, (28g)–(28h) can be recovered from (29a)–(29b) directly. Selecting $\delta = \min\{\delta_1, \delta_2\}$, it follows that for $\|\hat{d}_y\| \leq \delta$ and $\|\hat{d}^1_y\| \leq \delta$ with sufficiently small $\delta > 0$, the system consisting of (28a)–(28f) and (29a)–(29b) is locally equivalent to (28a)–(28h). From the results of parts (1) and (2), this system is locally equivalent to (11a)–(11d) and (12a)–(12d).

□

Remark 4.3 The first- and second-order directional derivatives are obtained via the quadratic subproblems in Theorem 2.1 and Theorem 4.1, respectively. The linear independence condition on the constraint gradients $\{\nabla_x g_i(x(y), y)\}_{i \in I}$ in their corresponding first- and second-order KKT systems does not affect these primal solutions; rather, it ensures that the multipliers in these systems correspond to extreme points of the multiplier sets. Whenever the active-set system (28a)–(28h) is feasible for nonzero vectors for some admissible index set $J(\mu) \subseteq I \subseteq I_g(y)$ and $\mu \in EM(y)$ such that $\{\nabla_x g_i(x(y), y)\}_{i \in I}$ is linearly independent, then from Theorem 4.1(1(b)), we have $\mu \in S^2(d_y, d^1_y) \cap EM(y)$. Furthermore, from Proposition 4.1, the first-order KKT system (11a)–(11d) is satisfied, which ensures that $d_\mu \in EM_\mu$.

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*

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

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 directionally differentiable functions around y. Then there does not exist $d_y^1 \in \mathbb{R}^m$ such that the following system is satisfied.

$$h_0'(y; d_y) + h_0''(y; d_y, d_y^1) < 0, \tag{32a}$$

$$h_i'(y; d_y) + h_i''(y; d_y, d_y^1) < 0, \quad i \in I_h(y), \tag{32b}$$

$$h_0'(y; d_y) \leq 0, \quad h_i'(y; d_y) \leq 0, \quad i \in I_h(y), \tag{32c}$$

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,

$$h_0(y^k) = h_0(y) + \alpha_k h_0'(y; d_y) + \alpha_k^2 h_0''(y; d_y, d_y^1) + o(\alpha_k^2), \ \forall k, \tag{33}$$

$$\text{and } h_i(y^k) = h_i(y) + \alpha_k h_i'(y; d_y) + \alpha_k^2 h_i''(y; d_y, d_y^1) + o(\alpha_k^2), \ \forall k. \tag{34}$$

Suppose, for the purpose of contradiction, that there exists d_y^1 such that (32a)-(32c) is satisfied. From the conditions (32b)-(32c), 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 $k \geq N_1$ from (34). Using the continuity of h_i for $i \in \Lambda_p \setminus I_h(y)$, there exists $N_2 \geq N_1$ such that $h_i(y^k) < 0$ for all $i \in \Lambda_p \setminus I_h(y)$ and $k \geq N_2$. Finally, using (32a), (32c), and (33) we can conclude in a similar manner that there is $N_3 \geq N_2$ such that $h_0(y^k) < h_0(y)$ for all $k \geq N_3$. Thus, $h_0(y^k) < h_0(y)$ for $h_i(y^k) < 0$ for all $i \in \Lambda_p$ and $k \geq N_3$. 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 $\{y^k\}$ is 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

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

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

$$\begin{aligned} \tilde{F}'(y; d_y) &= \nabla_x F(x(y), y) x'(y; d_y) + \nabla_y F(x(y), y) d_y, \\ \tilde{F}''(y; d_y, d_y^1) &= \nabla_x F(x(y), y) x''(y; d_y, d_y^1) + \nabla_y F(x(y), y) d_y^1 \end{aligned} \tag{35a}$$

$$+\frac{1}{2}\left(x'(y;d_y)^T \;\; d_y^T\right)\nabla^2_{zz}F(x(y),y)\begin{pmatrix}x'(y;d_y)\\ d_y\end{pmatrix}, \tag{35b}$$

respectively. Similarly, for each active index $i \in I_G(y)$, the first- and second-order directional derivatives of $\tilde{G}_i(y)$ are given by

$$\tilde{G}_i'(y;d_y) = \nabla_x G_i(x(y),y)x'(y;d_y) + \nabla_y G_i(x(y),y)d_y, \tag{36a}$$

$$\begin{aligned}\tilde{G}_i''(y;d_y,d_y^1) &= \nabla_x G_i(x(y),y)x''(y;d_y,d_y^1) + \nabla_y G_i(x(y),y)d_y^1\\ &\quad+\frac{1}{2}\left(x'(y;d_y)^T \;\; d_y^T\right)\nabla^2_{zz}G_i(x(y),y)\begin{pmatrix}x'(y;d_y)\\ d_y\end{pmatrix},\end{aligned} \tag{36b}$$

respectively.

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

$$v := \min\left\{v(\mu,I) \;:\; \begin{matrix}\mu \in EM(y),\;\; J(\mu)\subseteq I \subseteq I_g(y),\\ \left(\nabla_x g_i(x(y),y)\right)_{i\in I} \text{ is linearly independent}\end{matrix}\right\} \geq 0, \tag{37}$$

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

$$\begin{aligned}(P(\mu,I)): \min_{d_z,d_z^1,d_\mu,d_\mu^1,\alpha}\quad & \alpha\\ s.t.\quad & \nabla_x F(x(y),y)(d_x+d_x^1)+\nabla_y F(x(y),y)\,(d_y+d_y^1)\\ &\quad +\tfrac{1}{2}d_z^T\nabla^2_{zz}F(x(y),y)d_z \leq \alpha,\\ & \nabla_x G_i(x(y),y)\,(d_x+d_x^1)+\nabla_y G_i(x(y),y)\,(d_y+d_y^1)\\ &\quad +\tfrac{1}{2}d_z^T\nabla^2_{zz}G_i(x(y),y)d_z \leq \alpha,\;\; i\in I_G(y),\\ & \nabla_x F(x(y),y)\,d_x+\nabla_y F(x(y),y)\,d_y \leq 0,\\ & \nabla_x G_i(x(y),y)d_x+\nabla_y G_i(x(y),y)\,d_y \leq 0,\quad i\in I_G(y),\\ & (28a)-(28f)\ and\ (29a)-(29b)\ hold,\\ & \|d_y\|\leq 1,\quad \|d_y^1\|\leq 1.\end{aligned}$$

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. Therefore, BOP can be written in an implicit form in the neighborhood of y as

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

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.

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

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

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

Substituting the values for $\tilde{F}'(y;d_y)$, $\tilde{F}''(y;d_y,d_y^1)$, $\tilde{G}_i'(y;d_y)$ and $\tilde{G}_i''(y;d_y,d_y^1)$ from (35a)-(36b), we obtain

$$(P):\quad \min_{\alpha,d_y,d_y^1}\quad \alpha$$

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

Replacing $x'(y;d_y)=d_x$ and $x''(y;d_y,d_y^1)=d_x^1$ and incorporating the system (28a)-(28f) and (29a)-(29b) 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 (37) 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

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

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. □

To derive the second-order sufficient optimality condition, one typically needs to solve an auxiliary subproblem with respect to the second-order tangent direction, parameterized by each nonzero first-order tangent direction (see Theorem 22, Liu et al. [18], Condition 2.5, Rückmann and Shapiro [20]). Because verifying this condition pointwise for every direction can be difficult, we instead formulate a subproblem

that uses both the normalized first-order direction d_y and the second-order direction d_y^1 simultaneously. Specifically, for the case without upper-level constraints, we propose the following combinatorial quadratic problem $P'(\mu,\mathcal{I})$ with the index $\mathcal{I}=(I_1,I_2,I_3)$ with $J(\mu)\subseteq I_3\subseteq I_1\subseteq I_2\subseteq I_g(y)$.

$$
\begin{aligned}
&(P'(\mu,\mathcal{I})):\\
\min_{d_z,d_z^1,d_\mu,d_\mu^1}\quad & \nabla_x F(x(y),y)(d_x+d_x^1)+\nabla_y F(x(y),y)(d_y+d_y^1)+\frac{1}{2}d_z^T\nabla_{zz}^2F(x(y),y)d_z\\
\text{s.t.}\quad & \nabla_x F(x(y),y)d_x+\nabla_y F(x(y),y)d_y\le 0, && \text{(38a)}\\
& \text{System (28a)–(28f) holds for } I=I_1, && \text{(38b)}\\
& \nabla_x g_i(x(y),y)\,d_x^1+\nabla_y g_i(x(y),y)d_y^1+\frac{1}{2}d_z^T\nabla_{zz}^2 g_i(x(y),y)d_z\le 0,\quad i\in I_2\setminus I_1, && \text{(38c)}\\
& d_{\mu_i}^1\ge 0,\quad i\in I_1\setminus I_3,\quad d_{\mu_i}^1=0,\quad i\notin I_1, && \text{(38d)}\\
& \|d_y\|=1. && \text{(38e)}
\end{aligned}
$$

Denote the optimal value of $P'(\mu,\mathcal{I})$ as $v'(\mu,\mathcal{I})$ for later use. To establish the second-order sufficient condition, we shall need the concept of a second-order tangent set for the feasible set of the bilevel problem BOP.

Definition 5.1 (Liu et al. [18], Bonnans and Shapiro [24]) Given $C\subseteq\mathbb{R}^n\times\mathbb{R}^m$ and $z^*\in C$, the tangent cone to C at z^* is defined by

$$T_C(z^*):=\left\{d_z\in\mathbb{R}^n\times\mathbb{R}^m:\begin{array}{l}\exists\, t_k\downarrow 0,\ d_z^k\to d_z\\ \text{with } z^*+t_kd_z^k\in C\end{array}\right\}.$$

For $z^*\in C$ and $d_z\in T_C(z^*)$, the outer second-order tangent set to C at z^* in the direction d_z is defined by

$$T_C^2(z^*;d_z):=\left\{d_z^1\in\mathbb{R}^n\times\mathbb{R}^m:\begin{array}{l}\exists\, t_k\downarrow 0,\ d_z^{1,k}\to d_z^1\\ \text{with } z^*+t_kd_z+t_k^2d_z^{1,k}\in C\end{array}\right\}.$$

To derive the second-order sufficient optimality conditions for the nonsmooth problem with set constraint, Liu et al. [18] (Definition 2), and Bonnans and Shapiro [24](Definition 3.85) have used the outer second-order regularity assumption. It is used here on the feasible set of the bilevel problem BOP.

Definition 5.2 (Outer second-order regularity) Let C be a closed subset of $\mathbb{R}^n\times\mathbb{R}^m$, $z\in C$, and $d_z\in T_C(z)$. We say that C is *outer second-order regular* at z in the direction d_z if, for any sequence $\{z^k\}\subseteq C$ of the form

$$z^k:=z+t_kd_z+t_k^2d_z^{1,k},$$

where $\{t_k\}\downarrow 0$ and sequence $\{d_z^{1,k}\}$ satisfies $\{t_kd_z^{1,k}\}\to 0$, the following condition holds,

$$\lim_{k\to\infty}\operatorname{dist}\Big(d_z^{1,k},\,T_C^2(z;d_z)\Big)=0,$$

where $\operatorname{dist}(u,S):=\inf_{v\in S}\|u-v\|$ denotes the Euclidean distance from a point u to a set S.

Lemma 5.2 *Let $\chi\colon\mathbb{R}^m\to\mathbb{R}^n$ be locally Lipschitz continuous around $y\in\mathbb{R}^m$. Then, the following statements hold:*

1. (Bonnans and Shapiro [24, Proposition 2.49]) *If χ is directionally differentiable at y, then it is directionally differentiable at y in the Hadamard sense, and the first-order directional derivative mapping $\chi'(y;\cdot)\colon\mathbb{R}^m\to\mathbb{R}^n$ is Lipschitz continuous with modulus L on $\mathbb{R}^m$, that is,*

$$\|\chi'(y;d_1)-\chi'(y;d_2)\|\le L\|d_1-d_2\|\quad\forall d_1,d_2\in\mathbb{R}^m.$$

2. (Liu et al. [18, Proposition 2.6]) *Let $d_y\in\mathbb{R}^m$. If χ is second-order directionally differentiable at y in direction d_y, then χ is second-order directionally differentiable at y in direction d_y in the Hadamard sense, and the second-order directional derivative mapping $\chi''(y;d_y,\cdot)\colon\mathbb{R}^m\to\mathbb{R}^n$ is Lipschitz continuous with modulus L, that is,*

$$\|\chi''(y;d_y,w)-\chi''(y;d_y,v)\|\le L\|w-v\|\quad\forall w,v\in\mathbb{R}^m.$$

Under Assumptions $(\mathcal{H}1)$-$(\mathcal{H}5)$, the lower-level problem admits a unique strict local minimizer $x(y)$ for each parameter y near the reference point. Hence, the set-valued mapping Ψ reduces locally to a single-valued function $x(\cdot)$, and the graph of Ψ can be locally represented as

$$C := \mathrm{gph}(\Psi) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \ : \ x = x(y)\}. \tag{39}$$

By Theorem 3.2, the solution function $x(\cdot)$ is second-order directionally differentiable along parabolic paths. The first- and second-order tangent sets to C at $(x(y), y)$ are characterized as follows. The proof takes use of Lemma 5.2.

Proposition 5.1 *Suppose Assumptions $(\mathcal{H}1)$-$(\mathcal{H}5)$ hold at $(x(y), y) \in C$.*

1. *The tangent cone to C at $(x(y), y)$ is given by*
$$\mathcal{T}_C(x(y), y) = \{(d_x, d_y) \in \mathbb{R}^n \times \mathbb{R}^m \ : \ d_x = x'(y; d_y)\}.$$
2. *For any tangent direction $(d_x, d_y) \in \mathcal{T}_C(x(y), y)$ satisfying $d_x = x'(y; d_y)$, the outer second-order tangent set to C at $(x(y), y)$ in direction (d_x, d_y) is given by*
$$\mathcal{T}_C^2\big((x(y), y); (d_x, d_y)\big) = \Big\{(d_x^1, d_y^1) \in \mathbb{R}^n \times \mathbb{R}^m \ : \ d_x^1 = x''(y; d_y, d_y^1)\Big\}.$$

Proof Let $(d_x, d_y) \in \mathcal{T}_C(x(y), y)$. Hence, by definition, there exist sequences $\{t_k\} \downarrow 0$ and $\{(d_x^k, d_y^k)\} \to (d_x, d_y)$ such that $(x(y) + t_k d_x^k, y + t_k d_y^k) \in C$, implying

$$x(y) + t_k d_x^k = x(y + t_k d_y^k).$$

Since $x(\cdot)$ is Hadamard directionally differentiable from Lemma 5.2 (1), it follows that

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

Conversely, for any $d_y \in \mathbb{R}^m$ and $d_x = x'(y; d_y)$, let $\{t_k\} \downarrow 0$ be an arbitrary sequence, and set

$$(d_x^k, d_y^k) := \left(\frac{x(y + t_k d_y) - x(y)}{t_k}, \ d_y\right).$$

Then $(x(y) + t_k d_x^k, \ y + t_k d_y^k) = (x(y + t_k d_y), \ y + t_k d_y) \in C$ for all k. By directional differentiability, $\{(d_x^k, d_y^k)\} \to (x'(y; d_y), d_y) = (d_x, d_y)$, which verifies that $(d_x, d_y) \in \mathcal{T}_C(x(y), y)$. This proves the first part.

Next, let $(d_x^1, d_y^1) \in \mathcal{T}_C^2\big((x(y), y); (d_x, d_y)\big)$ with $d_x = x'(y; d_y)$. By the definition of the outer second-order tangent set, there exist sequences $\{t_k\} \downarrow 0$ and $\{(d_x^{1,k}, d_y^{1,k})\} \to (d_x^1, d_y^1)$ such that

$$\left(x(y) + t_k d_x + t_k^2 d_x^{1,k}, \ y + t_k d_y + t_k^2 d_y^{1,k}\right) \in C.$$

This implies

$$x(y) + t_k x'(y; d_y) + t_k^2 d_x^{1,k} = x(y + t_k d_y + t_k^2 d_y^{1,k}).$$

Since $x(\cdot)$ is second-order directionally differentiable in the sense of Hadamard from Lemma 5.2 (2), dividing by t_k^2 and taking $k \to \infty$ gives

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

The reverse inclusion follows by selecting $d_x^{1,k} = \frac{x(y + t_k d_y + t_k^2 d_y^1) - x(y) - t_k x'(y; d_y)}{t_k^2}$ and $d_y^{1,k} = d_y^1$, which completes the proof. □

Theorem 5.2 (Second-Order Sufficient Optimality Condition) *Let $(x(y), y)$ be a feasible point of BOP without upper-level constraints. Suppose that*

(i) Assumptions $(\mathcal{H}1)$-$(\mathcal{H}5)$ hold;

(ii) $F \in C^2$ in a neighborhood of $(x(y), y)$;

(iii)
$$\nabla_x F(x(y), y) x'(y; d_y) + \nabla_y F(x(y), y) d_y \geq 0 \quad \text{for all } \|d_y\| = 1; \tag{40}$$

(iv) *The graph* $C := \mathrm{gph}(\Psi)$ *is outer second-order regular at* $(x(y), y)$ *in direction* $(x'(y; d_y), d_y)$ *for all* d_y *with* $\|d_y\| = 1$*;*

(v) *For index sets* $\mathcal{I} = (I_1, I_2, I_3)$ *satisfying* $J(\mu) \subseteq I_3 \subseteq I_1 \subseteq I_2 \subseteq I_g(y)$ *such that* $\{\nabla_x g_i(x(y), y)\}_{i \in I_1}$ *is linearly independent,*

$$v' := \min \left\{ v'(\mu, \mathcal{I}) \ : \ \begin{array}{l} \mu \in EM(y), \text{ and the optimal solution } (d_z, d_z^1, d_\mu, d_\mu^1) \text{ to } P'(\mu, \mathcal{I}) \\ \text{satisfies } I_2 = I_g(y; d_y) \text{ and } I_3 = J(\mu; d_\mu). \end{array} \right\} > 0.$$

Then $(x(y), y)$ *is a strict local minimum of BOP, and there exist constants* $\varepsilon > 0$ *and* $c > 0$ *such that*

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

for all (x', y') *satisfying* $\|y' - y\| \leq \varepsilon$ *and* $x' \in \Psi(y')$.

Proof Let $(d_z, d_\mu, d_z^1, d_\mu^1)$ be the solution to $P'(\mu, \mathcal{I})$ with $J(\mu) \subseteq I_3 \subseteq I_1 \subseteq I_2 \subseteq I_g(y)$ such that $\{\nabla_x g_i(x(y), y)\}_{i \in I_1}$ is linearly independent, and $I_2 = I_g(y; d_y)$ and $I_3 = J(\mu; d_\mu)$. Then, by Proposition 4.1, in System (38b)–(38e) we have $d_x = x'(y; d_y)$ and $d_x^1 = x''(y; d_y, d_y^1)$. Since $v' > 0$, we have

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

for all (d_y, d_y^1) satisfying

$$\nabla_x F(x(y), y) x'(y; d_y) + \nabla_y F(x(y), y) d_y \leq 0, \quad \|d_y\| = 1.$$

In view of (40), we obtain

$$\nabla_x F(x(y), y) x'(y; d_y) + \nabla_y F(x(y), y) d_y = 0. \tag{41}$$

Suppose, to the contrary, that there exists a sequence $\{(x(y^k), y^k)\}$ with $x(y^k) \in \Psi(y^k)$ and $y^k \neq y$ such that $\{y^k\} \to y$ and

$$F(x(y^k), y^k) < F(x(y), y) + o(\|y^k - y\|^2). \tag{42}$$

Using Assumptions $(\mathcal{H}1)$-$(\mathcal{H}3)$, it follows from the continuity of $x(\cdot)$ that $\{x(y^k)\} \to x(y)$. Setting $\alpha_k := \|y^k - y\|$, the sequence $\{d^k\} := \left\{ \frac{y^k - y}{\alpha_k} \right\}$ converges, possibly over a subsequence, to some directional vector d_y with $\|d_y\| = 1$. Hence, for the sequence $\{d_y^{1,k}\} := \left\{ \frac{d^k - d_y}{\alpha_k} \right\}$, there exists a subsequence of $\{y^k\}$ such that $\{\alpha_k d_y^{1,k}\} \to 0$. Thus, $y^k = y + \alpha_k d^k = y + \alpha_k (d_y + \alpha_k d_y^{1,k}) = y + \alpha_k d_y + \alpha_k^2 d_y^{1,k}$.
Let

$$d_x^{1,k} := \frac{x(y^k) - x(y) - \alpha_k x'(y; d_y)}{\alpha_k^2}.$$

Since $x(\cdot)$ is second-order gph-regular at y along the sequence $y^k = y + \alpha_k d_y + \alpha_k^2 d_y^{1,k}$, it satisfies

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

which implies that

$$d_x^{1,k} = x''(y; d_y, d_y^{1,k}) + \frac{o(\alpha_k^2)}{\alpha_k^2}.$$

Because $x(\cdot)$ is piecewise smooth, by the selection arguments in Theorem 3.3, for each k there exists an index $j_k \in \{1, \ldots, N\}$ such that

$$x''(y; d_y, d_y^{1,k}) = x''_{j_k}(y; d_y, d_y^{1,k}) = \nabla_y x_{j_k}(y) d_y^{1,k} + \frac{1}{2} d_y^T \nabla^2_{yy} x_{j_k}(y) d_y.$$

Since the index set $\{1, \ldots, N\}$ is finite and each $x_j \in C^2$, the families $\{\nabla_y x_j(y)\}$ and $\{\nabla^2_{yy} x_j(y)\}$ are bounded. Thus,

$$\max_{1 \leq j \leq N} \left\{ \|\nabla_y x_j(y)\|, \ \|\nabla^2_{yy} x_j(y)\| \right\} < \infty.$$

Using $\{\alpha_k\} \to 0$ and $\lim_{k \to \infty} \alpha_k d_y^{1,k} = 0$, we obtain

$$\begin{aligned} \lim_{k \to \infty} \alpha_k d_x^{1,k} &= \lim_{k \to \infty} \alpha_k x''_{j_k}(y; d_y, d_y^{1,k}) + \lim_{k \to \infty} \frac{o(\alpha_k^2)}{\alpha_k} \\ &= \lim_{k \to \infty} \left(\nabla_y x_{j_k}(y)(\alpha_k d_y^{1,k}) + \frac{1}{2} \alpha_k d_y^T \nabla^2_{yy} x_{j_k}(y) d_y \right) \end{aligned}$$

$$= 0.$$

Because C is outer second-order regular at $(x(y), y)$ in direction $(x'(y; d_y), d_y)$, by Definition 5.2 we have

$$\lim_{k\to\infty} \operatorname{dist}\Big((d_x^{1,k}, d_y^{1,k}),\ \mathcal{T}_C^2\big((x(y), y); (x'(y; d_y), d_y)\big)\Big) = 0.$$

Therefore, there exists a sequence $\{(w_x^k, w_y^k)\} \subseteq \mathcal{T}_C^2\big((x(y), y); (x'(y; d_y), d_y)\big)$ such that

$$\lim_{k\to\infty} \|(d_x^{1,k}, d_y^{1,k}) - (w_x^k, w_y^k)\| = 0.$$

Recall the implicit objective function $\tilde{F}(y) := F(x(y), y)$. By Lemma 5.2, $\tilde{F}''(y; d_y, \cdot)$ is Lipschitz continuous with some modulus $L > 0$. Then, it follows that

$$\tilde{F}''(y; d_y, d_y^{1,k}) \geq \tilde{F}''(y; d_y, w_y^k) - L\|d_y^{1,k} - w_y^k\|.$$

From Proposition 5.1, $(w_x^k, w_y^k) \in \mathcal{T}_C^2\big((x(y), y); (x'(y; d_y), d_y)\big)$ satisfies $w_x^k = x''(y; d_y, w_y^k)$ for each k. Let $\mu \in S^2(d_y, w_y^k) \cap EM(y)$. Using Proposition 4.1(2), for some suitable index set I_1 satisfying $J(\mu) \subseteq I_1 \subseteq I_g(y)$, $(d_z, d_\mu, w_z^k, d_\mu^{1,k})$ satisfies (28a)-(28h), where $d_z = (x'(y; d_y), d_y)$ and $w_z^k = (w_x^k, w_y^k)$ for appropriate $d_\mu, d_\mu^{1,k}$. Then, by Proposition 4.1, for $I_2 = I_g(y; d_y)$ and $I_3 = J(\mu; d_\mu)$, $(d_z, d_\mu, w_z^k, d_\mu^{1,k})$ is feasible for system (38a)–(38e) for each k. Hence, $\tilde{F}''(y; d_y, w_y^k) \geq v' > 0$ for all k.
Expanding $F \in C^2$ around $(x(y), y)$ and applying (41), we obtain

$$\begin{aligned} F(x(y^k), y^k) &= F(x(y), y) + \alpha_k^2 \tilde{F}''(y; d_y, d_y^{1,k}) + o(\alpha_k^2) \\ &\geq F(x(y), y) + \alpha_k^2 \left(v' - L\|d_y^{1,k} - w_y^k\|\right) + o(\alpha_k^2). \end{aligned}$$

Since $\|d_y^{1,k} - w_y^k\| \to 0$ as $k \to \infty$ and $v' > 0$, there exists $v'' > 0$ such that for all sufficiently large k,

$$\frac{F(x(y^k), y^k) - F(x(y), y)}{\alpha_k^2} \geq v'' + \frac{o(\alpha_k^2)}{\alpha_k^2},$$

which contradicts (42). Hence, there exist constants $c > 0$ and $\varepsilon > 0$ such that

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

for all (x', y') satisfying $\|y' - y\| \leq \varepsilon$ and $x' \in \Psi(y')$. Under Assumption ($\mathcal{H}5$), Ψ is single-valued and continuous; therefore, $(x(y), y)$ is a strict local minimum of BOP. □

Notice that we do not impose a norm bound on the second-order direction d_y^1. In second-order sufficient conditions, the analysis uses an unbounded direction vector d_y^1. If the objective goes to $-\infty$, the optimal value is negative, which directly violates the condition $v' > 0$. Thus, the condition on the optimal value $v' > 0$ rules out unbounded descent.

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

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

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 $(3, 1)$ 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, $(3, 1)$ is the global optimal solution to this bilevel problem. Next, it remains to verify that the second-order sufficient optimality condition of Theorem 5.2 holds at $(3, 1)$, 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 $(3,1)$ as $\nabla_y F(3,1) = \nabla_x F(3,1) = 0$. Observe that Assumptions $\mathcal{H}1$-$\mathcal{H}5$ hold at $(3,1)$. 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

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

at $(3,1)$. Thus, at $(x,y) = (3,1)$, $M(y) = \text{conv}\{(4,0,0)^T, (0,0,2)^T\}$ and $EM(y) = \{(4,0,0)^T, (0,0,2)^T\}$. At $(x,y) = (3,1)$, the active index set is $I_g(y) = \{\text{i}, \text{iii}\}$. For the extreme multipliers in $EM(y)$, for $\mu = (4,0,0)^T$, we have $J(\mu) = \{\text{i}\}$; For $\mu = (0,0,2)^T$, we have $J(\mu) = \{\text{iii}\}$. Consequently, the family of admissible index triples

$$\left\{\mathcal{I} = (I_1, I_2, I_3)\ :\ \begin{array}{l} J(\mu) \subseteq I_3 \subseteq I_1 \subseteq I_2 \subseteq I_g(y), \quad \mu \in EM(y),\\ \{\nabla_x g_i(x(y), y)\}_{i \in I_1} \text{ is linearly independent} \end{array}\right\}$$

evaluates to

$$\begin{aligned}
&\mathcal{I}_1 = \big(\{\text{i}\}, \{\text{i}\}, \{\text{i}\}\big),\ \mathcal{I}_2 = \big(\{\text{i}\}, \{\text{i}, \text{iii}\}, \{\text{i}\}\big),\\
&\mathcal{I}_3 = \big(\{\text{iii}\}, \{\text{i}, \text{iii}\}, \{\text{iii}\}\big),\ \ \mathcal{I}_4 = \big(\{\text{iii}\}, \{\text{iii}\}, \{\text{iii}\}\big).
\end{aligned}$$

(a) For $\mathcal{I}_1 = \big(\{\text{i}\}, \{\text{i}\}, \{\text{i}\}\big)$,

$$\begin{aligned}
P'((4,0,0)^T, \mathcal{I}_1):\quad \min_{d_z, d_z^1, d_\mu, d_\mu^1}\ & d_x^2 + d_y^2\\
\text{s.t.}\quad & 2d_x + d_{\mu_1} + 2d_{\mu_3} = 0,\\
& 2d_x^1 + d_{\mu_1}^1 + 2d_{\mu_3}^1 = 0,\\
& d_x - 2d_y = 0,\quad d_x^1 - 2d_y^1 = 0,\\
& 2d_x + d_y \leq 0,\\
& \|d_y\| = 1.
\end{aligned}$$

The solution to $P'((4,0,0)^T, \mathcal{I}_1)$ is $(d_x, d_y, d_{\mu_1}, d_{\mu_3}, d_x^1, d_y^1, d_{\mu_1}^1, d_{\mu_3}^1) = (-2, -1, 4, 0, 0, 0, 0, 0)$ with the optimal value $v'((4,0,0)^T, \mathcal{I}_1) = 5$. Here $I_2 = I_g(y; d_y) = \{\text{i}\}$ and $I_3 = J(\mu; d_\mu) = \{\text{i}\}$.
(b) For $\mathcal{I}_2 = \big(\{\text{i}\}, \{\text{i}, \text{iii}\}, \{\text{i}\}\big)$,

$$\begin{aligned}
P'((4,0,0)^T, \mathcal{I}_2):\quad \min_{d_z, d_z^1, d_\mu, d_\mu^1}\ & d_x^2 + d_y^2\\
\text{s.t.}\quad & 2d_x + d_{\mu_1} + 2d_{\mu_3} = 0,\\
& 2d_x^1 + d_{\mu_1}^1 + 2d_{\mu_3}^1 = 0,\\
& d_x - 2d_y = 0,\quad d_x^1 - 2d_y^1 = 0,\\
& 2d_x + d_y \leq 0,\quad 2d_x^1 + d_y^1 \leq 0,\\
& \|d_y\| = 1.
\end{aligned}$$

The solution to $P'((4,0,0)^T, \mathcal{I}_2)$ is $(d_x, d_y, d_{\mu_1}, d_{\mu_3}, d_x^1, d_y^1, d_{\mu_1}^1, d_{\mu_3}^1) = (-2, -1, 4, 0, 0, 0, 0, 0)$ with the optimal value $v'((4,0,0)^T, \mathcal{I}_2) = 5$. Here, $I_g(y; d_y) = \{\text{i}\} \neq I_2$.

(c) For $\mathcal{I}_3 = \big(\{\text{iii}\}, \{\text{i}, \text{iii}\}, \{\text{iii}\}\big)$,

$$
\begin{aligned}
P'((0,0,2)^T, \mathcal{I}_3) \ : \quad & \min_{d_z, d_z^1, d_\mu, d_\mu^1} \quad d_x^2 + d_y^2 \\
& \text{s.t.} \quad\ \ 2d_x + d_{\mu_1} + 2d_{\mu_3} = 0, \\
& \qquad\quad\ \ 2d_x^1 + d_{\mu_1}^1 + 2d_{\mu_3}^1 = 0, \\
& \qquad\quad\ \ 2d_x + d_y = 0, \quad 2d_x^1 + d_y^1 = 0, \\
& \qquad\quad\ \ d_x - 2d_y \le 0, \quad d_x^1 - 2d_y^1 \le 0, \\
& \qquad\quad\ \ \|d_y\| = 1.
\end{aligned}
$$

The solution to $P'((0,0,2)^T, \mathcal{I}_3)$ is $(d_x, d_y, d_{\mu_1}, d_{\mu_3}, d_x^1, d_y^1, d_{\mu_1}^1, d_{\mu_3}^1) = (-\frac{1}{2}, 1, 0, \frac{1}{2}, 0, 0, 0, 0)$ with the optimal value $v'((0,0,2)^T, \mathcal{I}_3) = 1.25$. Here, $I_g(y; d_y) = \{\text{iii}\} \neq I_2$.
(d) For $\mathcal{I}_4 = \big(\{\text{iii}\}, \{\text{iii}\}, \{\text{iii}\}\big)$,

$$
\begin{aligned}
P'((0,0,2)^T, \mathcal{I}_4) \ : \quad & \min_{d_z, d_z^1, d_\mu, d_\mu^1} \quad d_x^2 + d_y^2 \\
& \text{s.t.} \quad\ \ 2d_x + d_{\mu_1} + 2d_{\mu_3} = 0, \\
& \qquad\quad\ \ 2d_x^1 + d_{\mu_1}^1 + 2d_{\mu_3}^1 = 0, \\
& \qquad\quad\ \ 2d_x + d_y = 0, \quad 2d_x^1 + d_y^1 = 0, \\
& \qquad\quad\ \ d_x - 2d_y \le 0, \\
& \qquad\quad\ \ \|d_y\| = 1.
\end{aligned}
$$

The solution to $P'((0,0,2)^T, \mathcal{I}_4)$ is $(d_x, d_y, d_{\mu_1}, d_{\mu_3}, d_x^1, d_y^1, d_{\mu_1}^1, d_{\mu_3}^1) = (-\frac{1}{2}, 1, 0, \frac{1}{2}, 0, 0, 0, 0)$ with the optimal value $v'((0,0,2)^T, \mathcal{I}_4) = 1.25$. Here, $I_2 = I_g(y; d_y) = \{\text{iii}\}$ and $I_3 = J(\mu; d_\mu) = \{\text{iii}\}$.
Taking the minimum over all feasible admissible triples that satisfy $I_2 = I_g(y; d_y)$ and $I_3 = J(\mu; d_\mu)$:

$$
v' = \min\Big\{v'((4,0,0)^T, \mathcal{I}_1),\ v'((0,0,2)^T, \mathcal{I}_4)\Big\} = \min\{5, 1.25\} = 1.25 > 0.
$$

Since $v' > 0$, the second-order sufficient optimality condition of Theorem 5.2 is satisfied at $(3,1)$.

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 derivative 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 a local optimal solution called a bi-local solution for the bilevel problems with nonconvex lower-level problems. The bi-local 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 solution and a standard local optimal solution are equivalent. Extending the main results of Section 5 to characterize optimality conditions for bi-local solutions of bilevel optimization problems with nonconvex lower-level problems constitutes another direction for future research.

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. Springer, New York (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. Applied Mathematics & Optimization **94**(2), 45 (2026)

[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.: Directional derivative of the marginal function in nonlinear programming. In: Fiacco, A.V. (ed.) Sensitivity, Stability and Parametric Analysis. Mathematical Programming Studies, vol. 21, pp. 110–126. Springer, Berlin, Heidelberg (1984). 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 derivatives of the optimal solution function in parametric programming problems. The Journal of Analysis, 1–37 (2026) https://doi.org/10.1007/s41478-026-01124-7

[33] Fan, K.: Fixed-point and minimax theorems in locally convex topological linear spaces. Proceedings of the National Academy of Sciences **38**(2), 121–126 (1952) https://doi.org/10.1073/pnas.38.2.121

[34] Kim, W.K., Kum, S.: On a non-compact generalization of fan's minimax theorem. Taiwanese Journal of Mathematics **14**(2), 347–358 (2010)

Appendix

Consider the lower-level value function $\varphi(y') := \min_x\{f(x,y') \mid g(x,y') \le 0\}$.

Lemma .1 *Suppose Assumptions* $(\mathcal{H}1)$-$(\mathcal{H}5)$ *hold at* $(x(y),y)$. *Then, for any direction vectors* $d_y, d_y^1 \in \mathbb{R}^m$, *the first-order directional derivative* $\varphi'(y;d_y)$ *and the parabolic second-order directional derivative* $\varphi''(y;d_y,d_y^1)$ *of the optimal value function* $\varphi(y) := \min_x\{f(x,y) \mid g(x,y) \le 0\}$ *exist and are given by*

$$\varphi'(y;d_y) = \max_{\mu\in M(y)} \nabla_y L(x(y),y,\mu)d_y,$$

and

$$\varphi''(y;d_y,d_y^1) = \max_{\mu\in S^1(d_y)} \left\{\nabla_y L(x(y),y,\mu)d_y^1 + \frac{1}{2}d_z^T\nabla^2_{zz}L(x(y),y,\mu)d_z\right\},$$

where $d_z := \begin{pmatrix} x'(y;d_y) \\ d_y \end{pmatrix}$.

Proof Under Assumptions $(\mathcal{H}1)$-$(\mathcal{H}3)$ and $(\mathcal{H}5)$, it follows from Theorem 4.1 and relation (4.7) of Shapiro [28] that the first-order directional derivative of $\varphi(\cdot)$ along the direction d_y exists and satisfies

$$\varphi'(y;d_y) = \max_{\mu\in EM(y)} \nabla_y L(x(y),y,\mu)d_y.$$

Since this is a linear programming problem in μ over the nonempty, compact polyhedral set $M(y)$, the maximum over $M(y)$ is attained at an extreme point in $EM(y)$. Therefore,

$$\varphi'(y;d_y) = \max_{\mu\in EM(y)} \nabla_y L(x(y),y,\mu)d_y = \max_{\mu\in M(y)} \nabla_y L(x(y),y,\mu)d_y.$$

This proves the first assertion. Next, by Theorem 4.2 of Shapiro [28], the parabolic second-order directional derivative $\varphi''(y;d_y,d_y^1)$ exists and is given by the minimax problem

$$\varphi''(y;d_y,d_y^1) = \min_{d_x\in \bigcup_{\mu\in M(y)} K^1_\mu(y;d_y)} \max_{\mu\in S^1(d_y)} \nabla_y L(x(y),y,\mu)d_y^1 + \frac{1}{2}\begin{pmatrix} d_x^T & d_y^T \end{pmatrix}\nabla^2_{zz}L(x(y),y,\mu)\begin{pmatrix} d_x \\ d_y \end{pmatrix}. \tag{43}$$

By Lemma 2.2 of Dempe [29], the critical cone $K^1_\mu(y;d_y)$ is nonempty if and only if $\mu \in S^1(d_y)$. Furthermore, by Theorem 2 of Ralph and Dempe [22], under Assumption $(\mathcal{H}4)$ the set $K^1_\mu(y;d_y)$ is invariant for all $\mu \in S^1(d_y)$. Denoting this common polyhedral set by $K^1(y;d_y)$, the union reduces to

$$\bigcup_{\mu\in M(y)} K^1_\mu(y;d_y) = \bigcup_{\mu\in S^1(d_y)} K^1_\mu(y;d_y) = K^1(y;d_y).$$

Hence, (43) can be rewritten as

$$\varphi''(y;d_y,d_y^1) = \min_{d_x\in K^1(y;d_y)} \max_{\mu\in S^1(d_y)} \left\{\nabla_y L(x(y),y,\mu)d_y^1 + \frac{1}{2}\begin{pmatrix} d_x^T & d_y^T \end{pmatrix}\nabla^2_{zz}L(x(y),y,\mu)\begin{pmatrix} d_x \\ d_y \end{pmatrix}\right\}.$$

Under Assumption $(\mathcal{H}3)$, the objective function is strictly convex in d_x over the closed convex set $K^1(y;d_y)$, and affine (hence concave and continuous) in μ over the compact polyhedral set $S^1(d_y)$. Therefore, applying Fan's minimax theorem [33] and Kim and Kum [34], we can interchange min-max operator to max-min operator, and we obtain

$$\begin{aligned}
\varphi''(y;d_y,d_y^1) &= \max_{\mu\in S^1(d_y)} \min_{d_x\in K^1(y;d_y)} \left\{\nabla_y L(x(y),y,\mu)d_y^1 + \frac{1}{2}\begin{pmatrix} d_x^T & d_y^T \end{pmatrix}\nabla^2_{zz}L(x(y),y,\mu)\begin{pmatrix} d_x \\ d_y \end{pmatrix}\right\} \\
&= \max_{\mu\in S^1(d_y)} \left\{\nabla_y L(x(y),y,\mu)d_y^1 + \frac{1}{2}d_y^T\nabla^2_{yy}L(x(y),y,\mu)d_y \right. \\
&\qquad \left. + \min_{d_x\in K^1(y;d_y)} \left(\frac{1}{2}d_x^T\nabla^2_{xx}L(x(y),y,\mu)d_x + d_x^T\nabla^2_{xy}L(x(y),y,\mu)d_y\right)\right\}.
\end{aligned}$$

By Theorem 2.1(ii), for every $\mu \in S^1(d_y)$, the inner quadratic program over $K^1(y;d_y)$ is uniquely solved by $d_x = x'(y;d_y)$. Substituting $d_x = x'(y;d_y)$ into the objective yields

$$\varphi''(y;d_y,d_y^1) = \max_{\mu\in S^1(d_y)} \left\{\nabla_y L(x(y),y,\mu)d_y^1 + \frac{1}{2}d_z^T\nabla^2_{zz}L(x(y),y,\mu)d_z\right\},$$

which completes the proof. □

Proof of Lemma 4.2(3):

Proof By definition, since $S^1(d_y)$ is the optimal solution set of (P^1) with the constraint set $M(y)$ and $S^2(d_y, d_y^1)$ is the optimal solution set of (P^2) with the constraint set $S^1(d_y)$, the inclusions $S^2(d_y, d_y^1) \subseteq S^1(d_y) \subseteq M(y)$ hold directly. Furthermore, under Assumption $(\mathcal{H}2)$(MFCQ), the multiplier set-valued mapping $M(\cdot)$ is upper semi-continuous and compact-valued in a neighborhood of y. Thus, the parabolic limiting set $V(d_y, d_y^1)$ is a nonempty, compact subset of $M(y)$.

The proof of Step 1 below holds from Lemma 2.3 of Dempe [29] using the first-order Taylor approximation of the marginal function φ. We provide the proof for the sake of completeness.

Step 1: Proof of the inclusion $V(d_y, d_y^1) \subseteq S^1(d_y)$.

Arguing by contradiction, suppose there exists $\mu^0 \in V(d_y, d_y^1) \setminus S^1(d_y)$. Since $S^1(d_y)$ is the solution set of (P^1), there exist constants $\delta_1 > 0$ and $\varepsilon_1 > 0$ such that

$$\nabla_y L(x(y), y, \mu) d_y < \max_{v \in M(y)} \nabla_y L(x(y), y, v) d_y - \delta_1, \quad \forall \mu \in M(y) \cap N_{\varepsilon_1}(\mu^0), \tag{44}$$

where $N_{\varepsilon_1}(\mu^0) := \{\mu \in \mathbb{R}^q : \|\mu - \mu^0\| < \varepsilon_1\}$.

Since $\mu^0 \in V(d_y, d_y^1)$, by definition there exist a sequence of positive real numbers $\{\alpha_k\} \downarrow 0$ and a sequence of multipliers $\{\mu^k\}$ with $\mu^k \in M(y^k)$, where $y^k = y + \alpha_k d_y + \alpha_k^2 d_y^1$, such that $\lim_{k\to\infty} \mu^k = \mu^0$. Along a suitable subsequence, we have

$$\lim_{k\to\infty} \alpha_k^{-1}\big((x(y^k), \mu^k) - (x(y), \mu^0)\big) = (d_x, d_\mu), \tag{45}$$

where $d_x = x'(y; d_y)$ by directional differentiability of the solution function (Theorem 3.2).

Under Assumptions $(\mathcal{H}1)$-$(\mathcal{H}3)$, φ is directionally differentiable at y along d_y, and from Lemma .1

$$\varphi'(y; d_y) = \lim_{\alpha_k \downarrow 0} \frac{\varphi(y^k) - \varphi(y)}{\alpha_k} = \max_{\mu \in S^1(d_y)} \nabla_y L(x(y), y, \mu) d_y. \tag{46}$$

Using Taylor's expansion of the Lagrangian about $(x(y), y, \mu)$, and observing that $\varphi(y^k) = L(x(y^k), y^k, \mu^k)$ and $\varphi(y) = L(x(y), y, \mu_0)$ by complementary slackness and relation (46) yields

$$\begin{aligned}
\varphi(y^k) &= L(x(y^k), y^k, \mu^k) \\
&= L(x(y), y, \mu_0) + \alpha_k \Bigg\{ \nabla_x L(x(y), y, \mu_0) \frac{x(y^k) - x(y)}{\alpha_k} \\
&\qquad + \nabla_y L(x(y), y, \mu^0) \left(d_y + \alpha_k d_y^1\right) + \nabla_\mu L(x(y), y, \mu_0) \frac{\mu^k - \mu_0}{\alpha_k} \Bigg\} + o(\alpha_k) \\
&= L(x(y), y, \mu_0) + \alpha_k \max_{v \in M(y)} \nabla_y L(x(y), y, \mu_0) d_y + o(\alpha_k).
\end{aligned} \tag{47}$$

Since $\mu_0 \in M(y)$, KKT stationarity gives $\nabla_x L(x(y), y, \mu_0) = 0$. Dividing by α_k and comparing both sides of (47), it follows from (44) and (45) that for all sufficiently large k

$$\nabla_y L(x(y), y, \mu_0) d_y + \nabla_\mu L(x(y), y, \mu_0) \frac{\mu^k - \mu_0}{\alpha_k} + \frac{o(\alpha_k)}{\alpha_k} \geq \nabla_y L(x(y), y, \mu_0) d_y + \delta_1. \tag{48}$$

On the other hand, note that $\nabla_\mu L(x(y), y, \mu_0) = g(x(y), y)$. For sufficiently large k, it follows that $I_g(y^k) \subseteq I_g(y)$. Because $\mu_i^k = 0$ for all $i \notin I_g(y^k)$ and $g_i(x(y), y) = 0$ for all $i \in I_g(y)$, we obtain

$$\nabla_\mu L(x(y), y, \mu_0)(\mu^k - \mu_0) = g(x(y), y)^T (\mu^k - \mu_0) = 0$$

for all large k. Consequently, the term $\nabla_\mu L(x(y), y, \mu_0) \frac{\mu^k - \mu_0}{\alpha_k}$ vanishes identically, which contradicts relation (48) as $k \to \infty$ since $\delta_1 > 0$. Hence, $V(d_y, d_y^1) \subseteq S^1(d_y)$.

Step 2: Proof of the inclusion $V(d_y, d_y^1) \subseteq S^2(d_y, d_y^1)$.

We again use the argument of contradiction. Suppose there exists $\mu^0 \in V(d_y, d_y^1) \setminus S^2(d_y, d_y^1)$. From Step 1, we have $\mu^0 \in S^1(d_y)$. Since $\mu^0 \notin S^2(d_y, d_y^1)$ and $S^2(d_y, d_y^1)$ is the set of optimal solutions to the linear program (P^2) in (13), there exists a scalar $\delta_2 > 0$ such that

$$\begin{aligned}
&\sum_{i \in I_g(y)} \mu_i^0 \left(\nabla_y g_i(x(y), y) d_y^1 + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z \right) \\
&\qquad < \max_{\mu \in S^1(d_y)} \sum_{i \in I_g(y)} \mu_i \left(\nabla_y g_i(x(y), y) d_y^1 + \frac{1}{2} d_z^T \nabla_{zz}^2 g_i(x(y), y) d_z \right) - \delta_2,
\end{aligned} \tag{49}$$

where $d_z = \begin{pmatrix} x'(y;d_y) \\ d_y \end{pmatrix}$. Under Assumptions ($\mathcal{H}1$)-($\mathcal{H}5$) using Lemma .1, we obtain

$$\varphi''(y;d_y,d_y^1) = \max_{\mu\in S^1(d_y)} \left\{ \nabla_y L(x(y),y,\mu)d_y^1 + \frac{1}{2} d_z^T \nabla^2_{zz} L(x(y),y,\mu) d_z \right\}.$$

Adding the constant quantity $\nabla_y f(x(y),y)d_y^1 + \frac{1}{2} d_z^T \nabla^2_{zz} f(x(y),y) d_z$ to both sides of (49), we obtain

$$\nabla_y L(x(y),y,\mu^0)d_y^1 + \frac{1}{2} d_z^T \nabla^2_{zz} L(x(y),y,\mu^0) d_z + \delta_2 < \varphi''(y;d_y,d_y^1). \tag{50}$$

Now, along the sequence $y^k = y + \alpha_k d_y + \alpha_k^2 d_y^1$, consider the sequence $\{\mu^k\}$ such as $\mu^k \in M(y^k)$ converging to $\mu^0 \in V(d_y,d_y^1)$. By Theorem 3.2, $x(\cdot)$ admits the second-order directional Taylor expansion:

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

Setting $d_\mu := \lim_{k\to\infty} \frac{\mu^k - \mu^0}{\alpha_k}$ along a suitable subsequence, we can write $\mu^k = \mu^0 + \alpha_k d_\mu + o(\alpha_k)$. Without loss of generality, that subsequence is $\{y^k\}$. Expanding the Lagrangian $L(x(y^k),y^k,\mu^k) = f(x(y^k),y^k) + (\mu^k)^T g(x(y^k),y^k)$ around $(x(y),y,\mu^0)$ up to second order in α_k gives

$$\begin{aligned} L(x(y^k),y^k,\mu^k) &= L(x(y),y,\mu^0) \\ &\quad + \alpha_k \Big(\nabla_x L(x(y),y,\mu^0) x'(y;d_y) + \nabla_y L(x(y),y,\mu^0) d_y + g(x(y),y)^T d_\mu \Big) \\ &\quad + \alpha_k^2 \Bigg(\nabla_x L(x(y),y,\mu^0) x''(y;d_y,d_y^1) + \nabla_y L(x(y),y,\mu^0) d_y^1 \\ &\qquad + \frac{1}{2} \begin{pmatrix} d_z^T & d_\mu^T \end{pmatrix} \widehat{\nabla}^2 L(x(y),y,\mu^0) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} + g(x(y),y)^T \left(\lim_{k\to\infty} \frac{\mu^k - \mu^0 - \alpha_k d_\mu}{\alpha_k^2} \right) \Bigg) \\ &\quad + o(\alpha_k^2). \end{aligned} \tag{51}$$

Since $\mu^0 \in M(y)$, $\nabla_x L(x(y),y,\mu^0) = 0$. For active indices $i \in I_g(y)$, $g_i(x(y),y) = 0$, while for inactive indices $i \notin I_g(y)$, $\mu_i^k = \mu_i^0 = 0$ for all large k. Thus,

$$g(x(y),y)^T d_\mu = 0 \quad \text{and} \quad g(x(y),y)^T \left(\lim_{k\to\infty} \frac{\mu^k - \mu^0 - \alpha_k d_\mu}{\alpha_k^2} \right) = 0.$$

Furthermore, from the quadratic form $\widehat{\nabla}^2 L(x(y),y,\mu^0)$,

$$\begin{aligned} \frac{1}{2} \begin{pmatrix} d_z^T & d_\mu^T \end{pmatrix} \widehat{\nabla}^2 L(x(y),y,\mu^0) \begin{pmatrix} d_z \\ d_\mu \end{pmatrix} &= \frac{1}{2} d_z^T \nabla^2_{zz} L(x(y),y,\mu^0) d_z + d_\mu^T \nabla^2_{\mu z} L(x(y),y,\mu^0) d_z + \frac{1}{2} d_\mu^T \nabla^2_{\mu\mu} L(x(y),y,\mu^0) d_\mu \\ &= \frac{1}{2} d_z^T \nabla^2_{zz} L(x(y),y,\mu^0) d_z + \sum_{i=1}^q d_{\mu_i} \Big(\nabla_x g_i(x(y),y) x'(y;d_y) + \nabla_y g_i(x(y),y) d_y \Big), \end{aligned}$$

where $\nabla^2_{\mu\mu} L(x(y),y,\mu^0) = 0$ because L is affine in μ. For $i \notin I_g(y)$, $d_{\mu_i} = 0$. For $i \in J(\mu^0)$, the first-order condition (11b) gives $\nabla_x g_i(x(y),y) x'(y;d_y) + \nabla_y g_i(x(y),y) d_y = 0$, while for $i \in I_g(y) \setminus J(\mu^0)$, the first-order complementary condition (11d) ensures that $d_{\mu_i} \big(\nabla_x g_i(x(y),y) x'(y;d_y) + \nabla_y g_i(x(y),y) d_y \big) = 0$. Hence,

$$d_\mu^T \nabla^2_{\mu z} L(x(y),y,\mu^0) d_z = \sum_{i=1}^q d_{\mu_i} \Big(\nabla_x g_i(x(y),y) x'(y;d_y) + \nabla_y g_i(x(y),y) d_y \Big) = 0.$$

By complementary condition $\mu^k g(x(y^k),y^k) = \mu^0 g(x(y),y) = 0$, we have $\varphi(y^k) = L(x(y^k),y^k,\mu^k)$ and $\varphi(y) = L(x(y),y,\mu^0)$. Substituting into (51) gives

$$\varphi(y^k) - \varphi(y) = \alpha_k \nabla_y L(x(y),y,\mu^0) d_y + \alpha_k^2 \left(\nabla_y L(x(y),y,\mu^0) d_y^1 + \frac{1}{2} d_z^T \nabla^2_{zz} L(x(y),y,\mu^0) d_z \right) + o(\alpha_k^2).$$

Since $\mu^0 \in S^1(d_y)$, we have $\varphi'(y;d_y) = \nabla_y L(x(y),y,\mu^0) d_y$ using Lemma .1. Subtracting $\alpha_k \varphi'(y;d_y)$, dividing by α_k^2, and taking the limit as $k \to \infty$, we obtain:

$$\varphi''(y;d_y,d_y^1) = \nabla_y L(x(y),y,\mu^0) d_y^1 + \frac{1}{2} d_z^T \nabla^2_{zz} L(x(y),y,\mu^0) d_z.$$

This contradicts (50) since $\delta_2 > 0$. Therefore, $V(d_y,d_y^1) \subseteq S^2(d_y,d_y^1)$.
Since $V(d_y,d_y^1)$ is nonempty and $V(d_y,d_y^1) \subseteq S^2(d_y,d_y^1)$, it follows immediately that

$$\emptyset \neq V(d_y,d_y^1) = V(d_y,d_y^1) \cap S^2(d_y,d_y^1).$$

This completes the proof. □