# GENERIC PROPERTY OF THE PARTIAL CALMNESS CONDITION FOR BILEVEL PROGRAMMING PROBLEMS[*]

RONGZHU KE[†], WEI YAO[‡], JANE J. YE[§], AND JIN ZHANG[¶]

**Abstract.** The partial calmness for the bilevel programming problem (BLPP) is an important condition which ensures that a local optimal solution of BLPP is a local optimal solution of a partially penalized problem where the lower level optimality constraint is moved to the objective function and hence a weaker constraint qualification can be applied. In this paper we propose a sufficient condition in the form of a partial error bound condition which guarantees the partial calmness condition. We analyse the partial calmness for the combined program based on the Bouligand (B-) and the Fritz John (FJ) stationary conditions from a generic point of view. Our main result states that the partial error bound condition for the combined programs based on B and FJ conditions are generic for an important setting with applications in economics and hence the partial calmness for the combined program is not a particularly stringent assumption. Moreover we derive optimality conditions for the combined program for the generic case without any extra constraint qualifications and show the exact equivalence between our optimality condition and the one by Jongen and Shikhman given in implicit form. Our arguments are based on Jongen, Jonker and Twilt's generic (five type) classification of the so-called generalized critical points for one-dimensional parametric optimization problems and Jongen and Shikhman's generic local reductions of BLPPs.

**Key words.** partial calmness, bilevel programming, generic property, partial error bound, optimality conditions, generalized equation

**AMS subject classifications.** 90C26, 90C30, 90C31, 90C33, 90C46, 49J52, 91A65

**1. Introduction.** In this paper we consider the following bilevel programming problem (BLPP):

$$\text{(BLPP)} \qquad \begin{aligned} &\min_{x,y} F(x,y) \\ &\text{s.t. } y \in S(x), \ G(x,y) \le 0, \end{aligned}$$

where $S(x)$ denotes the solution set of the lower level program

$$L(x) \qquad \min_y f(x,y) \quad \text{s.t.} \quad g(x,y) \le 0.$$

For convenience, we denote the feasible set of the lower level program by

$$(1.1) \qquad Y(x) := \big\{ y \in \mathbb{R}^m : g(x,y) \le 0 \big\}.$$

[†]School of Economics and Research Center for China's Digital Economy, Zhejiang University, Hangzhou, Zhejiang 310058, China (rzke455@zju.edu.cn).

[‡]Department of Mathematics, Southern University of Science and Technology, Shenzhen, Guangdong, China (yaow@sustech.edu.cn).

[§]Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada V8W 2Y2 (janeye@uvic.ca).

[¶]Corresponding author. Department of Mathematics and SUSTech International Center for Mathematics, Southern University of Science and Technology, and National Center for Applied Mathematics, Shenzhen, Guangdong, China (zhangj9@sustech.edu.cn).

Here $x \in \mathbb{R}^n$ and the mappings $F, f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$. We assume that $F, G, f, g$ are continuously differentiable unless otherwise specified.

BLPPs have broad applications. For example, they appear in the principal-agent moral hazard problem [34], electricity markets and networks [5]. Recently, BLPP has been applied to hyperparameters optimization and and meta-learning in machine learning; see e.g. [29, 18, 30, 51]. More applications can be found in the monographs [45, 4, 10, 15], the survey on bilevel optimization [17] and the references therein.

The main difficulty in studying BLPPs is that the constraint $y \in S(x)$ is a global one. The classical and most often used approach to study BLPPs is to replace the lower level problem by the Karush-Kuhn-Tucker (KKT) condition and minimize over the original variables as well as multipliers. The resulting problem is the so-called mathematical program with complementarity/equilibrium constraints (MPCC/MPEC) ([32, 42]), which itself is a difficult problem if treated as a nonlinear programming problem (NLP) since the Mangasarian Fromovitz constraint qualification (MFCQ) is violated at any feasible point of MPEC, see Ye et al. [55]. In general, this approach is only applicable to BLPPs where the lower level problem is a convex optimization problem in variable $y$ since KKT condition is not sufficient for $y \in S(x)$ when the lower level program is not convex. As a matter of fact, it was pointed out by Mirrlees [34] that an optimal solution of BLPP may not even be a stationary point of the resulting single-level problem by KKT/MPEC approach (usually called the first order approach in economics). Moreover, since the reformulated problem involves additionally the multipliers of the lower level problem as variables, as pointed out by Dempe and Dutta in [11], it is possible that $(\bar{x}, \bar{y}, \bar{u})$ is a local optimal solution of the reformulated problem but $(\bar{x}, \bar{y})$ is not a local optimal solution of BLPP even when the lower level problem is convex.

To deal with BLPPs without convexity assumption on the lower level problem, the value function approach was proposed by Outrata [41] for numerical purpose and used by Ye and Zhu [53] for optimality conditions. By defining the value (or the marginal) function as $V(x) := \inf_y \{ f(x, y) : g(x, y) \leq 0 \}$, one can replace the original BLPP by the following equivalent problem:

(VP)
$$\min_{x,y} F(x, y)$$
$$\text{s.t. } f(x, y) - V(x) \leq 0, \ g(x, y) \leq 0, \ G(x, y) \leq 0.$$

The resulting problem (VP) is clearly equivalent to the original bilevel problem. However, since the value function constraint $f(x, y) - V(x) \leq 0$ is actually an equality constraint, the nonsmooth MFCQ for (VP) will never hold (cf. [53, Prop. 3.2], [49, Section 4]). To derive necessary optimality conditions for BLPPs, Ye and Zhu [53] proposed the partial calmness condition for (VP) under which the difficult constraint $f(x, y) - V(x) \leq 0$ is penalized to the objective function. Under the partial calmness condition for (VP), the usual constraint qualifications such as MFCQ can be satisfied for the partially penalized (VP). Consequently, by using a nonsmooth multiplier rule, one can derive a KKT type optimality condition for (VP) provided that the value function is Lipschitz continuous. We refer the reader to [53, 46, 47, 12, 16, 37] for more discussions.

Although it was proved in [53] that the partial calmness condition for (VP) holds automatically for the minmax problem and the bilevel program where the lower level problem is linear in both variables $x$ and $y$, the partial calmness condition for (VP) has been shown to be a celebrated but restrictive assumption (see [23, 49, 33]). To

improve the value function approach, Ye and Zhu [54] proposed a combination of the classical KKT and the value function approach. The resulting problem is the combined problem using KKT condition:

(CP)
$$\min_{x,y,u} \ F(x,y)$$
$$\text{s.t.} \ f(x,y) - V(x) \le 0,$$
$$\nabla_y f(x,y) + u\nabla_y g(x,y) = 0,$$
$$g(x,y) \le 0, \ u \ge 0, \ u^T g(x,y) = 0, \ G(x,y) \le 0,$$

where $u\nabla_y g(x,y) := \sum_{i=1}^p u_i \nabla g_i(x,y)$. Similarly to [53], to deal with the fact that the nonsmooth MFCQ also fails for (CP), the following partial calmness condition for (CP) was proposed in [54].

DEFINITION 1.1 (**Partial calmness for** (CP)). *Let $(\bar{x}, \bar{y}, \bar{u})$ be a feasible solution of* (CP). *We say that* (CP) *is partially calm at $(\bar{x}, \bar{y}, \bar{u})$ if there exists $\mu \ge 0$ such that $(\bar{x}, \bar{y}, \bar{u})$ is a local solution of the following partially penalized problem:*

(CP$_\mu$)
$$\min_{x,y,u} \ F(x,y) + \mu\big(f(x,y) - V(x)\big)$$
$$\text{s.t.} \ \nabla_y f(x,y) + u\nabla_y g(x,y) = 0,$$
$$g(x,y) \le 0, \ u \ge 0, \ u^T g(x,y) = 0, \ G(x,y) \le 0.$$

Since there are more constraints in (CP) than in (VP), the partial calmness for (CP) is a weaker condition than the one for (VP). Moreover, it was shown in [54, 48] that under the partial calmness condition and various MPEC variant constraint qualifications for (CP$_\mu$), a local optimal solution of (CP) must be a corresponding MPEC variant stationary point based on the value function provided the value function is Lipschitz continuous. See [54, 48, 49] for more details. In the same spirit, Nie et al. [39] proposed a numerical algorithm for globally solving polynomial bilevel programs by adding the generalized critical point condition for the lower level program. Recently, Nie et al. [40] expressed the Lagrange multipliers of the lower level polynomial problem as a function of the upper and lower level variables and hence eliminated the multiplier variables in the reformulation of BLPPs in [39].

The reformulation (CP), however, requires the existence of KKT condition at each optimal solution of the lower level program (see [54, Proposition 3.1]). In order to deal with the case where KKT condition may not hold at all the solutions of the lower level program, we propose the following combined program using the Fritz John (FJ) condition[1]:

(CPFJ)
$$\min_{x,y,u_0,u} \ F(x,y)$$
$$\text{s.t.} \ f(x,y) - V(x) \le 0,$$
$$u_0\nabla_y f(x,y) + u\nabla_y g(x,y) = 0, \ g(x,y) \le 0, \ (u_0, u) \ge 0,$$
$$u^T g(x,y) = 0, \ \|(u_0, u)\|_1 = 1, \ G(x,y) \le 0,$$

where the one-norm $\|(u_0, u)\|_1 := \sum_{i=0}^p |u_i| = \sum_{i=0}^p u_i$ when $(u_0, u) \ge 0$. A relationship between problems (CPFJ) and (BLPP) can be established by similar arguments

---

[1]FJ condition was firstly used by Allende and Still [2] for BLPPs, where they replaced the lower level program by FJ condition and the resulting problem becomes (CPFJ) without the value function constraint.

as in [54, Proposition 3.1]. Similarly to [54], we propose the partial calmness condition for (CPFJ).

DEFINITION 1.2 (**Partial calmness for** (CPFJ)). *Let $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ be a feasible solution of* (CPFJ). *We say that* (CPFJ) *is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ if there exists $\mu \geq 0$ such that $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ is a local solution of the partially penalized problem:*

$$
\begin{aligned}
&\min_{x,y,u_0,u} \ F(x,y) + \mu\big(f(x,y) - V(x)\big) \\
\text{(CPFJ}_\mu\text{)} \quad &\text{s.t. } u_0 \nabla_y f(x,y) + u\nabla_y g(x,y) = 0, \ g(x,y) \leq 0, \ (u_0, u) \geq 0, \\
&\qquad u^T g(x,y) = 0, \ \|(u_0, u)\|_1 = 1, \ G(x,y) \leq 0.
\end{aligned}
$$

Inspired by [1, 19, 20, 49] where the authors study the mathematical program with a generalized equation involving the regular normal cone, we consider the combined program with the Bouligand (B-) stationary condition for the lower level program:

$$
\begin{aligned}
&\min_{x,y} \ F(x,y) \\
\text{(CPB)} \quad &\text{s.t. } f(x,y) - V(x) \leq 0, \\
&\qquad 0 \in \nabla_y f(x,y) + \widehat{N}_{Y(x)}(y), \ G(x,y) \leq 0.
\end{aligned}
$$

Here $\widehat{N}_{Y(x)}(y)$ is the regular normal cone to $Y(x)$ at $y$ if $y \in Y(x)$ (see Definition 2.1) and is equal to the empty set if $y \notin Y(x)$. The combined problem (CPB) is equivalent to the original bilevel program since the condition

$$
\text{(1.2)} \qquad 0 \in \nabla_y f(x,y) + \widehat{N}_{Y(x)}(y)
$$

is a necessary optimality condition for $y \in S(x)$. Indeed, if $y$ is a solution to the lower level problem $L(x)$, then there is no descent direction starting from $y$ and hence B-condition holds which means that (1.2) must hold; see Section 2 for more details. Similarly, we propose the following partial calmness condition for (CPB).

DEFINITION 1.3 (**Partial calmness for** (CPB)). *Let $(\bar{x}, \bar{y})$ be a feasible solution of* (CPB). *We say that* (CPB) *is partially calm at $(\bar{x}, \bar{y})$ if there exists $\mu \geq 0$ such that $(\bar{x}, \bar{y})$ is a local solution of the following partially penalized problem:*

$$
\begin{aligned}
&\min_{x,y} \ F(x,y) + \mu\big(f(x,y) - V(x)\big) \\
\text{(CPB}_\mu\text{)} \quad &\text{s.t. } 0 \in \nabla_y f(x,y) + \widehat{N}_{Y(x)}(y), \ G(x,y) \leq 0.
\end{aligned}
$$

Note that there are other alternative combined programs. For example, we may consider the combined program

$$
\begin{aligned}
&\min_{x,y} \ F(x,y) \\
\text{(CP}_{FJ}\text{)} \quad &\text{s.t. } f(x,y) - V(x) \leq 0, \\
&\qquad (x,y) \in \Sigma_{FJ}, \ G(x,y) \leq 0,
\end{aligned}
$$

where $\Sigma_{FJ}$ is the set of FJ points defined as in the Fritz John condition (FJ) and the partial calmness condition can be similarly defined.

From the discussion above, we can define various types of combined programs and the corresponding partial calmness conditions based on different stationary conditions for the lower level program. But combined programs can be classified into two

types. One involves multipliers as extra variables such as (CP) and (CPFJ) and the other does not involve any extra variables such as (CPB) and $(\text{CP}_{FJ})$. These two types of combined programs have its own advantages and disadvantages. The one with multipliers such as (CP) and (CPFJ) are explicit and easier to handle while the one without multipliers are implicit and harder to handle. An advantage of using a combined program without multipliers such as (CPB) is that the constraint qualification for $(\text{CPB}_{\mu})$ is in general weaker than the one for $(\text{CP}_{\mu})$. By a counterexample in [1, Appendix], it is possible that a feasible solution of $(\text{CPB}_{\mu})$ satisfies the calmness condition at $(\bar{x}, \bar{y}, 0)$ while the one for the corresponding $(\text{CP}_{\mu})$ does not satisfy the calmness condition at $(\bar{x}, \bar{y}, u, 0)$ for any multiplier $u$. By [19, Proposition 5], we see that for the case where $Y(x)$ is independent of $x$, if the lower level multipliers are not unique, then the MPEC-MFCQ never holds for $(\text{CP}_{\mu})$ and even the very weak constraint qualification MPCC-GCQ can be violated for $(\text{CP}_{\mu})$ while the calmness condition can hold at $(\bar{x}, \bar{y}, 0)$ for $(\text{CPB}_{\mu})$.

**Questions.** The main goal of this paper is to investigate two questions:

(Q1)     How stringent is the partial calmness condition for combined program?

This question is at present far from being solved. It is not clear, a priori, whether the partial calmness for the combined program is a typical property for BLPPs.

The second question is:

(Q2)     Is there a generic sufficient condition to ensure
the partial calmness condition for the combined program?

Here a property is said to be generic if it holds for all problems generated by data in an open and dense subset or, more generally, in the intersection of countably many open and dense subsets. It is well known that the nondegeneracy of KKT points is a generic property for standard nonlinear programming problems (cf. [27]) and it has been also shown that a MPCC variant of Linearly Independence Constraint Qualification (LICQ) called MPEC-LICQ is a generic property for MPCC/MPEC (cf. [44, 2]). The partial calmness condition plays an important role in the theory and applications of BLPPs, particularly in deriving optimality conditions since the partially penalized problem does not have the difficult value function constraint in it; [35, Chapter 6], [36], and [49].

**Contributions.** To answer the above questions, by virtue of the exact penalty principle of Clarke, we propose a sufficient condition for the partial calmness in the form of a partial error bound condition in Definition 4.1. In the case where the upper level variable has dimension one, our main result (Theorems 4.6) shows that the partial error bound (PEB) and consequently the partial calmness condition for $(\text{CP}_{FJ})$ and (CPB) are generic for BLPPs. Moreover, by an equivalence result in Theorem 4.4, we have that the partial calmness condition for (CPFJ) is also generic. This gives an affirmative answer to (Q2) in this case. Hence, generically, all local optimal solutions of BLPP are local optimal solutions of $(\text{CPB}_{\mu})$ for some nonnegative constant $\mu \geq 0$. Similarly, generically, all local optimal solutions to (CPFJ) are local optimal solutions of $(\text{CPFJ}_{\mu})$ for some nonnegative constant $\mu \geq 0$. Combining these results and recent results in [19, 20] for mathematical program with a parametric generalized equation gives some new sharp necessary optimality conditions for BLPPs. For the generic cases, we are able to show that MPEC-LICQ holds for partially perturbed problem $(\text{CPFJ}_{\mu})$ without the upper level constraint. Consequently in Theorem 6.2, we derive necessary optimality conditions for the generic cases explicitly in problem data and

without any extra constraint qualifications. Further, in Theorem 6.3, we prove that our optimality conditions for the generic case are equivalent to the one derived by Jongen and Shikhman in [28] and hence partially answer a question raised in [28, Remark 4.2]. When there is no upper and lower level constraints in BLPPs, our result (Corollary 6.5) shows that the condition in [54, Theorem 2.1] is a generic one. This gives a rigorous proof of Mirrlees' arguments about $f$-critical points in [34].

**Technique.** By adopting a parametric programming point of view, our arguments to prove the main result (Theorems 4.6) are based on Jongen, Jonker and Twilt's generic (five type) classification of the so-called generalized critical points (g.c.) for one-dimensional parametric optimization problems [26], and Jongen and Shikhman's generic local reductions of BLPPs in the case where the dimension of the upper variable is one [28]. To obtain local reduction, Jongen and Shikhman [28] studied the structure of points being local minimizers for $L(x)$. In this paper, to study the combined problems (CP), (CPFJ), (CPB) and ($\text{CP}_{FJ}$), we study the structure of points being KKT, FJ, and B-points for $L(x)$, by using a similar analysis in [28]. The roadmap in Figure 4.2 is helpful at first glance and the details are left to Sections 4,5, 6. Note that although the classification is done under the assumption that the parameter is one dimensional, it gives some indications on the situation in the general case where the parameter is from a multi-dimensional space; see [27, 13, 28, 14] for more discussions.

The generic classification of g.c. points is studied within the scope of singularity theory (see e.g. [7]). The proofs of results in [26] are mainly based on the transversality theory. We refer the reader to [27] for a detailed account of the topological aspects of nonlinear optimization. It has proved that topology method is a powerful tool for studying the fine structure of optimization problems (e.g. [27, 28, 44, 2, 14] and the references given there). This work is intended as an attempt to further motivate the study of topological aspects of BLPPs. And it is of interest to bring together two areas in studying BLPPs. One is to study the generic structure of bilevel problems in a neighborhood of its solution set, cf. [13, 28, 14]. Another one is to study constraint qualifications and optimality conditions for BLPPs, see e.g. [54, 36, 49].

**Outline.** The remaining part of the paper is organized as following. In Section 2, we gather some preliminaries in variational analysis. Some motivational applications in economics will be described in Section 3. In Section 4, we introduce the main technique and state our main result whose proof is given in Section 5. As a consequence, we study optimality conditions for BLPPs in Section 6.

**Symbols and Notations.** Our notation is basically standard. By $0_m$, we mean it is the zero vector in $\mathbb{R}^m$. For a vector $z \in \mathbb{R}^d$, we denote its Euclidean norm by $\|z\|$. Simply by $\mathbb{B}$, we denote the closed unit ball centered at the origin of the space in question. For a matrix $A$, we denote by $A^T$ its transpose and for a symmetric matrix $A$ we denote by $\det A$ its determinant. The inner product of two vectors $x, y$ is denoted by $x^T y$ or $\langle x, y \rangle$. For $z \in \mathbb{R}^d$ and $\Omega \subseteq \mathbb{R}^d$, we denote by $\text{dist}_\Omega(z)$ the distance from $z$ to $\Omega$. By $z \xrightarrow{\Omega} \bar{z}$ we mean $z \to \bar{z}$ and $z \in \Omega$. The polar cone of a set $\Omega$ is $\Omega^\circ := \{v : \langle v, z \rangle \le 0 \ \forall z \in \Omega\}$. We also denote by $\text{co}\,\Omega$ the convex hull of $\Omega$. For a function $h : \mathbb{R}^d \to \mathbb{R}$, we denote the gradient vector and the Hessian of $h$ at $z$ by $\nabla h$ and $\nabla^2 h$, respectively. Given a differentiable one-dimensional function $\Psi : \mathbb{R} \to \mathbb{R}^\ell$, $D\Psi$ denotes its derivative. Let $\Phi : \mathbb{R}^d \rightrightarrows \mathbb{R}^s$ be a set-valued map. We denote its graph by $\text{gph}\Phi := \{(z, w) : w \in \Phi(z)\}$. Let $C^3(\mathbb{R}^d)$ denotes the space of three times continuously differentiable real-valued functions on $\mathbb{R}^d$. We denote by $C_s^3$ the strong (or Whitney) $C^3$-topology and refer the reader to [27, Section 6.1] for more details on

$C_s^3$-topology.

**2. Preliminaries on variational analysis.** In this section, we gather some preliminaries in variational analysis that will be needed. The reader is referred to Rockafellar-Wets [43] and the references therein for more discussions on the subject.

First we recall the definition for the tangent cone and the normal cones.

DEFINITION 2.1 (**Tangent and normal cones**). *[43, Definitions 6.1 and 6.3] Given a set $\Omega \subseteq \mathbb{R}^d$ with $\bar{z} \in \Omega$, the (Bouligand-Severi) tangent/contingent cone to $\Omega$ at $\bar{z}$ is a closed cone defined by*

$$T_\Omega(\bar{z}) := \limsup_{t \downarrow 0} \frac{\Omega - \bar{z}}{t} = \left\{ u \in \mathbb{R}^d : \exists t_k \downarrow 0, u_k \to u \text{ with } \bar{z} + t_k u_k \in \Omega, \ \forall k \right\}.$$

*The (Fréchet) regular normal cone and the (Mordukhovich) limiting/basic normal cone to $\Omega$ at $\bar{z} \in \Omega$ are closed cones defined by*

$$\widehat{N}_\Omega(\bar{z}) := \left\{ v \in \mathbb{R}^d : \limsup_{z \xrightarrow{\Omega} \bar{z}} \frac{\langle v, z - \bar{z} \rangle}{\|z - \bar{z}\|} \leq 0 \right\}$$

*and* $\quad N_\Omega(\bar{z}) := \left\{ v \in \mathbb{R}^d : \exists z_k \xrightarrow{\Omega} \bar{z}, \ v_k \in \widehat{N}_\Omega(z_k) \text{ with } v_k \to v \text{ as } k \to \infty \right\}.$

When the set $\Omega$ is convex, the tangent/contingent cone and the regular/limiting normal cone reduce to the derivable cone and the normal cone in the sense of convex analysis, respectively; see e.g., [43, Theorem 6.9].

It is well-known that a necessary condition for $\bar{z}$ to be locally optimal to the problem of minimizing a differentiable function $f_0$ over a set $\Omega \subseteq \mathbb{R}^d$ is

$$(2.1) \qquad \qquad \langle \nabla f_0(\bar{z}), u \rangle \geq 0 \quad \text{ for all } u \in T_\Omega(\bar{z}).$$

By the tangent-normal polarity ([43, Theorem 6.28]) $\widehat{N}_\Omega(\bar{z}) = T_\Omega(\bar{z})^\circ$, the condition (2.1) is the same as

$$(2.2) \qquad \qquad - \nabla f_0(\bar{z}) \in \widehat{N}_\Omega(\bar{z}) \quad \text{or} \quad 0 \in \nabla f_0(\bar{z}) + \widehat{N}_\Omega(\bar{z}).$$

Since the condition (2.1) is called the Bouligand (B-) stationary condition in the literature, and conditions (2.1) and (2.2) are equivalent, we also call the condition (2.2) B-condition and any $y$ satisfying B-condition a B-point in this paper.

We now review some concepts of stability of a set-valued map.

DEFINITION 2.2 (**Calmness [50, 43]**). *Let $\Gamma : \mathbb{R}^d \rightrightarrows \mathbb{R}^s$ be a set-valued map and $(\bar{z}, \bar{w}) \in \text{gph}\,\Gamma$. We say that $\Gamma$ is calm (or pseudo upper-Lipschitz continuous) at $(\bar{z}, \bar{w})$ or equivalently its inverse map $\Gamma^{-1}$ is metrically subregular at $(\bar{w}, \bar{z})$ if there exist a neighbourhood $\mathcal{Z}$ of $\bar{z}$, a neighborhood $\mathcal{W}$ of $\bar{w}$ and $\kappa \geq 0$ such that*

$$\Gamma(z) \cap \mathcal{W} \subseteq \Gamma(\bar{z}) + \kappa \|z - \bar{z}\| \mathbb{B}, \ \forall z \in \mathcal{Z}.$$

DEFINITION 2.3 (**Metric subregularity constraint qualification**). *Let $g(\bar{x}, \bar{y}) \leq 0$, where $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$. We say that the system $g(x, y) \leq 0$ satisfies the metric subregularity constraint qualification (MSCQ) at $(\bar{x}, \bar{y})$ if the perturbed set-valued map $\Gamma(z) := \{(x, y) : g(x, y) \leq z\}$ is calm at $(0, \bar{x}, \bar{y})$ or equivalently the set-valued map $\Gamma^{-1}(x, y) := -g(x, y) + \mathbb{R}_-^p$ is metrically subregular at $(\bar{x}, \bar{y}, 0)$.*

It is known that the linear CQ (when $g$ is affine) and MFCQ both imply MSCQ. A summary of more CQs which implies MSCQ is given in [52, Theorem 7.4].

**3. Motivating applications in economics.** Our research on BLPPs is motivated by a number of important applications in economics. Below we describe some of them.

• **The classic principal-agent moral hazard problem**

In the situation of the moral hazard principal-agent problem, the agent chooses an action $a \in A$ that is unobservable to a principal. The action $a$ influences the random outcome $s \in S$ through a parameterized probability measure $P(\cdot|a)$ (see [34]). The outcome $s$ leads to a monetary payoff $\pi(s)$, which accrues directly to the principal. Since the principal cannot monitor the agent's action but only the outcome, the principal will pay the agent conditional on the observed outcome. Thus the compensation contract specifies a wage of $w(s)$. If $S$ is a discrete subset of $\mathbb{R}$ with cardinality $k$, then the contract is a vector $w = (w_1, \ldots, w_k) \in \mathbb{R}^k$. The principal's utility for outcome $s$ is a function of the net value $\pi(s) - w(s)$ and is denoted by $v(\pi(s) - w(s))$. The agent receives the wage $w(s)$ and pays the action $a$, then he receives utility $u(w(s), a)$. Hence their expected utility functions are

$$V(w, a) = \int_S v(\pi(s) - w(s))dP(s|a), \quad \text{and} \quad U(w, a) = \int_S u(w(s), a)dP(s|a).$$

Therefore, the principal's problem in parameterized distribution terms can be stated:

(PA) $$\max_{w,a} V(w, a)$$

(IC) $$\text{s.t. } a \in \arg\max_{a' \in A} U(w, a'),$$

(IR) $$U(w, a) \geq \underline{U}.$$

The first constraint is the incentive compatibility constraint which means that the agent will only take actions that maximize his own expected utility. Here we consider optimistic formulation. That is, we assume implicitly that the agent will choose the action most beneficial to the principal if he is indifferent between several different actions. The second constraint is the individual rationality constraint (or participation constraint) which guarantees that the agent earns at least his reservation utility $\underline{U}$.

The difficulty to solve the above moral hazard problem is well-recognized (e.g., Conlon [9]). But according to our method developed in this paper, two important categories of problem can fall into the special class of genericty. The first special case is binary outcome distribution $k = 2$. The second special case is that the contract takes a linear form $w(s) = \alpha + \beta s$.

◇ **Binary Outcome**

Binary outcome model is useful in economic literature (cf. Grossman and Hart [21]).[2] Suppose that the principal is risk-neutral (i.e., $v(r) = r$) and assume that agent's utility is additively separable (i.e., $u(w, a) = u(w) - c(a)$). Consider the case $S = \{s_1, s_2\}$ and assume that $u$ is increasing. It is known that at the optimal solution, the IR constraint must be binding [21, Proposition 2]. Hence by setting

$$p_i(a) := P(\{s_i\}|a), \ \pi_i := \pi(s_i), \ w_i := u(w(s_i)), \ \text{and} \ x := w_1 - w_2, \ y := a,$$

---

[2]For example, in the case of social insurance, the outcome of job search is a binary outcome, unemployment or employment. In the case of coroperate finance, the outcome of investment project could be a binary outcome, success or not success.

the principal-agent problem becomes

$$\min_{x,y} F(x,y) \quad \text{s.t.} \quad y \in \arg\min_{y' \in A} \big\{ -p_1(y')x + c(y') \big\},$$

where

$$F(x,y) := p_1(y) \left[ h\big(\underline{U} + c(y) + p_2(y)x\big) - \pi_1 \right] + p_2(y) \left[ h\big(\underline{U} + c(y) - p_1(y)x\big) - \pi_2 \right],$$

and $h(\cdot) := u^{-1}(\cdot)$ is increasing. This shows that the case $k = 2$ in (PA) becomes the case $n = 1$ in (BLPP).

◇ **Linear Contract**

Linear contracts play an important role in contract theory. Not only the linear contract is intuitive and simple in practice, but also Holmström and Milgrom [24] provided a theoretical foundation for the optimality of linear contract in a dynamic setting where the agent with constant absolute risk aversion controls the drift of a Brownian motion.

Assume that the agent's utility is additively separable, we consider a linear contract $w(s) = \alpha + \beta s$, where the slope $\beta$ is the "incentive intensity". Then agent's expected utility function becomes

$$\int_S u(\alpha + \beta s) dP(s|a) - c(a).$$

Assuming that $u$ is an increasing function, we can solve $\alpha = \alpha(\beta, a)$ in terms of $\beta$ and $a$ by the binding IR constraint. Therefore, relabel the notation $x = \beta, y = a$, for every implementable $y = a$ (i.e., there exists some $x$ such that $(x,y)$ is bilevel feasible), we can write the original moral hazard problem as

$$\min_x - \int_S v\Big(\pi(s) - \big(\alpha(x,y) + xs\big)\Big) dP(s|y)$$

$$\text{s.t. } y \in \arg\min_{y' \in A} \Big\{ c(y') - \int_S u(\alpha(x,y) + xs) \, dP(s|y') \Big\},$$

which belongs to the class specifed by (BLPP). ³

• **Multitask principal-agent problems**

Beyond the single-task moral hazard problem, some multi-task problem (see Holmström and Milgrom [25] for the discussion of multi-task agency problem) can also be addressed by our approach. Consider a similar model of Bond and Gomes [6], a principal employs an agent to work on $m$ symmetric projects. Each project can result in either success or failure, and the total value of the output is $f_j$, a function of the number of successful projects, $j = 0, \ldots, m$. The agent's actions determine the probability distribution of success over these tasks: $p = (p_1, \ldots, p_m) \in [0,1]^m$. We can view $p$ itself as the agent's action and write $c(p)$ for the agent's cost of action. The limited attention constraint $\sum_{i=1}^m p_i \leq \bar{p}$ is assumed for the agent.

Since the agent's effort allocation is not observable but the number of successful tasks is, the principal rewards the agent based on the number of successes using an incentive scheme $w = (w_j)_{j=0}^m$ where $w_j$ is the agent's reward when there are $j$

---

³Note that here we do not include the lower level variable $y$ into the upper level optimization problem because $y$ is given. So we obtain a characterization of $x$ for every implementable $y$, which is slightly different. However, if the agent is risk-neutral or the contract is chosen to make $u(w(s)) = \alpha + \beta s$, we can obtain an exactly form of problem specified by (BLPP).

successful tasks. Let $A_{m,j}(p)$ be the probability of $j$ successes in the $m$ tasks under action $p$. Then the expected utility functions of the principal and agent are

$$V(w,p) = \sum_{j=0}^{m} v(f_j - w_j) A_{m,j}(p), \quad \text{and} \quad U(w,p) = \sum_{j=0}^{m} u(w_j) A_{m,j}(p) - c(p),$$

respectively, where $v(\cdot)$ and $u(\cdot)$ are the principal's and the agent's preferences over wealth, respectively. In summary, the principal's optimization problem is to select an incentive scheme $w$ and instructions $p$ for the agent so as to maximize his utility subject to the constraints,

$$\max_{w,p} V(w,p),$$

$$\text{s.t.} \quad p \in \arg \max_{p' \in [0,1]^m} \left\{ U(w,p') : \sum_{i=1}^{m} p'_i \leq \bar{p} \right\},$$

$$U(w,p) \geq \underline{U},$$

$$\bar{w} \geq w_j \geq 0, \ \forall j = 0, \ldots, m.$$

The first constraint is the incentive compatibility constraint, the second is the individual rationality constraint (or participation constraint), and the third is the minimum payment constraint and a budget constraint.

Note that the sequence of probabilities $A_{m,j}(p)$ is a Pólya frequency sequence in $j$ and it is shown in [6] that the sequence $A_{m,j}(p)$ satisfies the monotone likelihood ratio property (MLRP). When the agent is risk neutral, it is a standard result that the principal can maximize his payoff by using an incentive scheme of the form

$$(3.1) \qquad w_j = \begin{cases} 0 & \text{if } j < k, \\ (1-\mu)\bar{w} & \text{if } j = k, \\ \bar{w} & \text{if } j > k, \end{cases}$$

for some integer $k$ and $\mu \in [0,1]$, see, e.g., Proposition 1 in [6]. We denote this incentive scheme briefly by $w^{k,\mu}$. For each integer $k \leq m$, after relabeling $x = \mu$ and $y = p$, the principal's contracting problem reduces to (BLPP) where the upper variable $x$ is one dimensional.

**4. Main generic results and techniques.** In this section we will first propose sufficient conditions for the partial calmness conditions in Definitions 1.2-1.3 to hold. Then motivated by the applications discussed in Section 3, we prove that these conditions are generic for BLPPs under the following problem setting.

> **Problem Setting:**
> (1) The real valued mappings $f, g$ are three times continuously differentiable.
> (2) The upper variable $x$ is one dimensional ($n = 1$).

Let $J_0(\bar{x}, \bar{y}) := \{j : g_j(\bar{x}, \bar{y}) = 0\}$ be the active index set for $\bar{y} \in Y(\bar{x})$. We say that $\bar{y} \in Y(\bar{x})$ is a g.c. point, a FJ-point or a KKT point if

(g.c.)  $\exists (u_0, u) \neq 0$ s.t. $0 = u_0 \nabla_y f(\bar{x}, \bar{y}) + \sum_{j \in J_0(\bar{x}, \bar{y})} u_j \nabla_y g_j(\bar{x}, \bar{y}),$

(FJ)  $\exists (u_0, u) \neq 0$ s.t. $0 = u_0 \nabla_y f(\bar{x}, \bar{y}) + \sum_{j \in J_0(\bar{x}, \bar{y})} u_j \nabla_y g_j(\bar{x}, \bar{y}),$ and $(u_0, u) \geq 0,$

(KKT)  $\exists u \geq 0$ s.t. $0 = \nabla_y f(\bar{x}, \bar{y}) + \sum_{j \in J_0(\bar{x}, \bar{y})} u_j \nabla_y g_j(\bar{x}, \bar{y}),$

respectively. We call $(u_0, u)$ satisfying the g.c. condition a Lagrange vector and the one satisfying KKT condition a (Lagrange) multiplier. Note that a Lagrange vector is not assumed to be nonnegative. For each $(x, y)$, denote the set of FJ-multipliers at $y$ for problem $L(x)$ by

$$(4.1) \quad M_{\text{FJ}}(x, y) := \left\{ (u_0, u) \in \mathbb{R}^{1+p} : \begin{array}{l} u_0 \nabla_y f(x, y) + u \nabla_y g(x, y) = 0, \\ (u_0, u) \geq 0, \ \|(u_0, u)\|_1 = 1, \ u^T g(x, y) = 0 \end{array} \right\}.$$

Similarly, we denote the set of KKT-multipliers for $L(x)$ at $y$ as follows:

$$(4.2) \quad M_{\text{KKT}}(x, y) := \left\{ u : \nabla_y f(x, y) + u \nabla_y g(x, y) = 0, \ u \geq 0, \ u^T g(x, y) = 0 \right\}.$$

Define

$$\begin{aligned}
\Sigma_f &:= \{(x, y) \in \mathbb{R}^{n+m} : y \in Y(x)\}, \\
\Sigma_{gc} &:= \{(x, y) \in \mathbb{R}^{n+m} : y \text{ is a g.c. point for } L(x)\}, \\
\Sigma_{FJ} &:= \{(x, y) \in \mathbb{R}^{n+m} : y \text{ is a FJ point for } L(x)\}, \\
\Sigma_{KKT} &:= \{(x, y) \in \mathbb{R}^{n+m} : y \text{ is a KKT point for } L(x)\}, \\
\Sigma_B &:= \{(x, y) \in \mathbb{R}^{n+m} : y \text{ is a B-point for } L(x)\}, \\
M &:= \{(x, y) \in \mathbb{R}^{n+m} : y \in S(x)\}.
\end{aligned}$$

We can show that

$$(4.3) \quad M \subseteq \Sigma_B \subseteq \Sigma_{FJ} \subseteq \Sigma_{gc} \subseteq \Sigma_f, \quad \Sigma_{KKT} \subseteq \Sigma_B, \text{ and } \Sigma_B \overset{MSCQ}{=} \Sigma_{KKT}.$$

Indeed, as explained in Section 2, $M \subseteq \Sigma_B$ follows by the necessary optimality condition. Now we prove that the condition $0 \in \nabla_y f(x, y) + \widehat{N}_{Y(x)}(y)$ implies FJ condition. Indeed, if MFCQ holds, then it implies KKT condition (cf. [35, Exercise 2.53] or [22, Theorem 4.1]). Thus there exists a normal multiplier in FJ condition. Otherwise, by the Motzkin's transposition theorem [38], FJ condition with an abnormal multiplier holds. Hence

$$0 \in \nabla_y f(x, y) + \widehat{N}_{Y(x)}(y) \implies y \text{ is a FJ point of } L(x).$$

To show that $\Sigma_{KKT} \subseteq \Sigma_B$, it suffices to make the following observation. First, KKT condition for $L(\bar{x})$ at $\bar{y}$ can be rewritten as $-\nabla_y f(\bar{x}, \bar{y}) \in L_{Y(\bar{x})}(\bar{y})^\circ$, where $L_{Y(\bar{x})}(\bar{y}) := \{ d \in \mathbb{R}^m : \langle \nabla_y g_j(\bar{x}, \bar{y}), d \rangle \leq 0, \ \forall j \in J_0(\bar{x}, \bar{y}) \}$ is the linearized cone of $Y(\bar{x})$ at $\bar{y}$. Note that the linearized cone always contains the tangent cone, i.e., $L_{Y(\bar{x})}(\bar{y}) \supseteq T_{Y(\bar{x})}(\bar{y})$. Hence $L_{Y(\bar{x})}(\bar{y})^\circ \subseteq T_{Y(\bar{x})}(\bar{y})^\circ = \widehat{N}_{Y(\bar{x})}(\bar{y})$. This implies that $\Sigma_{KKT} \subseteq \Sigma_B$. The reverse inclusion $\Sigma_B \subseteq \Sigma_{KKT}$ usually requires some CQ. Indeed, by [19, Proposition 4], we have $\Sigma_B = \Sigma_{KKT}$ under MSCQ at each point of $(x, y) \in \Sigma_f$.

**4.1. Our main results.** We propose sufficient conditions to guarantee the partial calmness conditions in Definitions 1.2-1.3 and show that both of them are generic in Problem Setting. Using the terminology initiated in [31], in fact we can define the PEB condition based on other stationary (or feasible) conditions as below.

DEFINITION 4.1 (**Partial error bound**). *We say that the partial error bound (PEB) over B, KKT, FJ, cp, feasible (f)-condition is satisfied at a feasible solution*

$(\bar{x}, \bar{y})$ of (BLPP) *if the partially perturbed mapping*

$$K_B(v) := \big\{(x,y) : f(x,y) - V(x) \le v, \, 0 \in \nabla_y f(x,y) + \widehat{N}_{Y(x)}(y)\big\},$$
$$K_{KKT}(v) := \big\{(x,y) : f(x,y) - V(x) \le v, \; y \text{ is a KKT point of } L(x)\big\},$$
$$K_{FJ}(v) := \big\{(x,y) : f(x,y) - V(x) \le v, \; y \text{ is a FJ point of } L(x)\big\},$$
$$K_{cp}(v) := \big\{(x,y) : f(x,y) - V(x) \le v, \; y \text{ is a g.c. point of } L(x)\big\},$$
$$K_f(v) := \big\{(x,y) : f(x,y) - V(x) \le v, \, y \in Y(x)\big\},$$

*is calm at* $(0, \bar{x}, \bar{y})$, *respectively. That is, there exist neighbourhoods* $\mathcal{V}$ *of* $0$ *and* $\mathcal{U}$ *of* $(\bar{x}, \bar{y})$, *and a constant* $L \ge 0$ *such that*

$$\text{dist}_{K(0)}(x,y) \le L|v|, \text{ for all } (x,y) \in \mathcal{U} \cap K(v) \text{ and for all } v \in \mathcal{V},$$

*where* $K(v) := K_B(v), K_{KKT}(v), K_{FJ}(v), K_{cp}(v), K_f(v)$, *respectively.*

Here when we say $y$ is a KKT point, we mean that KKT condition holds at $y$. Note that PEB over $f$ is the so-called value function constraint qualification introduced by Henrion and Surowiec [23] for BLPPs with convex lower level problems.

DEFINITION 4.2 (**Uniform weak sharp minimum**). *We say that the uniform weak sharp minimum (UWSM) over B, KKT, FJ, cp, feasible (f)-condition is satisfied at a feasible solution* $(\bar{x}, \bar{y})$ *of* (BLPP) *if there exists a neighbourhood* $\mathcal{U}$ *of* $(\bar{x}, \bar{y})$, *and a constant* $L \ge 0$ *such that*

$$\text{dist}_{S(x)}(y) \le L(f(x,y) - V(x)), \text{ for all } (x,y) \in \mathcal{U} \cap \Sigma,$$

*where* $\Sigma = \Sigma_B, \Sigma_{KKT}, \Sigma_{FJ}, \Sigma_{cp}, \Sigma_f$, *respectively.*

Note that UWSM over $f$ reduces to the UWSM introduced by Ye and Zhu in [53]. By definition, it is easy to see that the PEB condition is weaker than the UWSM. Furthermore, as pointed out by Henrion and Surowiec in [23, Example 3.9], the PEB condition may be strictly weaker than the UWSM.

We now prove that under PEB condition, the corresponding combined program is partially calm at $(\bar{x}, \bar{y})$ as stated in the following theorem. For the case $\Sigma = \Sigma_f$, the result was proved in [23, Proposition 3.11].

THEOREM 4.3. *Let* $(\bar{x}, \bar{y})$ *be a local optimal solution of* (BLPP). *If PEB over B, KKT, FJ, cp, f-condition is satisfied at* $(\bar{x}, \bar{y})$, *then the corresponding combined program is partially calm at* $(\bar{x}, \bar{y})$, *respectively. That is, there is a constant* $\mu \ge 0$ *such that* $(\bar{x}, \bar{y})$ *is a local optimal solution of the partially penalized problem*

$$\min_{x,y} F(x,y) + \mu\big(f(x,y) - V(x)\big)$$
$$\text{s.t. } (x,y) \in \Sigma, \; G(x,y) \le 0,$$

*where* $\Sigma = \Sigma_B, \Sigma_{KKT}, \Sigma_{FJ}, \Sigma_{cp}, \Sigma_f$, *respectively.*

*Proof.* Since $(\bar{x}, \bar{y})$ is a local solution of (BLPP) and $F$ is continuously differentiable, without loss of generality, we assume that $F$ is Lipschitz of rank $L_f > 0$ on $S := \Omega \cap \mathcal{U}$, where $\Omega := \{(x,y) : G(x,y) \le 0\}$ and $\mathcal{U}$ is the (bounded) neighbourhood of $(\bar{x}, \bar{y})$ defined in the PEB condition, and $F(x,y)$ attains a minimum over $C := M \cap \Omega \cap \mathcal{U}$ at $(\bar{x}, \bar{y})$. By virtue of the exact penalty principle of Clarke ([8, Proposition 2.4.3]), $F(x,y) + L_f \text{dist}_C(x,y)$ attains a minimum over $S$. But for all

$(x, y) \in \Omega \cap \mathcal{U}$, $\text{dist}_C(x, y) = \text{dist}_{K(0)}(x, y)$. Hence we have

$$
\begin{aligned}
F(\bar{x}, \bar{y}) &\leq F(x, y) + L_f \text{dist}_C(x, y) & \forall (x, y) \in \Omega \cap \mathcal{U} \\
&= F(x, y) + L_f \text{dist}_{K(0)}(x, y) & \forall (x, y) \in \Omega \cap \mathcal{U} \\
&\leq F(x, y) + L_f L |v| & \forall (x, y) \in \Omega \cap \mathcal{U} \cap K(v) \text{ and } \forall v \in \mathcal{V} \\
&= F(x, y) + L_f L(f(x, y) - V(x)) & \forall (x, y) \in \Omega \cap \Sigma \cap \mathcal{U},
\end{aligned}
$$

where the last equality follows from the fact that $f(x, y) - V(x) \geq 0$ for $y \in Y(x)$. The desired result follows from taking $\mu := L_f L$. □

By (4.3), it is easy to see that

PEB over $f \implies$ PEB over g.c. condition $\implies$ PEB over $FJ \implies$ PEB over $B$.

If MSCQ holds, by virtue of (4.3) again, we have

PEB over $KKT \iff$ PEB over $B$.

Hence, for the partial calmness condition, the PEB over $B$ is the weakest sufficient condition while PEB over $f$ is the strongest sufficient condition.

Note that problem $(\text{CP}_{FJ})$ is not a practical problem to solve. However we can show the equivalence between the partial calmness conditions for (CPFJ) and $(\text{CP}_{FJ})$, and use it to study the generic property of the partial calmness condition for (CPFJ).

THEOREM 4.4 (**Equivalence of partial calmness for** (CPFJ) **and** $(\text{CP}_{FJ})$). *Let* $\bar{y} \in S(\bar{x})$. *Then problem* $(\text{CP}_{FJ})$ *is partially calm at* $(\bar{x}, \bar{y})$ *if and only if problem* (CPFJ) *is partially calm at* $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ *for each FJ-multiplier* $(\bar{u}_0, \bar{u}) \in M_{\text{FJ}}(\bar{x}, \bar{y})$.

*Proof.* Suppose that $(\text{CP}_{FJ})$ is partially calm at $(\bar{x}, \bar{y})$. Then there exist a calmness modulus $\bar{\mu} \geq 0$ and a neighborhood $U(\bar{x}, \bar{y})$ of $(\bar{x}, \bar{y})$ such that

$$(4.4) \qquad F(\bar{x}, \bar{y}) \leq F(x, y) + \bar{\mu}\big(f(x, y) - V(x)\big), \ \forall (x, y) \in \Sigma_{FJ} \cap U(\bar{x}, \bar{y}).$$

Let $(x', y', u_0', u')$ be a feasible solution for problem $(\text{CPFJ}_\mu)$ with $(x', y')$ lying in the neighborhood $U(\bar{x}, \bar{y})$. Then $(x', y') \in \Sigma_{FJ} \cap U(\bar{x}, \bar{y})$. Hence (4.4) implies that problem (CPFJ) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ for any FJ-multiplier $(\bar{u}_0, \bar{u}) \in M_{\text{FJ}}(\bar{x}, \bar{y})$ with the same calmness modulus $\mu = \bar{\mu}$.

Conversely, suppose that problem (CPFJ) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ for each FJ-multiplier $(\bar{u}_0, \bar{u}) \in M_{\text{FJ}}(\bar{x}, \bar{y})$. We first show that there exists $\bar{\mu} \geq 0$ such that (CPFJ) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ for all FJ-multipliers $(\bar{u}_0, \bar{u}) \in M_{\text{FJ}}(\bar{x}, \bar{y})$ with the same $\mu = \bar{\mu}$. To the contrary, suppose that there exists $(\bar{u}_0^k, \bar{u}^k) \in M_{\text{FJ}}(\bar{x}, \bar{y})$ such that problem (CPFJ) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0^k, \bar{u}^k)$ and as $k \to \infty$,

$$\mu^k := \inf\left\{\mu \geq 0 : \text{problem (CPFJ) is partially calm at } (\bar{x}, \bar{y}, \bar{u}_0^k, \bar{u}^k) \text{ with } \mu\right\} \to \infty$$

Since $M_{\text{FJ}}(\bar{x}, \bar{y})$ is compact, up to a subsequence, the sequence $(\bar{u}_0^k, \bar{u}^k)$ has a limit $(\bar{u}_0^*, \bar{u}^*)$ in $M_{\text{FJ}}(\bar{x}, \bar{y})$. By assumption, (CPFJ) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0^*, \bar{u}^*)$. Hence, there is $\mu^* \geq 0$ and a neighborhood $U(\bar{x}, \bar{y}, \bar{u}_0^*, \bar{u}^*)$ of $(\bar{x}, \bar{y}, \bar{u}_0^*, \bar{u}^*)$ such that

$$(4.5) \ F(\bar{x}, \bar{y}) \leq F(x, y) + \mu^*\big(f(x, y) - V(x)\big), \ \forall (x, y, u_0, u) \in \mathcal{F}_{\text{CPFJ}} \cap U(\bar{x}, \bar{y}, \bar{u}_0^*, \bar{u}^*),$$

where $\mathcal{F}_{\text{CPFJ}}$ is the feasible region of $(\text{CPFJ}_\mu)$. Since $(\bar{u}_0^k, \bar{u}^k) \to (\bar{u}_0^*, \bar{u}^*)$ in $M_{\text{FJ}}(\bar{x}, \bar{y})$, for sufficiently large $k$, we can take a neighborhood $U(\bar{x}, \bar{y}, \bar{u}_0^k, \bar{u}^k)$ of $(\bar{x}, \bar{y}, \bar{u}_0^k, \bar{u}^k)$

such that $U(\bar{x}, \bar{y}, \bar{u}_0^k, \bar{u}^k) \subseteq U(\bar{x}, \bar{y}, \bar{u}_0^*, \bar{u}^*)$. Hence (4.5) holds with $U(\bar{x}, \bar{y}, \bar{u}_0^*, \bar{u}^*)$ replaced with $U(\bar{x}, \bar{y}, \bar{u}_0^k, \bar{u}^k)$. This implies that problem (CPFJ) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0^k, \bar{u}^k)$ with $\mu = \mu^*$ for all $k$ sufficiently large. Thus $\mu^k \leq \mu^*$ for sufficiently large $k$, a contradiction with $\mu^k \to \infty$ as $k \to \infty$.

Now we show that $(\text{CP}_{FJ})$ is partially calm at $(\bar{x}, \bar{y})$ with $\mu = \bar{\mu}$. Suppose not, then there exists $(x^k, y^k) \in \Sigma_{FJ}$ satisfying $(x^k, y^k) \to (\bar{x}, \bar{y})$ and

$$(4.6) \qquad F(\bar{x}, \bar{y}) > F(x^k, y^k) + \bar{\mu}\big(f(x^k, y^k) - V(x^k)\big), \quad \forall k.$$

Take $(u_0^k, u^k) \in M_{\text{FJ}}(x^k, y^k)$. Since $M_{\text{FJ}}(x^k, y^k) \subseteq \big\{(u_0, u) : \|(u_0, u)\|_1 = 1\big\}$, the sequence $(u_0^k, u^k) \to (\bar{u}_0, \bar{u})$, up to a subsequence. It follows that $(\bar{u}_0, \bar{u}) \in M_{\text{FJ}}(\bar{x}, \bar{y})$ due to the continuous differentiability of $f(x, y)$ and $g(x, y)$. Hence, $(x^k, y^k, u_0^k, u^k) \to (\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ and (4.6) holds. Therefore, $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ is not a local minimum of $(\text{CPFJ}_\mu)$ with $\mu = \bar{\mu}$, a contradiction. □

The results in Theorem 4.4 can be improved if an extra assumption is imposed as below.

COROLLARY 4.5. *Let $\bar{y} \in S(\bar{x})$. Assume that the constant rank constraint qualification (CRCQ) for the lower level problem holds at $(\bar{x}, \bar{y})$, i.e., there exists an open neighborhood $U(\bar{x}, \bar{y})$ of $(\bar{x}, \bar{y})$ such that for each index set $I \subseteq J_0(\bar{x}, \bar{y})$ the family of gradients $\{\nabla_y g_j(x, y) : j \in I\}$ has the same rank on $U(\bar{x}, \bar{y})$. Then the combined program $(\text{CP}_{FJ})$ is partially calm at $(\bar{x}, \bar{y})$ if problem (CPFJ) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ for all $(\bar{u}_0, \bar{u})$ which are vertices of $M_{\text{FJ}}(\bar{x}, \bar{y})$.*

*Proof.* The proof is similar to the argument of the proof of [11, Corollary 3.3]. Indeed, FJ-multipliers $(u_0^k, u^k)$ in the proof of Theorem 4.4 can without assumptions be taken as vertices of $M_{\text{FJ}}(x^k, y^k)$. Then CRCQ implies that the sequence $(u_0^k, u^k)$ converges to a vertex of $M_{\text{FJ}}(\bar{x}, \bar{y})$. Hence, the desired result follows. □

Through the equivalence between the partial calmness for (CPFJ) and $(\text{CP}_{FJ})$, we can study the generic property of the partial calmness condition for (CPFJ).

We can now formulate our main results.

THEOREM 4.6 (**Generic property**). *Under Problem Setting, we have*
(1) *Let $\mathcal{P}$ denote the set of functions $(f, g) \in [C^3(\mathbb{R}^{n+m})]^{1+p} \cap \mathcal{O}$ such that the partial error bound over FJ-condition defined in Definition 4.1 holds at all points $(\bar{x}, \bar{y}) \in M$. Then $\mathcal{P}$ is $C_s^3$-open and $C_s^3$-dense in $[C^3(\mathbb{R}^{n+m})]^{1+p} \cap \mathcal{O}$. Consequently, the partial error bound over FJ-condition is generic.*
(2) *Let $\mathcal{Q}$ denote the set of functions $(f, g) \in [C^3(\mathbb{R}^{n+m})]^{1+p} \cap \mathcal{O}$ such that the partial calmness for $(\text{CP}_{FJ})$ holds at all local optimal solutions of (BLPP). Then $\mathcal{Q}$ is $C_s^3$-open and $C_s^3$-dense in $[C^3(\mathbb{R}^{n+m})]^{1+p} \cap \mathcal{O}$. Consequently, the partial calmness condition for (CPB) and $(\text{CP}_{FJ})$ are generic.*
(3) *Let $\mathcal{R}$ denote the set of functions $(f, g) \in [C^3(\mathbb{R}^{n+m})]^{1+p} \cap \mathcal{O}$ such that the partial calmness for (CPFJ) holds at all local optimal solutions of (CPFJ). Then $\mathcal{R}$ is $C_s^3$-open and $C_s^3$-dense in $[C^3(\mathbb{R}^{n+m})]^{1+p} \cap \mathcal{O}$. Consequently, the partial calmness condition for (CPFJ) is generic.*
*Here $\mathcal{O}$ stands for the set of $(f, g)$ such that*

$$(4.7) \qquad\qquad B_{f,g}(\bar{x}, c) \text{ is compact for all } (\bar{x}, c) \in \mathbb{R}^n \times \mathbb{R},$$

*where $B_{f,g}(\bar{x}, c) := \big\{(x, y) : \|x - \bar{x}\| \leq 1, f(x, y) \leq c, g(x, y) \leq 0\big\}$.*

The proof for part (1) will be given in Section 5. Assuming Part (1) is true, part (2) follows from the fact that the partial calmness condition for (CPB) is weaker

than the one for $(\mathrm{CP}_{FJ})$. Part (3) follows from the equivalence between the partial calmness for (CPFJ) and $(\mathrm{CP}_{FJ})$ in Theorem 4.4.

*Remark* 4.7. It is worth pointing out that the generic analysis can not be done with the set $\Sigma_f$. Hence the value function CQ of Henrion and Surowiec [23] is not generic and neither is the partial calmness for (VP) (i.e., the partial calmness over $f$).

The condition (4.7) (also in [28]) implies the inf-compactness condition: for all $\bar{x} \in \mathbb{R}^n$ there exist $\alpha > 0$, $\delta > 0$ and a bounded set $C$ such that $\alpha > V(\bar{x})$ and

$$\left\{y : \|x - \bar{x}\| \leq \delta,\, f(x,y) \leq \alpha,\, g(x,y) \leq 0\right\} \subseteq C.$$

Since $\bar{y} \in S(\bar{x})$ is a global minimizer for $L(\bar{x})$, the inf-compactness condition is sufficient to avoid asymptotical effects for the feasible set of BLPPs. Note also that $[C^3(\mathbb{R}^{n+m})]^{1+p} \cap \mathcal{O}$ is $C_s^3$-open.

**4.2. Techniques and Roadmap of Analysis.** It is clear that the main goal of Theorem 4.6 is to verify that the partial error bound over FJ-condition defined in Definition 4.1 holds generically. The main difficulty is to deal with the global constraint $y \in S(x)$ or the value function constraint $f(x,y) - V(x) \leq 0$. Hence the generic structure of $M = \{(x,y) : y \in S(x)\}$ is important.

To describe the generic structure of $M$, it is natural to adopt a parametric $(\mathrm{CP}_{FJ})$ming point of view, using the upper level variable as a parameter. This issue has been studied by Jongen and Shikhman [28] when the upper level variable is one dimensional. The main result in [28] states that–generically–the set $M$ is the union of $C^2$ curves with boundary points and kinks which can be parametrized by means of the upper variable. The appearance of the boundary points and kinks is due to certain degeneracies of optimal solutions for the lower level problem (e.g., the failure of strict complementarity (SC) or LICQ or the second order condition (SOC), as well as the change from local to global solutions).

The main idea in [28] is to exploit a generic classification of g.c. points for one dimensional parametric optimization problems in [26]. When $n = 1$, it was shown in [26] that for an open and dense subset of problem data (i.e., generically) the set $\Sigma_{gc}$ is pieced together from one-dimensional manifolds. Moreover, for those problem data, the g.c. points for $L(x)$ can be divided into five (characteristic) types. A point $(\bar{x}, \bar{y}) \in \Sigma_{gc}$ is of type 1 if it is a nondegenerate critical point for $L(\bar{x})$ which means that the following conditions hold:

ND1: LICQ is satisfied at $(\bar{x}, \bar{y})$ and hence there exists a unique multiplier $\bar{u}$.

ND2: SC holds at $(\bar{x}, \bar{y})$, i.e., $\bar{u}_j \neq 0$, $j \in J_0(\bar{x}, \bar{y})$.

ND3: SOC is satisfied, i.e., the matrix

$$(4.8) \qquad (\nabla_{yy}^2(f + g^T\bar{u})(\bar{x}, \bar{y}))|_{T_{\bar{y}}Y(\bar{x})} := V^T\nabla_{yy}^2(f + g^T\bar{u})(\bar{x}, \bar{y})V$$

is nonsingular where $V$ is some matrix whose columns form a basis for the tangent space

$$T_{\bar{y}}Y(\bar{x}) := \{\xi \in \mathbb{R}^m : \nabla_y g_j(\bar{x}, \bar{y})^T\xi = 0,\, j \in J_0(\bar{x}, \bar{y})\}.$$

Other types originate from degeneracies of the generalized critical points. A degeneracy describes the failure of at least one of the nondegeneracy properties ND1-ND3. For cleanness of the paper, we do not repeat the precise characterization of other types here and introduce them in Table 4.1 for brevity. For more details we refer the reader to [26, 28].

|  | Characteristic |
|---|---|
| **Type 1** | LICQ, SC, and SOC are satisfied |
| **Type 2** | LICQ, SOC are satisfied, SC is violated and exactly one of $\bar{u}_j, j \in J_0$ vanishes |
| **Type 3** | LICQ, SC are satisfied, and exactly one eigenvalue of the matrix (4.8) vanishes |
| **Type 4** | LICQ violated, the rank of $\{\nabla_y g_j(\bar{x}, \bar{y}), j \in J_0\} = |J_0| - 1 < m$, and KKT condition fails |
| **Type 5-1** | LICQ violated, the rank of $\{\nabla g_j(\bar{x}, \bar{y}), j \in J_0\} = |J_0| = m + 1$, MFCQ fails but KKT condition holds |
| **Type 5-2** | LICQ violated, the rank of $\{\nabla g_j(\bar{x}, \bar{y}), j \in J_0\} = |J_0| = m + 1$, and MFCQ holds |

TABLE 4.1

*The Five Types in Parametric Optimization ($n = 1$). Here $J_0 := J_0(\bar{x}, \bar{y})$.*

We next briefly review the main result of Jongen and Shikhman [28] for BLPPs. To this end, we first recall the definition of simplicity of bilevel problems in [28].

DEFINITION 4.8 (**[28, Definition 4.1]**).  *A bilevel programming problem (with $n = 1$) is called simple at $(\bar{x}, \bar{y}) \in M$ if one of the following cases occurs:*
*Case I: $S(\bar{x}) = \{\bar{y}\}$ and $(\bar{x}, \bar{y})$ is of Type 1, 2, 4, 5-1 or 5-2.*
*Case II: $S(\bar{x}) = \{\bar{y}_1, \bar{y}_2\}$ and $(\bar{x}, \bar{y}_1)$, $(\bar{x}, \bar{y}_2)$ are both of Type 1. In addition, it holds that*

$$(4.9) \qquad \alpha := \frac{d}{dx}[f(x, y_2(x)) - f(x, y_1(x))]\Big|_{x=\bar{x}} \neq 0,$$

*where $y_1(x)$, $y_2(x)$ are unique local minimizers for $L(x)$ in a neighborhood of $\bar{x}$ with $y_1(\bar{x}) = \bar{y}_1$, $y_2(\bar{x}) = \bar{y}_2$ according to Type 1.*

THEOREM 4.9 (**[28, Theorem 4.1]**).
*Let $\mathcal{F}$ be the set of defining functions $(F, f, g) \in C^3(\mathbb{R}^{1+m}) \times \left([C^3(\mathbb{R}^{1+m})]^{1+p} \cap \mathcal{O}\right)$ such that the corresponding BLPP is simple at all its feasible points $(\bar{x}, \bar{y}) \in M$. Then $\mathcal{F}$ is $C_s^3$-open and $C_s^3$-dense in $C^3(\mathbb{R}^{1+m}) \times \left([C^3(\mathbb{R}^{1+m})]^{1+p} \cap \mathcal{O}\right)$.*

*Remark* 4.10. First, since the main goal of Jongen and Shikhman [28] is to describe the generic structure of the bilevel feasible set $M$, they focus on such parts of $\Sigma_{gc}$ which correspond to (local) minimizers, i.e.,

$$\Sigma_{min} := \left\{(x, y) \in \mathbb{R}^{n+m} : y \text{ is a local minimizer for } L(x)\right\}$$

in a neighborhood of $(\bar{x}, \bar{y}) \in M$. Hence, the cases in [28, Definition 4.1] are related to $\Sigma_{min}$; cf. [28, Section 3].

Our purpose is to study the partial calmness conditions for the combined programs (CPB) and (CP$_{FJ}$) which involve B- and FJ conditions. Note that $\Sigma_{min} \subseteq \Sigma_B \subseteq \Sigma_{FJ}$. Hence it is crucial whether [28, Theorem 4.1] is still true when the set $\Sigma_{min}$ in Definition 4.8 is replaced by $\Sigma_{FJ}$. The answer is affirmative. Actually, a stronger result is in fact true. The definition of simplicity, those analysis in [28, Section 3] and then [28, Theorem 4.1] can be generalized to a so-called 'Kuhn-Tucker subset' $\Sigma_{KT}$ of $\Sigma_{gc}$. This notation is studied in more details, in particular in connection with the (failure of) MFCQ, cf. [26, Section 4]. More precisely, the subset $\Sigma_{KT}$ of $\Sigma_{gc}$ is defined to be the closure of nondegenerate critical points (i.e., points of Type 1) with nonnegative Lagrange multipliers.

Now we show that $(x, y) \in \Sigma_{FJ}$ and $(x, y)$ is of Types 1-5 imply that $(x, y) \in \Sigma_{KT}$. Indeed, point $(x, y) \in \Sigma_{FJ}$ implies that the multipliers are nonnegative. Thus, if

further $(x, y)$ is of Type 1, then it is clear that $(x, y) \in \Sigma_{KT}$. If $(x, y)$ is of Type 2, then $y$ is a nondegenerate critical point for the problem $\widetilde{L}(\bar{x})$, which differs from $L(\bar{x})$ by omitting the inequality constraint with vanishing Lagrange multiplier. Hence, we get a unique $C^2$-map $\widetilde{y}(x)$. But only one of its branches from $\bar{x}$ satisfies the omitting inequality constraint with strictly sign, and then belongs to the set of Type 1 points with positive Lagrange multipliers. Hence $(\bar{x}, \bar{y})$ belongs to $\Sigma_{KT}$. Similar arguments apply to the left cases. Since $\Sigma_{min} \subseteq \Sigma_{FJ}$, Theorem 4.9 is still true when the set $\Sigma_{min}$ in the definition 4.8 is replaced by $\Sigma_{FJ}$.

It is worth pointing out that the proof of [28, Theorem 4.1] does not use any property of the upper level objective function. In other words, the generic structures of $\Sigma_B, \Sigma_{KKT}, \Sigma_{FJ}, \Sigma_{KT}, \Sigma_{gc}, \Sigma_{min}$ are properties of the lower level $(\mathrm{CP}_{FJ})$ only, independent of the upper level objective function. Hence, [28, Theorem 4.1] can be stated as follows.

THEOREM 4.11 (**Modified Theorem 4.1 in [28]**). *Let $\mathcal{F}$ denote the set of defining functions $(f, g) \in \left([C^3(\mathbb{R}^{1+m})]^{1+p} \cap \mathcal{O}\right)$ such that the corresponding bilevel programming problem is simple at all its feasible points $(\bar{x}, \bar{y}) \in M$. Then $\mathcal{F}$ is $C_s^3$-open and $C_s^3$-dense in $\left([C^3(\mathbb{R}^{1+m})]^{1+p} \cap \mathcal{O}\right)$.*

For the convenience of the reader, in each case one of the possibilities in Definition 4.8 is depicted in Figure 4.1, which can be seen as an extension of Fig. 1 in [28].
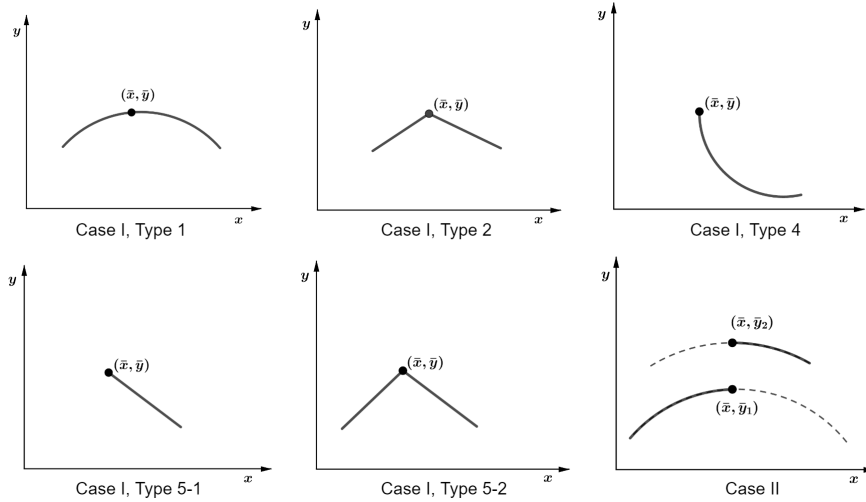


FIG. 4.1. *Generic structure of FJ points locally around global minimizers.*

To present our analysis more clearly, let us show the roadmap in Figure 4.2.

**5. Proof of the main result.** The main aim of this section is to prove Theorem 4.6(1). By Theorems 4.3 and 4.11, and the relationship between the partial error bounds over FJ and B conditions, it suffices to verify partial error bound over FJ-condition holds in Case I and Case II.

To verify partial error bound over FJ-condition, the following lemma is useful.

LEMMA 5.1. *Assume that both $f(x, y)$ and $g(x, y)$ are continuous. Let $\bar{y} \in S(\bar{x})$ and $x_i \to \bar{x}$ as $i \to \infty$. Suppose that there exist $\bar{y}_i \in Y(x_i)$ such that $\bar{y}_i \to \bar{y}$. Then $y_0 \in S(\bar{x})$ if $y_i \in S(x_i)$ and $y_i \to y_0$.*

> Generic classification of generalized critical point of
> one-dimensional parametric optimization problems
> ([26, Theorem 2.1])

> Generic structure of FJ points locally around
> global minimizers for the lower level problems
> (Generalization of [28, Theorem 4.1], cf. Definition 4.8 and Remark 4.10)

> Partial error bound over FJ-condition for BLPPs
> (Section 5: Lemmas 5.3 and 5.4)

> Partial error bound over B-condition for BLPPs
> (Definition 4.1and Theorem 4.6(1))

> Partial calmness for CPB and (CPFJ)
> (Definitions 1.3-1.2 and Theorem 4.6(2) and (3))

partially penalized problem CPFJ$_\mu$ plus constraint qualifications
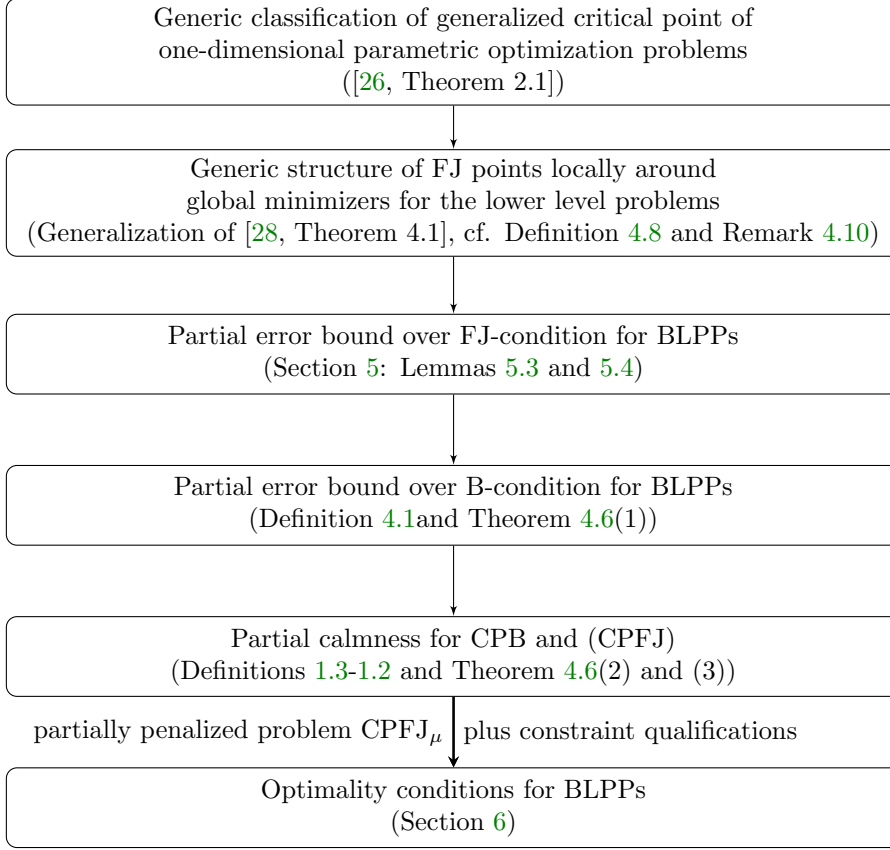
> Optimality conditions for BLPPs
> (Section 6)

FIG. 4.2. *Roadmap to study the genericity result.*

*Proof.* Let $y_i \in S(x_i)$ and $\lim_{i \to \infty} y_i = y_0$. Since $y_i \in S(x_i) \subseteq Y(x_i)$, we have $g(x_i, y_i) \leq 0$. Thus, by the continuity of $g(x, y)$,

$$g(\bar{x}, y_0) = \lim_{i \to \infty} g(x_i, y_i) \leq 0.$$

Hence $y_0 \in Y(\bar{x})$. Next we show that $f(\bar{x}, y_0) \leq f(\bar{x}, \bar{y})$. Indeed, since $y_i \in S(x_i)$ and $\bar{y}_i \in Y(x_i)$, we have $f(x_i, y_i) \leq f(x_i, \bar{y}_i)$. Hence, by the continuity of $f(x, y)$,

$$f(\bar{x}, y_0) = \lim_{i \to \infty} f(x_i, y_i) \leq \lim_{i \to \infty} f(x_i, \bar{y}_i) = f(\bar{x}, \bar{y}).$$

Therefore, it is clear that $y_0 \in S(\bar{x})$ since $\bar{y} \in S(\bar{x})$. □

*Remark* 5.2. The above lemma is important enough to be stated separately since we will apply it to verify UWSM over FJ-condition in Case I and PEB over FJ-condition in Case II. Its principal significance is that it allows us to show that some branch of $\Sigma_{FJ}$ become global minimizers under some condition (e.g., $|S(\bar{x})| \leq 2$).

**5.1. Verification of UWSM over FJ-condition in Case I.** In fact in Case I, we can show the stronger condition UWSM over FJ-condition holds and the corresponding constant $L = 0$.

LEMMA 5.3. *Let* (BLPP) *be simple at* $(\bar{x}, \bar{y}) \in M$ *in Case I. Then there exists a neighbourhood* $\mathcal{U}$ *of* $(\bar{x}, \bar{y})$ *such that* $(x, y) \in \mathcal{U}$ *and* $y$ *is a FJ-point for* $L(x)$ *imply that* $y$ *minimizes* $L(x)$, *i.e.*,

$$(5.1) \qquad \Sigma_{FJ} \cap \mathcal{U} = M \cap \mathcal{U}.$$

*Hence for all* $(x, y) \in \Sigma_{FJ} \cap \mathcal{U}$,

$$(5.2) \qquad \operatorname{dist}_{S(x)}(y) = 0.$$

*This implies that the UWSM over FJ condition holds at* $(\bar{x}, \bar{y})$ *and the corresponding constant* $L = 0$.

*Proof.* The proof falls naturally into four parts.

**Part 1: Case I, Type 1.** In this case, a stronger result is in fact true. We claim that there exists a neighbourhood $\mathcal{U}$ of $(\bar{x}, \bar{y})$ such that

$$(5.3) \qquad \Sigma_{gc} \cap \mathcal{U} = M \cap \mathcal{U}.$$

The proof will be divided into two steps.

**Step 1**: There exists a neighbourhood $\mathcal{U} := \mathcal{U}(\bar{x}) \times \mathcal{U}(\bar{y})$ of $(\bar{x}, \bar{y})$ such that for some continuous function $y(x)$,

$$\Sigma_{gc} \cap \mathcal{U} = \big\{ (x, y(x)) : x \in \mathcal{U}(\bar{x}) \big\}.$$

This result is standard, cf. [26, Section 3.1]. Indeed, it is well-known that conditions ND1-ND3 allow us to apply the implicit function theorem and obtain a unique $C^2$-mapping $y(x)$ in an open neighborhood of $\bar{x}$ such that we can parametrize the set $\Sigma_{gc}$ by means of a unique $C^2$-mapping $x \mapsto (x, y(x))$ in an open neighbourhood of $(\bar{x}, \bar{y})$. As a consequence, the set $Y(x) \neq \emptyset$ for all $x \in \mathcal{U}(\bar{x})$. It then follows from the definition of $\mathcal{O}$, the solution set $S(x) \neq \emptyset$ for all $x \in \mathcal{U}(\bar{x})$.

**Step 2**: There exists a small neighbourhood $\mathcal{U}(\bar{x})$ of $\bar{x}$ such that $y(x)$ minimizes $L(x)$ for all $x \in \mathcal{U}(\bar{x})$.

Suppose not, then there exists $x_i \to \bar{x}$ such that $y(x_i) \notin S(x_i)$. Taking $\bar{y}_i = y(x_i)$, by the continuity of $y(x)$, we have $\bar{y}_i \to \bar{y}$. Taking $y_i \in S(x_i)$ and $y_i \to y_0$, by Lemma 5.1, we have $y_0 \in S(\bar{x})$. Recall that $S(\bar{x}) = \{\bar{y}\}$ in Case I. Hence $y_0 = \bar{y}$ and then $y_i \to \bar{y}$. By the fact that optimal solution $y_i$ must be a g.c. point for $L(x_i)$. Hence, when $i$ is sufficient large, by Step 1, it is immediate that $y_i = y(x_i)$ and then $y(x_i) \in S(x_i)$ since $y_i \in S(x_i)$. This contradicts our assumption that $y(x_i) \notin S(x_i)$.

**Part 2: Case I, Type 2.** In this case, LICQ is satisfied and the only degeneracy is the vanishing of exactly one Lagrange multiplier corresponding to an active inequality constraint, indexed by $j_0$. Then $\bar{y}$ is a nondegenerate critical point for the problems $\widetilde{L}(\bar{x})$ and $\widehat{L}(\bar{x})$, where $\widetilde{L}(\bar{x})$ differs from $L(\bar{x})$ by omitting the inequality constraint $g_{j_0}$ and whereas in $\widehat{L}(\bar{x})$ the constraint $g_{j_0}$ is regarded as an equality constraint. Hence, the set $\Sigma_{gc}$ around $(\bar{x}, \bar{y})$ consists of the feasible part of two curves, cf. [26, Section 3.2].

Let the KT subset $\Sigma_{KT}$ of $\Sigma_{gc}$ be the closure of the set of Type 1 points with nonnegative Lagrange multipliers. Because $\bar{y} \in S(\bar{x})$ implies that $\bar{y}$ is a local minimizer for $L(\bar{x})$, similar to those analysis in [28, Section 3.2], by Morse relations, we can parameterize $\Sigma_{KT} \cap \mathcal{U}$ by the $x$-variable. Hence Step 1 in Part 1 is still true for some continuous function $y(x)$ when $\Sigma_{gc}$ is replaced by $\Sigma_{KT}$. Note that the continuity of

$y(x)$ is enough in the proof of Step 2 in Part 1. Therefore, by the same argument in Part 1, we can prove

$$(5.4) \qquad \Sigma_{KT} \cap \mathcal{U} = M \cap \mathcal{U}.$$

Now we show that there exists a neighbourhood $\mathcal{U}$ of $(\bar{x}, \bar{y})$ such that $(x, y) \in \mathcal{U}$ and $y$ is a FJ-point for $L(x)$ imply that $y$ minimizes $L(x)$. By the argument in the proof of Step 2 in Part 1 and [26, Theorem 2.1], since an optimal solutin must be a FJ point, it suffices to show that $(x, y) \in \Sigma_{FJ}$ and $(x, y)$ is of Types 1-5 imply that $(x, y) \in \Sigma_{KT}$. In fact, this result has been shown in Remark 4.10.

**Part 3: Case I, Type 4.** In this case, LICQ is violated and the number of active constraints is less than $m + 1$. Moreover, $(\bar{x}, \bar{y})$ is no longer a KKT-point. But in an open neighborhood of $(\bar{x}, \bar{y})$, apart from $(\bar{x}, \bar{y})$, the points of $\Sigma_{gc}$ are nondegenerate critical points. Let the parameter $x$ viewed as a function on $\Sigma_{gc}$, the set $\Sigma_{gc}$ can be locally approximated by means of a parabola, cf. [26, Section 3.4].

Because $\bar{y}$ is a local minimizer for $L(\bar{x})$ when $\bar{y} \in S(\bar{x})$, similar to those analysis in [28, Section 3.4], it can be shown that only one branch of $\Sigma_{gc}$ starting from $\bar{y}$ left in $\Sigma_{KT}$. Hence we can prove (5.4) by the same argument in Part 2 and then obtain the desired result.

**Part 4: Case I, Type 5-1 or 5-2.** In this case, LICQ is violated, but in contract to Type 4, the number of active constraints equals $m + 1$. Then, by the characteristic conditions in Type 5 (cf. [26, Section 3.5]), around $(\bar{x}, \bar{y})$, the set $\Sigma_{gc}$ consists of exactly $m + 1$ (half) curves, each of them either emanating from $(\bar{x}, \bar{y})$ or ending at $(\bar{x}, \bar{y})$. Moreover, there exists an open neighbourhood $\mathcal{U}$ of $(\bar{x}, \bar{y})$ such that either only one branch of $\Sigma_{KT} \cap \mathcal{U}$ left (referred to this case as Type 5-1) or one emanating from $(\bar{x}, \bar{y})$ with another one ending at $(\bar{x}, \bar{y})$ in $\Sigma_{KT}$ (referred to this case as Type 5-2), cf. [26, Section 3.5]. Furthermore, it is shown in [28, Section 3.2] that we can parameterize $\Sigma_{KT} \cap \mathcal{U}$ by some continuous function of the $x$-variable. Hence we can prove (5.4) by the argument in Part 1. Finally, by the same result in Part 2 and $M \subseteq \Sigma_{FJ}$, the desired result (5.1) follows. $\qquad \square$

**5.2. Verification of PEB over FJ-condition in Case II.**

LEMMA 5.4. *Let* (BLPP) *be simple at* $(\bar{x}, \bar{y}) \in M$ *in Case II. Then PEB over FJ-condition holds at* $(\bar{x}, \bar{y})$ *with* $L > 0$.

*Proof.* In Case II, by definition, $S(\bar{x}) = \{\bar{y}_1, \bar{y}_2\}$ and there exist $C^2$-mappings $y_1(x), y_2(x)$ being unique local minimizers for $L(x)$ in a neighborhood of $\bar{x}$ with $y_1(\bar{x}) = \bar{y}_1$, $y_2(\bar{x}) = \bar{y}_2$ according to Type 1.

Without loss of generality we assume that $\bar{y} = \bar{y}_1$. Now we verify PEB over FJ-condition at $(\bar{x}, \bar{y})$. By definition, for all $v$ in $\mathcal{V}$, a small neighbourhood of $0$,

$$K_{FJ}(v) \subseteq \Sigma_{FJ} \subseteq \Sigma_{gc}.$$

Since $(\bar{x}, \bar{y}) = (\bar{x}, \bar{y}_1)$ is of Type 1, by taking a small neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{y})$, points $(x, y) \in \mathcal{U} \cap K_{FJ}(v)$ implies that $y = y_1(x)$. Next, by (4.9), in a neighborhood of $\bar{x}$,

$$(5.5) \qquad \left| \frac{d}{dx}[f(x, y_2(x)) - f(x, y_1(x))] \right| \geq \frac{|\alpha|}{2} > 0.$$

Clearly, we can without loss of generality assume that the above property holds on $\mathcal{U} \cap K_{FJ}(v)$ and $v \in \mathcal{V}$. We claim that for some large constant $L > 0$,

$$(5.6) \qquad \operatorname{dist}_{K_{FJ}(0)}(x, y) \leq L|v|, \quad \forall (x, y) \in \mathcal{U} \cap K_{FJ}(v), \ v \in \mathcal{V},$$

which means that the PEB over FJ-condition holds at $(\bar{x}, \bar{y})$. Indeed, The proof will be divided into two cases.

**Case 1**: $f(x, y) - V(x) = 0$.

Clearly $f(x, y) - V(x) = 0$ implies that $(x, y) \in K_{FJ}(0)$. Thus $\text{dist}_{K_{FJ}(0)}(x, y) = 0$ and the inequality (5.6) holds with $L = 0$.

**Case 2**: $f(x, y) - V(x) \neq 0$.

On one hand, since $y = y_1(x)$ for $(x, y) \in \mathcal{U} \cap K_{FJ}(v), v \in \mathcal{V}$ and $(\bar{x}, y_1(\bar{x})) = (\bar{x}, \bar{y}_1) \in K_{FJ}(0)$, for some positive constant $C$, we obtain

$$\text{dist}_{K_{FJ}(0)}(x, y) = \text{dist}_{K_{FJ}(0)}(x, y_1(x)) \leq \|x - \bar{x}\| + \|y_1(x) - y_1(\bar{x})\|$$
$$(5.7) \qquad\qquad\qquad\qquad\qquad \leq C\|x - \bar{x}\|,$$

since $y_1(x)$ is a $C^2$ mapping. On the other hand, since $f(x, y_1(x)) - V(x) \neq 0$, by Lemma 5.1 and $S(\bar{x}) = \{\bar{y}_1, \bar{y}_2\}$, the argument in Part 1 of the proof of Lemma 5.3 implies that $V(x) = f(x, y_2(x))$. Thus, for all $(x, y) \in \mathcal{U} \cap K_{FJ}(v), v \in \mathcal{V}$,

$$f(x, y) - V(x) = f(x, y_1(x)) - f(x, y_2(x))$$
$$= [f(x, y_1(x)) - f(x, y_2(x))] - [f(\bar{x}, y_1(\bar{x})) - f(\bar{x}, y_2(\bar{x}))]$$
$$= \frac{d}{dx}[f(x, y_1(x)) - f(x, y_2(x))]\Big|_{x=\widetilde{x}}(x - \bar{x}),$$

where $\widetilde{x}$ is some point between $x$ and $\bar{x}$ obtained by the mean value theorem. Hence by (5.5), since $f(x, y) - V(x) \geq 0$, we have

$$(5.8) \qquad\qquad f(x, y) - V(x) \geq \frac{|\alpha|}{2}\|x - \bar{x}\|, \text{ if } f(x, y) - V(x) \neq 0.$$

Hence combining (5.7) and (5.8), for positive constant $L = \frac{2C}{|\alpha|} > 0$, we have

$$\text{dist}_{K_{FJ}(0)}(x, y) \leq L\left(f(x, y) - V(x)\right),$$

which proves (5.6) and then completes the proof. $\qquad\qquad\qquad\qquad\qquad$ □

**6. Application to optimality conditions.** When the lower level constraint $Y(x) = Y$ is independent of $x$, and the partial calmness constant $\mu = 0$ or $\mu > 0$ but the value function $V(x)$ is continuously differentiable at the point of interest, for problem (CPB$_\mu$), we can use the new constraint qualification given in [19, Theorem 5] and the new sharp necessary optimality condition given in [20, Theorem 6] to derive a necessary optimality condition. Similarly to the discussion in [20], this necessary optimality condition would be sharper than the usual M-stationary condition based on the value function [48, Theorem 4.1] for the combined program (CP), provided that LICQ holds for the lower level program at $\bar{y}$.

Now we derive optimality conditions for the general bilevel programming problem (BLPP) with no upper level constraints which are simple at $(\bar{x}, \bar{y}) \in M$. First we prove that MPCC-LICQ is satisfied for the partially penalized problem in all cases.

PROPOSITION 6.1. *Let* (BLPP) *be simple at* $(\bar{x}, \bar{y}) \in M$.
(1) *If* $S(\bar{x}) = \{\bar{y}\}$ *and KKT condition holds at* $\bar{y}$ *for* $L(\bar{x})$, *then for any KKT-multiplier* $\bar{u}$, *MPCC-LICQ is satisfied at* $(\bar{x}, \bar{y}, \bar{u})$ *for program* (CP$_\mu$) *with* $\mu = 0$.
(2) *If* $S(\bar{x}) = \{\bar{y}\}$ *and KKT condition does not hold at* $\bar{y}$ *for* $L(\bar{x})$, *then for any FJ-multiplier* $(\bar{u}_0, \bar{u}) \in \mathbb{R} \times \mathbb{R}^p$ *at* $\bar{y}$ *for* $L(\bar{x})$, *where* $\bar{u}_0 = 0$, *MPCC-LICQ is satisfied at* $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ *for program* (CPFJ$_\mu$) *with* $\mu = 0$.

(3) If $S(\bar{x}) = \{\bar{y}_1, \bar{y}_2\}$, where $\bar{y}_1 = \bar{y}$, then for the unique KKT-multiplier $\bar{u}$ of $\bar{y}$ for $L(\bar{x})$, MPCC-LICQ is satisfied at $(\bar{x}, \bar{y}, \bar{u})$ for program $(\mathrm{CP}_\mu)$.

*Proof.* Define $\mathcal{J} := \{j : g_j(\bar{x}, \bar{y}) = 0\}$, $\mathcal{K} := \{k : \bar{u}_k = 0\}$. Then by [44], MPCC-LICQ holds at $(\bar{x}, \bar{y}, \bar{u})$ for program $(\mathrm{CP}_\mu)$ with $\mu = 0$ if and only if the matrix

$$M(\mathcal{J}, \mathcal{K}, \bar{x}, \bar{y}, \bar{u}) := \begin{pmatrix} (\nabla_{yx}^2 f + \nabla_{yx}^2 g^T \bar{u})(\bar{x}, \bar{y}) & \nabla_x g_{\mathcal{J}}(\bar{x}, \bar{y}) \\ (\nabla_{yy}^2 f + \nabla_{yy}^2 g^T \bar{u})(\bar{x}, \bar{y}) & \nabla_y g_{\mathcal{J}}(\bar{x}, \bar{y}) \\ \nabla_y g_{\mathcal{K}^c}(\bar{x}, \bar{y})^T & 0 \end{pmatrix}$$

has full column rank. Here $g_{\mathcal{J}}$ denotes the functions with components $g_j$, $j \in \mathcal{J}$ and $\mathcal{K}^c$ denotes the complement of $\mathcal{K}$. For simplicity we omit the dependence on $(\bar{x}, \bar{y})$ in the proof if it is clear from the context. The proof falls naturally into four parts.

**Part 1: Case I, Type 1.**  In this case, by SC, we have $\mathcal{J} = \mathcal{K}^c$. Moreover, by LICQ, $\nabla_y g_{\mathcal{J}}(\bar{x}, \bar{y})$ has full column rank, and $(\nabla_{yy}^2 f + \nabla_{yy}^2 g^T u)(\bar{x}, \bar{y})|_{T_{\bar{y}} Y(\bar{x})}$ is nonsingular by using SOC and $T_{\bar{y}} Y(\bar{x}) = \ker([\nabla_y g_{\mathcal{J}}]^T)$. Hence

$$\begin{pmatrix} \nabla_{yy}^2 f + \nabla_{yy}^2 g^T \bar{u} & \nabla_y g_{\mathcal{J}} \\ [\nabla_y g_{\mathcal{K}^c}]^T & 0 \end{pmatrix} = \begin{pmatrix} \nabla_{yy}^2 f + \nabla_{yy}^2 g^T \bar{u} & \nabla_y g_{\mathcal{J}} \\ [\nabla_y g_{\mathcal{J}}]^T & 0 \end{pmatrix} \text{ is nonsingular,}$$

which implies $M(\mathcal{J}, \mathcal{K}, \bar{x}, \bar{y}, \bar{u})$ has full column rank. Since $\mathcal{J} = \mathcal{K}^c$, MPCC-LICQ reduces to LICQ in this case.

**Part 2: Case I, Type 2.**  In this case, by A2 and A3 in [28], $\mathcal{J} = \{1, \dots, p\}$, $p \geq 1$ and $|\mathcal{K}^c| = p - 1$. As in [28], after renumbering we assume $\mathcal{K}^c = \{1, \dots, p-1\}$. As in Part I, by A1, A5 in [28], it is easy to obtain that the matrix

$$C := \begin{pmatrix} \nabla_{yy}^2 f + \nabla_{yy}^2 g^T \bar{u} & \nabla_y g_{\mathcal{K}^c} \\ [\nabla_y g_{\mathcal{K}^c}]^T & 0 \end{pmatrix} \text{ is nonsingular.}$$

Note that

$$\begin{pmatrix} \nabla_{yx}^2 f + \nabla_{yx}^2 g^T \bar{u} & \nabla_x g_{\mathcal{J}} \\ \nabla_{yy}^2 f + \nabla_{yy}^2 g^T \bar{u} & \nabla_y g_{\mathcal{J}} \\ [\nabla_y g_{\mathcal{K}^c}]^T & 0 \end{pmatrix} = \begin{pmatrix} \nabla_{yx}^2 f + \nabla_{yx}^2 g^T \bar{u} & \nabla_x g_{\mathcal{K}^c} & \nabla_x g_p \\ \nabla_{yy}^2 f + \nabla_{yy}^2 g^T \bar{u} & \nabla_y g_{\mathcal{K}^c} & \nabla_y g_p \\ [\nabla_y g_{\mathcal{K}^c}]^T & 0 & 0 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where the matrix $C$ is nonsingular. To show that $M(\mathcal{J}, \mathcal{K}, \bar{x}, \bar{y}, \bar{u})$ has full column rank, it suffices to prove that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = 0 \implies \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = 0.$$

By the direct computation,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} A\lambda + B\mu \\ C\lambda + D\mu \end{pmatrix} = 0 \implies A\lambda + B\mu = 0 \text{ and } C\lambda + D\mu = 0.$$

Since $C$ is nonsingular, $\lambda = -C^{-1} D\mu$. Thus $(B - AC^{-1}D)\mu = 0$. Hence $\mu = 0$ and then $\lambda = 0$ if $B - AC^{-1}D$ is nonsingular. Note that $B - AC^{-1}D \in \mathbb{R}$ and $B - AC^{-1}D \neq 0$ can be derived by A6 in [28] and hence MPCC-LICQ holds.

**Part 3: Case I, Type 4.** In this case KKT-condition does not hold. For any FJ-multiplier $(\bar{u}_0, \bar{u}) \in \mathbb{R} \times \mathbb{R}^p$ at $\bar{y} \in S(\bar{x})$, where $\bar{u}_0 = 0$, MPCC-LICQ is satisfied for program $(\text{CPFJ}_\mu)$ if and only if the matrix

$$M(\mathcal{J}, \mathcal{K}, \bar{x}, \bar{y}, \bar{u}_0, \bar{u}) := \begin{pmatrix} \bar{u}_0 \nabla_{yx}^2 f + \nabla_{yx}^2 g^T \bar{u} & \nabla_x g_{\mathcal{J}} & 0 & 0 \\ \bar{u}_0 \nabla_{yy}^2 f + \nabla_{yy}^2 g^T \bar{u} & \nabla_y g_{\mathcal{J}} & 0 & 0 \\ [\nabla_y g_{\mathcal{K}^c}]^T & 0 & 1_{\mathcal{K}^c} & 0 \\ [\nabla_y f]^T & 0 & 0 & 1 \end{pmatrix}$$

has full column rank. Hence MPCC-LICQ holds if and only if the matrix

$$\begin{pmatrix} u_0 \nabla_{yx}^2 f + \nabla_{yx}^2 g^T \bar{u} & \nabla_x g_{\mathcal{J}} & 0 \\ u_0 \nabla_{yy}^2 f + \nabla_{yy}^2 g^T \bar{u} & \nabla_y g_{\mathcal{J}} & 0 \\ [\nabla_y g_{\mathcal{K}^c}]^T & 0 & 1_{\mathcal{K}^c} \end{pmatrix}$$

has full column rank.

In this case, by C1 and C4 in [28], $\mathcal{K}^c = \mathcal{J} = \{1, \ldots, p\}$, $p \geq 1$. Let us define the following map $\Phi : \mathbb{R} \times \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}$:

$$\Phi(x, y, u_0, u) = \begin{pmatrix} u_0 \nabla_y f(x, y) + u \nabla_y g(x, y) \\ g(x, y) \\ \sum_{j=0}^p u_j - 1 \end{pmatrix}.$$

Note that the map $\Phi$ has the same property of the map $\Psi$ defined in [28]. The only difference is that: the authors in [28] normalize $\bar{u}_j$'s ($\bar{\mu}_j$'s in [28]) by setting $\bar{\mu}_p = 1$; we normalize $\bar{u}_j$'s by setting $\sum_{j=0}^p u_j = 1$. Since $\bar{u}_j \neq 0$ by C4 in [28], they are equivalent. Due to C5, $D_{x,y,u}\Psi(\bar{x}, \bar{y}, 0, \bar{u})$ and then $D_{x,y,u}\Phi(\bar{x}, \bar{y}, 0, \bar{u})$ is nonsingular. Note that

$$D_{x,y,u}\Phi(\bar{x}, \bar{y}, 0, \bar{u}) = \begin{pmatrix} \bar{u}_0 \nabla_{yx}^2 f + \nabla_{yx}^2 g^T \bar{u} & \nabla_x g_{\mathcal{J}} & 0 \\ \bar{u}_0 \nabla_{yy}^2 f + \nabla_{yy}^2 g^T \bar{u} & \nabla_y g_{\mathcal{J}} & 0 \\ [\nabla_y g_{\mathcal{K}^c}]^T & 0 & 1_{\mathcal{K}^c} \end{pmatrix}.$$

Hence MPCC-LICQ holds. Moreover, since $\mathcal{J} = \mathcal{K}^c$, it reduces to LICQ in this case.

**Part 4: Case I, Type 5.** In this case, by D1, $\mathcal{J} = \{1, \ldots, p\}$, $p = m + 1 \geq 2$. Although MFCQ may does not hold in Type 5, KKT condition holds. We show that MPCC-LICQ holds for all Lagrange multipliers $\bar{u}$. To this end, we write the matrix

$$\begin{pmatrix} \nabla_{yx}^2 f + \nabla_{yx}^2 g^T \bar{u} & \nabla_x g_{\mathcal{J}} \\ \nabla_{yy}^2 f + \nabla_{yy}^2 g^T \bar{u} & \nabla_y g_{\mathcal{J}} \\ [\nabla_y g_{\mathcal{K}^c}]^T & 0 \end{pmatrix} = \begin{pmatrix} A & B \\ C & 0 \end{pmatrix},$$

where $B := \nabla g_{\mathcal{J}}$ and $C := [\nabla_y g_{\mathcal{K}^c}]^T$. To prove the matrix $M(\mathcal{J}, \mathcal{K}, \bar{x}, \bar{y}, \bar{u})$ has full column rank, it suffices to show that both $B$ and $C$ have full column rank. By D1 and D2 in [28], the matrix $B$ has full column rank and then it is nonsingular. This implies that $B$ is also full row rank. In particular, $(\nabla_y g_{\mathcal{J}})$ has full row rank. Since $\mathcal{K}^c \subseteq \mathcal{J}$, the matrix $C$ has full column rank. Therefore, $M(\mathcal{J}, \mathcal{K}, \bar{x}, \bar{y}, \bar{u})$ has full column rank and then MPCC-LICQ holds in this case. □

Since we have shown that the partial calmness holds and MPCC-LICQ is satisfied for the partially penalized problem of the combined program in the simple case, we can now write down the optimality conditions for the combined program and hence obtain the optimality conditions in the following theorem.

THEOREM 6.2 (**Optimality conditions for the simple BLPPs**).
*Let problem* (BLPP) *be simple at its local minimizer* $(\bar{x}, \bar{y}) \in M$ *and denote by $J_0 := J_0(\bar{x}, \bar{y})$. Then the following necessary optimality conditions hold.*
(1) *Suppose $S(\bar{x}) = \{\bar{y}\}$ and $(\bar{x}, \bar{y})$ is of Type 1. Let $\bar{u}$ be the unique KKT-multiplier for $L(\bar{x})$ at $\bar{y}$. Then $(\bar{x}, \bar{y}, \bar{u})$ is a local solution of the following standard nonlinear programming problem (NLP):*

$$
\begin{aligned}
\min_{x,y,u}\ & F(x,y) \\
\text{s.t.}\ & \nabla_y f(x,y) + \sum_{j \in J_0} u_j \nabla_y g_j(x,y) = 0,\ \ g_j(x,y) = 0,\ \ \forall j \in J_0.
\end{aligned}
\tag{6.1}
$$

*And KKT condition holds for problem* (6.1) *at $(\bar{x}, \bar{y}, \bar{u})$, i.e., there exist $w \in \mathbb{R}^m$, $\xi \in \mathbb{R}^p$ such that*

$$0 = \nabla_x F(\bar{x}, \bar{y}) - \left[\nabla^2_{xy}(f + \bar{u}^T g)(\bar{x}, \bar{y})\right] w + \nabla_x(\xi^T g)(\bar{x}, \bar{y}), \tag{6.2}$$

$$0 = \nabla_y F(\bar{x}, \bar{y}) - \left[\nabla^2_{yy}(f + \bar{u}^T g)(\bar{x}, \bar{y})\right] w + \nabla_y(\xi^T g)(\bar{x}, \bar{y}), \tag{6.3}$$

$$\xi_i = 0,\ \ \forall i \notin J_0, \tag{6.4}$$

$$\nabla_y g_j(\bar{x}, \bar{y})^T w = 0,\ \ \forall j \in J_0.$$

(2) *Suppose $S(\bar{x}) = \{\bar{y}\}$ and $(\bar{x}, \bar{y})$ is of Type 2. Let $\bar{u}$ be the unique KKT-multiplier for $L(\bar{x})$ at $\bar{y}$. Then there exists a unique $q \in J_0$ such that $\bar{u}_q = 0$, $\bar{u}_j > 0$ for $j \in J_0 \setminus \{q\}$, and $(\bar{x}, \bar{y}, \bar{u})$ is a local minimizer of the following MPCC:*

$$
\begin{aligned}
\min_{x,y,u}\ & F(x,y) \\
\text{s.t.}\ & \nabla_y f(x,y) + \sum_{j \in J_0} u_j \nabla_y g_j(x,y) = 0,\ \ g_j(x,y) = 0,\ \ \forall j \in J_0 \setminus \{q\}, \\
& g_q(x,y) \le 0,\ \ u_q \ge 0,\ \ u_q g_q(x,y) = 0.
\end{aligned}
\tag{6.5}
$$

*And the Strong (S-) stationary condition holds for problem* (6.5) *at $(\bar{x}, \bar{y}, \bar{u})$, i.e., there exists $(w, \xi) \in \mathbb{R}^m \times \mathbb{R}^p$ such that* (6.2)*-*(6.4) *and the following condition holds:*

$$\nabla_y g_j(\bar{x}, \bar{y})^T w = 0,\ \ \forall j \in J_0(\bar{x}, \bar{y}) \setminus \{q\},\ \ \nabla_y g_q(\bar{x}, \bar{y})^T w \le 0,\ \ \xi_q \ge 0. \tag{6.6}$$

(3) *Suppose $S(\bar{x}) = \{\bar{y}\}$ and $(\bar{x}, \bar{y})$ is of Type 4. Then $M_{\mathrm{FJ}}(\bar{x}, \bar{y}) = \{(\bar{u}_0, \bar{u})\}$ with $\bar{u}_0 = 0$ and $\bar{u}_j > 0$, $\forall j \in J_0$. Moreover, $(\bar{x}, \bar{y}, 0, \bar{u})$ is a local minimizer of the following standard NLP:*

$$
\begin{aligned}
\min_{x,y,u_0,u}\ & F(x,y) \\
\text{s.t.}\ & u_0 \nabla_y f(x,y) + \sum_{j \in J_0} u_j \nabla_y g_j(x,y) = 0, \\
& g_j(x,y) = 0,\ \ \forall j \in J_0,\ \ u_0 \ge 0,\ \ u_0 + \sum_{j \in J_0} u_j = 1.
\end{aligned}
\tag{6.7}
$$

*And KKT condition holds for problem (6.7) at $(\bar{x}, \bar{y}, 0, \bar{u})$, i.e., there exist $w \in \mathbb{R}^m$, $\xi \in \mathbb{R}^p$ such that (6.2)-(6.4) without the terms involving $f$ and the following condition holds:*

$$(6.8) \qquad \nabla_y f(\bar{x}, \bar{y})^T w \leq 0, \ \nabla_y g_j(\bar{x}, \bar{y})^T w = 0, \ \forall j \in J_0(\bar{x}, \bar{y}).$$

(4) *Suppose $S(\bar{x}) = \{\bar{y}\}$ and $(\bar{x}, \bar{y})$ is of Type 5-1. Then there exists a unique KKT-multiplier $\bar{u}$ and an index $q \in J_0$ with $\bar{u}_q = 0$, $\bar{u}_j > 0$, $\forall j \in J_0 \setminus \{q\}$, $\bar{\mu}_j > 0$, $\forall j \in J_0$ such that the set of KKT-multipliers have the representation*

$$(6.9) \qquad\qquad\qquad M_{\mathrm{KKT}}(\bar{x}, \bar{y}) = \bar{u} + \mathbb{R}_+ \bar{\mu}.$$

*Point $(\bar{x}, \bar{y}, \bar{u})$ is a local solution of MPCC (6.5) and S-stationary condition holds for problem (6.5) at $(\bar{x}, \bar{y}, \bar{u})$ with $w = 0$. Furthermore, $(\bar{x}, \bar{y})$ is a local solution of the following standard NLP:*

$$(6.10) \qquad \begin{aligned} \min_{x,y} \ & F(x, y) \\ \text{s.t. } & g_j(x, y) = 0, \ \forall j \in J_0(\bar{x}, \bar{y}) \setminus \{q\}, \ g_q(x, y) \leq 0, \end{aligned}$$

*and KKT condition for the above NLP holds i.e., there exists $\xi \in \mathbb{R}^p$ such that*

$$(6.11) \qquad \begin{aligned} 0 &= \nabla F(\bar{x}, \bar{y}) + \nabla(\xi^T g)(\bar{x}, \bar{y}), \\ \xi_i &= 0, \ \forall i \notin J_0(\bar{x}, \bar{y}), \ \xi_q \geq 0. \end{aligned}$$

(5) *Suppose $S(\bar{x}) = \{\bar{y}\}$ and $(\bar{x}, \bar{y})$ is of Type 5-2. Then there exists a unique pair of $q, r \in J_0$, $q \neq r$ such that there are $\bar{u}^1, \bar{u}^2 \in M_{\mathrm{KKT}}(\bar{x}, \bar{y})$ satisfying*

$$\bar{u}_q^1 = 0, \ \bar{u}_j^1 > 0, \ \forall j \in J_0 \setminus \{q\}; \quad \bar{u}_r^2 = 0, \ \bar{u}_j^2 > 0, \ \forall j \in J_0 \setminus \{r\},$$

*and $M_{\mathrm{KKT}}(\bar{x}, \bar{y})$ is the line segment jointing $\bar{u}^1$ and $\bar{u}^2$. Moreover, for any $\bar{u} \in M_{\mathrm{KKT}}(\bar{x}, \bar{y})$, $(\bar{x}, \bar{y}, \bar{u})$ is a local solution of the following MPCC:*

$$(6.12) \qquad \begin{aligned} \min_{x,y,u} \ & F(x, y) \\ \text{s.t. } & \nabla_y f(x, y) + \sum_{j \in J_0} u_j \nabla_y g_j(x, y) = 0, \ g_j(x, y) = 0, \ \forall j \in J_0 \setminus \{q, r\}, \\ & g_k(x, y) \leq 0, \ u_k \geq 0, \ u_k g_k(x, y) = 0, \ \forall k \in \{q, r\}. \end{aligned}$$

*And S-stationary condition holds for problem (6.12) at $(\bar{x}, \bar{y}, \bar{u})$. Moreover, S-stationary condition reduces to the existence of $\xi \in \mathbb{R}^p$ such that*

$$\begin{aligned} 0 &= \nabla F(\bar{x}, \bar{y}) + \nabla(\xi^T g)(\bar{x}, \bar{y}), \\ \xi_i &= 0, \ \forall i \notin J_0(\bar{x}, \bar{y}), \ \xi_q \geq 0 \text{ if } \bar{u} = \bar{u}^1, \ \xi_r \geq 0 \text{ if } \bar{u} = \bar{u}^2. \end{aligned}$$

(6) *Case II: $S(\bar{x}) = \{\bar{y}_1, \bar{y}_2\}$. Let $y_1(x), y_2(x)$ be the unique local minimizers for $L(x)$ in a neighborhood of $\bar{x}$ with $y_k(\bar{x}) = \bar{y}_k$, and $\bar{u}^k$ be the unique KKT-multiplier for $L(\bar{x})$ at $\bar{y}_k$, $k = 1, 2$. Then the value function can be written as*

$$V(x) = \min\left\{ f(x, y_1(x)), f(x, y_2(x)) \right\}$$

*for all $x$ lying in a neighbourhood of $\bar{x}$. Let $\bar{u} := \bar{u}^1$ be the unique KKT-multiplier for $L(\bar{x})$ at $\bar{y} = \bar{y}_1$. Then $\bar{u}_j > 0$ for all $j \in J_0(\bar{x}, \bar{y})$ and $(\bar{x}, \bar{y}, \bar{u})$ is a local solution of the following nonsmooth NLP for some $\mu > 0$:*

$$(6.13) \qquad \begin{aligned} \min_{x,y,u} \; & F(x,y) + \mu\big(f(x,y) - V(x)\big) \\ \text{s.t. } & \nabla_y f(x,y) + u\nabla_y g(x,y) = 0, \; g_j(x,y) = 0, \; \forall j \in J_0(\bar{x}, \bar{y}). \end{aligned}$$

*And the nonsmooth KKT condition holds for problem (6.13) at $(\bar{x}, \bar{y}, \bar{u})$, that is, there exist $\lambda^k \geq 0$ with $\sum_{k=1}^2 \lambda^k = 1$, $w \in \mathbb{R}^m$, $\xi \in \mathbb{R}^p$, and $\mu > 0$ such that*

$$\begin{aligned} 0 = \nabla_x F(\bar{x}, \bar{y}) + \mu \sum_{k=1}^2 \lambda^k \big(\nabla_x f(\bar{x}, \bar{y}) - \nabla_x f(\bar{x}, \bar{y}_k) - \bar{u}^k \nabla_x g(\bar{x}, \bar{y}_k)\big) \\ - \big[\nabla_{xy}^2 (f + \bar{u}^T g)(\bar{x}, \bar{y})\big] w + \nabla_x (\xi^T g)(\bar{x}, \bar{y}), \\ 0 = \nabla_y F(\bar{x}, \bar{y}) + \mu \nabla_y f(\bar{x}, \bar{y}) - \big[\nabla_{yy}^2 (f + \bar{u}^T g)(\bar{x}, \bar{y})\big] w + \nabla_y (\xi^T g)(\bar{x}, \bar{y}), \\ \xi_i = 0, \; \forall i \notin J_0(\bar{x}, \bar{y}), \; \nabla_y g_j(\bar{x}, \bar{y})^T w = 0, \; \forall j \in J_0(\bar{x}, \bar{y}). \end{aligned}$$

*Proof.* For Case I, by Lemma 5.3, $\Sigma_{FJ} = M$ in a neighbourhood of $(\bar{x}, \bar{y})$. Hence problem $(\mathrm{CP}_{FJ})$ is partially calm at $(\bar{x}, \bar{y})$ with $\mu = 0$. By Theorem 4.4, problem (CPFJ) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ for each FJ-multiplier $(\bar{u}_0, \bar{u}) \in M_{\mathrm{FJ}}(\bar{x}, \bar{y})$. When $\bar{u}_0 \neq 0$, this means that problem (CP) is partially calm at $(\bar{x}, \bar{y}, \bar{u})$ for each KKT-multiplier $\bar{u} \in M_{\mathrm{KKT}}(\bar{x}, \bar{y})$ with the same calmness modulus $\mu = 0$. For Case II, by Lemma 5.4 and the Lipschitz continuity of the objective function $F(x,y)$, problem (CP) is partially calm at $(\bar{x}, \bar{y}, \bar{u})$ with a calmness modulus $\mu > 0$.

By dropping inactive inequalities, it is easy to see that locally, the partially penalized problem can be reduced to corresponding optimization problem in each case.

Since MPCC-LICQ (which reduces to a standard constraint qualification if the problem is a NLP) is proved in Proposition 6.1, S-stationary condition hold for each optimization problem and so no extra constraint qualifications are required.

Next, we consider the structure of FJ and KKT-multipliers. Except Type 5, the FJ and KKT-multipliers are unique. In Case I, Type 5-1, since $|J_0(\bar{x}, \bar{y})| = m + 1$ by (D1) in [28] and any $m$ elements of $\{\nabla_y g_j(\bar{x}, \bar{y}), \; j \in J_0\}$ is linearly independent by (D3) in [28], CRCQ holds. But MFCQ does not hold since all $\mu_j$, $j \in J_0$ in [28] have the same sign. Since $\mu_j$ in [28] is unique up to a common multiple, there exists a unique KKT-multiplier $\bar{u}$ and an index $q \in J_0$ such that (6.9) holds. On the other hand, the optimality condition (6.11) follows from (6.2)-(6.4) and (6.6) since

$$(6.14) \qquad \nabla_y g_j(\bar{x}, \bar{y})^T w = 0, \; \forall j \in J_0(\bar{x}, \bar{y}) \setminus \{q\} \implies w = 0,$$

because $|J_0(\bar{x}, \bar{y}) \setminus \{q\}| = m$ and $\{\nabla_y g_j(\bar{x}, \bar{y}), \; j \in J_0 \setminus \{q\}\}$ is linearly independent. In the same manner we can see that $w = 0$ in Case I, Type 5-2. We could say more on Case I, Type 5-1. In this case, since $\bar{u}_j > 0$, $\forall j \in J_0 \setminus \{q\}$ for all KKT-multiplier $\bar{u} \in M_{\mathrm{KKT}}(\bar{x}, \bar{y})$ and $\{\nabla_y g_j(\bar{x}, \bar{y}), \; j \in J_0 \setminus \{q\}\}$ is linearly independent in $\mathbb{R}^m$ with $|J_0(\bar{x}, \bar{y}) \setminus \{q\}| = m$, by the implicit function theorem, the curve $y^q(x)$ defined by $y^q(\bar{x}) = \bar{y}$ and $g_j(x, y^q(x)) = 0$, $\forall j \in J_0 \setminus \{q\}$ satisfies $(x, y^q(x)) \in M$ in a neighbourhood of $\bar{x}$. Hence, $(\bar{x}, \bar{y})$ is actually a local solution of NLP (6.10). In Case I, Type 5-2, both CRCQ and MFCQ hold since any $m$ elements of $\{\nabla_y g_j(\bar{x}, \bar{y}), \; j \in J_0\}$ is linearly independent by (D1-D3) in [28] and $\mu_j$, $j \in J_0$ have different signs. Hence, by a similar argument of the proof of [3, Lemma 5.2], there exists a unique pair of $q, r \in J_0$, $q \neq r$ such that there are $\bar{u}^1, \bar{u}^2 \in M_{\mathrm{KKT}}(\bar{x}, \bar{y})$ satisfying (6.2) and $M_{\mathrm{KKT}}(\bar{x}, \bar{y})$ is the line segment jointing $\bar{u}^1$ and $\bar{u}^2$. $\qquad\square$

By using the implicit function theorem, we can show that Theorem 6.2 and the necessary optimality conditions derived in [28, Theorem 4.3] are equivalent in the following sense. Due to the page limit we omit the proof.

THEOREM 6.3 (**Equivalence of optimality conditions derived in Theorem 4.3 of [28] and Theorem 6.2**).

**Case I, Type 1:** *The optimality condition in Theorem 6.2 (1) holds if and only if there exist $C^2$-mappings $y(x)$, $u_j(x)$, $j \in \mathcal{J} := J_0(\bar{x}, \bar{y})$, satisfying*

$$(6.15) \qquad\qquad y(\bar{x}) = \bar{y}, \quad u(\bar{x}) = \bar{u},$$

*and the following equations in a neighborhood of $\bar{x}$:*

$$(6.16) \qquad \begin{cases} \nabla_y f(x, y(x)) + \sum_{j \in \mathcal{J}} u_j(x) \nabla_y g_j(x, y(x)) = 0, \\ g_j(x, y(x)) = 0, \ \forall j \in \mathcal{J}, \end{cases}$$

*such that $\nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x y(\bar{x}) = 0$.*

**Case I, Type 2:** *The optimality condition in Theorem 6.2 (2) holds if and only if there exist $C^2$-mappings $\widetilde{y}(x)$, $\widetilde{u}_j(x)$, $j \in J_0(\bar{x}, \bar{y}) \setminus \{q\}$ and $\widehat{y}(x)$, $\widehat{u}_j(x)$, $j \in J_0(\bar{x}, \bar{y})$ satisfying (6.15) and (6.16) with $\mathcal{J} = J_0(\bar{x}, \bar{y}) \setminus \{q\}$ and $\mathcal{J} = J_0(\bar{x}, \bar{y})$, respectively, such that*

$$\begin{cases} \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x \widetilde{y}(\bar{x}) \leq 0, \\ \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x \widehat{y}(\bar{x}) \geq 0, \end{cases} \text{if } \left. \frac{dg_q(x, \widetilde{y}(x))}{dx} \right|_{x=\bar{x}} > 0,$$

$$\begin{cases} \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x \widetilde{y}(\bar{x}) \geq 0, \\ \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x \widehat{y}(\bar{x}) \leq 0, \end{cases} \text{if } \left. \frac{dg_q(x, \widetilde{y}(x))}{dx} \right|_{x=\bar{x}} < 0.$$

**Case I, Type 4:** *The optimality condition in Theorem 6.2 (3) holds if and only if there exist $C^2$-mappings $x(u_0)$, $y(u_0)$, $u_j(u_0)$, $j \in \mathcal{J} := J_0(\bar{x}, \bar{y})$ satisfying $x(0) = \bar{x}$, $y(0) = \bar{y}$, $u(0) = \bar{u}$ and the following system in an open neighborhood of $\bar{u}_0$:*

$$\begin{cases} u_0 \nabla_y f(x(u_0), y(u_0)) + \sum_{j \in \mathcal{J}} u_j(u_0) \nabla_y g_j(x(u_0), y(u_0)) = 0, \\ u_0 + \sum_{j \in \mathcal{J}} u_j(u_0) = 1, \\ g_j(x(u_0), y(u_0)) = 0, \ \forall j \in \mathcal{J}, \end{cases}$$

*such that*

$$\nabla_x F(\bar{x}, \bar{y}) D_{u_0} x(0) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_{u_0} y(0) \geq 0.$$

**Case I, Type 5-1:** *The optimality condition in Theorem 6.2 (4) holds if and only if there exist $C^2$-mappings $y^q(x)$, $u_j^q(x)$, $j \in J_0(\bar{x}, \bar{y}) \setminus \{q\}$ satisfying (6.15) and (6.16) with $\mathcal{J} = J_0(\bar{x}, \bar{y}) \setminus \{q\}$ such that*

$$\begin{cases} \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x y^q(\bar{x}) \leq 0 & \text{if } \left. \frac{dg_q(x, y^q(x))}{dx} \right|_{x=\bar{x}} > 0, \\ \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x y^q(\bar{x}) \geq 0 & \text{if } \left. \frac{dg_q(x, y^q(x))}{dx} \right|_{x=\bar{x}} < 0. \end{cases}$$

**Case I, Type 5-2:** *The optimality condition in Theorem 6.2 (5) holds if and only if there exist $C^2$-mappings curves $y^q(x), u_j^q(x)$, $j \in J_0(\bar{x}, \bar{y}) \setminus \{q\}$, $y^r(x), u_j^r(x)$,*

$j \in J_0(\bar{x}, \bar{y}) \setminus \{r\}$ satisfying (6.15) with $\bar{u} = \bar{u}^1, \bar{u}^2$ and (6.16) with $\mathcal{J} = J_0(\bar{x}, \bar{y}) \setminus \{q\}$ and $\mathcal{J} = J_0(\bar{x}, \bar{y}) \setminus \{r\}$, respectively, such that

$$(6.17) \qquad \begin{cases} \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x y^q(\bar{x}) \leq 0, \\ \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x y^r(\bar{x}) \geq 0, \end{cases} \text{ if } \gamma_q > 0 \text{ (i.e. } \gamma_r < 0),$$

$$(6.18) \qquad \begin{cases} \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x y^q(\bar{x}) \geq 0, \\ \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x y^r(\bar{x}) \leq 0, \end{cases} \text{ if } \gamma_q < 0 \text{ (i.e. } \gamma_r > 0),$$

where $\gamma_q := \text{sign}\left(\frac{dg_q(x, y^q(x))}{dx}\Big|_{x=\bar{x}}\right)$ and $\gamma_r := \text{sign}\left(\frac{dg_r(x, y^r(x))}{dx}\Big|_{x=\bar{x}}\right)$.

**Case II:** *The optimality condition in Theorem* 6.2 *(6) holds if and only if there exist* $C^2$-*mappings* $y_k(x)$, $u_j^k(x)$, $j \in J_0(\bar{x}, \bar{y}_k)$, $k = 1, 2$, *satisfying* (6.15) *with* $\bar{y} = \bar{y}_k, \bar{u} = \bar{u}^k$ *and* (6.16) *with* $\mathcal{J} = J_0(\bar{x}, \bar{y}_k)$ *such that*

$$\begin{cases} \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x y_1(\bar{x}) \geq 0 & \text{if } \frac{d[f(x, y_2(x)) - f(x, y_1(x))]}{dx}\Big|_{x=\bar{x}} > 0, \\ \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \cdot D_x y_1(\bar{x}) \leq 0 & \text{if } \frac{d[f(x, y_2(x)) - f(x, y_1(x))]}{dx}\Big|_{x=\bar{x}} < 0. \end{cases}$$

*Remark* 6.4. This result is related to an interesting and involved issue in [28, Remark 4.2]: the comparison of optimality conditions derived in the literature [28, Theorem 4.3] and from the literature [11, 53, 54] by using KKT, value function, and the combined approaches. From our point of view, the underlying reason for the equivalence in Theorem 6.3 is that the structure of FJ-multipliers can distinguish the local structure of bilevel problems. The generic structure of FJ-multipliers locally around global minimizers are given in the following Figure 6.1. It is worth noting that Figure 6.1 distinguishes all of the cases.
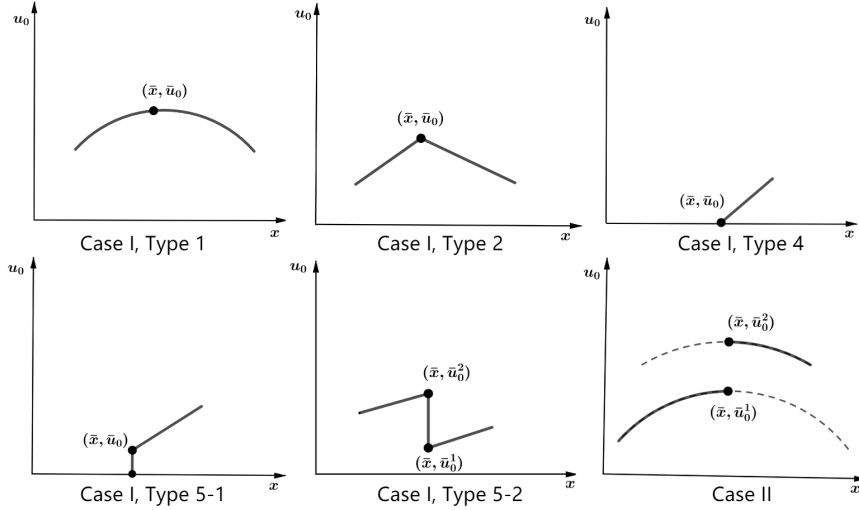


FIG. 6.1. *Generic structure of FJ-multipliers locally around global minimizers.*

The following example is a typical example of Case I, type 4. Let us demonstrate our results by deriving optimality conditions for this example.

EXAMPLE 6.1 ([28, Example 2.2]).

$$\min_{x,y} x + \sum_{j=1}^{m} y_j$$

$$\text{s.t. } y \in S(x) := \arg\min_{y} \Big\{ -y_1 : \sum_{j=1}^{m} y_j^2 - x \leq 0 \Big\}.$$

*It is easy to see that $S(x) = \{(\sqrt{x}, 0_{m-1})\}$ if $x \geq 0$ and $S(x) = \emptyset$ if $x < 0$. The optimal solution of the bilevel problem is $(0, 0_m)$. The lower level problem violates MFCQ at $0_m$ and $0_m$ is not a KKT-point. One may consider whether we can use the value function approach. When $m = 1$, it is not difficulty to check that $L(x)$ has a uniformly weak sharp minimum over $f$ around $(0,0)$ and hence the value function approach can be used if $m = 1$. However since MFCQ fails, the value function $V(x) = \sqrt{x}$ is not Lipschitz continuous at $\bar{x} = 0$ and it is difficult to obtain a necessary optimality condition for problem (VP). Moreover for $m \geq 2$, the partial calmness for (VP) (i.e., partial calmness over $f$) does not hold at $(0, 0_m)$. It is worth noting that the violation of the partial calmness for (VP) cannot be rectified by a small perturbation of the data of the problem. However, it is easy to check the only FJ multiplier for problem $L(\bar{x})$ is $(\bar{u}_0, \bar{u}) = (0, 1)$. The optimality condition from Theorem 6.2(2) is verified with $\xi = 1$ and $w = \frac{1}{2}(1, \ldots, 1)^T$.*

Next we consider a special case where there is no upper and lower level constraints. That is, we consider the unconstrained BLPPs:

(UBLPP) $$\min_{x,y} F(x,y) \quad \text{s.t.} \quad y \in \arg\min_{y} f(x,y).$$

It is clear that only Case I, Type 1 and Case II will happen if (UBLPP) is simple at $(\bar{x}, \bar{y}) \in M$. Hence the following corollary follows from Theorem 6.2.

COROLLARY 6.5. *Let the unconstrained bilevel programming problem (UBLPP) be simple at its local minimizer $(\bar{x}, \bar{y}) \in M$, i.e., Case I type 1 or Case II, then there exist $\mu \geq 0$ and $w \in \mathbb{R}^m$ such that*

$$0 = \nabla_x F(\bar{x}, \bar{y}) + \mu(\nabla_x f(\bar{x}, \bar{y}) - \nabla_x f(\bar{x}, \bar{y}_2)) + \nabla_{yx}^2 f(\bar{x}, \bar{y})w,$$
$$0 = \nabla_y F(\bar{x}, \bar{y}) + \nabla_{yy}^2 f(\bar{x}, \bar{y})w,$$

*where $\mu = 0$ in Case I and $\bar{y}_2$ is the distinct minimizer of $f(\bar{x}, \cdot)$ in Case II.*

We claim that the points in Case II are $f$-critical by the notation of Mirrlees [34]. Indeed, according to [34], $(\bar{x}, \bar{y})$ is $f$-critical if $\bar{y} \in S(\bar{x})$, but there is no neighborhood $\mathcal{N}$ of $(\bar{x}, \bar{y})$ such that $(x', y') \in \mathcal{N}$, $\nabla_y f(x', y') = 0$ imply that $y' \in S(x')$. Since $\alpha \neq 0$ in (4.9), it is easily seen that $(\bar{x}, \bar{y})$ is $f$-critical if $(\bar{x}, \bar{y})$ is in Case II. On the other hand, if $(\bar{x}, \bar{y})$ is not $f$-critical, then there is a neighborhood $\mathcal{N}$ of $(\bar{x}, \bar{y})$ such that $(x', y') \in \mathcal{N}$, $\nabla_y f(x', y') = 0$ imply that $y' \in S(x')$. Thus $\Sigma_{gc} \cap \mathcal{U} = M \cap \mathcal{U}$ and then $(\bar{x}, \bar{y})$ is in Case I. This implies that $f$-critical points–generically–are in Case II under our problem setting by Theorem 4.11. Clearly, this verifies Mirrlees' claim: "Normally, a $V$-critical point $(a', z')$ is such that there are two distinct maxima, $z'$ and $z''$, of $V(a, \cdot)$, there being no others" on page 16 of [34]. Here $V$ is the lower level objective function and $a, z$ are the upper and lower level variables, respectively. Moreover, under this claim, Mirrlees gave a form of KKT condition for (UBLPP) in (52) on page 17 of [34] which coincides with the necessary optimality condition in [54,

Theorem 2.1] with a rigorous proof. Hence the condition in [54, Theorem 2.1] (the one in Corollary 6.5) is, of course, a generic one.

## REFERENCES

[1] L. Adam, R. Henrion and J.V. Outrata, *On M-stationarity conditions in MPECs and the associated qualification conditions*, Math. Program., 168 (2018), pp. 229–259.

[2] G. Allende and G. Still, *Solving bilevel programs with the KKT-approach*, Math. Program., 138 (2013), pp. 309-332.

[3] A. Baccari and A. Trad, *On the classical necessary second-order optimality conditions in the presence of equality and inequality constraints*, SIAM J. Optim., 15 (2005), pp. 394-408.

[4] J. Bard, *Practical Bilevel Optimization: Algorithms and Applications*, Dordrecht: Kluwer Academic Publishers, 1998.

[5] M. Bjørndal and K. Jørnsten, *The deregulated electricity market viewed as a bilevel programming problem*, J. Global Optim., 33 (2005), pp. 465-475.

[6] P. Bond and A. Gomes, *Multitask principal-agent problems: optimal contracts, fragility, and effort misallocation*, J. Econ. Theory, 144 (2009), pp. 175-211.

[7] P. Bröcker and L. Lander, *Differentiable Germs and Catastrophes*, Cambridge University Press, Cambridge, 1975.

[8] F. Clarke, *Optimization and Nonsmooth Analysis*, Wiley Interscience, New York, 1983.

[9] J.R. Conlon, *Two new conditions supporting the first-order approach to multisignal principal-agent problems*, Econometrica, 77 (2009), pp. 249-278.

[10] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, 2002.

[11] S. Dempe and J. Dutta, *Is bilevel programming a special case of a mathematical program with complementarity constraints?*, Math. Program., 131 (2012), pp. 37-48.

[12] S. Dempe, J. Dutta and B.S. Mordukhovich, *New necessary optimality conditions in optimistic bilevel programming*, Optimization, 56 (2007), pp. 577-604.

[13] S. Dempe, H. Günzel and H.Th. Jongen, *On reducibility in bilevel problems*, SIAM J. Optim., 20 (2009), pp. 718-727.

[14] D. Dominik, H.Th. Jongen and V. Shikhman, *On Intrinsic Complexity of Nash Equilibrium Problems and Bilevel Optimization*, J. Optim. Theory Appl., 159 (2013), pp. 606-634.

[15] S. Dempe, V. Kalashnikov, G. Prez-Valds and N. Kalashnykova, *Bilevel Programming Problems, Energy Systems*, Springer, Berlin. 2015.

[16] S. Dempe and A.B. Zemkoho, *The bilevel programming problems: reformulations, constraint qualifications and optimality conditions*, Math. Program., 138 (2013), pp. 447-473.

[17] S. Dempe and A.B. Zemkoho, *Bilevel optimization: advances and next challenges.* Springer Optimization and its Applications, vol. 161, 2020.

[18] L. Franceschi, P. Frasconi, S. Salzo, R. Grazzi and M. Pontil, *Bilevel programming for hyper-parameter optimization*, International Conference on Machine Learning (2018), pp. 1563-1572.

[19] H. Gfrerer and J.J. Ye, *New constraint qualifications for mathematical programs with equilibrium constraints via variational analysis*, SIAM J. Optim., 27 (2017), pp. 842–865.

[20] H. Gfrerer and J.J. Ye, *New sharp necessary optimality conditions for mathematical programs with equilibrium constraints*, Set-Valued and Variational Analysis, 28 (2020), pp. 395-426.

[21] S.J. Grossman and O.D. Hart, *An analysis of the principal-agent problem*, Econometrica, 51 (1983), pp. 7-45.

[22] R. Henrion, A. Jourani and J.V. Outrata, *On the calmness of a class of multifunctions*, SIAM J. Optim., 26 (2002), pp. 603-618.

[23] R. Henrion and T.M. Surowiec, *On calmness conditions in convex bilevel programming*, Appl. Anal., 90 (2011), pp. 951-970.

[24] B. Holmström and P. Milgrom, *Aggregation and Linearity in the Provision of Intertemporal Incentives*, Econometrica, 55 (1987), pp. 303-328.

[25] B. Holmström and P. Milgrom, *Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design*, J. Law, Econ., Organ., 7 (1991), pp. 24–52.

[26] H.Th. Jongen, P. Jonker and F. Twilt, *Critical sets in parametric optimization*, Math. Program., 34 (1986), pp. 333-353.

[27] H.Th. Jongen, P. Jonker and F. Twilt, *Nonlinear Optimization in Finite Dimensions*, Kluwer Academic, Dordrecht, 2000.

[28] H.Th. Jongen and V. Shikhman, *Bilevel optimization: on the structure of the feasible set*, Math. Program., 136 (2012), pp. 65-89.

[29] G. Kunapuli, K. Bennett, J. Hu and J-S. Pang, *Classification model selection via bilevel pro-*

*gramming*, Optim. Meth. Softw., 23 (2008), pp. 475-489.

[30] R. Liu, P. Mu, X. Yuan, S. Zeng and J. Zhang, *A generic first-order algorithmic Framework for bi-level programming beyond lower-level singleton*, International Conference on Machine Learning (2020).

[31] G.S. Liu, J.J. Ye and J.P. Zhu, *Partial exact penaty for mathematical programs with equilibrium constraints*, Set-Valued Anal., 16 (2008), pp. 785-804.

[32] Z.-Q. Luo, J-S. Pang, and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, United Kingdom, 1996.

[33] P. Mehlitz, L.I. Minchenko, and A.B. Zemkoho, *A note on partial calmness for bilevel optimization problems with linearly structured lower level*, Optimization Letters (2020), pp. 1-15.

[34] J. Mirrlees, *The theory of moral hazard and unobservable behaviour– part I*, Review of Economic Studies, 66 (1999), pp. 3-22.

[35] B.S. Mordukhovich, *Variational Analysis and Applications*, Springer, Cham, 2018.

[36] B.S. Mordukhovich, *Bilevel Optimization and Variational Analysis*, Bilevel Optimization: Advances and Next Challenges, ch. 7, Springer Optimization and its Applications vol. 161, 2020.

[37] B.S. Mordukhovich, N.M. Namand and H.M. Phan, *Variational analysis of marginal functions with applications to bilevel programming*, J. Optim. Theory Appl., 152 (2012), pp. 557-586.

[38] T.S. Motzkin, *Two Consequences of the Transposition Theorem on Linear Inequalities*, Econometrica, 19 (1951), pp. 184-185.

[39] J. Nie, L. Wang and J.J. Ye, *Bilevel polynomial programs and semidefinite relaxation methods*, SIAM J. Optim., 27 (2017), pp. 1728-1757.

[40] J. Nie, L. Wang, J.J. Ye and S. Zhong, *A Lagrange multiplier expression method for bilevel polynomial programs*, to apprear in SIAM J. Optim., arXiv:2007.07933v1.

[41] J.V. Outrata, *On the numerical solution of a class of Stackelberg problems*, ZOR-Math. Methods Oper. Res., 34 (1990), pp. 255-277.

[42] J.V. Outrata, M. Kocvara and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equlilibrium Constraints: Theory, Applications, and Numerical Results*, Kluwer Academic Publisher, Dordrect, The Netherlands, 1998.

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

[44] S. Scholtes and M. Stöhr, *How stringent is the linear independence assumption for mathematical programs with complementarity constraints?*, Math. Oper. Res., 26 (2001), pp. 851–863.

[45] K. Shimizu, Y. Ishizuka and J. Bard, *Nondifferentiable and Two-level Mathematical Programming*, Kluwer Academic, 1997.

[46] J.J. Ye, *Nondifferentiable multiplier rules for optimization and bilevel optimization problems*, SIAM J. Optim., 15 (2004), pp. 252-274.

[47] J.J. Ye, *Constraint qualifications and KKT conditions for bilevel programming problems*, Math. Oper. Res., 31 (2006), pp. 811-824.

[48] J.J. Ye, *Necessary optimality conditions for multiobjective bilevel programming problems*, Math. Oper. Res., 36 (2011), pp. 165-184.

[49] J.J. Ye, *Constraint qualifications and optimality conditions in bilevel optimization*, Bilevel Optimization: Advances and Next Challenges, ch. 8, Springer Optimization and its Applications vol. 161, 2020.

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

[51] J.J. Ye, X.M. Yuan, S. Zeng and J. Zhang, *Difference of convex algorithms for bilevel programs with applications in hyperparameter selection*, preprint, arXiv 2102.09006.

[52] J.J. Ye and J.C. Zhou, *Verifiable sufficient conditions for the error bound property of second-order cone complementarity problems*, Math. Program., 171 (2018), pp. 361-395.

[53] J.J. Ye and D.L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization, 33 (1995), pp. 9-27.

[54] J.J. Ye and D.L. Zhu, *New necessary optimality conditions for bilevel programs by combining the MPEC and value function approaches*, SIAM J. Optim., 20 (2010), pp. 1885-1905.

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