

Combined Approach with Second-Order Optimality Conditions for Bilevel Programming Problems

Xiaoxiao Ma*

*Department of Mathematics and Statistics, University of Victoria, Canada
xiaoxiaoma@uvic.ca*

Wei Yao

*Department of Mathematics, Southern University of Science and Technology,
and: National Center for Applied Mathematics, Shenzhen, Guangdong, China
yaow@sustech.edu.cn*

Jane J. Ye

*Department of Mathematics and Statistics, University of Victoria, Canada
janeye@uvic.ca*

Jin Zhang[†]

*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*

Dedicated to Roger J-B Wets on the occasion of his 85th birthday.

Received: February 10, 2022

Accepted: February 2, 2023

We propose a combined approach with second-order optimality conditions of the lower level problem to study constraint qualifications and optimality conditions for bilevel programming problems. The new method is inspired by the combined approach developed by Ye and Zhu in 2010, where the authors combined the classical first-order and the value function approaches to derive new necessary optimality conditions. In our approach, we add a second-order optimality condition to the combined program as a new constraint. We show that when all known approaches fail, adding the second-order optimality condition as a constraint makes the corresponding partial calmness condition and the resulting necessary optimality condition easier to hold. We also give some discussions on advantages and disadvantages of the combined approaches with the first-order and the second-order information.

Keywords: Partial calmness, bilevel program, optimality condition, second-order optimality condition.

2010 Mathematics Subject Classification: 90C26, 90C30, 90C31, 90C33, 90C46, 49J52, 91A65.

* The alphabetical order of the authors indicates the equal contribution to the paper. The research of Ye is supported by NSERC. Zhang's work is supported by National Science Foundation of China (No. 12222106), Guangdong Basic and Applied Basic Research Foundation (No. 2022B1515020082) and Shenzhen Science and Technology Program (No. RCYX20200714114700072).

[†] Corresponding author.

1. Introduction

In this paper we consider the following bilevel programming problem (BLPP):

$$(BLPP) \quad \min_{x,y} F(x,y) \quad \text{s.t. } y \in S(x), \quad G(x,y) \leq 0,$$

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

$$P(x) \quad \min_y f(x,y) \quad \text{s.t. } g(x,y) \leq 0.$$

For convenience, we denote the feasible set of $P(x)$ by

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

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

The bilevel programming problem has many applications including the principal-agent moral hazard problem [40], hyperparameter optimization and meta-learning in machine learning [20, 31, 33, 50]. More applications can be found in [5, 13, 15, 44]. For a comprehensive review, we refer to [17] and the references therein.

It is well known that optimality conditions of the lower level program are very useful in the reformulation of BLPPs both theoretically and computationally. The classical Karush-Kuhn-Tucker (KKT) approach is to replace the lower level program by its KKT condition and minimize over the original variables as well as multipliers. When the lower level program has inequality constraints, the KKT reformulation has been studied in the framework of the mathematical program with equilibrium constraints (MPEC). However there are some problems associated with the KKT approach. First, the KKT condition may only be necessary but not sufficient. In this case the KKT approach will enlarge the feasible region and hence the resulting single level problem may not be equivalent to the original bilevel program. Moreover an example in Mirrlees [40] shows that the solution set of the equivalent single level reformulation may not even include solutions of the original bilevel program. Second, even in the case where the KKT conditions are necessary and sufficient for $y \in S(x)$, treating multipliers of the lower level as extra variables can still make the resulting single level reformulation different from the original BLPPs in the sense of local optimality; see [14]. Recently [8] has discussed the issue of equivalence for more general problems for which some reformulations may include implicit variables with the BLPP as an example. Recently reformulations using the Bouligand (B-) stationary condition for the lower level program:

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

where $\widehat{N}_C$ denotes the regular normal cone to set C to replace the lower level program have been investigated; see [1, 25, 26, 30]. Calmness properties for the KKT reformulation and the B-stationarity reformulation have been compared in [1, 25] in the context of MPECs and it was discovered that usually the B-stationarity reformulation is easier to satisfy the calmness condition than the KKT reformulation.

Note that extra assumptions (at least the smoothness of the objective function of the lower level program) are always required for a reformulation using optimality conditions for the lower level program.

In contrast to any reformulation using optimality conditions for the lower level program, the value function approach proposed by Outrata [43] for numerical purpose and used by Ye and Zhu [52] for optimality conditions does not require any extra assumptions. By this approach, one defines the value function as an extended real-valued function

$$V(x) := \inf_y \{f(x, y) : g(x, y) \leq 0\},$$

and replaces the original BLPP by the following equivalent problem:

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

However, since the value function constraint $f(x, y) - V(x) \leq 0$ is actually an equality constraint, the nonsmooth Mangasarian-Fromovitz constraint qualification (MFCQ) for (VP) will never hold [52, Proposition 3.2]. To derive necessary optimality conditions for BLPPs, Ye and Zhu [52, Definition 3.1 and Proposition 3.3] proposed the partial calmness condition for (VP) under which the difficult constraint $f(x, y) - V(x) \leq 0$ was added as a penalty term to the objective function.

Although it was proved in [52] that the partial calmness condition for (VP) holds automatically for the minmax problem and the bilevel program where the lower level program is linear in both upper and lower level variables, the partial calmness condition for (VP) is celebrated but has been shown to be restrictive (cf. [16, 29, 37, 41, 48]). To improve the value function approach, Ye and Zhu [51] proposed a combination of the classical KKT and the value function approach. The resulting problem is the combined problem using KKT condition:

$$(CP) \quad \begin{aligned} & \min_{x,y,u} F(x, y) \\ & \text{s.t.} \quad f(x, y) - V(x) \leq 0, \quad \nabla_y f(x, y) + \nabla_y g(x, y)^T u = 0, \\ & \quad g(x, y) \leq 0, \quad u \geq 0, \quad u^T g(x, y) = 0, \quad G(x, y) \leq 0. \end{aligned}$$

Problem (CP) is equivalent to problem (BLPP) (in the sense of [51, Proposition 3.1]) when the KKT condition holds at each optimal solution of the lower level program. Similar to [52], to deal with the fact that the nonsmooth MFCQ also fails for (CP), the corresponding partial calmness condition for (CP) was proposed in [51, Definition 3.1].

Note that the reformulation (CP) requires the validity of the KKT conditions at each optimal solution of the lower level program. To deal with the case where the KKT condition may not hold at all the solutions of the lower level program, the Fritz John (FJ) condition was considered. In [2], Allende and Still replaced the lower-level program of BLPPs with the FJ condition (without a value function constraint). Regarding the combined approach, Ke et al. [30] proposed the following combined program using the FJ condition:

$$\begin{aligned}
(\text{CPFJ}) \quad & \min_{x,y,u_0,u} F(x,y) \\
& \text{s.t. } f(x,y) - V(x) \leq 0, \quad u_0 \nabla_y f(x,y) + \nabla_y g(x,y)^T u = 0, \\
& g(x,y) \leq 0, \quad (u_0, u) \geq 0, \quad u^T g(x,y) = 0, \quad \sum_{i=0}^p u_i = 1, \quad G(x,y) \leq 0.
\end{aligned}$$

The equivalence of problem (CPFJ) and problem (BLPP) (for both local and global optimal points) is a special case of [8, Theorem 4.5].

Similar to Ye and Zhu's work in [51], Ke et al. proposed the following partial calmness condition for (CPFJ) in [30].

Definition 1.1. (Partial calmness for (CPFJ)) Let $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ be a local 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:

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

Moreover, they analyzed the partial calmness for the combined program based on FJ conditions from a generic point of view and proved that the partial calmness for (CPFJ) is generic when the upper level variable has dimension one.

The following two combined problems and the corresponding partial calmness conditions are also discussed in [30].

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

where $\Sigma_{FJ} := \{(x,y) \in \mathbb{R}^{n+m} : y \text{ satisfies the FJ condition for } P(x)\}$, and the combined program with the B-stationary condition for the lower level program:

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

Although the partial calmness for the combined program may hold quite often, there are still cases where it does not hold; see e.g. Examples 3.1, 4.7, 4.8 in this paper. The main goal of this paper is to investigate the following question:

(Q) *How to derive necessary optimality conditions for bilevel problems when necessary optimality conditions for problems (VP) and (CP) do not hold?*

Contributions. To answer (Q), we propose to use second-order optimality conditions of the lower level program on top of the value function constraint and the first-order optimality constraint. The key point of adding the second order condition is to increase the freedom of choosing multipliers. Since each extra redundant constraint is associated with a multiplier, the more redundant constraints we add, the weaker is the optimality condition and hence easier for the resulting necessary optimality condition to hold.

In this sense, the resulting necessary optimality condition by adding the second-order condition is much more likely to hold than the one adding the first-order condition only, i.e., (CP), which is in term more likely to hold than the one without adding any optimality condition, i.e., (VP).

To illustrate our approach, consider the following KKT combined program:

$$(\text{KKTCP}) \quad \min_{x,y} F(x,y) \quad \text{s.t.} \quad f(x,y) - V(x) \leq 0, \quad (x,y) \in \Sigma_{\text{KKT}}, \quad G(x,y) \leq 0,$$

where

$$\Sigma_{\text{KKT}} := \left\{ (x,y) : \begin{array}{l} \exists u \text{ s.t. } \nabla_y f(x,y) + \nabla_y g(x,y)^T u = 0, \\ g(x,y) \leq 0, \quad u \geq 0, \quad u^T g(x,y) = 0 \end{array} \right\}$$

and its partially penalized problem:

$$(\text{KKTCP}_\mu) \quad \min_{x,y} F(x,y) + \mu(f(x,y) - V(x)) \quad \text{s.t.} \quad (x,y) \in \Sigma_{\text{KKT}}, \quad G(x,y) \leq 0.$$

Note that the combined program (CP) is a relaxed problem of (KKTCP) in the sense that the minimization is also performed on multipliers in problem (CP). To use the second-order information, we propose the following second-order combined problem:

$$(\text{SOCP}) \quad \min_{x,y} F(x,y) \quad \text{s.t.} \quad f(x,y) - V(x) \leq 0, \quad (x,y) \in \Sigma_{\text{SOC}}, \quad G(x,y) \leq 0,$$

where

$$\Sigma_{\text{SOC}} := \left\{ (x,y) \in \mathbb{R}^{n+m} : y \text{ satisfies a second-order optimality condition for } P(x) \right\},$$

and its partially penalized problem:

$$(\text{SOCP}_\mu) \quad \min_{x,y} F(x,y) + \mu(f(x,y) - V(x)) \quad \text{s.t.} \quad (x,y) \in \Sigma_{\text{SOC}}, \quad G(x,y) \leq 0.$$

When both the KKT condition and a certain second-order optimality condition hold for $y \in S(x)$, one has

$$\text{gph } S := \{(x,y) \in \mathbb{R}^{n+m} : y \in S(x)\} \subseteq \Sigma_{\text{SOC}} \subseteq \Sigma_{\text{KKT}}. \quad (1)$$

In general, the inclusions above are strict. If the second inclusion is strict, i.e., the set Σ_{KKT} is strictly larger than the set Σ_{SOC} , then obviously it is easier for a local optimal solution of (BLPP) to be a solution to (SOCP_μ) than to (KKTCP_μ) . This means that the partial calmness for the combined program with second-order optimality conditions is more likely to hold than the one for the combined program with first-order optimality conditions.

For the bilevel programming problem where the lower level is unconstrained, when we add the second-order optimality condition, the partially penalized problem becomes a nonlinear semidefinite programming problem. For the general (BLPP) where the lower level problem is a constrained optimization problem, there are several different second-order optimality conditions. We propose the corresponding combined program with each second-order optimality condition. Similar to the KKT approach where one minimizes over the original variables and the multipliers, we also propose some relaxed version of these second-order combined programs where multipliers are used as variables.

Another difficulty of the value function or the combined approach is that the value function is usually nonsmooth and implicit. Since the set of second-order stationary points Σ_{SOC} is in general smaller than the set of first-order stationary points Σ_{KKT} , it is more likely that the set of second-order stationary points coincides with $\text{gph } S$. In particular, if it happens that $\Sigma_{\text{SOC}} = \text{gph } S$, then the value function constraint $f(x, y) - V(x) \leq 0$ can be removed from (SOCP) and so the partial calmness of the problem (SOCP) holds with penalty parameter $\mu = 0$. Consequently, the resulting necessary optimality condition is much easier to obtain and does not involve the value function. This is another advantage of using the combined program with second-order optimality conditions.

Outline. The remaining part of the paper is organized as follows. In Section 2, we gather some preliminaries and preliminary results that will be used later. An illustrative example will be given in Section 3. In Section 4, we introduce the combined problems with different kinds of second-order optimality conditions and the relaxed problems, discuss the partial calmness conditions and optimality conditions, and also give some examples. Conclusions are given in Section 5.

Symbols and Notations. Our notation is basically standard. For a matrix A , we denote by A^T its *transpose*. The *inner product* of two vectors x, y is denoted by $x^T y$ or $\langle x, y \rangle$. We denote by $\mathbb{S}^m$ the set of symmetric $m \times m$ matrices equipped with the inner product $\langle A, B \rangle := \text{tr}(AB)$, $A, B \in \mathbb{S}^m$, where $\text{tr}(A)$ denotes the *trace* of the matrix A . The notation $A \succeq 0$ ($A \preceq 0$) means that A is a symmetric positive (negative) semidefinite matrix. The set of symmetric positive semidefinite $m \times m$ matrices is denoted by $\mathbb{S}_+^m$. For $z \in \mathbb{R}^d$ and $\Omega \subseteq \mathbb{R}^d$, we denote by $\text{dist}(z; \Omega)$ the *distance* from z to Ω . For a differentiable mapping $h : \mathbb{R}^d \rightarrow \mathbb{R}^l$, we denote its *Jacobian matrix* by $Dh(z) \in \mathbb{R}^{l \times d}$ and by $\nabla h(z) := Dh(z)^T$ the transpose of the Jacobian, and in the case where $l = 1$, its *gradient vector* by $\nabla h(z)$ and its *Hessian* by $\nabla^2 h(z)$. For a nonsmooth function $g : \mathbb{R}^d \rightarrow \mathbb{R}$, we denote the *Clarke generalized gradient* of g at z by $\partial^c g(z)$. For a set-valued mapping $\Gamma : K_1 \rightrightarrows K_2$ where K_1 and K_2 are Euclidean spaces, we denote the *domain* of Γ with $\text{dom } \Gamma := \{x \in K_1 | \Gamma(x) \neq \emptyset\}$.

2. Preliminaries and preliminary results

In this section, we review and obtain some results that are needed in this paper.

2.1. Second-order optimality conditions for the lower level program

In this subsection, we review some results on second-order optimality conditions for the lower level program of (BLPP).

We first recall some second-order optimality conditions for the following nonlinear optimization problem:

$$(\text{NLP}) \quad \min_t f(t) \text{ s.t. } g(t) \leq 0, \quad h(t) = 0,$$

where $f : \mathbb{R}^m \rightarrow \mathbb{R}$, $g : \mathbb{R}^m \rightarrow \mathbb{R}^{p_1}$, $h : \mathbb{R}^m \rightarrow \mathbb{R}^{p_2}$ are twice continuously differentiable. Denote the *Lagrangian function* for (NLP) by

$$L(t, u, v) := f(t) + \sum_{i=1}^{p_1} u_i g_i(t) + \sum_{i=1}^{p_2} v_i h_i(t), \text{ for } (t, u, v) \in \mathbb{R}^m \times \mathbb{R}_+^{p_1} \times \mathbb{R}^{p_2},$$

and the generalized Lagrangian function for (NLP) by

$$\mathcal{L}_0(t, u_0, u, v) := u_0 f(t) + \sum_{i=1}^{p_1} u_i g_i(t) + \sum_{i=1}^{p_2} v_i h_i(t), \text{ for } (t, u_0, u, v) \in \mathbb{R}^m \times \mathbb{R}_+ \times \mathbb{R}_+^{p_1} \times \mathbb{R}^{p_2}.$$

Given a feasible point t of problem (NLP), we denote the set of KKT multipliers at t as follows:

$$M^1(t) := \left\{ (u, v) \in \mathbb{R}^{p_1} \times \mathbb{R}^{p_2} : \nabla_t L(t, u, v) = 0, u \geq 0, \sum_{i=1}^{p_1} u_i g_i(t) = 0 \right\}.$$

We define the *critical cone* at t as follows:

$$\mathcal{C}(t) := \left\{ d \in \mathbb{R}^m : Df(t)d \leq 0, Dg_j(t)d \leq 0, \forall j \in J_0(t), Dh_i(t)d = 0, i = 1, \dots, p_2 \right\},$$

where $J_0(t) := \{j : g_j(t) = 0\}$ denotes the set of indices of active inequalities at t . When $(u, v) \in M^1(t)$, by using the KKT condition, the critical cone can be written as

$$\mathcal{C}(t) = \left\{ d : \begin{array}{l} Dg_j(t)d = 0 \text{ if } u_j > 0, Dg_j(t)d \leq 0 \text{ if } u_j = 0, \forall j \in J_0(t), \\ Dh_i(t)d = 0, i = 1, \dots, p_2 \end{array} \right\}. \quad (2)$$

Another important set is the critical subspace given by

$$\mathcal{S}(t) := \left\{ d \in \mathbb{R}^m : Dg_j(t)d = 0, \forall j \in J_0(t), Dh_i(t)d = 0, i = 1, \dots, p_2 \right\}. \quad (3)$$

Note that when $M^1(t) \neq \emptyset$, the critical subspace $\mathcal{S}(t)$ is the lineality space of the critical cone $\mathcal{C}(t)$ (i.e., the largest linear space contained in $\mathcal{C}(t)$) and then $\mathcal{S}(t) = \mathcal{C}(t) \cap (-\mathcal{C}(t))$. If the strict complementarity holds, i.e., $u_j > 0, \forall j \in J_0(t)$, we have $\mathcal{S}(t) = \mathcal{C}(t)$.

Now we review some classical second-order conditions.

Definition 2.1. Let t be a feasible point of problem (NLP). If $M^1(t) \neq \emptyset$, we say that

- (i) the *basic second-order optimality condition* (BSOC) holds at t , if $\forall d \in \mathcal{C}(t)$, there exists $(u, v) \in M^1(t)$ such that $d^T \nabla_{tt}^2 L(t, u, v) d \geq 0$;
- (ii) the *weak second-order optimality condition* (WSOC) holds at t , if there exists $(u, v) \in M^1(t)$ such that $d^T \nabla_{tt}^2 L(t, u, v) d \geq 0, \forall d \in \mathcal{S}(t)$;
- (iii) the *strong second-order optimality condition* (SSOC) holds at t , if there exists $(u, v) \in M^1(t)$ such that $d^T \nabla_{tt}^2 L(t, u, v) d \geq 0, \forall d \in \mathcal{C}(t)$.

Note that when the linear independence constraint qualification (LICQ) holds at a feasible point t , there is a unique multiplier, i.e., the set $M^1(t)$ is a singleton. Hence, BSOC is equivalent to SSOC under LICQ. All KKT type second-order optimality conditions such as BSOC, WSOC and SSOC hold at (local) minimizers only if certain constraint qualifications are valid. BSOC requires a fairly weak constraint qualification. In classical results, MFCQ was required for BSOC to hold, c.f., [9, Proposition 5.48]. Recently under a much weaker constraint qualification called the directional metrical subregularity condition [24, Theorem 5.2], it was shown that BSOC holds. However, WSOC and SSOC require much stronger constraint qualifications.

It is known that SSOC (and hence WSOC) holds under the relaxed constant-rank constraint qualification (RCRCQ) [39, Theorem 6] and it is known that MFCQ was shown to be not enough for SSOC to hold [3, page 1350]. Another condition called the critical regularity condition, which is not stronger than RCRCQ, is enough to give SSOC at local minimizers [38, Theorem 2.1]. Recently, it was shown that WSOC holds under MFCQ plus the weak constant rank property [6, Theorem 3.1]. Even when no constraint qualification is assumed, a Fritz John second-order optimality condition (FJSOC) always holds at a local minimizer.

Theorem 2.2. [9, Proposition 5.48] *Suppose t is a local minimizer of (NLP). Then, for all $d \in \mathcal{C}(t)$, there is a Fritz John multiplier (u_0, u, v) such that*

$$d^T \nabla_{tt}^2 \mathcal{L}_0(t, u_0, u, v) d \geq 0.$$

Since we will use the above concepts for the lower level program of BLPPs frequently, we give the following notation. For fixed upper variable x of (BLPP), we denote the Lagrangian function and the generalized Lagrangian function for the lower level program by $L(y, u; x)$ and $\mathcal{L}_0(y, u_0, u; x)$, respectively. For any $y \in S(x)$, we denote the set of KKT multipliers for the lower level program $P(x)$ at y by $M^1(y; x)$. For any $u \in M^1(y; x)$, we call (y, u) a KKT pair of program $P(x)$. We use $\mathcal{C}(y; x)$ and $\mathcal{S}(y; x)$ to denote the critical cone and the critical subspace at y for fixed x , respectively.

Since it is difficult to deal with the set of indices of active inequalities in the definition of the critical cone, we introduce slack variables $z := (z_1, \dots, z_p)^T \in \mathbb{R}^p$ for the lower level program, and obtain

$$\tilde{P}(x) \quad \min_{y, z} f(x, y) \quad \text{s.t.} \quad g(x, y) + z^2 = 0.$$

Here, $z^2 := (z_1^2, \dots, z_p^2)^T$. The above problem is equivalent to $P(x)$ in the following sense. For fixed x , if y^* is a global (local) optimal solution of $P(x)$, then there exists z^* such that (y^*, z^*) is a global (local) optimal solution of $\tilde{P}(x)$. Conversely, if (y^*, z^*) is a global (local) optimal solution of $\tilde{P}(x)$, then y^* is a global (local) optimal solution of $P(x)$.

Let (y, z) be a feasible point of problem $\tilde{P}(x)$. By definition, we say that u is a multiplier and (y, z, u) is a KKT triple of problem $\tilde{P}(x)$ provided that

$$\nabla_{(y, z)} L(y, z, u; x) = 0,$$

where $L(y, z, u; x) := f(x, y) + \sum_{i=1}^p u_i [g_i(x, y) + z_i^2]$. That is,

$$\nabla_y f(x, y) + \sum_{i=1}^p u_i \nabla_y g_i(x, y) = 0, \quad u_i z_i = 0, \quad g_i(x, y) + z_i^2 = 0, \quad i = 1, \dots, p.$$

Note that, different from the KKT multipliers in $M^1(y; x)$, the multipliers u_i above are not necessarily nonnegative.

Since problem $\tilde{P}(x)$ has only equality constraints, if the KKT condition holds, then the critical cone and the critical subspace of problem $\tilde{P}(x)$ are equal and given by

$$\mathcal{C}(y, z; x) = \mathcal{S}(y, z; x) := \left\{ (d, \nu) \in \mathbb{R}^m \times \mathbb{R}^p : D_y g_i(x, y) d + 2z_i \nu_i = 0, \forall i \right\}. \quad (4)$$

As an optimization problem with equality constraints, WSOC and SSOC for problem $\tilde{P}(x)$ coincide and hence we call it SOC. Let (y, z, u) be a KKT triple of problem $\tilde{P}(x)$. We say that SOC holds at (y, z, u) if

$$(d, \nu)^T \nabla_{(y,z)}^2 L(y, z, u; x)(d, \nu) \geq 0, \quad \forall (d, \nu) \in \mathcal{C}(y, z; x). \quad (5)$$

Note that
$$\nabla_{(y,z)}^2 L(y, z, u; x) = \begin{pmatrix} \nabla_{yy}^2 L(y, u; x) & 0 \\ 0 & 2\text{diag}(u) \end{pmatrix},$$

where $\text{diag}(u)$ denotes the $p \times p$ diagonal matrix with the elements of vector u on the main diagonal. Thus

$$(d, \nu)^T \nabla_{(y,z)}^2 L(y, z, u; x)(d, \nu) = d^T \nabla_{yy}^2 L(y, u; x)d + 2 \sum_{i=1}^p u_i \nu_i^2. \quad (6)$$

It is a simple matter to show that if (y^*, u) is a KKT pair of $P(x)$ then there exists z^* such that (y^*, z^*, u) is a KKT triple of $\tilde{P}(x)$. Moreover suppose that (y^*, u) satisfies WSOC for $P(x)$. Then

$$d^T \nabla_{yy}^2 L(y, u; x)d \geq 0 \quad \forall d \in \mathcal{S}(y; x).$$

By (4) and (3), we have $(d, \nu) \in \mathcal{S}(y, z; x) \implies d \in \mathcal{S}(y; x)$.

Since $u \geq 0$ for KKT pair (y^*, u) of $P(x)$, by (6), the following result is valid.

Proposition 2.3. *Let (y^*, u) be a KKT pair of $P(x)$. Then there exists z^* such that (y^*, z^*, u) is a KKT triple of $\tilde{P}(x)$. Furthermore, if (y^*, u) satisfies WSOC for $P(x)$, then (y^*, z^*, u) satisfies SOC (5) for $\tilde{P}(x)$.*

But the converse is not always true, that is, even if (y^*, z^*, u) is a KKT triple of $\tilde{P}(x)$, (y^*, u) is not necessarily a KKT pair of $P(x)$. In fact, the condition $u \geq 0$, concerning the sign of the multiplier, may not hold. For a counterexample, we refer the reader to [21, Example 3.2]. Under the second-order sufficient conditions and some constraint qualification, it has been proved that KKT points of the original $P(x)$ and the reformulated $\tilde{P}(x)$ problems are essentially equivalent, cf. [21, Proposition 3.6]. Moreover in the final remarks of [21], the authors asked if there are other conditions which guarantee equivalence of the KKT points. In the next result, we answer this question by showing that the converse holds under the second-order necessary condition and hence improve the result of [21, Proposition 3.6].

Proposition 2.4. *Let (y^*, z^*, u^*) be a KKT triple of $\tilde{P}(x)$. Assume that (y^*, z^*, u^*) satisfies SOC (5). Then $u_i^* \geq 0$ for all $i = 1, \dots, p$. Hence (y^*, u^*) is a KKT pair of $P(x)$ satisfying WSOC.*

Proof. First, since (y^*, z^*, u^*) is a KKT triple of $\tilde{P}(x)$, we have $u_i^* z_i^* = 0$ for all $i = 1, \dots, p$. Thus $u_i^* g_i(x, y^*) = -u_i^* (z_i^*)^2 = 0$, which implies that $u_i^* = 0$ if $z_i^* \neq 0$ or equivalently $g_i(x, y^*) \neq 0$.

Now we consider the index j such that $g_j(x, y^*) = 0 = z_j^*$. Let us prove that in this case $u_j^* \geq 0$. Taking $d^* = 0$, $\nu_i^* = 0$ for $i \neq j$ and $\nu_j^* = 1$, by the formula for

$\mathcal{S}(y^*, z^*; x)$ in (4), we have $(d^*, \nu^*) \in \mathcal{S}(y^*, z^*; x)$. By (6), we have

$$0 \leq (d^*, \nu^*)^T \nabla_{(y,z)}^2 L(y^*, z^*, u^*; x)(d^*, \nu^*) = 2u_j^*,$$

which implies that $u_j^* \geq 0$. Hence, we conclude that (y^*, u^*) is a KKT pair of $P(x)$.

Next we show that (y^*, u^*) satisfies WSOC. For every $d \in \mathcal{S}(y^*; x)$, we obviously have $D_y g_j(x, y^*)d = 0$ for all $j \in J_0(y^*; x)$. For $i \notin J_0(y^*; x)$, i.e., $z_i^* \neq 0$, we take $\nu_i = -D_y g_i(x, y^*)d/(2z_i^*)$. For all $j \in J_0(y^*; x)$, take $\nu_j = 0$. Then it is obvious that $(d, \nu) \in \mathcal{S}(y^*, z^*; x)$. Hence by (6)

$$\begin{aligned} 0 &\leq (d, \nu)^T \nabla_{(y,z)}^2 L(y^*, z^*, u^*; x)(d, \nu) = d^T \nabla_{yy}^2 L(y^*, u^*; x)d + 2 \sum_{i=1}^p u_i^* \nu_i^2 \\ &= d^T \nabla_{yy}^2 L(y^*, u^*; x)d \end{aligned}$$

since $u_i^* = 0$ for all $i \notin J_0(y^*; x)$ and $\nu_j = 0$ for all $j \in J_0(y^*; x)$. Therefore, (y^*, u^*) satisfies WSOC. $\square$

2.2. Lipschitz continuity of the value function and the upper estimate of the Clarke subdifferential of the value function

For convenience, we quote the original result obtained by Gauvin-Dubeau in [23] below. For results under weaker assumptions and sharper upper estimates, the reader is referred to [28, Corollary 4.8] and [48, Proposition 2]. Note that under extra assumptions, the convex hull operation in the formula below can be removed and a tighter bound for the subdifferential can be obtained; see e.g. [48, Proposition 1] for the case where the lower level program is linear, and [42, Section 5] for the case where the solution map S is V -inner semicontinuous at the point of interest. Note that in the last case, the uniform boundedness of Y assumption can be removed.

Proposition 2.5. [23, Theorem 5.3] *Assume that the set-valued map Y is uniformly bounded around $\bar{x}$, i.e., there exists a neighborhood $U(\bar{x})$ of $\bar{x}$ such that $\cup_{x \in U(\bar{x})} Y(x)$ is bounded. Suppose that MFCQ holds at each $y \in S(\bar{x})$. Then the value function V is Lipschitz continuous near $\bar{x}$ and the Clarke subdifferential of V at $\bar{x}$ has the following upper estimate:*

$$\partial^c V(\bar{x}) \subseteq \text{co}\{\nabla_x f(\bar{x}, y') + \nabla_x g(\bar{x}, y')^T u' : y' \in S(\bar{x}), u' \in M^1(y'; \bar{x})\},$$

where $\text{co}C$ denotes the convex hull of the set C .

2.3. Constraint qualifications and optimality conditions for the combined problem

As discussed in the introduction, there are various reformulations of (BLPP). To simplify the discussion, in this section we consider the following general combined problem:

$$\begin{aligned} (\text{GCP}) \quad & \min_{x, y, u, w} F(x, y) \\ \text{s.t.} \quad & f(x, y) - V(x) \leq 0, \quad g(x, y) \leq 0, \quad u \geq 0, \quad u^T g(x, y) = 0, \\ & H_1(x, y, u) := \nabla_y f(x, y) + \nabla_y g(x, y)^T u = 0, \quad H_2(x, y, u, w) \in C, \end{aligned}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $u \in \mathbb{R}^p$, $w \in \mathbb{R}^l$ and the mappings $F, f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^p$, $H_2 : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^l \rightarrow K$ are continuously differentiable, K is a Euclidean space, and C is a nonempty convex subset of K . Here the constraint $H_2(x, y, u, w) \in C$ represents a constraint which comes from a second-order optimality condition. Note that for simplicity we have omitted the upper level constraint in this section.

We define the partial calmness for (GCP) as follows.

Definition 2.6. (Partial calmness for (GCP)) Let $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ be a local solution of (GCP). We say that (GCP) is *partially calm* at $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ if there exists $\mu \geq 0$ such that $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ is a local solution of the following partially penalized problem:

$$\begin{aligned} (\text{GCP}_\mu) \quad & \min_{x, y, u, w} F(x, y) + \mu(f(x, y) - V(x)) \\ & \text{s.t. } g(x, y) \leq 0, \quad u \geq 0, \quad u^T g(x, y) = 0, \\ & H_1(x, y, u) = 0, H_2(x, y, u, w) \in C. \end{aligned}$$

Since the second-order condition $H_2(x, y, u, w) \in C$ is redundant, the feasible region of (GCP_μ) must include in the feasible region of the corresponding partially penalized problem (CP_μ) , and thus the partial calmness condition for problem (GCP) is easier to hold than that for problem (CP).

Proposition 2.7. Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution to problem (CP). Suppose that (CP) is partially calm at $(\bar{x}, \bar{y}, \bar{u})$ and there is $\bar{w}$ such that $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ is a local optimal solution for (GCP), then (GCP) is also partially calm at $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$.

We now study the constraint qualification and optimality condition for problem (GCP). If the value function is Lipschitz continuous and $K = \mathbb{R}^s$, then problem (GCP) is an MPEC with Lipschitz continuous problem data. Due to the value function constraint, the nonsmooth MFCQ fails to hold at any feasible solution of the above problem [52, Proposition 3.2].

Recall that in MPEC literature, one usually defines a Mordukhovich (M-) or a Strong (S-) stationarity condition based on whether the multipliers are taken from the limiting normal cone or the regular normal cone of the complementarity set respectively. Similar to [47, Definition 4.2], we define M-/S- stationarity condition based on the value function for (GCP). Given a feasible vector $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ of problem (GCP), we define the following index sets:

$$\begin{aligned} I_g &= I_g(\bar{x}, \bar{y}, \bar{u}, \bar{w}) := \{j : g_j(\bar{x}, \bar{y}) = 0, \bar{u}_j > 0\}, \\ I_u &= I_u(\bar{x}, \bar{y}, \bar{u}, \bar{w}) := \{j : g_j(\bar{x}, \bar{y}) < 0, \bar{u}_j = 0\}, \\ I_0 &= I_0(\bar{x}, \bar{y}, \bar{u}, \bar{w}) := \{j : g_j(\bar{x}, \bar{y}) = 0, \bar{u}_j = 0\}. \end{aligned}$$

Definition 2.8. (Stationary conditions for (GCP) based on the value function)

Let $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ be a feasible solution to (GCP).

- (i) We say that $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ is an M-stationary point based on the value function if there exist $\mu \geq 0$, $\lambda^g \in \mathbb{R}^p$, $\lambda^u \in \mathbb{R}^p$ and $\lambda_1^H \in \mathbb{R}^m$, $\lambda_2^H \in \mathbb{R}^s$ such that

$$\begin{aligned} 0 \in & [\nabla F(\bar{x}, \bar{y}) + \mu(\nabla f(\bar{x}, \bar{y}) - \partial^c V(\bar{x}) \times \{0\}) + \nabla g(\bar{x}, \bar{y})^T \lambda^g] \times \{(0, 0)\} \\ & - (0, 0, \lambda^u, 0) + \nabla H_1(\bar{x}, \bar{y}, \bar{u})^T \lambda_1^H + \nabla H_2(\bar{x}, \bar{y}, \bar{u}, \bar{w})^T \lambda_2^H, \end{aligned} \quad (7)$$

$$\lambda_j^g = 0, \forall j \in I_u, \quad \lambda_j^u = 0, \forall j \in I_g, \quad \lambda_2^H \in N_C(H_2(\bar{x}, \bar{y}, \bar{u}, \bar{w})), \quad (8)$$

and either $\lambda_j^g > 0, \lambda_j^u > 0$, or $\lambda_j^g \lambda_j^u = 0, \forall j \in I_0$,

where N_C denotes the normal cone to the convex set C .

- (ii) We say that $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ is an *S-stationary point* based on the value function if there exist $\mu \geq 0, \lambda^g \in \mathbb{R}^p, \lambda^u \in \mathbb{R}^p$ and $\lambda_1^H \in \mathbb{R}^m, \lambda_2^H \in \mathbb{R}^s$ such that (7)–(8) and the following condition hold:

$$\lambda_j^g \geq 0, \lambda_j^u \geq 0, \forall j \in I_0.$$

By Definition 2.8, an M-/S- stationary point of problem (CP) must correspond to an M-/S- stationary point of problem (GCP), but the converse is not true since the multiplier λ_2^H can be nonzero. When the multiplier λ_2^H is nonzero, the combined approach of using only the first-order condition fails but the one using the second-order condition is useful.

To obtain M-stationary conditions, we reformulate problem (GCP) equivalently as the following optimization problem:

$$\begin{aligned} \min_{x,y,u,w} \quad & F(x, y) \\ \text{s.t.} \quad & f(x, y) - V(x) \leq 0, \quad (g(x, y), -u) \in \Omega_{\text{CS}}^p, \\ & H_1(x, y, u) = 0, \quad H_2(x, y, u, w) \in C, \end{aligned} \quad (9)$$

where $\Omega_{\text{CS}}^p := \{(a, b) \in \mathbb{R}^p \times \mathbb{R}^p : a \leq 0, b \leq 0, \langle a, b \rangle = 0\}$ is the negative complementarity set. Denote the set of feasible solutions for problem (9) by $\mathcal{F}$ and the perturbed feasible map by

$$\mathcal{F}(r_1, r_2, r_3, P_1, P_2) := \left\{ (x, y, u, w) : \begin{aligned} & f(x, y) - V(x) + r_1 \leq 0, \\ & (g(x, y) - r_2, -u + r_3) \in \Omega_{\text{CS}}^p, \\ & H_1(x, y, u) + P_1 = 0, \\ & H_2(x, y, u, w) + P_2 \in C \end{aligned} \right\}. \quad (10)$$

We now define the Clarke calmness for problem (GCP) as the one for its equivalent reformulation (9) as follows.

Definition 2.9. (Clarke calmness for problem (GCP)) Let $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ be a local optimal solution of (GCP). We say that (GCP) is *Clarke calm* at $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ if there exist $\epsilon > 0$ and $\mu \geq 0$ such that, for all $(r_1, r_2, r_3, P_1, P_2)$ in $B(0, \epsilon)$, for all $(x, y, u, w) \in B((\bar{x}, \bar{y}, \bar{u}, \bar{w}), \epsilon) \cap \mathcal{F}(r_1, r_2, r_3, P_1, P_2)$, one has

$$F(x, y) - F(\bar{x}, \bar{y}) + \mu \|(r_1, r_2, r_3, P_1, P_2)\| \geq 0,$$

where $B(z, \epsilon)$ denotes the open ball centered at z with radius ϵ .

Similar to Burke [11, Theorem 1.1], the Clarke calmness defined in Definition 2.9 is equivalent to the exact penalization, i.e., (GCP) is Clarke calm at $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ if and only if there exists $\mu \geq 0$ such that it is a local solution of the penalized problem:

$$\begin{aligned} \min_{x,y,u,w} \quad & F(x, y) + \mu (f(x, y) - V(x) + \sum_{i=1}^m |H_1^i(x, y, u)| \\ & + \text{dist}((g(x, y), -u); \Omega_{\text{CS}}^p) + \text{dist}(H_2(x, y, u, w); C)). \end{aligned}$$

It is well-known that the calmness of the perturbed feasible map (10) or equivalently the existence of a local error bound for the feasible region $\mathcal{F}$ is a sufficient condition for Clarke calmness; see e.g. [18, Proposition 2.2]. Moreover many classical constraint qualifications can be used to guarantee the Clarke calmness at a local minimizer; see e.g. [18, Proposition 2.3].

Similar to the equivalence between the Clarke calmness and exact penalization, it was pointed out in [52, Proposition 3.3] that there is an equivalence between the partial calmness and partial exact penalization. The Clarke calmness condition is in general stronger than the partial calmness. The partial calmness condition plus a usual constraint qualification for the partially penalized problem implies the Clarke calmness condition [52, Theorem 3.1]. One may derive sufficient condition for the calmness for the general combined program using the results on the relaxed constant positive linear dependence constraint qualification (RCPLD) [36, Theorem 3.2], [46, Theorem 3.2].

We can now state the optimality conditions for the general combined program below. In fact, one can also apply the directional calmness and optimality conditions in [4, Theorem 3.1], which was developed using the directional approach to variational analysis in [7], to the general combined problem. To obtain S-stationary condition, we introduce the following constraint qualification.

Definition 2.10. (MPEC LICQ) Let $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ be a feasible solution to problem (GCP_μ) . We say that MPEC LICQ holds at $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ if the following non-degeneracy condition holds:

$$\begin{cases} 0 = \sum_{j \in J_0(\bar{x}, \bar{y})} \lambda_j^g \nabla g_j(\bar{x}, \bar{y}) \times \{(0, 0)\} - \left\{ \left(0, 0, \sum_{j \in I_u \cup I_0} \lambda_j^u e_j, 0 \right) \right\} \\ \quad + \nabla H_1(\bar{x}, \bar{y}, \bar{u})^T \lambda_1^H \times \{0\} + \nabla H_2(\bar{x}, \bar{y}, \bar{u}, \bar{w})^T \lambda_2^H, \\ \lambda_2^H \in \text{span } N_C(H_2(\bar{x}, \bar{y}, \bar{u}, \bar{w})), \\ \Rightarrow (\lambda^g, \lambda^u, \lambda_1^H, \lambda_2^H) = (0, 0, 0, 0), \end{cases}$$

where $e_j \in \mathbb{R}^p$ denotes the vector whose j -th component is 1, and others are all zero, and $\text{span}(\Pi)$ denotes the affine hull of the set Π .

Theorem 2.11. Let $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ be a local optimal solution to (GCP). Suppose that (GCP) is Clarke calm at $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$, then $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ is an M-stationary point based on the value function. If (GCP) is partially calm at $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$, either $\mu = 0$ or the value function is smooth, and MPEC LICQ holds, then $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ is an S-stationary point based on the value function.

Proof. Since (GCP) is equivalent to (9), by [18, Theorem 2.1] and the expression for the limiting normal cone of the complementarity set, we get the result for the M-stationary point. Similarly, by Corollary 6 in [27] and the expression for the regular normal cone of the complementarity set, we get the result for the S-stationary point. Alternatively, if the set C is a polyhedral set, then we can also use the [35, Theorem 3.8] to derive the desired result. $\square$

Remark 2.12. By definition, it is easy to see that if (GCP) is partially calm and problem (GCP_μ) is Clarke calm at $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$, then the Clarke calmness for (GCP) holds.

3. An illustrative example

To illustrate the difficulties of BLPPs and our approach, we consider the following example for which all known approaches fail.

Example 3.1.

$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2 \text{ s.t. } y \in S(x) := \arg \min_y \left\{ \frac{1}{4}y^4 - \frac{1}{2}xy^2 : y \in \mathbb{R} \right\}. \quad (11)$$

The first-order necessary condition for optimality of the lower level objective function with respect to y is $y^3 - xy = 0$, which is equivalent to saying that $y = 0$ or $x = y^2$. Its graph is shown in Figure 3.1.

Since the objective of the lower level program is not convex in lower level variable y , for each fixed x , not all corresponding y 's lying on the curve are global optimal solutions of the lower level program. The true global optimal solutions for the lower level problem are shown in Figure 3.2. It is easy to see that

$$S(x) = \begin{cases} \{\pm\sqrt{x}\} & \text{if } x > 0, \\ \{0\} & \text{if } x \leq 0, \end{cases} \quad V(x) = \begin{cases} -\frac{1}{4}x^2 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

and $(\bar{x}, \bar{y}) = (0, 0)$ is the unique global optimal solution.

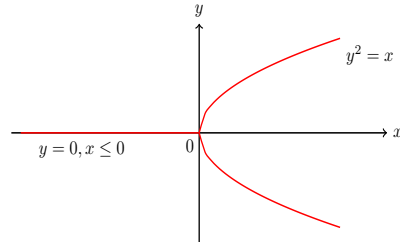
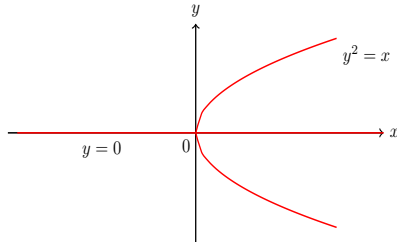


Figure 3.1: Feasible set of problem (12) Figure 3.2: Feasible set of problem (15) (also the graph of $S(\cdot)$)

Now we claim that the partial calmness for (CP) does not hold at $(0, 0)$. Indeed, the associated partially penalized problem is given by

$$\min_{x,y} \left\{ F_\mu(x, y) := \left(x - \frac{1}{2}\right)^2 + y^2 + \mu \left(\frac{1}{4}y^4 - \frac{1}{2}xy^2 - V(x)\right) : y^3 - xy = 0 \right\}. \quad (12)$$

Take any $\mu \geq 0$. For the objective values, we find

$$F_\mu\left(\frac{1}{k}, 0\right) = k^{-2} - k^{-1} + \frac{\mu}{4}k^{-2} + \frac{1}{4}, \quad F_\mu(0, 0) = \frac{1}{4}. \quad (13)$$

Thus, for $k > 1 + \mu/4$, $F_\mu\left(\frac{1}{k}, 0\right) < F_\mu(0, 0)$ holds, and this shows that $(\bar{x}, \bar{y}) = (0, 0)$ is not a local minimizer of the associated partially penalized problem (12). Hence the partial calmness for (CP) does not hold at $(0, 0)$. Moreover since (CP) is a standard nonlinear program, it is easy to check that the KKT condition does not hold at $(0, 0)$.

To explain our new approach, we now consider the following optimization problem in which we add the first and the second-order conditions to the value function reformulation of problem (11):

$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2 \quad \text{s.t.} \quad f(x,y) - V(x) \leq 0, \quad y^3 - xy = 0, \quad 3y^2 - x \geq 0. \quad (14)$$

Since both the first and the second-order conditions for the lower level program hold at $y \in S(x)$ without any further assumption, the constraints $y^3 - xy = 0$ and $3y^2 - x \geq 0$ are redundant. Hence $(\bar{x}, \bar{y}) = (0, 0)$ is still the optimal solution to the above problem.

From the graph in Figure 3.2, we can see that any point (x, y) satisfying the first and the second-order conditions together lies in the graph of the solution mapping $S(\cdot)$. This means that the value function constraint can be removed and hence $(0, 0)$ is a (local) minimizer of the following partially penalized problem with $\mu = 0$:

$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2 + \mu(f(x,y) - V(x)) \quad \text{s.t.} \quad y^3 - xy = 0, \quad 3y^2 - x \geq 0. \quad (15)$$

Problem (15) is a one-level optimization problem. Furthermore, it is easy to check that its KKT condition holds at $(0, 0)$.

Next we present a geometric explanation for Example 3.1.

For Example 3.1, the partial calmness for (CP) at $(\bar{x}, \bar{y}) = (0, 0)$ means that for some $\mu \geq 0$, $(\bar{x}, \bar{y})$ is still the optimal solution of the associated partially penalized problem (12), whose feasible set is given by Figure 3.1. But by (13), this is violated by taking points $\{(\frac{1}{k}, 0)\}_{k=1}^{\infty}$ on the line $\{(x, y) : x > 0, y = 0\}$ in the feasible set.

To fix the above issue, we add the second-order necessary optimality condition of the lower level program in the combined problem (14). The advantage of using the second-order necessary optimality condition is that the feasible set of the new associated partially penalized problem (15) ruled out all of the points on the line $\{(x, y) : x > 0, y = 0\}$ which are actually local maxima for the lower level objective function with $x > 0$ (see Figure 3.2).

4. Combined with second-order optimality conditions

A natural idea that comes from Example 3.1 is to add the second-order necessary optimality conditions of the lower level program in the combined problem. In this section, we consider combined problems with different kinds of second-order optimality conditions.

4.1. Unconstrained case

For the unconstrained bilevel programming problem

$$(\text{UBLPP}) \quad \min_{x,y} F(x,y) \quad \text{s.t.} \quad y \in \arg \min_y f(x,y), \quad G(x,y) \leq 0,$$

we propose the following combined program using the second-order necessary optimality condition:

$$\begin{aligned}
(\text{CPSOC}) \quad & \min_{x,y} F(x,y) \quad \text{s.t.} \quad f(x,y) - V(x) \leq 0, \\
& \nabla_y f(x,y) = 0, \quad \nabla_{yy}^2 f(x,y) \in \mathbb{S}_+^m, \quad G(x,y) \leq 0.
\end{aligned}$$

We denote the corresponding partially penalized problem for (CPSOC) (as in Definition 2.6) by (CPSOC_μ) . The problem (CPSOC_μ) is a nonlinear semidefinite optimization problem. To derive an optimality condition for it, we may apply some constraint qualification, e.g., Robinson's constraint qualification (or a generalized MFCQ) of nonlinear semidefinite optimization problems.

Theorem 4.1. *Let $(\bar{x}, \bar{y})$ be a local optimal solution to (UBLPP). Suppose that the partial calmness for (UBLPP) holds with either $\mu = 0$ or with $\mu > 0$ and the value function V is Lipschitz continuous near $\bar{x}$. Then under some constraint qualification, there exist $\Omega \in \mathbb{S}_+^m$, $\mu \geq 0$, $\alpha \in \mathbb{R}^m$, and $\beta \in \mathbb{R}_+^q$ such that*

$$\begin{aligned}
0 \in & \nabla F(\bar{x}, \bar{y}) + \mu(\nabla f(\bar{x}, \bar{y}) - \partial^c V(\bar{x}) \times \{0\}) + \nabla(\nabla_y f)(\bar{x}, \bar{y})^T \alpha \\
& - D\nabla_{yy}^2 f(\bar{x}, \bar{y})^* \Omega + \nabla G(\bar{x}, \bar{y})^T \beta, \\
& \langle \nabla_{yy}^2 f(\bar{x}, \bar{y}), \Omega \rangle = 0, \quad \beta^T G(\bar{x}, \bar{y}) = 0,
\end{aligned}$$

where $D\nabla_{yy}^2 f(\bar{x}, \bar{y})^* \Omega := \left(\left\langle \frac{\partial}{\partial x_1} \nabla_{yy}^2 f(\bar{x}, \bar{y}), \Omega \right\rangle, \dots, \left\langle \frac{\partial}{\partial x_m} \nabla_{yy}^2 f(\bar{x}, \bar{y}), \Omega \right\rangle \right)^T$.

4.2. Constrained case

In the constrained case, as we reviewed in Section 2, there are four kinds of second-order optimality conditions: FJSOC, BSOC, SSOC, and WSOC.

4.2.1. Combined with the Fritz John second-order optimality condition

We say that $y \in Y(x)$ is an FJSOC-point if for all $d \in \mathcal{C}(y; x)$, there exists $(u_0, u) \neq 0$ such that

$$\begin{aligned}
& u_0 \nabla_y f(x, y) + \nabla_y g(x, y)^T u = 0, \\
& g(x, y) \leq 0, \quad (u_0, u) \geq 0, \quad \sum_{i=0}^p u_i = 1, \quad u^T g(x, y) = 0, \\
& d^T \nabla_{yy}^2 \mathcal{L}_0(y, u_0, u; x) d \geq 0.
\end{aligned} \tag{16}$$

By Theorem 2.2, if $y \in S(x)$ then y is an FJSOC-point for $P(x)$.

Now we define $\Sigma_{\text{FJSOC}} := \left\{ (x, y) \in \mathbb{R}^{n+m} : \exists (u_0, u) \neq 0 \text{ s.t. (16) hold.} \right\}$, and consider the following combined problem with FJSOC:

$$(\text{FJSOCP}) \quad \min_{x,y} F(x,y) \quad \text{s.t.} \quad f(x,y) - V(x) \leq 0, \quad (x,y) \in \Sigma_{\text{FJSOC}}, \quad G(x,y) \leq 0.$$

Since it is not easy dealing with the set of indices of active inequalities in the critical cone, we propose to use the following set to relax the critical cone:

$$\{d \in \mathbb{R}^m : D_y f(x, y)d \leq 0, \quad u_j D_y g_j(x, y)d \leq 0, \quad \forall j = 1, \dots, p\} \supseteq \mathcal{C}(y; x), \tag{17}$$

where (u_0, u) is an FJ-multiplier. Under the strict complementarity, “ $\supseteq$ ” becomes “ $=$ ” in the above relationship. Hence $y \in S(x)$ implies that there are (u_0, u, d) such that the following relaxed FJ system holds:

$$\begin{aligned} u_0 \nabla_y f(x, y) + \nabla_y g(x, y)^T u &= 0, \\ g(x, y) &\leq 0, \quad (u_0, u) \geq 0, \quad \sum_{i=0}^p u_i = 1, \quad u^T g(x, y) = 0, \\ d^T \nabla_{yy}^2 \mathcal{L}_0(y, u_0, u; x) d &\geq 0, \\ D_y f(x, y) d &\leq 0, \quad u_j D_y g_j(x, y) d \leq 0, \quad \forall j = 1, \dots, p. \end{aligned} \quad (18)$$

$$\text{Define} \quad K(x, y) := \{(u_0, u, d) \in \Xi(x, y) \mid u_0 \nabla_y f(x, y) + \nabla_y g(x, y)^T u = 0\}, \quad (19)$$

where

$$\Xi(x, y) := \left\{ (u_0, u, d) : \begin{aligned} &(u_0, u) \geq 0, \quad \sum_{i=0}^p u_i = 1, \quad u^T g(x, y) = 0, \\ &d^T \nabla_{yy}^2 [u_0 f(x, y) + \sum_{i=1}^p u_i g_i(x, y)] d \geq 0, \\ &D_y f(x, y) d \leq 0, \quad u_j D_y g_j(x, y) d \leq 0, \quad \forall j = 1, \dots, p \end{aligned} \right\}. \quad (20)$$

Then problem (FJSOCP) can be reformulated as the following problem equivalently.

$$\begin{aligned} \text{(FJSOCP-2)} \quad & \min_{x, y} F(x, y) \\ \text{s.t.} \quad & f(x, y) - V(x) \leq 0, \quad g(x, y) \leq 0, \quad G(x, y) \leq 0, \\ & 0 \in \bigcup_{(u_0, u, d) \in \Xi(x, y)} \{u_0 \nabla_y f(x, y) + \nabla_y g(x, y)^T u\}. \end{aligned}$$

Problem (FJSOCP-2) can be considered as an optimization problem with implicit variables as studied in recent paper [8]. Since problem (FJSOCP-2) is still not practical to solve, we consider its explicit version – the relaxed combined problem with FJ second-order condition:

$$\begin{aligned} \text{(R-FJSOCP)} \quad & \min_{x, y, u_0, u, d} F(x, y) \\ \text{s.t.} \quad & f(x, y) - V(x) \leq 0, \quad u_0 \nabla_y f(x, y) + \nabla_y g(x, y)^T u = 0, \quad g(x, y) \leq 0, \\ & (u_0, u) \geq 0, \quad \sum_{i=0}^p u_i = 1, \quad u^T g(x, y) = 0, \quad d^T \nabla_{yy}^2 [u_0 f(x, y) + \sum_{i=1}^p u_i g_i(x, y)] d \geq 0, \\ & D_y f(x, y) d \leq 0, \quad u_j D_y g_j(x, y) d \leq 0, \quad \forall j = 1, \dots, p, \quad G(x, y) \leq 0. \end{aligned}$$

Recall that a set-valued map K is inner semicompact at $(\bar{x}, \bar{y})$ with respect to $\text{dom} K$ if for each sequence $\{(x_k, y_k)\}_{k \in \mathbb{N}} \subseteq \text{dom} K$ such that $(x_k, y_k) \rightarrow (\bar{x}, \bar{y})$, there is a convergent sequence $\{(u_0^\zeta, u^\zeta, d^\zeta)\}_{\zeta \in \mathbb{N}}$ and a subsequence $\{(x_k^\zeta, y_k^\zeta)\}_{\zeta \in \mathbb{N}}$ such that $(u_0^\zeta, u^\zeta, d^\zeta) \in K(x_k^\zeta, y_k^\zeta)$ holds for all $\zeta \in \mathbb{N}$; see e.g. [8, p. 7]. Note that the mapping K defined in (19) is automatically inner semicompact since the set of FJ multipliers (u_0, u) are uniformly bounded, and the sequence d^ζ in the definition can

always be taken as zero. Hence by [8, Theorems 4.3 and 4.5] we have the following equivalence between the problem with implicit variables and its explicit form.

Proposition 4.2. *Let $(\bar{x}, \bar{y})$ be a local (global) optimal solution to (BLPP). Then for each $(u_0, u, d) \in K(\bar{x}, \bar{y})$, $(\bar{x}, \bar{y}, u_0, u, d)$ is a local (global) optimal solution of (R-FJSOCP). Conversely, let $(\bar{x}, \bar{y}, u_0, u, d)$ be a global optimal solution to problem (R-FJSOCP). Then $(\bar{x}, \bar{y})$ is a global solution of (BLPP). Moreover if for each $(u_0, u, d) \in K(\bar{x}, \bar{y})$, $(\bar{x}, \bar{y}, u_0, u, d)$ is a local optimal solution to (R-FJSOCP), then $(\bar{x}, \bar{y})$ is a local solution of (BLPP).*

As in Definition 2.6, we can define partial calmness for (FJSOCP) and partial calmness for (R-FJSOCP), and denote the corresponding partially penalized problems by (FJSOCP_μ) and (R-FJSOCP_μ) , respectively.

Different from the relation between the partial calmness condition for (CP_{FJ}) and the partial calmness condition for (CPFJ) in [30, Theorem 4.4], the partial calmness condition for (FJSOCP) could not imply the partial calmness condition for (R-FJSOCP) directly because the critical cone has been relaxed in (R-FJSOCP). But as we will show in Proposition 4.3, the partial calmness condition for (CP_{FJ}) implies the partial calmness condition for (R-FJSOCP). Note that this relation coincides with the result in Proposition 2.7. On the other hand, since $\Sigma_{\text{FJSOC}} \subseteq \Sigma_{\text{FJ}}$, it is immediate that

$$\text{partial calmness for } (\text{CP}_{FJ}) \implies \text{partial calmness for (FJSOCP)}, \quad (21)$$

where Σ_{FJ} denotes the set of points which satisfy the Fritz John condition.

In the following proposition, we show that the partial calmness for (R-FJSOCP) at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u}, \bar{d})$ with $\bar{d} \neq 0$ is not stronger than the one for (CPFJ) at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$. Hence when the critical cone $\mathcal{C}(\bar{y}; \bar{x}) \neq \{0\}$, one can always take a nonzero critical direction $\bar{d}$ to obtain a combined program with weaker partial calmness condition.

Proposition 4.3. *Let $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ be a local solution of (CPFJ). Suppose that the partial calmness condition for (CPFJ) holds at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$. Then, for each $\bar{d} \in \mathcal{C}(\bar{y}; \bar{x})$, $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u}, \bar{d})$ is a local optimal solution of problem (R-FJSOCP) where the partial calmness condition holds. Conversely, suppose that problem (R-FJSOCP) is partially calm at a local solution $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u}, 0)$ and $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ is a local solution of problem (CPFJ). Then problem (CPFJ) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$.*

Proof. Adding an additional variable d and a superfluous constraint $d \in \mathbb{R}^m$ to problem (CPFJ), the first assertion follows from Proposition 2.7.

Now suppose that problem (R-FJSOCP) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u}, 0)$. Then there exist $\mu \geq 0$ and a neighborhood $U(\bar{x}, \bar{y}, \bar{u}_0, \bar{u}, 0)$ of $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u}, 0)$ such that

$$F(\bar{x}, \bar{y}) \leq F(x, y) + \mu(f(x, y) - V(x)), \quad \forall (x, y, u_0, u, d) \in \mathcal{F}_R \cap U(\bar{x}, \bar{y}, \bar{u}_0, \bar{u}, 0),$$

where $\mathcal{F}_R$ is the feasible region of problem (R-FJSOCP_μ) .

Let $(x, y, u_0, u) \in U(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ be a feasible solution of problem (CPFJ_μ) . Then $(x, y, u_0, u, 0)$ is feasible to problem (R-FJSOCP_μ) . Hence it follows that the problem (CPFJ) is partially calm at $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$. $\square$

Note that problem (R-FJSOCP) and its partially penalized problem (R-FJSOCP_μ) are MPECs. Based on the discussions on the partial calmness condition, Theorem 2.11 can be used to derive S-/M- type necessary optimality conditions for problem (R-FJSOCP) under appropriate constraint qualifications.

4.2.2. Combined with the basic second-order optimality condition

As reviewed in Section 2, under certain constraint qualifications, $M^1(y; x) \neq \emptyset$ for $y \in S(x)$ and one of the second-order optimality conditions BSOC, WSOC, and SSOC holds. In this subsection, we study the combined problem with BSOC. We say that y is a BSOC, WSOC, or SSOC point of $P(x)$ respectively if Definition 2.1(i), 2.1(ii), or 2.1(iii) holds respectively. Now we define

$$\begin{aligned}\Sigma_{\text{BSOC}} &:= \left\{ (x, y) \in \mathbb{R}^{n+m} : y \text{ is a BSOC-point for } P(x) \right\}, \\ \Sigma_{\text{WSOC}} &:= \left\{ (x, y) \in \mathbb{R}^{n+m} : y \text{ is a WSOC-point for } P(x) \right\}, \\ \Sigma_{\text{SSOC}} &:= \left\{ (x, y) \in \mathbb{R}^{n+m} : y \text{ is an SSOC-point for } P(x) \right\}.\end{aligned}$$

It is easily seen that

$$\Sigma_{\text{SSOC}} \subseteq \Sigma_{\text{BSOC}}, \Sigma_{\text{SSOC}} \subseteq \Sigma_{\text{WSOC}}, \text{ and } \Sigma_{\text{SSOC}} \stackrel{\text{LICQ}}{=} \Sigma_{\text{BSOC}}. \quad (22)$$

Similar to the combined problem (FJSOCP), we consider the combined problem with basic (weak, strong) second-order optimality conditions (SOCP) where $\Sigma_{\text{SOC}} = \Sigma_{\text{BSOC}}, \Sigma_{\text{WSOC}}, \Sigma_{\text{SSOC}}$, respectively. Different from FJSOC, none of BSOC, WSOC, and SSOC is necessary without extra constraint qualifications.

Thus this reformulation requires that BSOC, WSOC, and SSOC hold at the optimal solution of the lower level program. At least it requires that the KKT conditions hold at the optimal solutions of the lower level program (i.e., $M^1(y; x) \neq \emptyset$).

Since it is difficult to express the set of indices of active inequalities directly in the combined problem (SOCP) with $\Sigma_{\text{SOC}} = \Sigma_{\text{BSOC}}$ such that it is still an optimization problem with equality and inequality constraints, we relax the critical cone (2) as

$$\begin{aligned}\mathcal{C}(y; x) &\subseteq \left\{ d \in \mathbb{R}^m : D_y f(x, y)d \leq 0, u_j D_y g_j(x, y)d \leq 0, \forall j = 1, \dots, p \right\} \\ &= \left\{ d \in \mathbb{R}^m : u_j D_y g_j(x, y)d = 0, \forall j = 1, \dots, p \right\},\end{aligned} \quad (23)$$

where u is a KKT multiplier and (23) follows from

$$0 \geq D_y f(x, y)d = - \sum_{j \in J_0(y; x)} u_j D_y g_j(x, y)d \geq 0.$$

Hence we propose to consider the following relaxed problem for the combined problem (SOCP) with $\Sigma_{\text{SOC}} = \Sigma_{\text{BSOC}}$:

$$\begin{aligned}
(\text{R-BSOCP}) \quad & \min_{x,y,u,d} F(x,y) \\
\text{s.t.} \quad & f(x,y) - V(x) \leq 0, \quad \nabla_y f(x,y) + \nabla_y g(x,y)^T u = 0, \quad g(x,y) \leq 0, \\
& u \geq 0, \quad u^T g(x,y) = 0, \quad d^T \nabla_{yy}^2 [f(x,y) + \sum_{i=1}^p u_i g_i(x,y)] d \geq 0, \\
& u_j D_y g_j(x,y) d = 0, \quad \forall j = 1, \dots, p, \quad G(x,y) \leq 0.
\end{aligned}$$

Similar to Proposition 4.2, since there is the value function constraint, the combined problem (SOCP) and the relaxed combined problem (R-BSOCP) are both equivalent in global solutions to the original problem when the corresponding second-order optimality conditions hold [8, Theorem 4.3]. To state the relationship on local solutions, we define the following mapping:

$$\tilde{K}(x,y) := \{(u,d) \in \tilde{\Xi}(x,y) | \nabla_y f(x,y) + \nabla_y g(x,y)^T u = 0\}, \quad (24)$$

where

$$\tilde{\Xi}(x,y) := \left\{ (u,d) : \begin{aligned} & u \geq 0, \quad u^T g(x,y) = 0, \\ & d^T \nabla_{yy}^2 [f(x,y) + \sum_{i=1}^p u_i g_i(x,y)] d \geq 0, \\ & u_j D_y g_j(x,y) d = 0, \quad \forall j = 1, \dots, p \end{aligned} \right\}. \quad (25)$$

Proposition 4.4. *Let $(\bar{x}, \bar{y})$ be a local optimal solution to (BLPP). Suppose that the basic second-order optimality condition holds for the lower level problem $P(\bar{x})$ at $\bar{y}$. Then for each $(u,d) \in \tilde{K}(\bar{x}, \bar{y})$, $(\bar{x}, \bar{y}, u, d)$ is a local optimal solution of (R-BSOCP). Conversely, let $(\bar{x}, \bar{y}, u, d)$ be a local optimal solution to problem (R-BSOCP) for each $(u,d) \in \tilde{K}(\bar{x}, \bar{y})$. Furthermore, let $\tilde{K}$ be inner semicompact at $(\bar{x}, \bar{y})$ with respect to $\text{dom} \tilde{K}$. Then $(\bar{x}, \bar{y})$ is a local solution of (BLPP).*

Remark 4.5. The function $\tilde{K}$ defined in (24) is inner-semicompact when MFCQ for the lower level holds at $(\bar{x}, \bar{y})$. This is true because at the optimal solutions, MFCQ is equivalent to the property of the set of KKT multipliers being nonempty and bounded [22].

Next, we study the relation between the partial calmness for (R-BSOCP) and the partial calmness for (CP). Similar to Proposition 4.3, we can prove the following proposition.

Proposition 4.6. *Suppose that $(\bar{x}, \bar{y}, \bar{u}, 0)$ is a local solution of (R-BSOCP). Then the partial calmness for (R-BSOCP) holds at $(\bar{x}, \bar{y}, \bar{u}, 0)$ if and only if the partial calmness for (CP) holds at the local optimal solution $(\bar{x}, \bar{y}, \bar{u})$. Furthermore, if BSOC holds at $\bar{y}$ for the lower level problem $P(\bar{x})$, then for all $\bar{d} \in \mathcal{C}(\bar{y}; \bar{x})$, validity of the partial calmness condition for (CP) at the local optimal solution $(\bar{x}, \bar{y}, \bar{u})$ implies validity of the partial calmness condition for (R-BSOCP) at the local optimal solution $(\bar{x}, \bar{y}, \bar{u}, \bar{d})$.*

Finally, (R-BSOCP) and its partially penalized problem (R-BSOCP _{μ}) are MPECs. Theorem 2.11 can be used to derive S-/M- type necessary optimality conditions under appropriate constraint qualifications.

4.2.3. Combined with the weak second-order optimality condition

If WSOC holds at the lower level, we can consider the following combined problem with WSOC:

$$\begin{aligned}
 (\text{WSOCP}) \quad & \min_{x,y,u} F(x,y) \\
 \text{s.t.} \quad & f(x,y) - V(x) \leq 0, \quad \nabla_y f(x,y) + \nabla_y g(x,y)^T u = 0, \\
 & g(x,y) \leq 0, \quad u \geq 0, \quad u^T g(x,y) = 0, \quad G(x,y) \leq 0, \\
 & 0 \preceq \nabla_{yy}^2 [f(x,y) + \sum_{i=1}^p u_i g_i(x,y)] \Big|_{\mathcal{S}(y;x)},
 \end{aligned}$$

and propose the corresponding partial calmness condition. Here

$$0 \preceq \nabla_{yy}^2 [f(x,y) + \sum_{i=1}^p u_i g_i(x,y)] \Big|_{\Gamma},$$

with $\Gamma := \mathcal{S}(y;x)$ means that

$$d^T \nabla_{yy}^2 [f(x,y) + \sum_{i=1}^p u_i g_i(x,y)] d \geq 0, \quad \forall d \in \Gamma,$$

i.e., the matrix $\nabla_{yy}^2 [f(x,y) + \sum_{i=1}^p u_i g_i(x,y)]$ is a Γ -copositive matrix.

But the copositive matrix condition in (WSOCP) is not easy to tackle because the critical subspace $\mathcal{S}(y;x)$ involves the set of indices of active inequalities of $P(x)$. To cope with this difficulty, the equivalence between the KKT points satisfying WSOC of the original problem $P(x)$ and the reformulated problem $\tilde{P}(x)$ by introducing the squared slack variables is very useful. Indeed, by Propositions 2.3 and 2.4, problem (WSOCP) is equivalent to the following reformulated problem by introducing the squared slack variables:

$$\begin{aligned}
 (\text{WSOCPZ}) \quad & \min_{x,y,z,\lambda} F(x,y) \\
 \text{s.t.} \quad & f(x,y) - V(x) \leq 0, \quad \nabla_y f(x,y) + \sum_{i=1}^p \lambda_i \nabla_y g_i(x,y) = 0, \\
 & g_i(x,y) + z_i^2 = 0, \quad \lambda_i z_i = 0, \quad \forall i = 1, \dots, p, \quad G(x,y) \leq 0, \\
 & 0 \preceq \nabla_{(y,z)}^2 L(y,z,\lambda;x) \Big|_{\mathcal{S}(y,z;x)}.
 \end{aligned}$$

Now it is worth noting that the critical subspace

$$\mathcal{S}(y,z;x) = \left\{ (d,\nu) \in \mathbb{R}^m \times \mathbb{R}^p : D_y g_i(x,y)d + 2z_i \nu_i = 0, \forall i \right\}$$

does not involve the set of indices of active inequalities of $P(x)$.

Note that in contrast to all the other reformulations derived before, (WSOCPZ) is not a complementarity – but a generalized copositive programming problem with switching constraints [32, 34] which has slightly different properties than an MPEC.

4.2.4. Combined with the strong second-order optimality condition

If SSOC holds at the lower level for each $y \in S(x)$, we can consider the following combined problem:

$$\begin{aligned}
(\text{SSOCP}) \quad & \min_{x,y,u} F(x,y) \\
& \text{s.t. } f(x,y) - V(x) \leq 0, \quad \nabla_y f(x,y) + \nabla_y g(x,y)^T u = 0, \\
& g(x,y) \leq 0, \quad u \geq 0, \quad u^T g(x,y) = 0, \quad G(x,y) \leq 0, \\
& 0 \preceq \nabla_{yy}^2 [f(x,y) + \sum_{i=1}^p u_i g_i(x,y)] \Big|_{\mathcal{C}(y;x)}.
\end{aligned}$$

Recall that for a closed convex cone Γ , the class of all Γ -copositive matrices is the dual cone of the convex hull of $\{dd^T \in \mathbb{S}_+^m : d \in \Gamma \subseteq \mathbb{R}^m\}$ [19, Lemma 2.28]. This provides a natural generalization of the constraint $\nabla_{yy}^2 f(x,y) \in \mathbb{S}_+^m$ in the unconstrained case. The problem (SSOCP) can be viewed as generalized semi-infinite programming problem [45, 49] or generalized copositive programming problem (set-semidefinite optimization) [10, 19]. Similarly as in section 4.2.3, to deal with the difficulty of the active index set, one can also use the squared-slack-variable trick here.

4.3. Examples and summary

In this section, we have discussed different types of combined problems with second-order optimality conditions, called (FJSOCP), (SOCP), (SSOCP) and (WSOCP). To address the issue caused by the set of indices of active inequalities, we come up with the related relaxed problems, called (R-FJSOCP) and (R-BSOCP), and also the problem with squared slack variables (WSOCPZ). All of the combined and relaxed problems are equivalent to the original (BLPP) under some mild and necessary assumptions.

Similarly to [30, 52, 51], we have proposed various partial calmness conditions based on the combined problems above. We summarize the relationships between various partial calmness conditions in Figure 4.1 on the next page.

Under the validity of the KKT condition, we can even establish the relationship between partial calmness for the FJ and the KKT type combined programs when the FJ multiplier considered satisfies $\bar{u}_0 = 0$. For example, even if the partial calmness for (CPFJ) holds for $(\bar{x}, \bar{y}, \bar{u}_0, \bar{u})$ with $\bar{u}_0 = 0$, if the set of multiplier $M^1(\bar{y}; \bar{x})$ is not empty and $\tilde{u} \in M^1(\bar{y}; \bar{x})$, it can be shown that the partial calmness for (CP) holds for $(\bar{x}, \bar{y}, \tilde{u} + k\bar{u})$ when $k > 0$ is sufficiently large. This comes from the fact that $1 + \sum_{j=1}^p (\tilde{u}_j + k\bar{u}_j) = k + 1 + \sum_{j=1}^p \tilde{u}_j$, and $(1, \tilde{u} + k\bar{u}) / (1 + \sum_{j=1}^p (\tilde{u}_j + k\bar{u}_j)) \rightarrow (0, \bar{u})$ as $k \rightarrow +\infty$.

Next, we use some nonconvex BLPPs to illustrate the combined approach with second-order optimality conditions and the necessary optimality conditions.

We first give an example for which the combined approach in [30, 51] fails, but the partial calmness and the necessary optimality condition will hold if one adds the basic second-order optimality condition for the lower level program in the associated combined problem.

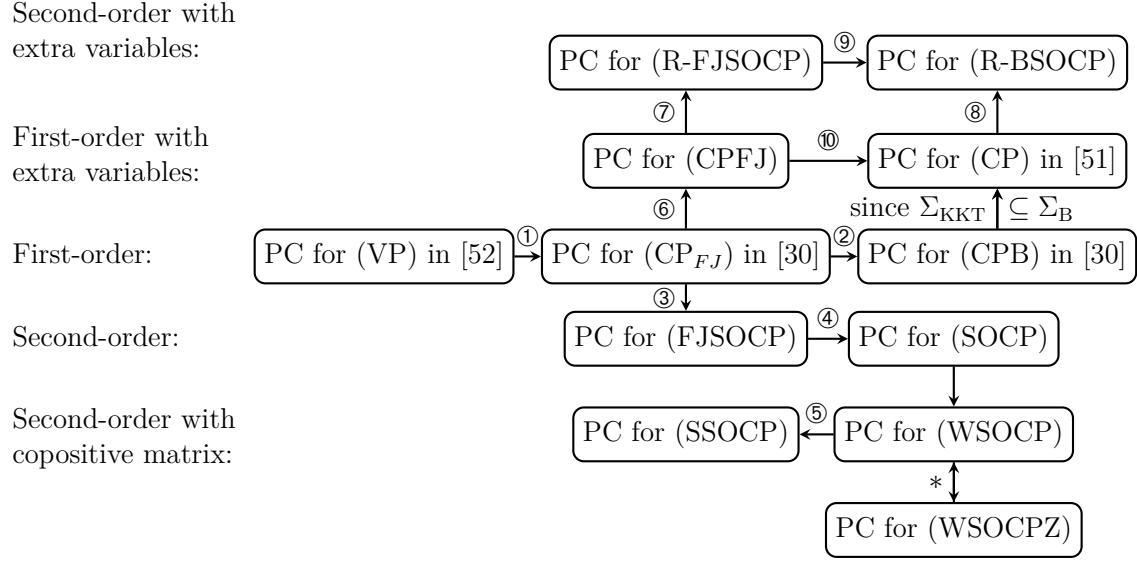


Figure 4.1: Relationship between various partial calmness conditions. Here we denote “partial calmness” briefly by PC. By Proposition 2.7, we have relations ①–⑤. For relations ⑥, ⑦, and ⑧, we refer the reader to [30, Theorem 4.4], Proposition 4.3, and Proposition 4.6, respectively. One may prove other relations by a similar argument of the proof of Proposition 4.3. The equivalent relation $*$ follows from Propositions 2.3 and 2.4. An arrow between two PCs means one implies the other under certain constraint qualifications. Specifically, both the relations ⑨ and ⑩ require the validity of the KKT condition of $P(x)$ for $y \in S(x)$.

Example 4.7.

$$\min_{x, y \in \mathbb{R}} y^2 - x \quad \text{s.t.} \quad -1 \leq x \leq 1, \quad y \in S(x) := \arg \min_y \left\{ \frac{1}{4}y^4 - \frac{1}{2}xy^2 : 0 \leq y \leq \sqrt{2} \right\}. \quad (26)$$

Claim: In this example, we will show that

- the partial calmness for (CP) does not hold at $(\bar{x}, \bar{y}, \bar{u}) = (0, 0, 0)$;
- the partial calmness for (SOCP) with $\Sigma_{\text{SOC}} := \Sigma_{\text{BSOC}} = \Sigma_{\text{SSOC}}$ holds at $(\bar{x}, \bar{y}) = (0, 0)$;
- the partial calmness for (R-BSOCP) holds at $(\bar{x}, \bar{y}, \bar{u}, \bar{d})$ for any $\bar{d} \neq 0$;
- the partial calmness for (WSOCP) does not hold at $(\bar{x}, \bar{y}, \bar{u}) = (0, 0, 0)$;
- necessary optimality conditions fail to hold for (CP);
- the S-stationarity condition holds for (R-BSOCP).

It is easy to see that

$$S(x) = \begin{cases} \{\sqrt{2}\} & \text{if } x > 2, \\ \{\sqrt{x}\} & \text{if } 0 < x \leq 2, \\ \{0\} & \text{if } x \leq 0, \end{cases} \quad V(x) = \begin{cases} 1 - x & \text{if } x > 2, \\ -\frac{1}{4}x^2 & \text{if } 0 < x \leq 2, \\ 0 & \text{if } x \leq 0, \end{cases} \quad (27)$$

and $(\bar{x}, \bar{y}) = (0, 0)$ is a global optimal solution. Moreover, $M^1(0; 0) = \{0\}$.

Now we show that the partial calmness for (CP) does not hold at $(0, 0, 0)$. Indeed, the associated partially penalized problem is given by

$$\begin{aligned}
\min_{x,y,u} F_\mu(x,y) &:= y^2 - x + \mu \left(\frac{1}{4} y^4 - \frac{1}{2} xy^2 - V(x) \right) \\
\text{s.t. } &y^3 - xy - u_1 + u_2 = 0, \quad u_1 \geq 0, \quad -u_1 y = 0, \\
&u_2 \geq 0, u_2(y - \sqrt{2}) = 0, \quad 0 \leq y \leq \sqrt{2}, \quad -1 \leq x \leq 1.
\end{aligned} \tag{28}$$

Note that when $x = \frac{1}{k}$ ($k > 0$), $V(x) = -\frac{1}{4}x^2$. For any fixed μ , the objective function value $F_\mu(\frac{1}{k}, 0) = -k^{-1} + (\mu/4)k^{-2} < 0 = F_\mu(0, 0)$ when $k > \mu/4$. Hence $(\bar{x}, \bar{y}, \bar{u}) = (0, 0, 0)$ is not a local minimizer of the associated partially penalized problem (28) and the partial calmness for (CP) does not hold at $(0, 0, 0)$.

Let us consider adding the second-order optimality conditions. The critical cone is given by

$$\mathcal{C}(y; x) = \begin{cases} \mathbb{R}_+ & \text{if } x \in \mathbb{R}, y = 0, \\ \mathbb{R} & \text{if } x \in (0, 2), y = \sqrt{x}, \\ \{0\} & \text{if } x \geq 2, y = \sqrt{2}. \end{cases} \tag{29}$$

Since LICQ holds, BSOC coincides with SSOC and hence $\Sigma_{\text{BSOC}} = \Sigma_{\text{SSOC}}$. Problem (SOCP) is given by

$$\min_{x,y} y^2 - x \quad \text{s.t.} \quad \frac{1}{4} y^4 - \frac{1}{2} xy^2 - V(x) \leq 0, \quad (x, y) \in \Sigma_{\text{SSOC}}, \quad -1 \leq x \leq 1. \tag{30}$$

Suppose $(x, y) \in \Sigma_{\text{KKT}}$. Then it must satisfy the KKT condition

$$y^3 - xy - u_1 + u_2 = 0, \quad u_1 \geq 0, \quad -u_1 y = 0, \quad u_2 \geq 0, \quad u_2(y - \sqrt{2}) = 0, \quad 0 \leq y \leq \sqrt{2},$$

with a unique multiplier u . It follows that

$$\Sigma_{\text{KKT}} = \left\{ (x, 0) : x \in \mathbb{R} \right\} \cup \left\{ (x, \sqrt{x}) : x \in (0, 2) \right\} \cup \left\{ (x, \sqrt{2}) : x \geq 2 \right\}. \tag{31}$$

But SSOC states that $d^2(3y^2 - x) \geq 0, \forall d \in \mathcal{C}(y; x)$, which is equivalent to saying that $3y^2 - x \geq 0$. This means that the point (x, y) with $x > 0$ and $y = 0$ does not satisfy SSOC and hence is not included in the set Σ_{SSOC} . By the expression for the solution set (27), we have $\Sigma_{\text{SSOC}} = \{(x, y) : y \in S(x)\}$. Hence the value function constraint in problem (30) holds for all $(x, y) \in \Sigma_{\text{SSOC}}$. We therefore can remove the value function constraint from problem (30). This means that the partial calmness for (SOCP) with $\Sigma_{\text{SOC}} = \Sigma_{\text{SSOC}}$ holds at $(\bar{x}, \bar{y}) = (0, 0)$ with $\mu = 0$. Now consider the (R-BSOCP):

$$\begin{aligned}
\min_{x,y,u,d} &y^2 - x \\
\text{s.t. } &\frac{1}{4} y^4 - \frac{1}{2} xy^2 - V(x) \leq 0, \quad y^3 - xy - u_1 + u_2 = 0, \quad 0 \leq y \leq \sqrt{2}, \\
&-1 \leq x \leq 1, \quad u_1 \geq 0, \quad -u_1 y = 0, \quad -u_1 d = 0, \quad u_2 \geq 0, \\
&u_2(y - \sqrt{2}) = 0, \quad u_2 d = 0, \quad (3y^2 - x)d^2 \geq 0.
\end{aligned} \tag{32}$$

Let $\bar{d} \neq 0$. Then for any d sufficiently close to $\bar{d}$, condition $(3y^2 - x)d^2 \geq 0$ is equivalent to $3y^2 - x \geq 0$. So similar to the analysis for the partial calmness for (SOCP) with $\Sigma_{\text{SOC}} = \Sigma_{\text{SSOC}}$, the value function constraint can be removed. Then the partial calmness for problem (R-BSOCP) holds at $(\bar{x}, \bar{y}, \bar{u}, \bar{d}) = (0, 0, 0, \bar{d})$ with $\mu = 0$.

Recall that $\Sigma_{\text{SSOC}} \subseteq \Sigma_{\text{WSOC}}$ and the partial calmness with the larger set Σ_{WSOC} would be harder to hold. By the expression for the critical cone in (29), we can obtain the expression for the critical subspace of the problem (26)

$$\mathcal{S}(y; x) = \begin{cases} \{0\} & \text{if } x \in \mathbb{R}, y = 0 \text{ or } \sqrt{2}, \\ \mathbb{R} & \text{if } x \in \mathbb{R}, 0 < y < \sqrt{2}. \end{cases}$$

WSOC states that $d^2(3y^2 - x) \geq 0, \quad \forall d \in \mathcal{S}(y; x).$

Since when $x > 0, y = 0$, $d \in \mathcal{S}(y; x)$ is taken as zero, these points are still in the set Σ_{WSOC} and hence $\Sigma_{\text{WSOC}} = \Sigma_{\text{KKT}}$. Since $\Sigma_{\text{WSOC}} = \Sigma_{\text{KKT}}$, for this example, the partial calmness for (SOCP) with $\Sigma_{\text{SOC}} = \Sigma_{\text{WSOC}}$ does not hold at $(\bar{x}, \bar{y})$ and the partial calmness for (WSOCP) does not hold at $(\bar{x}, \bar{y}, \bar{u})$.

Point $(\bar{x}, \bar{y}, \bar{u}) = (0, 0, 0)$ does not satisfy the stationary conditions for (CP) based on the value function as in Theorem 2.11. Indeed, there do not exist $\mu \geq 0$, β , η^g and η^G such that

$$0 \in \nabla F(0, 0) + \mu (\nabla f(0, 0) - \partial^c V(0) \times \{0\}) + \nabla_{(x,y)} (\nabla_y L)(0, 0; 0)^T \beta \\ + \nabla g(0, 0)^T \eta^g + \nabla G(0, 0)^T \eta^G$$

since $\nabla F(0, 0) = (-1, 0)^T$, $\nabla g_1(0, 0) = (0, -1)^T$ and other terms are all zero.

Problem (32) is an MPEC. The S-stationary condition based on the value function (Definition 2.8) holds at $(\bar{x}, \bar{y}, \bar{u}, \bar{d}) = (0, 0, 0, 1)$. Indeed, since $\nabla_{(x,y)} (\nabla_{yy}^2 L)(0, 0; 0) = (-1, 0)^T$, there exists $\gamma = 1$ (let other multipliers be all zero) such that

$$0 = \nabla F(0, 0) - \nabla_{(x,y)} (\nabla_{yy}^2 L)(0, 0; 0)^T \gamma.$$

In the following example, we show that the partial calmness for (WSOCP) may hold.

Example 4.8. $\min_{x,y} y^2 - x$
s.t. $-1 \leq x \leq 1, y \in S(x) := \arg \min_y \left\{ \frac{1}{4} y^4 - \frac{1}{2} xy^2 : -1 \leq y \leq 1 \right\}.$ (33)

It is easy to see that

$$S(x) = \begin{cases} \{\pm 1\} & \text{if } x > 1, \\ \{\pm \sqrt{x}\} & \text{if } 0 < x \leq 1, \\ \{0\} & \text{if } x \leq 0, \end{cases} \quad (34)$$

$$\Sigma_{\text{KKT}} = \left\{ (x, 0) : x \in \mathbb{R} \right\} \cup \left\{ (x, \pm \sqrt{x}) : x \in (0, 1) \right\} \cup \left\{ (x, \pm 1) : x \geq 1 \right\},$$

and $(\bar{x}, \bar{y}) = (0, 0)$ is a global optimal solution. Moreover, $M^1(0; 0) = \{0\}$. Similarly to Examples 3.1 and 4.7, we can show that

- the partial calmness for (CP) and necessary optimality conditions do not hold at $(\bar{x}, \bar{y}, \bar{u}) = (0, 0, 0)$;
- the partial calmness for (SOCP) with $\Sigma_{\text{SOC}} := \Sigma_{\text{BSOC}} = \Sigma_{\text{SSOC}}$ holds at $(\bar{x}, \bar{y}) = (0, 0)$;
- the partial calmness for (R-BSOCP) and necessary optimality conditions hold at $(\bar{x}, \bar{y}, \bar{u}, \bar{d})$ for any $\bar{d} \neq 0$.

However, different from Example 4.7, we can show that the partial calmness for (WSOCP) also holds at $(\bar{x}, \bar{y}, \bar{u}) = (0, 0, 0)$. In fact, for problem (33),

$$\mathcal{S}(y; x) = \begin{cases} \{0\} & \text{if } x \in \mathbb{R}, y = \pm 1, \\ \mathbb{R} & \text{if } x \in \mathbb{R}, y \in (-1, 1). \end{cases}$$

But WSOC states that $d^2(3y^2 - x) \geq 0, \forall d \in \mathcal{S}(y; x)$. Since points $(x, 0)$ with $x > 0$ do not satisfy the above WSOC, we have $\Sigma_{\text{WSOC}} = \text{gph } S$ (see Figure 3.2). Hence the partial calmness for (WSOCP) holds at $(\bar{x}, \bar{y}, \bar{u}) = (0, 0, 0)$ with $\mu = 0$.

We compare the results for the two examples in the following table.

Examples	CP	SOCP _B	SOCP _S	R-BSOCP	WSOCP
Example 4.7	No	Yes	Yes	Yes	No
Example 4.8	No	Yes	Yes	Yes	Yes

Table 1: Comparison in the examples. Here we denote (SOCP) with $\Sigma_{\text{SOC}} = \Sigma_{\text{BSOC}}$ or Σ_{SSOC} by **SOCP_B**, **SOCP_S**, respectively. “Yes” or “No” answers the question “Does the partial calmness for the combined problem hold?”

5. Conclusions

In this paper, we demonstrate that although the partial calmness condition is generic for a combined program with a first condition information, there are still cases where the partial calmness condition and the corresponding necessary optimality conditions do not hold. To deal with these cases, we propose to add both the first-order and the second-order optimality conditions of the lower level problem as constraints. There are several advantages in this approach. First, by adding extra constraints to the first-order combined problem, the new partial calmness condition and the resulting necessary optimality condition are easier to hold. Second, by adding second-order optimality conditions, it may be possible that the graph of the solution set to the lower level problem coincides with the set of second-order stationary points and hence the difficult value function constraint can be removed. However there are also some drawbacks to our second-order approach. First, by using a second-order optimality condition, we may need to introduce more extra variables other than the multipliers. Then similar to the difficulty in dealing with the reformulation involving with the KKT condition, the problem is no longer equivalent to the original bilevel program in the sense of local optimality. Second, since there are second-order optimality conditions in the feasible region of the partially penalized problem, constraint qualifications may be harder to verify for the partially penalized problem. Third, numerically the combined program with the second-order condition may sometimes be harder to solve than the combined program with the first-order condition. In our future work, we will try to utilize the advantages and address the drawbacks.

Acknowledgments. We thank the guest editor Tim Hoheisel and the two anonymous referees for their suggestions, which have helped to improve the presentation of this paper.

References

- [1] L. Adam, R. Henrion, J. V. Outrata: *On M -stationarity conditions in MPECs and the associated qualification conditions*, Math. Programming 168 (2018) 229–259.
- [2] G. Allende, G. Still: *Solving bilevel programs with the KKT-approach*, Math. Programming 138 (2013) 309–332.
- [3] A. V. Arutyunov: *Perturbations of extremum problems with constraints and necessary optimality conditions*, J. Sov. Math. 54 (1991) 1342–1400.
- [4] K. Bai, J. J. Ye: *Directional necessary optimality conditions for bilevel programs*, Math. Oper. Res. 47 (2022) 1169–1191.
- [5] J. Bard: *Practical Bilevel Optimization: Algorithms and Applications*, Kluwer Academic, Dordrecht (1998).
- [6] R. Behling, G. Haeser, A. Ramos, D. S. Viana: *On a conjecture in second-order optimality conditions*, J. Optim. Theory Appl. 176 (2018) 625–633.
- [7] M. Benko, H. Gfrerer, J. V. Outrata: *Calculus for directional limiting normal cones and subdifferentials*, Set-Valued Var. Analysis 27 (2019) 713–745.
- [8] M. Benko, P. Mehrlitz: *On implicit variables in optimization theory*, J. Nonsmooth Anal. Optim. 2 (2021), art. no. 7215.
- [9] J. F. Bonnans, A. Shapiro: *Perturbation Analysis of Optimization Problems*, Springer, Berlin (2000).
- [10] S. Burer, H. Dong: *Representing quadratically constrained quadratic programs as generalized copositive programs*, Oper. Res. Letters 40 (2012) 203–206.
- [11] J. V. Burke: *Calmness and exact penalization*, SIAM J. Control Optimization 29 (1991) 493–497.
- [12] F. H. Clarke: *Optimization and Nonsmooth Analysis*, Wiley Interscience, New York (1983).
- [13] S. Dempe: *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht (2002).
- [14] S. Dempe, J. Dutta: *Is bilevel programming a special case of a mathematical program with complementarity constraints?*, Math. Programming 131 (2012) 37–48.
- [15] S. Dempe, V. Kalashnikov, G. Prez-Valds, N. Kalashnykova: *Bilevel Programming Problems: Theory, Algorithms and Applications to Energy Systems*, Springer, Berlin (2015).
- [16] S. Dempe, A. B. Zemkoho: *The bilevel programming problems: reformulations, constraint qualifications and optimality conditions*, Math. Programming 138 (2013) 447–473.
- [17] S. Dempe, A. B. Zemkoho (eds.): *Bilevel Optimization: Advances and Next Challenges*, Springer Optimization and Applications Vol. 161, Springer, Cham (2020).
- [18] C. Ding, D. F. Sun, J. J. Ye: *First order optimality conditions for mathematical programs with semidefinite cone complementarity constraints*, Math. Programming 147 (2014) 539–579.
- [19] G. Eichfelder, J. Jahn: *Set-semidefinite optimization*, J. Convex Analysis 15 (2008) 767–801.
- [20] L. Franceschi, P. Frasconi, S. Salzo, R. Grazzi, M. Pontil: *Bilevel programming for hyperparameter optimization*, ICML (2018) 1563–1572.

- [21] E. H. Fukuda, M. Fukushima, *A note on the squared slack variables technique for nonlinear optimization*, J. Oper. Res. Soc. Japan 60 (2017) 262–270.
- [22] J. Gauvin: *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Math. Programming 12 (1977), 136–138.
- [23] J. Gauvin, F. Dubeau: *Differential properties of the marginal function in mathematical programming*, Math. Programming Study 18 (1982) 101–119.
- [24] H. Gfrerer: *On directional metric subregularity and second-order optimality conditions for a class of nonsmooth mathematical programs*, SIAM J. Optimization 23 (2013) 632–665.
- [25] H. Gfrerer, J. J. Ye: *New constraint qualifications for mathematical programs with equilibrium constraints via variational analysis*, SIAM J. Optimization 27 (2017) 842–865.
- [26] H. Gfrerer, J. J. Ye: *New sharp necessary optimality conditions for mathematical programs with equilibrium constraints*, Set-Valued Var. Analysis 28 (2020) 395–426.
- [27] H. Gfrerer, J. J. Ye, J. C. Zhou: *Second-order optimality conditions for non-convex set-constrained optimization problems*, Math. Oper. Res. 47 (2022) 2344–2365.
- [28] L. Guo, G. H. Lin, J. J. Ye, J. Zhang: *Sensitivity analysis of the value function for parametric mathematical programs with equilibrium constraints*, SIAM J. Optimization 24 (2014) 1206–1237.
- [29] R. Henrion, T. M. Surowiec: *On calmness conditions in convex bilevel programming*, Appl. Analysis 90 (2011) 951–970.
- [30] R. Z. Ke, W. Yao, J. J. Ye, J. Zhang: *Generic property of the partial calmness condition for bilevel programming problems*, SIAM J. Optimization 32 (2022) 604–634.
- [31] G. Kunapuli, K. Bennett, J. Hu, J. S. Pang: *Classification model selection via bilevel programming*, Optim. Meth. Software 23 (2008) 475–489.
- [32] Y. C. Liang, J. J. Ye: *Optimality conditions and exact penalty for mathematical programs with switching constraints*, J. Optim. Theory Appl. 190 (2021) 1–31.
- [33] R. Liu, P. Mu, X. Yuan, S. Z. Zeng, J. Zhang: *A generic first-order algorithmic framework for bi-level programming beyond lower-level singleton*, ICML (2020) 6305–6315.
- [34] P. Mehlitz: *Stationarity conditions and constraint qualifications for mathematical programs with switching constraints*, Math. Programming 181 (2020) 149–186.
- [35] P. Mehlitz: *On the linear independence constraint qualification in disjunctive programming*, Optimization 69 (2020) 2241–2277.
- [36] P. Mehlitz, L. I. Minchenko: *R-regularity of set-valued mappings under the relaxed constant positive linear dependence constraint qualification with applications to parametric and bilevel optimization*, Set-Valued Analysis 30 (2022) 179–205.
- [37] P. Mehlitz, L. I. Minchenko, A. B. Zemkoho: *A note on partial calmness for bilevel optimization problems with linearly structured lower level*, Optimization Letters 15 (2021) 1277–1291.
- [38] L. Minchenko, A. Leschov: *On strong and weak second-order necessary optimality conditions for nonlinear programming*, Optimization 65 (2016) 1693–1702.
- [39] L. Minchenko, S. Stakhovski: *Parametric nonlinear programming problems under the relaxed constant rank condition*, SIAM J. Optimization 21 (2011) 314–332.
- [40] J. Mirrlees: *The theory of moral hazard and unobservable behaviour. Part I*, Rev. Econ. Studies 66 (1999) 3–22.

- [41] B. S. Mordukhovich: *Variational Analysis and Applications*, Springer, Cham (2018).
- [42] B. S. Mordukhovich, N. M. Nam, H. M. Phan: *Variational analysis of marginal functions with applications to bilevel programming*, J. Optim. Theory Appl. 152 (2012) 557–586.
- [43] J. V. Outrata: *On the numerical solution of a class of Stackelberg problems*, ZOR Math. Methods Oper. Res. 34 (1990) 255–277.
- [44] K. Shimizu, Y. Ishizuka, J. Bard: *Nondifferentiable and Two-Level Mathematical Programming*, Kluwer Academic Publishers, Dordrecht (1997).
- [45] O. Stein: *Bi-Level Strategies in Semi-Infinite Programming*, Springer, Cham (2013).
- [46] M. W. Xu, J. J. Ye: *Relaxed constant positive linear dependence constraint qualification and its application to bilevel programs*, J. Global Optimization 78 (2020) 181–205.
- [47] J. J. Ye: *Necessary optimality conditions for multiobjective bilevel programming problems*, Math. Oper. Res. 36 (2011) 165–184.
- [48] J. J. Ye: *Constraint qualifications and optimality conditions in bilevel optimization*, in: *Bilevel Optimization: Advances and Next Challenges*, S. Dempe et al. (eds.), Springer Optimization and its Applications Vol. 161, Springer, Cham (2020) 227–251.
- [49] J. J. Ye, S. Y. Wu: *First order optimality conditions for generalized semi-infinite programming problems*, J. Optim. Theory Appl. 137 (2008) 419–434.
- [50] J. J. Ye, X. M. Yuan, S. Z. Zeng, J. Zhang: *Difference of convex algorithms for bilevel programs with applications in hyperparameter selection*, Math. Programming B 198/2 (2023) 1583–1616.
- [51] J. J. Ye, D. L. Zhu: *New necessary optimality conditions for bilevel programs by combining the MPEC and value function approaches*, SIAM J. Optimization 20 (2010) 1885–1905.
- [52] J. J. Ye, D. L. Zhu: *Optimality conditions for bilevel programming problems*, Optimization 33 (1995) 9–27.