On Bilevel Programs and Minimax Problems

by

Xiaoxiao Ma
B.Sc., Tianjin University, 2017
M.Sc., Tianjin University, 2020

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Mathematics and Statistics

On Bilevel Programs and Minimax Problems

by

Xiaoxiao Ma
B.Sc., Tianjin University, 2017
M.Sc., Tianjin University, 2020

Supervisory Committee

Dr. Jane Ye, Supervisor
(Department of Mathematics and Statistics)

Dr. Julie Zhou, Departmental Member
(Department of Mathematics and Statistics)

Dr. Yang Shi, Outside Member
(Department of Mechanical Engineering)

ABSTRACT

Second-order optimality conditions usually offer more precise insights into local optimality compared to their first-order counterparts. Concurrently, there has been a growing prevalence of bilevel programs and minimax problems in recent years. In our research, we intricately explore second-order optimality conditions within the realm of bilevel programs and minimax problems.

First, we provide a comprehensive exploration of second-order combined approaches for bilevel problems. Building on the well-known first-order combined approach, the research introduces novel techniques that incorporate lower-level second-order information to overcome the difficulty of the constraint qualification for bilevel problems. By characterizing lower-level optimal solutions using both first and second-order necessary optimality conditions, together with the value function constraint, we give some new single-level reformulations for bilevel problems for which the important partial calmness condition can be more likely to hold.

We then focus on the introduction and analysis of calm local minimax points, which is an appropriate local notion for nonconvex-nonconcave nonsmooth minimax problems. We study the properties of calm local minimax points, establishing their strong connections with existing optimality concepts. We provide a comprehensive exploration of first-order and second-order sufficient and necessary optimality conditions for calm local minimax points.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

my supervisor, Dr. Jane J. Ye for her endless patience, dedicated support and mentorship, without which this dissertation would not have been achievable.

Dr. Jin Zhang, Dr. Wei Yao, and Dr. Shangzhi Zeng for their kind assistance and guidance on my research.

UVic and China Scholarship Council, for the fellowships and scholarships.

Last but not least, I would like to thank all my friends and family, both in China and Victoria, for their invaluable support and kindness.

Chapter 1

Introduction

This chapter introduces the main focus of our research, providing clarity on the key problems under investigation. A concise summary of the contributions made by this research is provided. Additionally, essential background materials are presented to establish a foundational context for the subsequent discussions.

1.1 Symbols and notations

Our notation is standard. We denote by $\mathbb{R}^r$ the r-dimensional Euclidean space, $\mathbb{R}^r_+$ ($\mathbb{R}^r_-$) the nonnegative (nonpositive) orthant in $\mathbb{R}^r$, and $\overline{\mathbb{R}} = [-\infty, \infty]$ the extended real line. We use $e_j \in \mathbb{R}^p$ to denote the unit vector whose j-th component is 1, while all other components are zero. For any $z \in \mathbb{R}^r$, $\|z\|$ denotes its Euclidean norm. For $z \in \mathbb{R}^r$ and $\epsilon > 0$, we denote by $\mathbb{B}_\epsilon(z) := \{z' \mid \|z' - z\| \le \epsilon\}$ the closed ball centered at z with radius ϵ and by $\mathbb{B}$ the closed unit ball centered at 0. For $z \in \mathbb{R}^r$, $B_\epsilon(z)$ denotes the open ball centered at z with radius ϵ. For any $z \in \mathbb{R}^r$ and $S \subseteq \mathbb{R}^r$, the distance of z to S is $\operatorname{dist}(z, S) := \inf_{z' \in S} \|z - z'\|$. For a set $S \subseteq \mathbb{R}^r$, we denote by $\delta_S : \mathbb{R}^r \to \overline{\mathbb{R}}$ the indicator function, i.e., $\delta_S(z) = 0$ when $z \in S$ and $\delta_S(z) = \infty$ when $z \notin S$. The notation $S^\perp := \{\alpha \in \mathbb{R}^r | \langle \alpha, z \rangle = 0, \forall z \in S\}$ denotes the orthogonal complement. We denote by $\operatorname{cl} S, \operatorname{co} S, S^\circ$ the closure, the convex hull, and the polar of S, and by $\sigma_S(\bar{z}) := \sup_{s \in S} \langle \bar{z}, s \rangle$ the support function of S at $\bar{z}$, respectively. For a set $S \subseteq \mathbb{R}^r$, a point $\bar{z} \in \mathbb{R}^r$, and a sequence z_k, the notation $z_k \xrightarrow{S} \bar{z}$ means that the sequence $z_k \in S$ goes to $\bar{z}$. 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 $\langle A, B \rangle := \operatorname{tr}(AB)$, $A, B \in \mathbb{S}^m$, where $\operatorname{tr}(A)$ denotes the trace of the matrix

A. For a matrix $A \in \mathbb{R}^{n\times m}$, A^T is its transpose. For a symmetric matrix $A \in \mathbb{R}^{r\times r}$, $A \prec 0$ ($\succeq 0, \preceq 0$) means that the matrix A is a negative definite (positive semidefinite, negative semidefinite) matrix and A^{-1} is the inverse matrix. The set of symmetric positive semidefinite matrices is denoted by $\mathbb{S}^m_+$. For an extended real-valued function $\varphi : \mathbb{R}^r \to [-\infty, \infty]$, the domain of φ is defined by $\operatorname{dom}\varphi := \{z \in \mathbb{R}^r \mid \varphi(z) < +\infty\}$. For a set-valued mapping $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, $\operatorname{dom}\Gamma := \{x \in \mathbb{R}^n | \Gamma(x) \neq \emptyset\}$ denotes the domain of Γ, $\operatorname{gph}\Gamma := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m | x \in \mathbb{R}^n, y \in \Gamma(x)\}$ denotes the graph of Γ. For a differentiable mapping $h : \mathbb{R}^r \to \mathbb{R}^l$, we denote its Jacobian matrix by $Dh(z) \in \mathbb{R}^{l\times r}$ 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 $\nabla^2 h(z)$. By $\varphi(t) = o(t)$, we mean that $\varphi(t)$ is a function such that $\lim\limits_{t\downarrow 0} \frac{\varphi(t)}{t} = 0$.

1.2 Research objects

Our research focuses on the following bilevel programming problem (BLPP):

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

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

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

Here $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and the mappings $F, f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$.

When we define $f(x,y) := -F(x,y)$, problem (BLPP) transforms into a specific type known as minimax problems. In Chapter 3, we will study local optimality for minimax problems. Specifically, we study minimax problems in the following general form.

$$\min_{x\in X} \max_{y\in Y} f(x,y),$$

where the objective function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is possibly a nonsmooth function, the nonempty sets $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^m$ are closed but may be nonconvex.

1.3 Main contributions

We summarize our contributions as follows.

In Chapter 2, we introduce the second-order combined approaches for bilevel problems. We characterize optimal solutions of the lower-level problem through the lower-level value function, first-order necessary optimality conditions, and second-order necessary optimality conditions. The original bilevel problems are then reformulated into a single-level problem, incorporating constraints involving the lower-level value function and its optimality conditions. Through examples, we illustrate that the partial calmness condition for second-order combined problems is more likely to hold compared to existing approaches such as the value function approach, first-order approach, and combined approach.

In Chapter 3, we study the local optimality of minimax problems with simple constraints, where the maximum constraint is independent of the minimum variable. To characterize local optimality for general nonconvex-nonconcave constrained minimax problems, we introduce the concept of calm local minimax points. We explore the relationships between calm local minimax points and existing minimax concepts, such as global minimax points, local saddle points (Nash equilibrium), and local minimax points. Subsequently, we give first- and second-order sufficient and necessary optimality conditions for calm local minimax points, demonstrating their advantages among existing ones through examples.

1.4 Background materials

In this section, we review some fundamental concepts in the field of variational analysis.

Definition 1.4.1 (tangent and normal cones [12, 73])**.** *Given $S \subseteq \mathbb{R}^r, \bar{z} \in S$, the tangent/contingent cone and inner tangent cone to S at $\bar{z}$ are defined by*

$$T_S(\bar{z}) \; := \; \left\{ w \in \mathbb{R}^r \,\middle|\, \exists\ t_k \downarrow 0,\ w_k \to w \ \text{ with } \ \bar{z} + t_k w_k \in S \right\},$$

$$T_S^i(\bar{z}) \; := \; \left\{ w \in \mathbb{R}^r \,\middle|\, \forall\ t_k \downarrow 0,\ \exists w_k \to w \ \text{ with } \ \bar{z} + t_k w_k \in S \right\},$$

respectively. For $w \in T_S(\bar{z})$, the outer second-order tangent set and the inner second-

order tangent set to S at $\bar{z}$ in direction w are defined by

$$T_S^2(\bar{z};w) := \left\{ \nu \in \mathbb{R}^r \,\middle|\, \exists\ t_k \downarrow 0, \nu_k \to \nu \quad \text{with} \quad \bar{z} + t_k w + \frac{1}{2} t_k^2 \nu_k \in S \right\},$$

$$T_S^{i,2}(\bar{z};w) := \left\{ \nu \in \mathbb{R}^r \,\middle|\, \forall\ t_k \downarrow 0, \exists \nu_k \to \nu \quad \text{with} \quad \bar{z} + t_k w + \frac{1}{2} t_k^2 \nu_k \in S \right\},$$

respectively.

The regular/Fréchet normal cone, the proximal normal cone, and the limiting/Mordukhovich normal cone to S at $\bar{z}$ are given, respectively, by

$$\begin{aligned} \widehat{N}_S(\bar{z}) &:= \left\{ z^* \in \mathbb{R}^r \,\middle|\, \langle z^*, z - \bar{z} \rangle \le o\big(\|z - \bar{z}\|\big)\ \forall z \in S \right\}, \\ N_S^p(\bar{z}) &:= \left\{ z^* \in \mathbb{R}^r \,\middle|\, \exists\ \gamma > 0 : \langle z^*, z - \bar{z} \rangle \le \gamma \|z - \bar{z}\|^2\ \forall z \in S \right\}, \\ N_S(\bar{z}) &:= \left\{ z^* \in \mathbb{R}^r \,\middle|\, \exists\ z_k \xrightarrow{S} \bar{z}, z_k^* \to z^* \text{ with } z_k^* \in \widehat{N}_S(z_k) \right\}. \end{aligned}$$

The regular normal cone to S at $\bar{z}$ [73, Proposition 6.5] can also be characterized by

$$\widehat{N}_S(\bar{z}) := \{ z^* \in \mathbb{R}^r \mid \langle z^*, w \rangle \le 0 \quad \forall w \in T_S(\bar{z}) \} = T_S(\bar{z})^\circ. \tag{1.4.1}$$

For a closed set S, one always has $N_S^p(\bar{z}) \subseteq \widehat{N}_S(\bar{z}) \subseteq N_S(\bar{z})$, where all the cones agree and reduce to the normal cone of convex analysis if S is convex.

A set S is said to be normally regular at $\bar{z} \in S$ if $\widehat{N}_S(\bar{z}) = N_S(\bar{z})$. It is said that a family of sets $\{\Omega_t\}, t > 0$, in $\mathbb{R}^r$ converges to a set $\Omega \subseteq \mathbb{R}^r$ as $t \downarrow 0$ if Ω is closed and

$$\lim_{t \downarrow 0} \operatorname{dist}(z, \Omega_t) = \operatorname{dist}(z, \Omega) \text{ for all } z \in \mathbb{R}^r.$$

A set $S \subseteq \mathbb{R}^r$ is geometrically derivable at $\bar{z} \in S$ if $T_S(\bar{z}) = \lim_{t \downarrow 0} \frac{S - \bar{z}}{t}$. Convex sets are geometrically derivable.

It is well-known that in the convex case, the normal cone and the tangent cone are polar to each other.

Definition 1.4.2. *(Subdifferentials) Let $\varphi : \mathbb{R}^n \to [-\infty, \infty]$ and $\bar{z} \in \operatorname{dom} \varphi$. The Fréchet (regular) subdifferential of φ at $\bar{z}$ is the set*

$$\widehat{\partial} \varphi(\bar{z}) := \{ \xi \in \mathbb{R}^n \mid \varphi(z) \ge \varphi(\bar{z}) + \langle \xi, z - \bar{z} \rangle + o(\|z - \bar{z}\|) \},$$

the limiting (Mordukhovich or basic) subdifferential of φ *at* $\bar{z}$ *is the set*

$$\partial\varphi(\bar{z}) := \left\{\xi \in \mathbb{R}^n \mid \exists z^k \to \bar{z}, \xi^k \to \xi \ s.t.\ \varphi\left(z^k\right) \to \varphi(\bar{z}), \xi^k \in \widehat{\partial}\varphi\left(z^k\right)\right\}.$$

Definition 1.4.3 (subderivatives, superderivatives and semiderivatives; Clarke generalized directional derivative and Clarke generalized gradient [73, Definitions 8.1 and 7.20], [16, page 10])**.** *Consider a function* $\psi : \mathbb{R}^r \to \overline{\mathbb{R}}$*, a point* $\bar{z}$ *with* $\psi(\bar{z})$ *finite, and* $w \in \mathbb{R}^r$*. The subderivative and the superderivative of* ψ *at* $\bar{z}$ *for* w *is defined by*

$$\begin{aligned} \mathrm{d}\psi(\bar{z})(w) &:= \liminf_{\substack{t\downarrow 0\\ w'\to w}} \frac{\psi(\bar{z}+tw')-\psi(\bar{z})}{t}, \\ \mathrm{d}^+\psi(\bar{z})(w) &:= \limsup_{\substack{t\downarrow 0\\ w'\to w}} \frac{\psi(\bar{z}+tw')-\psi(\bar{z})}{t}, \end{aligned}$$

respectively. When the limit

$$\mathrm{d}\psi(\bar{z})(w) = \mathrm{d}^+\psi(\bar{z})(w) = \lim_{\substack{t\downarrow 0\\ w'\to w}} \frac{\psi(\bar{z}+tw')-\psi(\bar{z})}{t}$$

exists, we say ψ *is semidifferentiable at* $\bar{z}$ *for* w *(or Hadamard differentiable at* $\bar{z}$ *in direction* w*). Further if* ψ *is semidifferentiable at* $\bar{z}$ *for every* w*, we say that* ψ *is semidifferentiable at* $\bar{z}$*. It is easy to see that if* ψ *is semidifferentiable at* $\bar{z}$ *for* w*, then*

$$\mathrm{d}(-\psi)(\bar{z})(w) = -\mathrm{d}\psi(\bar{z})(w). \tag{1.4.2}$$

When ψ *is Lipschitz continuous, the Clarke generalized directional derivative of* ψ *at* $\bar{z}$ *along the direction* w *is defined as*

$$\psi^\circ(\bar{z}; w) := \limsup_{\substack{t\downarrow 0\\ z'\to \bar{z}}} \frac{\psi(z'+tw)-\psi(z')}{t}.$$

Moreover, the Clarke generalized gradient is defined as

$$\partial^\circ\psi(\bar{z}) := \{\xi \in \mathbb{R}^r | \langle \xi, w\rangle \le \psi^\circ(\bar{z}; w) \quad \forall w \in \mathbb{R}^r\}.$$

We know from [16, page 10] that

$$\psi^\circ(\bar{z}; w) = \max_{\xi \in \partial^\circ\psi(\bar{z})} \langle \xi, w\rangle. \tag{1.4.3}$$

By [73, Theorem 7.21], if ψ is semidifferentiable at $\bar{z}$, then $\mathrm{d}\psi(\bar{z})(w)$ is finite for any $w \in \mathbb{R}^r$, ψ is continuous at $\bar{z}$ and $w \to \mathrm{d}\psi(\bar{z})(w)$ is homogeneous and continuous. When ψ is semidifferentiable at $\bar{z}$ for w, ψ is also directionally differentiable at $\bar{z}$ in direction w and the subderivative/superderivative coincides with the classical directional derivative at $\bar{z}$ in direction w. That is,

$$\mathrm{d}\psi(\bar{z})(w) = \mathrm{d}^+\psi(\bar{z})(w) = \psi'(\bar{z}; w) := \lim_{t \downarrow 0} \frac{\psi(\bar{z} + tw) - \psi(\bar{z})}{t}. \tag{1.4.4}$$

Moreover when ψ is Lipschitz around $\bar{z}$, the directional differentiability is equivalent to the semidifferentiability. Hence when ψ is Lipschitz and semidifferentiable around $\bar{z}$, we have

$$\psi^\circ(\bar{z}; w) := \limsup_{\substack{t \downarrow 0 \\ z \to \bar{z}}} \frac{\psi(z + tw) - \psi(z)}{t} \geq \limsup_{t \downarrow 0} \frac{\psi(\bar{z} + tw) - \psi(\bar{z})}{t} = \mathrm{d}\psi(\bar{z})(w).$$

If ψ is continuously differentiable at $\bar{z}$, it is semidifferentiable at $\bar{z}$ and for any w, one has $\mathrm{d}\psi(\bar{z})(w) = \psi'(\bar{z}; w) = \nabla\psi(\bar{z})^T w$ [73, Exercise 8.20].

Definition 1.4.4. *[second subderivatives, twice semiderivatives, and twice epi-derivatives, twice semidifferentiability, twice epi-differentiability [73, Definitions 13.3 and 13.6]] Let* $\psi : \mathbb{R}^r \to \overline{\mathbb{R}}$, $\psi(\bar{z})$ *be finite and* $\bar{v}, w \in \mathbb{R}^r$. *The second subderivative of* ψ *at* $\bar{z}$ *for* $\bar{v}$ *and* w *is*

$$\mathrm{d}^2\psi(\bar{z}; \bar{v})(w) := \liminf_{\substack{t \downarrow 0 \\ w' \to w}} \frac{\psi(\bar{z} + tw') - \psi(\bar{z}) - t\langle \bar{v}, w' \rangle}{\frac{1}{2}t^2}.$$

On the other hand, the second subderivative of ψ *at* $\bar{z}$ *for* w *(without mention of* $\bar{v}$*) is defined by*

$$\mathrm{d}^2\psi(\bar{z})(w) := \liminf_{\substack{t \downarrow 0 \\ w' \to w}} \frac{\psi(\bar{z} + tw') - \psi(\bar{z}) - t\mathrm{d}\psi(\bar{z})(w')}{\frac{1}{2}t^2}, \tag{1.4.5}$$

where the sum of ∞ *and* $-\infty$ *is interpreted as* ∞. *The function* ψ *is twice semidifferentiable at* $\bar{z}$ *if it is semidifferentiable at* $\bar{z}$ *and the "liminf" in (1.4.5) is replaced by the "lim" for any* $w \in \mathbb{R}^r$. *The function* ψ *is said to be twice epi-differentiable at* $\bar{z}$ *for* $\bar{v}$ *if for any* $w \in \mathbb{R}^r$ *and any sequence* $t_k \downarrow 0$ *there exists a sequence* $w_k \to w$ *such that*

$$\mathrm{d}^2\psi(\bar{z}; \bar{v})(w) = \lim_{k \to \infty} \frac{\psi(\bar{z} + t_k w_k) - \psi(\bar{z}) - t_k\langle \bar{v}, w_k \rangle}{\frac{1}{2}t_k^2}. \tag{1.4.6}$$

When ψ is twice semidifferentiable at $\bar{z}$ for w, ψ is also twice directionally differentiable at $\bar{z}$ in direction w and the second subderivative coincides with the classical second directional derivative at $\bar{z}$ in direction w. That is,

$$\mathrm{d}^2\psi(\bar{z})(w) = \psi''(\bar{z}; w) := \lim_{t\downarrow 0} \frac{\psi(\bar{z}+tw) - \psi(\bar{z}) - t\psi'(\bar{z}; w)}{\frac{1}{2}t^2}. \tag{1.4.7}$$

It is easy to see that if ψ is twice semidifferentiable at $\bar{z}$, then

$$\mathrm{d}^2(-\psi)(\bar{z})(w) = -\mathrm{d}^2\psi(\bar{z})(w). \tag{1.4.8}$$

With twice epi-differentiability, the function ψ is properly twice epi-differentiable at $\bar{z}$ for $\bar{v}$ if in addition the function $\mathrm{d}^2\psi(\bar{z}; \bar{v})$ is proper [73, Definition 13.6]. By [73, Exercise 13.7], if ψ is twice semidifferentiable at $\bar{z}$, then $\mathrm{d}^2\psi(\bar{z})(w)$ is finite for any $w \in \mathbb{R}^r$. If ψ is twice continuously differentiable at $\bar{z}$, then it is twice semidifferentiable at $\bar{z}$ and for $\bar{v} = \nabla\psi(\bar{z})$, and one has $\mathrm{d}^2\psi(\bar{z}; \bar{v})(w) = \mathrm{d}^2\psi(\bar{z})(w) = w^T\nabla^2\psi(\bar{z})w$ [73, Example 13.8].

For a function $\psi(x, y) : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ and $w = (u, h) \in \mathbb{R}^n \times \mathbb{R}^m$, we denote the subderivative of ψ at $(\bar{x}, \bar{y})$ with respect to x for u and y for h by $\mathrm{d}_x\psi(\bar{x}, \bar{y})(u)$ and $\mathrm{d}_y\psi(\bar{x}, \bar{y})(h)$, respectively. Similarly we denote the second subderivative of ψ at $(\bar{x}, \bar{y})$ with respect to x for u and y for h by $\mathrm{d}^2_{xx}\psi(\bar{x}, \bar{y})(u)$ and $\mathrm{d}^2_{yy}\psi(\bar{x}, \bar{y})(h)$, respectively. When ψ is twice semidifferentiable, it is also semidifferentiable and by (1.4.4) we have

$$\mathrm{d}\psi(\bar{x}, \bar{y})(0, h) = \lim_{t\downarrow 0} \frac{\psi(\bar{x}, \bar{y}+th) - \psi(\bar{x}, \bar{y})}{t} = \mathrm{d}_y\psi(\bar{x}, \bar{y})(h)$$

and by (1.4.7)

$$\begin{aligned} \mathrm{d}^2\psi(\bar{x}, \bar{y})(0, h) &= \lim_{t\downarrow 0} \frac{\psi(\bar{x}, \bar{y}+th) - \psi(\bar{x}, \bar{y}) - t\mathrm{d}\psi(\bar{x}, \bar{y})(0, h)}{\frac{1}{2}t^2} \\ &= \lim_{t\downarrow 0} \frac{\psi(\bar{x}, \bar{y}+th) - \psi(\bar{x}, \bar{y}) - t\mathrm{d}_y\psi(\bar{x}, \bar{y})(h)}{\frac{1}{2}t^2} \\ &= \mathrm{d}^2_{yy}\psi(\bar{x}, \bar{y})(h). \end{aligned} \tag{1.4.9}$$

Now, we consider the constraint system

$$S = \{z \in \mathbb{R}^r | g(z) \in \Sigma\}, \tag{1.4.10}$$

where $g : \mathbb{R}^r \to \mathbb{R}^q$ and $\Sigma \subseteq \mathbb{R}^q$ is closed.

Definition 1.4.5 (metric subregularity constraint qualification)**.** *Let $\bar{z} \in S$ where S is the constraint system defined by (1.4.10). We say that the metric subregularity constraint qualification (MSCQ) for S holds at $\bar{z}$ if there exist a neighborhood U of $\bar{z}$ and a constant $\rho > 0$ such that*

$$\operatorname{dist}(z, S) \leq \rho \operatorname{dist}(g(z), \Sigma) \qquad \forall z \in U.$$

Sufficient conditions for MSCQ of the equality and inequality system can be found in [86, Theorem 7.4], e.g., the first-order sufficient condition for metric subregularity (FOSCMS), the second-order sufficient condition for metric subregularity (SOSCMS), the Mangasarian-Fromovitz constraint qualification (MFCQ), and the linear constraint qualification, i.e., g is affine and Σ is the union of finitely many polyhedral convex sets.

Chapter 2

Combined approach with second-order optimality conditions for bilevel programming problems

In this chapter, we study new combined approaches for bilevel problems, and discuss constraint qualifications and optimality conditions for the combined problem. In Section 2.1, we introduce bilevel problems and state the motivation of our work. 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.

All work in this chapter has been published as a journal paper, see [51].

2.1 Introduction

In this chapter we consider the bilevel programming problem (BLPP). For convenience, we denote the feasible set of (P(x)) by

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

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 [60], hyperparameter optimization and meta-learning in machine learning [27, 43, 48, 82]. More applications can be found in [6, 18, 22, 74]. For a comprehensive review, we refer to [21] 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 [60] 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 [19]. Recently [9] 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:

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

to replace the lower level program have been investigated; see [1, 32, 33, 42]. Calmness properties for the KKT reformulation and the B-stationarity reformulation have been compared in [1, 32] 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.

Contrast to any reformulation using optimality conditions for the lower level program, the value function approach proposed by Outrata [69] for numerical purpose and used by Ye and Zhu [83] for optimality conditions does not require any extra assumptions. By this approach, one defines the value function as an extended real-valued function

$$V(x) := \inf_{y} \left\{ f(x,y) | g(x,y) \leq 0 \right\},$$

and replaces the original BLPP by the following equivalent problem:

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

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 [83, Proposition 3.2]. To derive necessary optimality conditions for BLPPs, Ye and Zhu [83, 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 [83] 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. [20, 38, 57, 62, 80]). To improve the value function approach, Ye and Zhu [84] proposed a combination of the classical KKT and the value function approach. The resulting problem is the combined problem using KKT condition:

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

Problem (CP) is equivalent to problem (BLPP) (in the sense of [84, Proposition 3.1]) when the KKT condition holds at each optimal solution of the lower level program. Similar to [83], to deal with the fact that the nonsmooth MFCQ also fails for (CP), the corresponding partial calmness condition for (CP) was proposed in [84, 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. [42] proposed the following combined program using the FJ condition:

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

The equivalence of problem (CPFJ) and problem (BLPP) (for both local and global optimal points) is a special case of [9, Theorem 4.5].

Similar to Ye and Zhu's work in [84], Ke et al. proposed the following partial calmness condition for (CPFJ) in [42].

Definition 2.1.1 (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* $\mu\ge 0$ *such that* $(\bar{x},\bar{y},\bar{u}_0,\bar{u})$ *is a local solution of the partially penalized problem:*

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

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 [42].

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

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

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

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 2.3.1, 2.4.1, 2.4.2 in this chapter. The main goal of this chapter is to investigate the following question:

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

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:

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

where

$$\Sigma_{\text{KKT}} := \left\{ (x,y) \;\middle|\; \begin{array}{ll} \exists u \text{ s.t.} & \nabla_y f(x,y)+\nabla_y g(x,y)u=0, \\ & g(x,y)\le 0,\ u\ge 0,\ u^T g(x,y)=0 \end{array} \right\}$$

and its partially penalized problem:

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

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:

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

where

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

and its partially penalized problem:

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

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

$$\operatorname{gph} S := \big\{(x,y) \in \mathbb{R}^{n+m} | y \in S(x)\big\} \subseteq \Sigma_{\text{SOC}} \subseteq \Sigma_{\text{KKT}}. \qquad (2.1.1)$$

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 $\operatorname{gph} S$. In particular, if it happens that $\Sigma_{\text{SOC}} = \operatorname{gph} S$, 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.

We organize this chapter as follows. In Section 2.2, we gather some preliminaries and preliminary results that will be used later. An illustrative example will be given in Section 2.3. In Section 2.4, 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 2.5.

2.2 Preliminaries and preliminary results

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

2.2.1 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:

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

where $f : \mathbb{R}^m \to \mathbb{R}$, $g : \mathbb{R}^m \to \mathbb{R}^{p_1}$, $h : \mathbb{R}^m \to \mathbb{R}^{p_2}$ 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:

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

We define the critical cone at t as follows:

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

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

$$\mathcal{C}(t) = \left\{ d \;\middle|\; \begin{array}{l} Dg_j(t)d = 0 \text{ if } u_j > 0,\ Dg_j(t)d \leq 0 \text{ if } u_j = 0,\ \forall\, j \in J_0(t), \\ Dh_i(t)d = 0, i = 1, ..., p_2 \end{array} \right\}. \tag{2.2.1}$$

Another important set is the critical subspace given by

$$\mathcal{S}(t) := \Big\{ d \in \mathbb{R}^m | Dg_j(t)d = 0,\ \forall\, j \in J_0(t), Dh_i(t)d = 0, i = 1, ..., p_2 \Big\}. \tag{2.2.2}$$

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 2.2.1. *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* $d^T \nabla^2_{tt} L(t, u, v) d \geq 0$;

(ii) the weak second-order optimality condition (WSOC) *holds at* t, *if there exists* $(u, v) \in M^1(t)$ *such that* $d^T \nabla^2_{tt} L(t, u, v) d \geq 0, \forall\, d \in \mathcal{S}(t)$;

(iii) the strong second-order optimality condition (SSOC) *holds at* t, *if there exists* $(u, v) \in M^1(t)$ *such that* $d^T \nabla^2_{tt} L(t, u, v) d \geq 0, \forall\, d \in \mathcal{C}(t)$.

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., [12, Proposition 5.48]. Recently under a much weaker constraint qualification called the directional metrical subregularity condition [31, 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) [59, 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 [58, Theorem 2.1]. Recently, it was shown that WSOC holds under MFCQ plus the weak constant rank property [7, 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 2.2.1. *[12, 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*

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

Since we will use the above concepts for the lower level program of BLPPs frequently, we give the following notation. For fixed upper level 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 $z := (z_1, \ldots, z_p)^T \in \mathbb{R}^p$ for the lower level program, and obtain

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

Here, $z^2 := (z_1^2, \ldots, z_p^2)^T$. 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 $(\widetilde{P}(x))$. Conversely, if (y^*, z^*) is a global (local) optimal solution of $(\widetilde{P}(x))$, then y^* is a global (local) optimal solution of (P(x)).

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

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

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

$$\begin{aligned} &\nabla_y f(x,y) + \sum_{i=1}^p u_i \nabla_y g_i(x,y) = 0, \\ &u_i z_i = 0, \ g_i(x,y) + z_i^2 = 0, \ i = 1, \ldots, p. \end{aligned}$$

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

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

$$\mathcal{C}(y,z;x) = \mathcal{S}(y,z;x) := \Big\{(d,\nu) \in \mathbb{R}^m \times \mathbb{R}^p | D_y g_i(x,y)d + 2z_i\nu_i = 0, \, \forall\, i\Big\}. \qquad (2.2.3)$$

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

$(\widetilde{P}(x))$. We say that SOC holds at (y,z,u) if

$$(d,\nu)^T \nabla^2_{(y,z)} L(y,z,u;x)(d,\nu) \geq 0, \quad \forall\,(d,\nu) \in \mathcal{C}(y,z;x). \tag{2.2.4}$$

Note that

$$\nabla^2_{(y,z)} L(y,z,u;x) = \begin{pmatrix} \nabla^2_{yy} L(y,u;x) & 0 \\ 0 & 2\mathrm{diag}(u) \end{pmatrix},$$

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

$$(d,\nu)^T \nabla^2_{(y,z)} L(y,z,u;x)(d,\nu) = d^T \nabla^2_{yy} L(y,u;x) d + 2\sum_{i=1}^{p} u_i \nu_i^2. \tag{2.2.5}$$

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

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

By (2.2.3) and (2.2.2), we have

$$(d,\nu) \in \mathcal{S}(y,z;x) \Longrightarrow d \in \mathcal{S}(y;x).$$

Since $u \geq 0$ for KKT pair (y^*,u) of $(\mathrm{P}(x))$, by (2.2.5), the following result is valid.

Proposition 2.2.1. *Let (y^*,u) be a KKT pair of* $(\mathrm{P}(x))$. *Then there exists z^* such that (y^*,z^*,u) is a KKT triple of $(\widetilde{P}(x))$. Furthermore, if (y^*,u) satisfies* WSOC *for* $(\mathrm{P}(x))$, *then (y^*,z^*,u) satisfies* SOC (2.2.4) *for $(\widetilde{P}(x))$.*

But the converse is not always true, that is, even if (y^*,z^*,u) is a KKT triple of $(\widetilde{P}(x))$, (y^*,u) is not necessarily a KKT pair of $(\mathrm{P}(x))$. In fact, the condition $u \geq 0$, concerning the sign of the multiplier, may not hold. For a counterexample, we refer the reader to [28, Example 3.2]. Under the second-order sufficient conditions and some constraint qualification, it has been proved that KKT points of the original $(\mathrm{P}(x))$ and the reformulated $(\widetilde{P}(x))$ problems are essentially equivalent, cf. [28, Proposition 3.6]. Moreover in the final remarks of [28], 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 [28, Proposition 3.6].

Proposition 2.2.2. *Let* (y^*, z^*, u^*) *be a KKT triple of* $(\widetilde{P}(x))$. *Assume that* (y^*, z^*, u^*) *satisfies SOC* (2.2.4). *Then* $u_i^* \geq 0$ *for all* $i = 1, \ldots, p$. *Hence* (y^*, u^*) *is a KKT pair of* (P(x)) *satisfying* WSOC.

Proof. First, since (y^*, z^*, u^*) is a KKT triple of $(\widetilde{P}(x))$, we have $u_i^* z_i^* = 0$ for all $i = 1, \ldots, p$. 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 (2.2.3), we have $(d^*, \nu^*) \in \mathcal{S}(y^*, z^*; x)$. By (2.2.5), we have

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

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 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 (2.2.5)

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

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. □

2.2.2 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 [30] below. For results under weaker assumptions and sharper upper estimates, the reader is referred to [37, Corollary 4.8] and [80, 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. [80, Proposition 1] for the case where the lower level program is linear, and [63, 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 2.2.3. *[30, 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:*

$$\partial^{\circ}V(\bar{x}) \subseteq \mathrm{co}\{\nabla_x f(\bar{x}, y') + \nabla_x g(\bar{x}, y')u' | y' \in S(\bar{x}), u' \in M^1(y'; \bar{x})\}.$$

2.2.3 Constraint qualifications and optimality conditions for the combined problem

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

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

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $u \in \mathbb{R}^p$, $w \in \mathbb{R}^l$ and the mappings $F, f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$, $H_2 : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^l \to K$ 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 2.2.2 (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 $\mu \ge 0$ such* *that* $(\bar{x}, \bar{y}, \bar{u}, \bar{w})$ *is a local solution of the following partially penalized problem:*

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

For any feasible point of problem (GCP$_\mu$), say (x, y, u, w), the triplet (x, y, u) always satisfies the KKT conditions of the lower-level problem, which means (x, y, u) must also be a feasible point of the partially penalized problem corresponding to (CP). Thus, the partial calmness condition for problem (GCP) is more likely to hold than that for problem (CP).

Proposition 2.2.4. *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 [83, Proposition 3.2]. Furthermore, the complementarity constraint remains present in the partially penalized problem, which contributes to the violation of the MFCQ, as discussed in [85, Proposition 1.1].

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 [79, 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 2.2.3 (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*

$$0 \in \Big[\nabla F(\bar{x}, \bar{y}) + \mu(\nabla f(\bar{x}, \bar{y}) - \partial^\circ V(\bar{x}) \times \{0\}) + \nabla g(\bar{x}, \bar{y})\lambda^g\Big] \times \big\{(0,0)\big\} \tag{2.2.6}$$

$$- (0, 0, \lambda^u, 0) + \nabla H_1(\bar{x}, \bar{y}, \bar{u})\lambda_1^H + \nabla H_2(\bar{x}, \bar{y}, \bar{u}, \bar{w})\lambda_2^H,$$

$$\lambda_j^g = 0,\ \forall j \in I_u, \quad \lambda_j^u = 0,\ \forall j \in I_g, \quad \lambda_2^H \in N_C(H_2(\bar{x}, \bar{y}, \bar{u}, \bar{w})), \tag{2.2.7}$$

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

(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* (2.2.6)−(2.2.7) *and the following condition hold:*

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

By Definition 2.2.3, 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:

$$\begin{aligned} \min_{x,y,u,w}\ & F(x, y) \\ \text{s.t. } & f(x, y) - V(x) \leq 0,\ (g(x, y), -u) \in \Omega_{\text{CS}}^p, \\ & H_1(x, y, u) = 0,\ H_2(x, y, u, w) \in C, \end{aligned} \tag{2.2.8}$$

where $\Omega_{\text{CS}}^p := \big\{(a, b) \in \mathbb{R}^p \times \mathbb{R}^p | a \leq 0, b \leq 0, \langle a, b\rangle = 0\big\}$ is the negative complementarity set.

Denote the set of feasible solutions for problem (2.2.8) by $\mathcal{F}$ and the perturbed feasible map by

$$\mathcal{F}(r_1, r_2, r_3, P_1, P_2) := \left\{ (x, y, u, w) \;\middle|\; \begin{array}{l} f(x, y) - V(x) + r_1 \leq 0, \\ (g(x, y) - r_2, -u + r_3) \in \Omega^p_{\mathrm{CS}}, \\ H_1(x, y, u) + P_1 = 0, \\ H_2(x, y, u, w) + P_2 \in C \end{array} \right\}. \quad (2.2.9)$$

We now define the Clarke calmness for problem (GCP) as the one for its equivalent reformulation (2.2.8) as follows.

Definition 2.2.4. *(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* $B_\epsilon(0)$*, for all* $(x, y, u, w) \in B_\epsilon(\bar{x}, \bar{y}, \bar{u}, \bar{w}) \cap \mathcal{F}(r_1, r_2, r_3, P_1, P_2)$*, one has*

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

Similar to Burke [14, Theorem 1.1], the Clarke calmness defined in Definition 2.2.4 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:

$$\begin{aligned} \min_{x,y,u,w} F(x, y) + \mu \; (f(x, y) - V(x) + \sum_{i=1}^{m} |H_{1,i}(x, y, u)| \\ + \operatorname{dist}((g(x, y), -u), \Omega^p_{\mathrm{CS}}) + \operatorname{dist}(H_2(x, y, u, w), C)). \end{aligned}$$

It is well-known that the calmness of the perturbed feasible map (2.2.9) 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. [24, Proposition 2.2]. Moreover many classical constraint qualifications can be used to guarantee the Clarke calmness at a local minimizer; see e.g. [24, Proposition 2.3].

Similar to the equivalence between the Clarke calmness and exact penalization, it was pointed out in [83, 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 [83, 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) [56, Theorem 3.2], [78, 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 [10], to the general combined problem. To obtain S-stationary condition, we introduce the following constraint qualification.

Definition 2.2.5 (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:*

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

where $\operatorname{span}(\Pi)$ *denotes the affine hull of the set* Π.

Theorem 2.2.2. *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 (2.2.8), by [24, 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 [34] 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 [55, Theorem 3.8] to derive the desired result. □

Remark 2.2.1. *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.*

2.3 An illustrative example

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

Example 2.3.1.

$$\min_{x,y} \left(x-\frac{1}{2}\right)^2 + y^2 \quad \text{s.t.} \quad y \in S(x) := \arg\min_y \left\{ \frac{1}{4}y^4 - \frac{1}{2}xy^2 \;\middle|\; y \in \mathbb{R} \right\}. \qquad (2.3.1)$$

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 2.1.

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 2.2. It is easy to see that

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

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

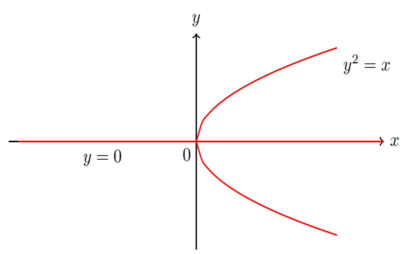

Figure 2.1: Feasible set of problem (2.3.2)

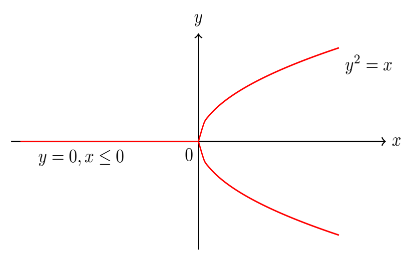

Figure 2.2: Feasible set of problem (2.3.5) (also the graph of $S(\cdot)$)

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

$$\min_{x,y} \left\{ F_\mu(x,y) := \left(x-\frac{1}{2}\right)^2 + y^2 + \mu\left(\frac{1}{4}y^4 - \frac{1}{2}xy^2 - V(x)\right) \;\middle|\; y^3 - xy = 0 \right\}. \qquad (2.3.2)$$

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

$$F_\mu\left(\frac{1}{k},0\right) = k^{-2} - k^{-1} + \frac{\mu}{4}k^{-2} + \frac{1}{4}, \quad F_\mu(0,0) = \frac{1}{4}. \tag{2.3.3}$$

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 (2.3.2). 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 (2.3.1):

$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2 \quad \text{s.t.} \quad f(x,y) - V(x) \leq 0,\ y^3 - xy = 0,\ 3y^2 - x \geq 0. \tag{2.3.4}$$

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 2.2, 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 $\mu = 0$:

$$\min_{x,y} \left(x - \frac{1}{2}\right)^2 + y^2 + \mu\big(f(x,y) - V(x)\big) \quad \text{s.t.} \quad y^3 - xy = 0,\ 3y^2 - x \geq 0. \tag{2.3.5}$$

Problem (2.3.5) 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 2.3.1.

For Example 2.3.1, 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 (2.3.2), whose feasible set is given by Figure 2.1. But by (2.3.3), 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 (2.3.4). The advantage of using the second-order necessary optimality condition is that the feasible set of the new associated partially penalized problem (2.3.5) 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 2.2).

2.4 Combined with second-order optimality conditions

A natural idea that comes from Example 2.3.1 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.

2.4.1 Unconstrained case

For the unconstrained bilevel programming problem

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

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

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

We denote the corresponding partially penalized problem for (CPSOC) (as in Definition 2.2.2) by (CPSOC$_\mu$). The problem (CPSOC$_\mu$) 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 2.4.1. *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*

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

where

$$D\nabla^2_{yy} f(\bar{x}, \bar{y})^* \Omega := \left(\left\langle \frac{\partial}{\partial x_1} \nabla^2_{yy} f(\bar{x}, \bar{y}), \Omega \right\rangle, \ldots, \left\langle \frac{\partial}{\partial y_m} \nabla^2_{yy} f(\bar{x}, \bar{y}), \Omega \right\rangle \right)^T.$$

2.4.2 Constrained case

In the constrained case, as we reviewed in Section 2.2, there are four kinds of second-order optimality conditions: FJSOC, BSOC, SSOC, and WSOC.

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) such that

$$\begin{aligned}
&u_0 \nabla_y f(x, y) + \nabla_y g(x, y) u = 0, \\
&g(x, y) \leq 0, \ (u_0, u) \geq 0, \ \sum_{i=0}^{p} u_i = 1, \ u^T g(x, y) = 0, \\
&d^T \nabla^2_{yy} \mathcal{L}_0(y, u_0, u; x) d \geq 0.
\end{aligned} \tag{2.4.1}$$

By Theorem 2.2.1, if $y \in S(x)$ then y is an FJSOC-point for (P(x)).

Now we define

$$\Sigma_{\text{FJSOC}} := \Big\{ (x, y) \in \mathbb{R}^{n+m} | \forall d \in \mathcal{C}(y; x), \exists (u_0, u) \text{ s.t. (2.4.1) hold.} \Big\},$$

and consider the following combined problem with FJSOC:

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

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:

$$\{d \in \mathbb{R}^m | D_y f(x,y)d \leq 0,\ u_j D_y g_j(x,y)d \leq 0,\ \forall\, j = 1,\ldots,p\} \supseteq \mathcal{C}(y;x), \tag{2.4.2}$$

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:

$$\begin{aligned}
&u_0 \nabla_y f(x,y) + \nabla_y g(x,y)u = 0,\\
&g(x,y) \leq 0,\ (u_0, u) \geq 0,\ \sum_{i=0}^{p} u_i = 1,\ u^T g(x,y) = 0,\\
&d^T \nabla^2_{yy} \mathcal{L}_0(y, u_0, u; x)d \geq 0,\\
&D_y f(x,y)d \leq 0,\ u_j D_y g_j(x,y)d \leq 0,\ \forall\, j = 1,\ldots,p.
\end{aligned} \tag{2.4.3}$$

Denote by

$$K(x,y) := \{(u_0, u, d) \in \Xi(x,y) | u_0 \nabla_y f(x,y) + \nabla_y g(x,y)u = 0\}, \tag{2.4.4}$$

where

$$\Xi(x,y) := \left\{ (u_0,u,d) \;\middle|\; \begin{aligned}
&(u_0, u) \geq 0,\ \sum_{i=0}^{p} u_i = 1,\ u^T g(x,y) = 0,\\
&d^T \nabla^2_{yy} \big[u_0 f(x,y) + \sum_{i=1}^{p} u_i g_i(x,y)\big] d \geq 0,\\
&D_y f(x,y)d \leq 0,\ u_j D_y g_j(x,y)d \leq 0,\ \forall\, j = 1,\ldots,p
\end{aligned} \right\}. \tag{2.4.5}$$

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

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

Problem (FJSOCP-2) can be considered as an optimization problem with implicit variables as studied in recent paper [9]. Since problem (FJSOCP-2) is still not practical to solve, we consider its explicit version—the relaxed combined problem with FJ second-order condition:

$$\begin{aligned}
\min_{x,y,u_0,u,d}\ & F(x,y)\\
\text{s.t. }\ & f(x,y)-V(x)\le 0,\ u_0\nabla_y f(x,y)+\nabla_y g(x,y)u=0,\ g(x,y)\le 0,\\
& (u_0,u)\ge 0,\ \sum_{i=0}^{p}u_i=1,\ u^Tg(x,y)=0,\ d^T\nabla^2_{yy}\big[u_0f(x,y)+\sum_{i=1}^{p}u_ig_i(x,y)\big]d\ge 0,\\
& D_yf(x,y)d\le 0,\ u_jD_yg_j(x,y)d\le 0,\ \forall\, j=1,\dots,p,\ G(x,y)\le 0.
\end{aligned} \tag{R-FJSOCP}$$

Recall that a set-valued map Γ is inner semicompact at $(\bar{x},\bar{y})$ with respect to domΓ if for each sequence $\{(x_k,y_k)\}_{k\in\mathbb{N}}\subseteq$ domΓ such that $(x_k,y_k)\to(\bar{x},\bar{y})$, there is a convergent sequence $\left\{(u_0^\zeta,u^\zeta,d^\zeta)\right\}_{\zeta\in\mathbb{N}}$ and a subsequence $\left\{(x_k^\zeta,y_k^\zeta)\right\}_{\zeta\in\mathbb{N}}$ such that $(u_0^\zeta,u^\zeta,d^\zeta)\in\Gamma\left(x_k^\zeta,y_k^\zeta\right)$ holds for all $\zeta\in\mathbb{N}$; see e.g. [9, page 7]. Note that the mapping Γ defined in (2.4.4) 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 [9, Theorems 4.3 and 4.5] we have the following equivalence between the problem with implicit variables and its explicit form.

Proposition 2.4.1. *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* (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 2.2.2, we can define partial calmness for (FJSOCP) and partial calmness for (R-FJSOCP), and denote the corresponding partially penalized problems by (FJSOCP$_\mu$) and (R-FJSOCP$_\mu$), respectively.

Different from the relation between the partial calmness condition for (CP$_{FJ}$) and the partial calmness condition for (CPFJ) in [42, 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 2.4.2, the partial calmness condition for (CPFJ) implies the partial calmness condition for (R-FJSOCP). Note that this relation coincides with the result in Proposition 2.2.4. On the other hand, since $\Sigma_{\text{FJSOC}}\subseteq\Sigma_{\text{FJ}}$, it is immediate that

$$\text{partial calmness for } (\text{CP}_{FJ}) \implies \text{partial calmness for (FJSOCP)}, \qquad (2.4.6)$$

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 2.4.2. *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. The first assertion can be obtained similarly to Proposition 2.2.4.

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

Note that (R-FJSOCP) and its partially penalized problem (R-FJSOCP$_\mu$) are MPECs. Based on the discussions on the partial calmness condition, Theorem 2.2.2 can be used to derive S-/M- type necessary optimality conditions for problem (R-FJSOCP) under appropriate constraint qualifications.

Combined with the basic second-order optimality condition

As reviewed in Section 2.2, 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 2.2.1(i), 2.2.1(ii), or 2.2.1(iii) holds respectively. Now we define

$$\Sigma_{\text{BSOC}} := \Big\{(x,y) \in \mathbb{R}^{n+m} | y \text{ is a BSOC-point for } P(x)\Big\},$$
$$\Sigma_{\text{WSOC}} := \Big\{(x,y) \in \mathbb{R}^{n+m} | y \text{ is a WSOC-point for } P(x)\Big\},$$
$$\Sigma_{\text{SSOC}} := \Big\{(x,y) \in \mathbb{R}^{n+m} | y \text{ is an SSOC-point for } P(x)\Big\}.$$

It is easily seen that

$$\Sigma_{\text{SSOC}} \subseteq \Sigma_{\text{BSOC}},\ \Sigma_{\text{SSOC}} \subseteq \Sigma_{\text{WSOC}},\ \text{and } \Sigma_{\text{SSOC}} \overset{\text{LICQ}}{=} \Sigma_{\text{BSOC}}. \tag{2.4.7}$$

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 $\Sigma_{SOC} = \Sigma_{BSOC}$ such that it is still an optimization problem with equality and inequality constraints, we relax the critical cone (2.2.1) as

$$\begin{aligned} \mathcal{C}(y;x) &\subseteq \Big\{d \in \mathbb{R}^m | D_y f(x,y)d \leq 0,\ u_j D_y g_j(x,y)d \leq 0,\ \forall\, j = 1,\dots,p\Big\} \\ &= \Big\{d \in \mathbb{R}^m | u_j D_y g_j(x,y)d = 0,\ \forall\, j = 1,\dots,p\Big\}, \end{aligned} \tag{2.4.8}$$

where u is a KKT multiplier and (2.4.8) 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 $\Sigma_{\text{SOC}} = \Sigma_{\text{BSOC}}$:

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

Similar to Proposition 2.4.1, 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 [9, Theorem 4.3]. To state the relationship on local solutions, we define the following mapping:

$$\widetilde{K}(x,y) := \{(u,d) \in \widetilde{\Xi}(x,y) | \nabla_y f(x,y) + \nabla_y g(x,y)u = 0\}, \tag{2.4.9}$$

where

$$\widetilde{\Xi}(x,y) := \left\{ (u,d) \;\middle|\; \begin{array}{l} u \ge 0,\ u^T g(x,y) = 0, \\ d^T \nabla^2_{yy}\big[f(x,y) + \sum_{i=1}^p u_i g_i(x,y)\big] d \ge 0, \\ u_j D_y g_j(x,y) d = 0,\ \forall\, j = 1,\dots,p \end{array} \right\}. \tag{2.4.10}$$

Proposition 2.4.3. *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 $(u,d) \in \widetilde{K}(\bar{x}, \bar{y}), (\bar{x}, \bar{y}, u, d)$ is a local optimal solution of* (R-BSOCP). *Conversely, let $(\bar{x}, \bar{y}, u, d)$ be a local optimal solution to* (R-BSOCP) *for each $(u,d) \in \widetilde{K}(\bar{x}, \bar{y})$, Furthermore, let $\widetilde{K}$ be inner semicompact at $(\bar{x}, \bar{y})$ with respect to* dom$\widetilde{K}$. *Then $(\bar{x}, \bar{y})$ is a local solution of* (BLPP).

Remark 2.4.1. *The function $\widetilde{K}$ defined in* (2.4.9) *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 [29].*

Next, we study the relation between the partial calmness for (R-BSOCP) and the partial calmness for (CP). Similar to Proposition 2.4.2, we can prove the following proposition.

Proposition 2.4.4. *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$_\mu$) are MPECs. Theorem 2.2.2 can be used to derive S-/M- type necessary optimality conditions under appropriate constraint qualifications.

Combined with the weak second-order optimality condition

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

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

and propose the corresponding partial calmness condition. Here $0 \preceq \nabla^2_{yy}\big[f(x,y) + \sum_{i=1}^p u_i g_i(x,y)\big]\Big|_\Gamma$, with $\Gamma := \mathcal{S}(y;x)$ means that

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

i.e., the matrix $\nabla^2_{yy}\big[f(x,y) + \sum_{i=1}^p u_i g_i(x,y)\big]$ 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 ($\widetilde{P}(x)$) by introducing the squared slack variables is very useful. Indeed, by Propositions 2.2.1 and 2.2.2, problem (WSOCP) is equivalent to the following reformulated problem by introducing the squared slack variables:

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

Now it is worth noting that the critical subspace

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

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 [44, 54] which has slightly different properties than an MPEC.

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:

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

Recall that for a closed convex cone Γ, the class of all Γ-copositive matrices is the dual cone of the convex hull of $\big\{dd^T \in \mathbb{S}^m_+ | d \in \Gamma \subseteq \mathbb{R}^m\big\}$ [25, Lemma 2.28]. This provides a natural generalization of the constraint $\nabla^2_{yy} f(x,y) \in \mathbb{S}^m_+$ in the unconstrained case. The problem (SSOCP) can be viewed as generalized semi-infinite programming problem [76, 81] or generalized copositive programming problem (set-semidefinite optimization) [13, 25]. Similarly as in section 2.4.2, to deal with the difficulty of the active index set, one can also use the squared-slack-variable trick here.

2.4.3 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 [42, 83, 84], we have proposed various partial calmness conditions based on the combined problems above. We summarize the relationships between various partial calmness conditions in Figure 2.3.

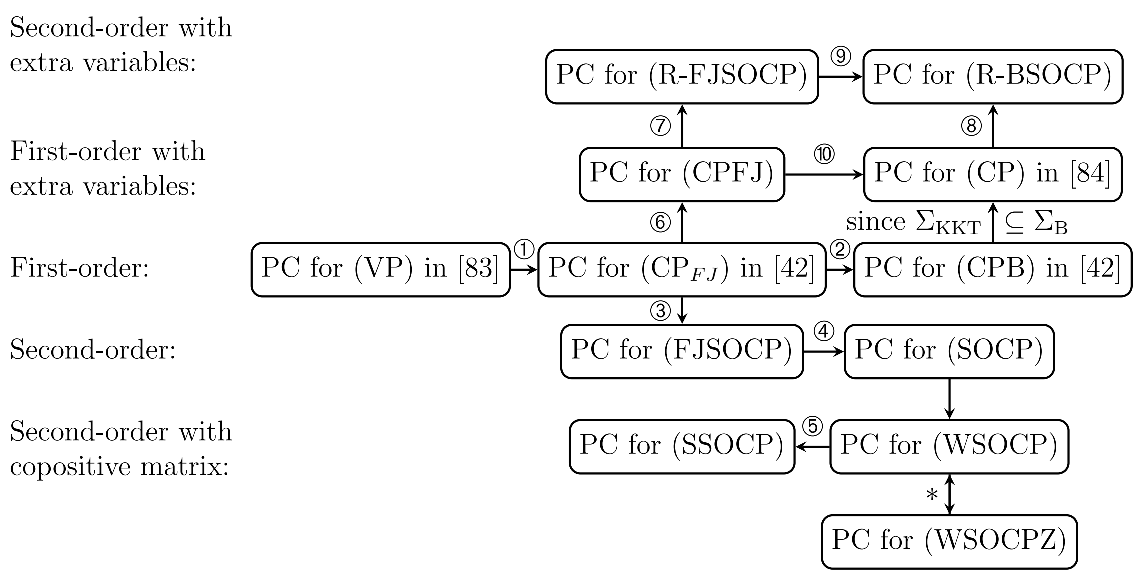

Figure 2.3: Relationship between various partial calmness conditions. Here we denote "partial calmness" briefly by PC. By Proposition 2.2.4, we have relations ①-⑤. For relations ⑥, ⑦, and ⑧, we refer the reader to [42, Theorem 4.4], Proposition 2.4.2, and Proposition 2.4.4, respectively. One may prove other relations by a similar argument of the proof of Proposition 2.4.2. The equivalent relation $*$ follows from Propositions 2.2.1 and 2.2.2. 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 $y \in S(x)$.

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 $\widetilde{u} \in M^1(\bar{y}; \bar{x})$, it can be shown that the partial calmness for (CP) holds for $(\bar{x}, \bar{y}, \widetilde{u} + k\bar{u})$ when $k > 0$ is sufficiently large. This comes from the fact that $1 + \sum_{j=1}^{p}(\widetilde{u}_j + k\bar{u}_j) = k + 1 + \sum_{j=1}^{p} \widetilde{u}_j$, $(1, \widetilde{u} + k\bar{u})/(1 + \sum_{j=1}^{p}(\widetilde{u}_j + k\bar{u}_j)) \to (0, \bar{u})$ as $k \to +\infty$.

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 [42, 84] 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.

Example 2.4.1.

$$\begin{aligned} &\min_{x,y \in \mathbb{R}} \; y^2 - x \\ &\text{s.t.} \; -1 \le x \le 1, \; y \in S(x) := \arg\min_{y} \left\{ \frac{1}{4}y^4 - \frac{1}{2}xy^2 \;\middle|\; 0 \le y \le \sqrt{2} \right\}. \end{aligned} \tag{2.4.11}$$

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

$$S(x) = \begin{cases} \{\sqrt{2}\} & \text{if } x > 2, \\ \{\sqrt{x}\} & \text{if } 0 < x \le 2, \\ \{0\} & \text{if } x \le 0, \end{cases} \qquad V(x) = \begin{cases} 1 - x & \text{if } x > 2, \\ -\frac{1}{4}x^2 & \text{if } 0 < x \le 2, \\ 0 & \text{if } x \le 0, \end{cases} \tag{2.4.12}$$

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*

$$\begin{aligned} \min_{x,y,u} \ & F_\mu(x, y) := y^2 - x + \mu\left(\frac{1}{4}y^4 - \frac{1}{2}xy^2 - V(x)\right) \\ \text{s.t. } & y^3 - xy - u_1 + u_2 = 0,\ u_1 \geq 0,\ -u_1 y = 0, \\ & \quad u_2 \geq 0, u_2(y - \sqrt{2}) = 0,\ 0 \leq y \leq \sqrt{2},\ -1 \leq x \leq 1. \end{aligned} \tag{2.4.13}$$

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 (2.4.13) 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

$$\mathcal{C}(y; x) = \begin{cases} \mathbb{R}_+ & \text{if } x \in \mathbb{R}, y = 0, \\ \mathbb{R} & \text{if } x \in (0, 2), y = \sqrt{x}, \\ \{0\} & \text{if } x \geq 2, y = \sqrt{2}. \end{cases} \tag{2.4.14}$$

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

$$\begin{aligned} \min_{x,y} \ & y^2 - x \\ \text{s.t. } & \frac{1}{4}y^4 - \frac{1}{2}xy^2 - V(x) \leq 0,\ (x, y) \in \Sigma_{\text{SSOC}},\ -1 \leq x \leq 1. \end{aligned} \tag{2.4.15}$$

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

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

with a unique multiplier u. *It follows that*

$$\Sigma_{\text{KKT}} = \Big\{(x, 0) | x \in \mathbb{R}\Big\} \cup \Big\{(x, \sqrt{x}) | x \in (0, 2)\Big\} \cup \Big\{(x, \sqrt{2}) | x \geq 2\Big\}. \tag{2.4.16}$$

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* (2.4.12), *we have* $\Sigma_{\text{SSOC}} = \{(x,y)|y \in S(x)\}$. *Hence the value function constraint in problem* (2.4.15) *holds for all* $(x,y) \in \Sigma_{\text{SSOC}}$. *We therefore can remove the value function constraint from problem* (2.4.15). *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):

$$\begin{aligned}
&\min_{x,y,u,d}\ y^2 - x\\
&\quad \text{s.t.} \frac{1}{4}y^4 - \frac{1}{2}xy^2 - V(x) \le 0, y^3 - xy - u_1 + u_2 = 0, 0 \le y \le \sqrt{2},\ -1 \le x \le 1,\\
&\qquad u_1 \ge 0,\ -u_1 y = 0,\ -u_1 d = 0, u_2 \ge 0,\ u_2(y-\sqrt{2}) = 0,\ u_2 d = 0,\\
&\qquad (3y^2 - x)d^2 \ge 0.
\end{aligned} \tag{2.4.17}$$

Let $\bar{d} \neq 0$. *Then for any* d *sufficiently close to* $\bar{d}$, *condition* $(3y^2 - x)d^2 \ge 0$ *is equivalent to* $3y^2 - x \ge 0$. *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 (2.4.14), we can obtain the expression for the critical subspace of the problem* (2.4.11)

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

WSOC states that

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

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 2.2.2. Indeed, there do not exist* $\mu \ge 0$, β, η^g

and η^G *such that*

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

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

Problem (2.4.17) *is an MPEC. The S-stationary condition based on the value function (Definition 2.2.3) holds at* $(\bar{x},\bar{y},\bar{u},\bar{d}) = (0,0,0,1)$. *Indeed, since* $\nabla_{(x,y)}(\nabla^2_{yy}L)(0,0;0)$ $= (-1,0)^T$, *there exists* $\gamma = 1$ *(let other multipliers be all zero) such that*

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

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

Example 2.4.2.

$$\begin{aligned} &\min_{x,y}\ y^2 - x \\ &\text{s.t. } -1 \le x \le 1, \quad y \in S(x) := \arg\min_y \left\{ \frac{1}{4}y^4 - \frac{1}{2}xy^2 \;\middle|\; -1 \le y \le 1 \right\}. \end{aligned} \tag{2.4.18}$$

It is easy to see that

$$S(x) = \begin{cases} \{\pm 1\} & \text{if } x > 1, \\ \{\pm\sqrt{x}\} & \text{if } 0 < x \le 1, \\ \{0\} & \text{if } x \le 0, \end{cases} \tag{2.4.19}$$

$$\Sigma_{\mathrm{KKT}} = \Big\{(x,0) | x \in \mathbb{R}\Big\} \cup \Big\{(x, \pm\sqrt{x}) | x \in (0,1)\Big\} \cup \Big\{(x,\pm 1) | x \ge 1\Big\},$$

and $(\bar{x},\bar{y}) = (0,0)$ *is a global optimal solution. Moreover,* $M^1(0;0) = \{0\}$. *Similarly to Examples 2.3.1 and 2.4.1, 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* $\Sigma_{\mathrm{SOC}} := \Sigma_{\mathrm{BSOC}} = \Sigma_{\mathrm{SSOC}}$ *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 2.4.1, 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* (2.4.18),

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

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 2.2). 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 2.4.1	No	Yes	Yes	Yes	No
Example 2.4.2	No	Yes	Yes	Yes	Yes

Table 2.1: Comparison in the examples. Here we denote (SOCP) with 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?"

2.5 Conclusions

In this chapter, 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.

Chapter 3

Calm local optimality for nonconvex-nonconcave minimax problems

Nonconvex-nonconcave minimax problems have found numerous applications in various fields including machine learning. However, questions remain about what is a good surrogate for local minimax optimum and how to characterize the minimax optimality. Recently Jin, Netrapalli, and Jordan (ICML 2020) introduced a concept of local minimax point and derived optimality conditions for the smooth and unconstrained case. In this chapter, we introduce the concept of calm local minimax point, which is a local minimax point with a calm radius function. With the extra calmness property we obtain first and second-order sufficient and necessary optimality conditions for a very general class of nonsmooth nonconvex-nonconcave minimax problem. Moreover we show that the calm local minimax optimality and the local minimax optimality coincide under a second-order sufficient condition for the inner maximization problem. This equivalence allows us to derive stronger optimality conditions under weaker assumptions for local minimax optimality.

All the content of this chapter has been submitted as a journal paper, see [52].

3.1 Introduction

In this chapter, we consider the following minimax problem

$$\min_{x\in X}\max_{y\in Y} f(x,y), \qquad \text{(Min-Max)}$$

where the objective function $f:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ is possibly a nonsmooth function, the nonempty sets $X\subseteq\mathbb{R}^n$, $Y\subseteq\mathbb{R}^m$ are closed but may be nonconvex. Throughout the chapter, we assume that for each $x\in X$, the inner maximization problem $\max_{y'\in Y} f(x,y')$ has an optimal solution.

There are two ways to understand the minimax problem. One way is nonsequential/simultaneous and leads to the concept of a Nash equilibrium. That is, given x, the function $f(x,y')$ is maximized over $y'\in Y$ while given y, the function $f(x',y)$ is minimized over $x'\in X$. Another way is sequential and leads to the concept of a Stackelberg equilibrium. That is, for any given x, the function $f(x,y')$ is maximized over $y'\in Y$ and the inner maximum value $\max_{y\in Y} f(x',y)$ is minimized over $x'\in X$. In a convex-concave case which means that $f(x,y)$ is convex in x and concave in y and the sets X,Y are convex, by the celebrated minimax theorem [75], the order between the minimization and maximization can be switched and these two concepts are equivalent in the sense that the resulting optimal value of the minimax problems for the two concepts are the same. However for the nonconvex-nonconcave minimax problem where the objective $f(x,y)$ may be nonconvex in x and nonconcave in y and X,Y may not be convex, the order between the minimization and maximization is crucial, these two concepts may be different.

Theoretical investigations and numerical algorithms for convex-concave minimax problems have been extensively studied [23, 47, 64, 65, 66, 75, 88]. Nevertheless, recent advances in machine learning, such as generative adversarial networks [35], adversarial training [53], and reinforcement learning [68], have introduced new challenges for studying nonconvex-nonconcave minimax optimization. In recent years, there are more and more works for nonconvex-nonconcave minimax problems (e.g., [40, 45, 46, 49, 77]).

In most applications of minimax problems, in particular those from the machine learning, the sequential/Stackelberg concept of minimax is required. Denote the value function of the maximization problem as $V(x):=\max_{y\in Y} f(x,y)$. Then the minimax problem is equivalent to minimizing $V(x)$ over $x\in X$, see e.g. [36] . But the value function is global which means one has to solve a nonconcave maximization problem globally. In practice for solving a nonconvex/nonconcave minimization/maximization problem, many algorithms aim to find stationary points or local optimal solutions as local surrogates for optimal solutions. What is an appropriate definition for local surrogates of the sequential/Stackelberg concept of the minimax problem? To answer this question, we should consider some concepts where both x and y are optimized locally. Since we consider problem (Min-Max) as a sequential game, i.e., the second player (the one selects y) can observe the action taken by the first player (the one selects x) and adjust her action accordingly, the radii of the local neighborhoods where maximization or minimization takes place can be different. Based on these considerations, Jin et al. [40] introduced the concept of local minimax points which is equivalent to saying that a point $(\bar{x}, \bar{y})$ is a local minimax point if there exists a radius function $\tau : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\tau(\delta) \to 0$ as $\delta \downarrow 0$ such that $\bar{y}$ is a maximum point of $f(\bar{x}, \cdot)$ on $Y \cap \mathbb{B}_\delta(\bar{y})$, and $\bar{x}$ is a minimum point of the localized optimal value function

$$V_{\tau(\delta)}(x) := \max_{y \in \mathbb{B}_{\tau(\delta)}(\bar{y}) \cap Y} f(x, y)$$

on $X \cap \mathbb{B}_\delta(\bar{x})$, for any small enough δ. They demonstrated that under mild conditions, all stable limit points of gradient descent ascent (GDA) are exactly local minimax points up to some degenerate points. Recently, extensions and optimality conditions for local minimax points for problem (Min-Max) have been studied in [17, 39, 40, 87]. Jin et al. [40] focus on optimality conditions for the smooth unconstrained case, i.e., f is twice continuously differentiable, $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. Dai and Zhang [17] study optimality conditions for the smooth minimax problems with equality and inequality constraints, i.e., f is twice continuously differentiable, X and Y are described by equality and inequality constraints. Jiang and Chen [39] derive optimality conditions for the nonsmooth and constrained case where X and Y are required to be convex. Zhang et al. [87] study the smooth constrained case where X and Y are sometimes assumed to be convex or even the whole space based on the local optimal value function $V_\epsilon(x) := \max_{y \in \mathbb{B}_\epsilon(\bar{y}) \cap Y} f(x, y)$.

3.1.1 Motivating example

In fair classification [67], the objective is to minimize the maximum loss over multiple categories. An example formulation is the problem

$$\min_{x \in \mathbb{R}^n} \max_{i \in \{1,\ldots,m\}} \ell_i(x),$$

where $\ell_i(x)$ represents the loss on category i with x denoting neural network parameters. A reformulation [67] of this problem where Y is a simplex in $\mathbb{R}^m$ is given by the zero-sum game

$$\min_{x \in \mathbb{R}^n} \max_{y \in Y} \sum_{i=1}^{m} y_i \ell_i(x).$$

We observe a subclass of local minimax points in the fair classification problem, whose radius function τ in Definition 3.3.4 has good properties.

Example 3.1.1. *Consider the fair classification problem*

$$\min_{x \in \mathbb{R}} \max_{y \in [0,1]} f(x, y) := y\ell_1(x) + (1-y)\ell_2(x),$$

where $\ell_1(x) = -x^3 + x$ *and* $\ell_2(x) = -x^3$*. Then* $(0,0)$ *is not a local Nash equilibrium (which means that* $\bar{x} = 0$ *is not a local minimizer of* $f(x, 0) = -x^3$ *or* $\bar{y} = 0$ *is not a local minimizer of* $f(0, y)$ *over* $y \in [0, 1]$*), since* $x = 0$ *is not a local minimizer of* $f(x, 0) = -x^3$*. On the other hand, for* $0 < \delta \leq 1$*, we have*

$$\max_{y' \in [0,\delta]} f(x, y') = \begin{cases} -x^3 + \delta x, & \text{if } x \geq 0, \\ -x^3, & \text{if } x < 0. \end{cases}$$

Thus, the point $(0,0)$ *is a local minimax point with* $\tau(\delta) := \delta$ *since for any* $0 < \delta \leq 1$ *and* $|x| \leq \delta \leq \sqrt{\delta}$*,*

$$f(0, y) \leq f(0, 0) \leq \max_{y' \in [0,\delta]} f(x, y').$$

Actually, we find that in many minimax examples presented in the literature, see e.g., [40, Figure 1], [39, Examples 3.21 and A.1], the radius function τ is calm at 0. Moreover, the strict local minimax point defined in [40, Proposition 20] and the differential Stackelberg equilibrium defined in [26, Definition 4] also turn out to be local minimax points with a calm radius function. Recall that a function $\phi(x) : \mathbb{R}^n \to \mathbb{R}$ is said to be calm at $\bar{x}$ if there is $\kappa > 0$ and U, a neighborhood of $\bar{x}$, such that $|\phi(x) - \phi(\bar{x})| \leq \kappa\|x - \bar{x}\|$ for all $x \in U$.

Motivated by the above observations, in this chapter, we introduce a new notion of local minimax point, which requires the radius function τ in the definition of a local minimax point to be calm at 0. Note that this is equivalent to imposing a bound on the ratio of the radii of the local neighborhoods where the maximization and minimization takes place respectively. It turns out that this extra calmness property is essential for us to develop optimality conditions for minimax problems.

3.1.2 Contributions

We summarize our main contributions as follows.

(i) We introduce the concept of the calm local minimax point, and analyze its relationship with various types of minimax points. In particular we give a general sufficient condition under which the calm local minimax optimality and the local minimax optimality coincide. The concept of calm local minimax points alleviates the issue of non-existence in local Nash equilibria and properly reflects the order of the minimization and maximization problems. Furthermore, since the calm local minimax optimality and the local minimax optimality coincide under the second-order sufficient condition for the inner maximization problem, our optimality conditions based on the new concept give a more precise and comprehensive characterization of local minimax points than the existing ones.

(ii) We derive first-order optimality conditions in primal and dual forms for nonsmooth constrained minimax problems, as well as the second-order sufficient and necessary optimality conditions. When reducing the problem to some special cases, e.g., when f is smooth, X and Y are set-constrained systems, we give explicit form for the optimality conditions. Our findings effectively capture the nested structure of minimax problems and offer explicit optimality conditions derived for the initial problem data. We have demonstrated that in the case of a smooth problem with some easily satisfied properties of X and Y and the fulfillment of a second-order sufficient condition for the inner maximization problem, our necessary conditions can be sharper, and our sufficient condition requires weaker assumptions compared to the existing results. In particular, for the smooth and unconstrained minimax problem, our second-order optimality condition recovers the one derived by Jin et al. [40] in the case where $\nabla^2_{yy} f(\bar{x}, \bar{y})$ is negative definite and is sharper otherwise; for the smooth and constrained case with equality and inequality constraints, our second-order optimality condition is sharper and our assumptions are weaker than the ones in Dai and Zhang [17]. Unlike those in Zhang et al. [87], all of our optimality conditions are based on the point $(\bar{x}, \bar{y})$ of concern and not requiring any other information such as the local optimal solutions of the maximization problem on Y. Since we do not require the convexity of the constraint sets X and Y, our optimality conditions apply to more general problems than those in Jiang and Chen [39]. By using an example, we show that even in the smooth and unconstrained case, our optimality condition can be used to rule out the possibility of a non-local minimax point while the one in Jiang and Chen [39] can not.

The remainder of this chapter is organized as follows. In Section 3.2, we introduce the necessary background, including techniques in variational analysis and obtain some preliminary results on optimality conditions for nonsmooth optimization problems. In Section 3.3, we discuss different types of minimax points and introduce a new notion for local optimality in the minimax problem – calm local minimax point. In Section 3.4, we give our new optimality conditions for the general case. In Section 3.5, we focus on some special cases and compare our results with existing results in the literature.

3.2 Preliminary results

In this section, we will first introduce some preliminary materials and results that will be instrumental in deriving optimality conditions for the minimax problem.

3.2.1 Variational analysis

Definition 3.2.1. *Let $\psi(x, y) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. We say that the separation property holds for the subderivative of ψ at $(\bar{x}, \bar{y})$ if for all $(u, h) \in \mathbb{R}^n \times \mathbb{R}^m$,*

$$\mathrm{d}\psi(\bar{x}, \bar{y})(u, h) = \mathrm{d}_x\psi(\bar{x}, \bar{y})(u) + \mathrm{d}_y\psi(\bar{x}, \bar{y})(h). \tag{3.2.1}$$

We say that the separation property holds for the second subderivative of ψ at $(\bar{x}, \bar{y})$ if the separation property holds for the subderivative of ψ at $(\bar{x}, \bar{y})$ and for all $(u, h) \in$

$\mathbb{R}^n \times \mathbb{R}^m$,

$$\mathrm{d}^2\psi(\bar{x},\bar{y})(u,h) = 2\mathrm{d}^2_{xy}\psi(\bar{x},\bar{y})(u,h) + \mathrm{d}^2_{xx}\psi(\bar{x},\bar{y})(u) + \mathrm{d}^2_{yy}\psi(\bar{x},\bar{y})(h), \tag{3.2.2}$$

where

$$\mathrm{d}^2_{xy}\psi(\bar{x},\bar{y})(u,h) := \liminf_{\substack{t\downarrow 0\\ u'\to u, h'\to h}} \frac{\mathrm{d}_y\psi(\bar{x}+tu',\bar{y})(h') - \mathrm{d}_y\psi(\bar{x},\bar{y})(h')}{t}.$$

The separation property will be useful in deriving the optimality conditions. Next, we give a sufficient condition for the separation property.

Proposition 3.2.1. *Suppose that ψ is semidifferentiable around $(\bar{x},\bar{y})$. Moreover suppose that either $\mathrm{d}_y\psi(x,\bar{y})(h)$ is a continuous function of x around $\bar{x}$ for all h or $\mathrm{d}_x\psi(\bar{x},y)(u)$ is a continuous function of y around $\bar{y}$ for all u. Then the separation property holds for the subderivative of ψ at $(\bar{x},\bar{y})$.*

Proof. By the definition, we have

$$\begin{aligned}
\mathrm{d}\psi(\bar{x},\bar{y})(u,h) &= \lim_{t\downarrow 0}\frac{\psi(\bar{x}+tu,\bar{y}+th)-\psi(\bar{x},\bar{y})}{t}\\
&= \lim_{t\downarrow 0}\frac{\psi(\bar{x}+tu,\bar{y}+th)-\psi(\bar{x}+tu,\bar{y})-\psi(\bar{x},\bar{y}+th)+\psi(\bar{x},\bar{y})}{t}\\
&\quad+\lim_{t\downarrow 0}\frac{\psi(\bar{x}+tu,\bar{y})-\psi(\bar{x},\bar{y})}{t}+\lim_{t\downarrow 0}\frac{\psi(\bar{x},\bar{y}+th)-\psi(\bar{x},\bar{y})}{t}\\
&= \mathrm{d}_x\psi(\bar{x},\bar{y})(u)+\mathrm{d}_y\psi(\bar{x},\bar{y})(h).
\end{aligned}$$

The last equality holds by [73, Theorem 7.21], i.e.,

$$\begin{aligned}
\psi(\bar{x}+tu,\bar{y}+th)-\psi(\bar{x}+tu,\bar{y}) &= \mathrm{d}_y\psi(\bar{x}+tu,\bar{y})(th)+o(t) = t\mathrm{d}_y\psi(\bar{x}+tu,\bar{y})(h)+o(t),\\
\psi(\bar{x},\bar{y}+th)-\psi(\bar{x},\bar{y}) &= \mathrm{d}_y\psi(\bar{x},\bar{y})(th)+o(t) = t\mathrm{d}_y\psi(\bar{x},\bar{y})(h)+o(t).
\end{aligned}$$

Similarly, if ψ is semidifferentiable around $(\bar{x},\bar{y})$ and $\mathrm{d}_x\psi(\bar{x},y)(u)$ is a continuous function of y, then (3.2.1) holds. □

Proposition 3.2.2. *Suppose that ψ is twice semidifferentiable around $(\bar{x},\bar{y})$, the separation property holds for the subderivative of ψ at $(\bar{x},\bar{y})$, and $\mathrm{d}^2_{xy}\psi(\bar{x},\bar{y})(u,h) = \mathrm{d}^2_{yx}\psi(\bar{x},\bar{y})(u,h)$. Suppose further that either $\mathrm{d}^2_{yy}\psi(x,\bar{y})(h)$ is a continuous function of x around $\bar{x}$ for all h or $\mathrm{d}^2_{xx}\psi(\bar{x},y)(u)$ is a continuous function of y around $\bar{y}$ for all u. Then the separation property holds for the second subderivative of ψ at $(\bar{x},\bar{y})$.*

Proof. By the definition, we have

$$
\begin{aligned}
&\mathrm{d}^2\psi(\bar{x},\bar{y})(u,h)\\
&= \lim_{t\downarrow 0} \frac{\psi(\bar{x}+tu,\bar{y}+th)-\psi(\bar{x},\bar{y})-t\mathrm{d}\psi(\bar{x},\bar{y})(u,h)}{\frac{1}{2}t^2}\\
&= \lim_{t\downarrow 0} \frac{\psi(\bar{x}+tu,\bar{y}+th)-\psi(\bar{x}+tu,\bar{y})-\psi(\bar{x},\bar{y}+th)+\psi(\bar{x},\bar{y})}{\frac{1}{2}t^2}\\
&\quad+\lim_{t\downarrow 0} \frac{\psi(\bar{x}+tu,\bar{y})-\psi(\bar{x},\bar{y})-t\mathrm{d}_x\psi(\bar{x},\bar{y})(u)}{\frac{1}{2}t^2}+\lim_{t\downarrow 0} \frac{\psi(\bar{x},\bar{y}+th)-\psi(\bar{x},\bar{y})-t\mathrm{d}_y\psi(\bar{x},\bar{y})(h)}{\frac{1}{2}t^2}\\
&= \lim_{t\downarrow 0} \frac{\mathrm{d}_y\psi(\bar{x}+tu,\bar{y})(th)+\frac{1}{2}t^2\mathrm{d}^2_{yy}\psi(\bar{x}+tu,\bar{y})(h)+o(t^2)-\mathrm{d}_y\psi(\bar{x},\bar{y})(th)-\frac{1}{2}t^2\mathrm{d}^2_{yy}\psi(\bar{x},\bar{y})(h)-o(t^2)}{\frac{1}{2}t^2}\\
&\quad+\mathrm{d}^2_{xx}\psi(\bar{x},\bar{y})(u)+\mathrm{d}^2_{yy}\psi(\bar{x},\bar{y})(h)\\
&= 2\mathrm{d}^2_{xy}\psi(\bar{x},\bar{y})(u,h)+\mathrm{d}^2_{xx}\psi(\bar{x},\bar{y})(u)+\mathrm{d}^2_{yy}\psi(\bar{x},\bar{y})(h).
\end{aligned}
$$

The third equality holds by [73, Exercise 13.7 and Proposition 13.5], i.e.,

$$
\begin{aligned}
\psi(\bar{x}+tu,\bar{y}+th)-\psi(\bar{x}+tu,\bar{y}) &= \mathrm{d}_y\psi(\bar{x}+tu,\bar{y})(th)+\frac{1}{2}t^2\mathrm{d}^2_{yy}\psi(\bar{x}+tu,\bar{y})(h)+o(t^2),\\
\psi(\bar{x},\bar{y}+th)-\psi(\bar{x},\bar{y}) &= \mathrm{d}_y\psi(\bar{x},\bar{y})(th)+\frac{1}{2}t^2\mathrm{d}^2_{yy}\psi(\bar{x},\bar{y})(h)+o(t^2).
\end{aligned}
$$

Similar discussions can be given under the assumption that $\mathrm{d}^2_{xx}\psi(\bar{x},y)(u)$ is a continuous function of y around $\bar{y}$. □

The condition $\mathrm{d}^2_{xy}\psi(\bar{x},\bar{y})(u,h)=\mathrm{d}^2_{yx}\psi(\bar{x},\bar{y})(u,h)$ required in the above proposition can be satisfied when ψ has a special structure. For instance, this holds when $\psi(x,y):=\phi(x)+\varphi(y)+c(x,y)$ where $\phi(\cdot)$ and $\varphi(\cdot)$ may be nonsmooth and $c(\cdot,\cdot)$ is twice continuously differentiable.

The computation of subderivatives is crucial for applying the optimality conditions. In cases where ψ is a composition of certain special functions, we provide the following chain rule for computing subderivatives which will be used in Example 3.5.1. In fact one can easily extend the result to the more general case where $\psi(x_1,\ldots,x_n):=\varphi(g(x_1),\ldots,g(x_n))$, where φ is a C^2 function.

Proposition 3.2.3. *Let $\psi(x,y):=\varphi(g(x),g(y))$ where $\varphi(\alpha,\beta):\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is C^2 and $g:\mathbb{R}^n\to\mathbb{R}$ is Lipschitz continuous and directionally differentiable around $\bar{x}$ and $\bar{y}$. Then ψ is Lipschitz continuous, twice semidifferentiable at $(\bar{x},\bar{y})$ in any direction (u,h). Moreover, the separation property holds for the subderivative of ψ at $(\bar{x},\bar{y})$,* *and*

$$\begin{aligned}
\mathrm{d}\psi(\bar{x},\bar{y})(u,h) &= \nabla\varphi(g(\bar{x}),g(\bar{y}))^T(g'(\bar{x};u),g'(\bar{y};h)),\\
\mathrm{d}_x\psi(\bar{x},\bar{y})(u) &= \nabla_\alpha\varphi(g(\bar{x}),g(\bar{y}))g'(\bar{x};u),\\
\mathrm{d}_y\psi(\bar{x},\bar{y})(h) &= \nabla_\beta\varphi(g(\bar{x}),g(\bar{y}))g'(\bar{y};h),
\end{aligned}$$

$$\begin{aligned}
\mathrm{d}^2\psi(\bar{x},\bar{y})(u,h) =&(g'(\bar{x};u),g'(\bar{y};h))^T\nabla^2\varphi(g(\bar{x}),g(\bar{y}))(g'(\bar{x};u),g'(\bar{y};h))\\
&+\nabla\varphi(g(\bar{x}),g(\bar{y}))^T(\mathrm{d}^2g(\bar{x})(u),\mathrm{d}^2g(\bar{y})(h)).
\end{aligned}$$

$$\mathrm{d}^2_{yy}\psi(\bar{x},\bar{y})(h) = g'(\bar{y};h)^T\nabla^2_{\beta\beta}\varphi(g(\bar{x}),g(\bar{y}))g'(\bar{y};h)+\nabla_\beta\varphi(g(\bar{x}),g(\bar{y}))\mathrm{d}^2g(\bar{y})(h).$$

Proof. We know that ψ is Lipshitz continuous, as it is a composition of Lipshitz functions. Since

$$\begin{aligned}
&\frac{\psi(\bar{x}+tu',\bar{y}+th')-\psi(\bar{x},\bar{y})}{t}\\
=\;&\frac{\varphi(g(\bar{x}+tu'),g(\bar{y}+th'))-\varphi(g(\bar{x}),g(\bar{y}))}{t}\\
=\;&\frac{\nabla\varphi(g(\bar{x}),g(\bar{y}))^T(g(\bar{x}+tu')-g(\bar{x}),g(\bar{y}+th'))-g(\bar{y}))}{t}+\frac{o(t)}{t},
\end{aligned}$$

the limit exists when $t\downarrow 0, u'\to u, h'\to h$, ψ is also directionally differentiable at $(\bar{x},\bar{y})$ and the formula for $\mathrm{d}\psi(\bar{x},\bar{y})(u,h)$ is obtained. Similarly for each x,y around $(\bar{x},\bar{y})$,

$$\begin{aligned}
\mathrm{d}_x\psi(\bar{x},y)(u) &= \nabla_\alpha\varphi(g(\bar{x}),g(y))g'(\bar{x};u),\\
\mathrm{d}_y\psi(x,\bar{y})(h) &= \nabla_\beta\varphi(g(x),g(\bar{y}))g'(\bar{y};h).
\end{aligned}$$

Hence $\mathrm{d}_x\psi(\bar{x},y)(u)$ and $\mathrm{d}_y\psi(x,\bar{y})(h)$ are continuous in y,x respectively. Therefore by Proposition 3.2.1, the separation property holds at $(\bar{x},\bar{y})$. Since

$$\begin{aligned}
&\frac{\psi(\bar{x}+tu',\bar{y}+th')-\psi(\bar{x},\bar{y})-t\mathrm{d}\psi(\bar{x},\bar{y})(u,h)}{\frac{1}{2}t^2}\\
=\;&\frac{\varphi(g(\bar{x}+tu'),g(\bar{y}+th'))-\varphi(g(\bar{x}),g(\bar{y}))-t\nabla\varphi(g(\bar{x}),g(\bar{y}))^T(g'(\bar{x};u),g'(\bar{y};h))}{\frac{1}{2}t^2}\\
=\;&\frac{\nabla\varphi(g(\bar{x}),g(\bar{y}))^T(g(\bar{x}+tu')-g(\bar{x})-tg'(\bar{x};u),g(\bar{y}+th')-g(\bar{y})-tg'(\bar{y};h))}{\frac{1}{2}t^2}+\frac{o(t^2)}{\frac{1}{2}t^2}\\
&+\frac{1}{2}\frac{(g(\bar{x}+tu')-g(\bar{x}),g(\bar{y}+th')-g(\bar{y}))^T\nabla^2\varphi(g(\bar{x}),g(\bar{y}))(g(\bar{x}+tu')-g(\bar{x}),g(\bar{y}+th')-g(\bar{y}))}{\frac{1}{2}t^2},
\end{aligned}$$

and the limit exists when $t \downarrow 0, u' \to u, h' \to h$, ψ is twice directionally differentiable at $(\bar{x}, \bar{y})$ and the formula for $\mathrm{d}^2\psi(\bar{x}, \bar{y})(u, h)$ is obtained. Similarly, we have the result for $\mathrm{d}^2_{yy}\psi(\bar{x}, \bar{y})(h)$. $\square$

By Definition 1.4.4, for a set $S \subseteq \mathbb{R}^r, \bar{z} \in S$, and $\bar{v}, w \in \mathbb{R}^r$, the second subderivative of the indicator function δ_S at $\bar{z}$ for $\bar{v}$ and w is

$$\mathrm{d}^2\delta_S(\bar{z}; \bar{v})(w) := \liminf_{\substack{t\downarrow 0 \\ w'\to w}} \frac{\delta_S(\bar{z} + tw') - t\langle \bar{v}, w'\rangle}{\frac{1}{2}t^2} = \liminf_{\substack{t\downarrow 0, w'\to w \\ \bar{z}+tw'\in S}} \frac{-2\langle \bar{v}, w'\rangle}{t}. \tag{3.2.3}$$

Proposition 3.2.4 ([11]). *Consider a closed set $S \subseteq \mathbb{R}^r, \bar{z} \in S$, and $\bar{v}, w \in \mathbb{R}^r$. The following statement hold:*

(i) If $w \notin T_S(\bar{z})$ or $\langle \bar{v}, w\rangle < 0$, then $\mathrm{d}^2\delta_S(\bar{z}; \bar{v})(w) = \infty$.

(ii) We have

$$\mathrm{d}^2\delta_S(\bar{z}; \bar{v})(w) \le -\sigma_{T^2_S(\bar{z};w)}(\bar{v})$$

if and only if $w \in T_S(\bar{z})$ and $\langle \bar{v}, w\rangle \ge 0$ or $T^2_S(\bar{z}; w) = \emptyset$, where $\sigma_{T^2_S(\bar{z};w)}(\bar{v})$ denotes the support function of the second-order tangent set $T^2_S(\bar{z}; w)$.

Consider a function $\psi : \mathbb{R}^r \to \mathbb{R}$. If ψ is differentiable, then by taking $\bar{v}$ as $\nabla\psi(\bar{z})$ in (3.2.3), we have

$$\mathrm{d}^2\delta_S(\bar{z}; \nabla\psi(\bar{z}))(w) = \liminf_{\substack{t\downarrow 0 \\ w'\to w}} \frac{\delta_S(\bar{z} + tw') - t\langle \nabla\psi(\bar{z}), w'\rangle}{\frac{1}{2}t^2} = \liminf_{\substack{t\downarrow 0, w'\to w \\ \bar{z}+tw'\in S}} \frac{-2\langle \nabla\psi(\bar{z}), w'\rangle}{t}.$$

In order to deal with a nonsmooth function ψ, we propose the following definition.

Definition 3.2.2. *Let $S \subseteq \mathbb{R}^r$, $\psi : \mathbb{R}^r \to \mathbb{R}$ be semidifferentiable at $\bar{z} \in S$. The second subderivative of δ_S at $\bar{z}$ for $\mathrm{d}\psi(\bar{z})$ and w is*

$$\mathrm{d}^2\delta_S(\bar{z}; \mathrm{d}\psi(\bar{z}))(w) := \liminf_{t\downarrow 0, w'\to w} \frac{\delta_S(\bar{z} + tw') - t\mathrm{d}\psi(\bar{z})(w')}{\frac{1}{2}t^2} = \liminf_{\substack{t\downarrow 0, w'\to w \\ \bar{z}+tw'\in S}} \frac{-2\mathrm{d}\psi(\bar{z})(w')}{t},$$

where the sum of ∞ and $-\infty$ is interpreted as ∞. We say that δ_S is twice epi-differentiable at $\bar{z} \in S$ for $\mathrm{d}\psi(\bar{z})$ if for any $w \in \mathbb{R}^r$ and any sequence $t_k \downarrow 0$ there exists a sequence $w_k \to w$ such that

$$\mathrm{d}^2\delta_S(\bar{z}; \mathrm{d}\psi(\bar{z}))(w) = \lim_{k\to\infty} \frac{\delta_S(\bar{z} + t_k w_k) - t_k\mathrm{d}\psi(\bar{z})(w_k)}{\frac{1}{2}t_k^2}.$$

Suppose that ψ is semidifferentiable at $\bar{z}$ for w, then we have $\lim_{w' \to w} \mathrm{d}\psi(\bar{z})(w') = \mathrm{d}\psi(\bar{z})(w)$. It follows that

$$\mathrm{d}^2\delta_S(\bar{z}; \mathrm{d}\psi(\bar{z}))(w) = \infty \qquad \text{if } w \notin T_S(\bar{z}) \text{ or } \mathrm{d}\psi(\bar{z})(w) < 0. \tag{3.2.4}$$

We now introduce parabolic properties of a set.

Definition 3.2.3 (parabolic properties of sets [61])**.** *Let $S \subseteq \mathbb{R}^r$ be nonempty. S is said to be parabolically derivable at $\bar{z} \in \mathbb{R}^r$ for $w \in \mathbb{R}^r$ if $T_S^2(\bar{z}; w) \neq \emptyset$ and for each $\nu \in T_S^2(\bar{z}; w)$ there exist a number $\varepsilon > 0$ and an arc $\xi : [0, \varepsilon] \to S$ such that $\xi(0) = \bar{z}, \xi'_{++}(0) = w$, and $\xi''_{++}(0) = \nu$ with*

$$\xi'_{++}(0) := \lim_{t \downarrow 0} \frac{\xi(t) - \xi(0)}{t}, \quad \xi''_{++}(0) := \lim_{t \downarrow 0} \frac{\xi(t) - \xi(0) - t\xi'_{++}(0)}{\frac{1}{2}t^2}.$$

S is said to parabolically regular at $\bar{z} \in S$ for $\bar{v} \in \mathbb{R}^r$ if for any $w \in \mathbb{R}^r$ with $\mathrm{d}^2\delta_S(\bar{z}; \bar{v})(w) < \infty$ there exist, among all the sequences $t_k \downarrow 0$ and $w_k \to w$ satisfying the condition

$$\frac{\delta_S(\bar{z} + t_k w_k) - \delta_S(\bar{z}) - t_k \langle \bar{v}, w_k \rangle}{\frac{1}{2}t^2} \to \mathrm{d}^2\delta_S(\bar{z}; \bar{v})(w) \text{ as } k \to \infty,$$

those with the additional property that $\limsup_{k \to \infty} \frac{\|w_k - w\|}{t_k} < \infty$.

The following results show that a set that has parabolic properties has useful properties.

Proposition 3.2.5 ([61, Theorems 3.3 and 3.6])**.** *Let S be a closed subset of $\mathbb{R}^r$ with $\bar{z} \in S$, and let $\bar{v} \in N_S^p(\bar{z})$. Assume further that S is parabolically derivable at $\bar{z}$ for every vector $w \in T_S(\bar{z}) \cap \{\bar{v}\}^\perp$. If S is parabolically regular at $\bar{z}$ for $\bar{v}$, then the indicator function δ_S is properly twice epi-differentiable at $\bar{z}$ for $\bar{v}$, its second subderivative is finite for all $w \in T_S(\bar{z}) \cap \{\bar{v}\}^\perp$ and is calculated by $\mathrm{d}^2\delta_S(\bar{z}; \bar{v})(w) = -\sigma_{T_S^2(\bar{z}; w)}(\bar{v})$.*

The following sets are parabolically derivable at $\bar{z}$ for every vector $w \in T_S(\bar{z}) \cap \{\bar{v}\}^\perp$ and are parabolically regular at $\bar{z}$ for any $\bar{v} \in N_S^p(\bar{z})$: the convex polyhedral set [61, Example 3.4], the disjunctive set [61, Remark 3.5], the second-order cone [61, Example 5.8], and the cone of positive semidefinite symmetric matrices [61, Theorem 6.2].

Now, we consider the constraint system

$$S = \{z \in \mathbb{R}^r | g(z) \in \Sigma\}, \tag{3.2.5}$$

where $g : \mathbb{R}^r \to \mathbb{R}^q$ and $\Sigma \subseteq \mathbb{R}^q$ is closed.

Under the MSCQ, the parabolic properties of the set S are guaranteed under the parabolic properties of the set Σ provided that Σ is convex. Note that a convex set satisfying the following conditions (i)-(iii) with $g(z) = z$ includes a polyhedral convex set, the second-order cone and a semidefinite matrix cone.

Proposition 3.2.6 ([61, Theorems 4.5 and 5.6, Propositions 4.2 and 5.2])**.** *Let $\bar{z} \in S$ where S is the constraint system defined by (3.2.5) and $\bar{v} \in N_S(\bar{z})$. Suppose that g is twice continuously differentiable, the MSCQ holds for S at $\bar{z}$, and*

(i) Σ is convex,

(ii) Σ is parabolically derivable at $g(\bar{z})$ for all vectors $\nabla g(\bar{z})\eta$ satisfying $\nabla g(\bar{z})\eta \in T_\Sigma(g(\bar{z}))$ and $\eta \in \{\bar{v}\}^\perp$,

(iii) Σ is parabolically regular at $g(\bar{z})$ for every $\lambda \in \Lambda(\bar{z}, \bar{v}) := \{\lambda \in N_\Sigma(g(\bar{z})) | \nabla g(\bar{z})\lambda = \bar{v}\}$.

Then, the set S is parabolically regular at $\bar{z}$ for $\bar{v}$ and is parabolically derivable at $\bar{z}$ for all vectors $w \in T_S(\bar{z}) \cap \{\bar{v}\}^\perp$. Moreover, $N^p_S(\bar{z}) = \widehat{N}_S(\bar{z}) = N_S(\bar{z})$,

$$T_S(\bar{z}) = \{w \in \mathbb{R}^r | \nabla g(\bar{z})w \in T_\Sigma(g(\bar{z}))\}, \tag{3.2.6}$$

and for any $w \in T_S(\bar{z}) \cap \{\bar{v}\}^\perp$, the second subderivative is finite and

$$\mathrm{d}^2\delta_S(\bar{z}; \bar{v})(w) = \max_{\lambda \in \Lambda(\bar{z},\bar{v})} \{\langle \lambda, \nabla^2 g(\bar{z})(w, w)\rangle + \mathrm{d}^2\delta_\Sigma(g(\bar{z}), \lambda)(\nabla g(\bar{z})w)\}. \tag{3.2.7}$$

Let $\bar{z} \in S$, $\Sigma := \mathbb{R}^p_- \times \{0\}^q$, and suppose that the MSCQ holds for S at $\bar{z}$. We have then that

$$T_S(\bar{z}) = \left\{ w \in \mathbb{R}^r \;\middle|\; \begin{array}{l} \nabla g_i(\bar{z})^T w = 0, i = 1, \ldots, q, \\ \nabla g_i(\bar{z})^T w \leq 0, i \in I(\bar{z}) \end{array} \right\}, \tag{3.2.8}$$

where

$$I(\bar{z}) := \{i | g_i(\bar{z}) = 0, i = 1, \ldots, p\}$$

denotes the set of inequality constraints active at $\bar{z}$. For $w \in T_S(\bar{z})$,

$$T_S^2(\bar{z}, w) = \left\{ \nu \in \mathbb{R}^r \;\middle|\; \begin{array}{l} \nabla g_i(\bar{z})^T \nu + w^T \nabla^2 g_i(\bar{z}) w = 0, i = 1, \dots, q, \\ \nabla g_i(\bar{z})^T \nu + w^T \nabla^2 g_i(\bar{z}) w \le 0, i \in I_1(\bar{z}, w) \end{array} \right\}, \tag{3.2.9}$$

where

$$I_1(\bar{z}, w) := \left\{ i \in I(\bar{z}) | \nabla g_i(\bar{z})^T w = 0 \right\}.$$

Let S be convex polyhedral. Consider $\bar{z} \in S$, $\bar{v} \in N_S(\bar{z})$, and $w \in T_S(\bar{z})$. By [73, Exercise 13.17], we have $\mathrm{d}^2 \delta_S(\bar{z}; \bar{v})(w) = 0$ for any $w \in T_S(\bar{z}) \cap \{\bar{v}\}^\perp$. By [73, Proposition 13.12],

$$T_S^2(\bar{z}; w) = T_{T_S(\bar{z})}(w) = T_S(\bar{z}) + \mathbb{R}w. \tag{3.2.10}$$

When the set S has the parabolic properties, the second subderivative of the indicator function δ_S can be given.

Proposition 3.2.7 (calculation of the second subderivative of the indicator function)**.** *Given $S \subseteq \mathbb{R}^r$, $\bar{z} \in S$, and $\bar{v} \in N_S(\bar{z})$.*

(i) Suppose that $S = \mathbb{R}^r$. Then, $\mathrm{d}^2 \delta_S(\bar{z}; \bar{v})(w) = 0$ for all $w \in \mathbb{R}^r$.

(ii) Suppose that S is convex polyheral. Then,

$$\mathrm{d}^2 \delta_S(\bar{z}; \bar{v})(w) = \begin{cases} 0 & \text{if } w \in T_S(\bar{z}) \cap \{\bar{v}\}^\perp, \\ \infty & \text{otherwise.} \end{cases}$$

(iii) Suppose that $S := \{z \in \mathbb{R}^r | g(z) \in \Sigma\}$ where S and Σ satisfy the conditions in Proposition 3.2.6. Then, for any $w \in T_S(\bar{z}) \cap \{\bar{v}\}^\perp$, $\mathrm{d}^2 \delta_S(\bar{z}; \bar{v})(w)$ is finite and

$$\begin{aligned} &\mathrm{d}^2 \delta_S(\bar{z}; \bar{v})(w) \\ &\quad = \begin{cases} \max_{\lambda \in \Lambda(\bar{z}, \bar{v})} \left\{ \langle \lambda, \nabla^2 g(\bar{z})(w, w) \rangle - \sigma_{T_\Sigma^2(g(\bar{z}), \nabla g(\bar{z}) w)}(\lambda) \right\} & \text{if } w \in T_S(\bar{z}) \cap \{\bar{v}\}^\perp, \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

Here, the tangent cone $T_S(\bar{z})$ can be characterized by (3.2.6)*.*

Proof. (i) and (ii) can be derived by [73, Exercise 13.17]. (iii) follows from Proposition 3.2.5 and (3.2.7). □

3.2.2 Optimality conditions for nonsmooth optimization problems

The following first-order conditions for optimality for unconstrained problems can be found in [12, Proposition 3.99].

Proposition 3.2.8 (first-order optimality conditions for unconstrained problems)**.** *Let $\psi : \mathbb{R}^r \to \overline{\mathbb{R}}$ and $\bar{z}$ be a point where $\psi(\bar{z})$ is finite. Then:*

(a) If $\bar{z}$ is a local minimum point of the function ψ, then

$$\mathrm{d}\psi(\bar{z})(w) \geq 0 \quad \forall w \in \mathbb{R}^r.$$

(b) If $\mathrm{d}\psi(\bar{z})(w) > 0$ for all $w \in \mathbb{R}^r \setminus \{0\}$, then $\bar{z}$ is a local minimum point of the function ψ.

The following result was given in [12, Proposition 3.100]; see also [73, Theorem 13.24].

Proposition 3.2.9 (second-order optimality conditions for unconstrained problems)**.** *Let $\psi : \mathbb{R}^r \to \overline{\mathbb{R}}$ and $\bar{z}$ be a point where $\psi(\bar{z})$ is finite. Then:*

(a) If $\bar{z}$ is a local minimum point of the function ψ, then

$$\mathrm{d}^2\psi(\bar{z};0)(w) \geq 0 \ \ \forall w \in \mathbb{R}^r.$$

(b) The condition $\mathrm{d}^2\psi(\bar{z};0)(w) > 0$ for all $w \in \mathbb{R}^r \setminus \{0\}$ holds if and only if the second-order growth condition holds at $\bar{z}$, i.e., there exist $\varepsilon > 0$ and $\eta > 0$ such that

$$\psi(z) \geq \psi(\bar{z}) + \varepsilon\|z - \bar{z}\|^2 \quad \text{when } \|z - \bar{z}\| \leq \eta,$$

which implies that $\bar{z}$ is a local minimum point of the function ψ.

Based on the aforementioned results, we can obtain optimality conditions for the constrained problems, which are essential for analyzing the optimality conditions of the minimax problem.

Proposition 3.2.10 (first-order optimality conditions for constrained problems)**.** *Let $\psi : \mathbb{R}^r \to \mathbb{R}$, $S \subseteq \mathbb{R}^r$ be nonempty and closed and $\bar{z} \in S$.*

(i) If $\bar{z}$ is a local minimizer of ψ on S, then $\mathrm{d}^+\psi(\bar{z})(w) \geq 0$ for any $w \in T_S(\bar{z})$. Suppose further that ψ is semidifferentiable at $\bar{z}$. Then, we have $\mathrm{d}\psi(\bar{z})(w) \geq 0$ for any $w \in T_S(\bar{z})$.

(ii) If $\mathrm{d}\psi(\bar{z})(w) > 0$ for any $w \in T_S(\bar{z}) \setminus \{0\}$, then $\bar{z}$ is a local minimizer of ψ on S.

Proof. (i) For any $w \in T_S(\bar{z})$,

$$\begin{aligned}\mathrm{d}^+\psi(\bar{z})(w) &= \limsup_{t\downarrow 0, w'\to w} \frac{\psi(\bar{z}+tw') - \psi(\bar{z})}{t} \\ &\geq \limsup_{\substack{t\downarrow 0, w'\to w \\ \bar{z}+tw'\in S}} \frac{\psi(\bar{z}+tw') - \psi(\bar{z})}{t} \geq 0.\end{aligned}$$

If f is semidifferentiable at $\bar{z}$, then $\mathrm{d}\psi(\bar{z})(w) = \mathrm{d}^+\psi(\bar{z})(w)$. The desired result holds.

(ii) Given $\psi : \mathbb{R}^r \to \mathbb{R}$ and $\bar{z} \in S \subseteq \mathbb{R}^r$. For any $w \in T_S(\bar{z})$ and $v \in \mathbb{R}^r$,

$$\begin{aligned}\mathrm{d}(\psi+\delta_S)(\bar{z})(w) &= \liminf_{t\downarrow 0, w'\to w} \frac{\psi(\bar{z}+tw') - \psi(\bar{z}) + \delta_S(\bar{z}+tw')}{t} \\ &= \liminf_{\substack{t\downarrow 0, w'\to w \\ \bar{z}+tw'\in S}} \frac{\psi(\bar{z}+tw') - \psi(\bar{z})}{t} \geq \mathrm{d}\psi(\bar{z})(w).\end{aligned} \tag{3.2.11}$$

For any $w \notin T_S(\bar{z})$,

$$\mathrm{d}(\psi+\delta_S)(\bar{z})(w) = \infty. \tag{3.2.12}$$

The desired result follows from Theorem 3.2.8 (b), (3.2.11) and (3.2.12). □

We now derive second-order optimality conditions for constrained minimization problems. This kind of results can be traced back to Penot [71, Theorems 1.2 and 1.7]. For the smooth problems, under more conditions on set S, second-order optimality conditions with $\mathrm{d}^2\delta_S(\bar{z}; -\nabla\psi(\bar{z}))(w)$ replaced by the support function of the second-order tangent set $\sigma_{T_S^2(\bar{z};w)}(\nabla\psi(\bar{z}))$ have been established in Bonnans and Shapiro [12, Sections 3.2.2 and 3.3.3]. The term $\sigma_{T_S^2(\bar{z};w)}(\nabla\psi(\bar{z}))$ is referred to a "sigma" term and it can not be dismissed in general if the set S has some curvature.

Proposition 3.2.11 (second-order optimality conditions for constrained problems)**.** *Let $\psi : \mathbb{R}^r \to \mathbb{R}$, $S \subseteq \mathbb{R}^r$ is nonempty and closed and $\bar{z} \in \mathbb{R}^r$. Suppose that ψ is twice semidifferentiable at $\bar{z}$.*

(i) If $\bar{z}$ is a local minimizer of ψ on S, then $\mathrm{d}\psi(\bar{z})(w) \geq 0$ for any $w \in T_S(\bar{z})$, and for any $w \in T_S(\bar{z}) \cap \{w' | \mathrm{d}\psi(\bar{z})(w') = 0\}$,

$$\mathrm{d}^2\psi(\bar{z})(w) + \mathrm{d}^2\delta_S(\bar{z}; -\mathrm{d}\psi(\bar{z}))(w) \geq 0.$$

(ii) Suppose that $\mathrm{d}\psi(\bar{z})(w) \geq 0$ for any $w \in T_S(\bar{z})$, and that for any $w \in T_S(\bar{z}) \cap \{w' | \mathrm{d}\psi(\bar{z})(w') = 0\} \setminus \{0\}$,

$$\mathrm{d}^2\psi(\bar{z})(w) + \mathrm{d}^2\delta_S(\bar{z}; -\mathrm{d}\psi(\bar{z}))(w) > 0. \tag{3.2.13}$$

Then, $\bar{z}$ is a local minimizer of ψ on S with the second-order growth condition, i.e., there exist $\varepsilon > 0$ and $\eta > 0$ such that

$$\psi(z) \geq \psi(\bar{z}) + \varepsilon\|z - \bar{z}\|^2 \quad \textit{when } z \in S \cap \mathbb{B}_\eta(\bar{z}).$$

Proof. (i) The first-order necessary optimality condition follows from Proposition 3.2.10. Let $w \in T_S(\bar{z}) \cap \{w' | \mathrm{d}\psi(\bar{z})(w') = 0\}$. Since ψ is twice semidifferentiable at $\bar{z} \in S$, we have

$$\begin{aligned} &\mathrm{d}^2(\psi + \delta_S)(\bar{z}; 0)(w) \\ &\quad = \liminf_{t \downarrow 0, w' \to w} \frac{\psi(\bar{z} + tw') - \psi(\bar{z}) + \delta_S(\bar{z} + tw')}{\frac{1}{2}t^2} \\ &\quad = \liminf_{t \downarrow 0, w' \to w} \frac{\psi(\bar{z} + tw') - \psi(\bar{z}) - t\mathrm{d}\psi(\bar{z})(w') + \delta_S(\bar{z} + tw') + t\mathrm{d}\psi(\bar{z})(w')}{\frac{1}{2}t^2} \\ &\quad = \mathrm{d}^2\psi(\bar{z})(w) + \mathrm{d}^2\delta_S(\bar{z}; -\mathrm{d}\psi(\bar{z}))(w). \end{aligned} \tag{3.2.14}$$

The desired result follows from (3.2.14) and Theorem 3.2.9 (a).

(ii) By assumption, $\mathrm{d}\psi(\bar{z})(w) \geq 0$ for any $w \in T_S(\bar{z})$. Let $w \in \mathbb{R}^r$. If $w \notin T_S(\bar{z})$ or $w \in T_S(\bar{z})$ but $\mathrm{d}\psi(\bar{z})(w) > 0$, then we have $\mathrm{d}^2\delta_S(\bar{z}; -\mathrm{d}\psi(\bar{z}))(w) = \infty$ by (3.2.4). Since ψ is twice semidifferentiable, $\mathrm{d}^2\psi(\bar{z})(w)$ is finite. By (3.2.14), it follows that

$$\mathrm{d}^2(\psi + \delta_S)(\bar{z}; 0)(w) = \infty > 0 \quad \text{either } w \notin T_S(\bar{z}) \text{ or } \mathrm{d}\psi(\bar{z})(w) > 0.$$

If $w \in T_S(\bar{z}) \cap \{w' | \mathrm{d}\psi(\bar{z})(w') = 0\} \setminus \{0\}$, (3.2.14) and (3.2.13) imply $\mathrm{d}^2(\psi + \delta_S)(\bar{z}; 0)(w) > 0$. Finally we obtain

$$\mathrm{d}^2(\psi + \delta_S)(\bar{z}; 0)(w) > 0 \quad \forall w \in \mathbb{R}^r \setminus \{0\}.$$

Applying Theorem 3.2.9 (b), we obtain that $\bar{z}$ is a local minimizer of ψ on S. □

3.3 Concepts of optimality for the minimax problem

In this section, we introduce a new notion for local optimality of the minimax problem (Min-Max). Before that, we first review several existing optimality concepts for the minimax problem.

Definition 3.3.1 (global minimax point/Stackelberg equilibrium). *A point* $(\bar{x}, \bar{y}) \in X \times Y$ *is a global minimax point or a Stackelberg equilibrium of problem* (Min-Max)*, if for any* (x, y) *in* $X \times Y$*,*

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq \max_{y' \in Y} f(x, y').$$

Define the (global) value function as $V(x) := \max_{y \in Y} f(x, y)$. Then $(\bar{x}, \bar{y})$ is a global minimax point if and only if $\bar{y}$ is a global maximum point of $f(\bar{x}, \cdot)$ on Y, and $\bar{x}$ is a global minimum point of $V(x)$ on X.

Definition 3.3.2 (saddle point/Nash equilibrium). *A point* $(\bar{x}, \bar{y}) \in X \times Y$ *is a saddle point or a Nash equilibrium of problem* (Min-Max)*, if for any* (x, y) *in* $X \times Y$*,*

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq f(x, \bar{y}).$$

Definition 3.3.3 (local saddle point/local Nash equilibrium). *A point* $(\bar{x}, \bar{y}) \in X \times Y$ *is a local saddle point or a local Nash equilibrium of problem* (Min-Max) *if there exists* $\delta > 0$ *such that for any* $x \in X \cap \mathbb{B}_\delta(\bar{x}), y \in Y \cap \mathbb{B}_\delta(\bar{y})$*,*

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq f(x, \bar{y}).$$

Definition 3.3.4 (local minimax point [40, Definition 14]). *A point* $(\bar{x}, \bar{y}) \in X \times Y$ *is a local minimax point of problem* (Min-Max)*, if there exist a* $\delta_0 > 0$ *and a radius function* $\tau : \mathbb{R}_+ \to \mathbb{R}_+$ *satisfying* $\tau(\delta) \to 0$ *as* $\delta \downarrow 0$*, such that for any* $\delta \in (0, \delta_0]$ *and any* $x \in X \cap \mathbb{B}_\delta(\bar{x}), y \in Y \cap \mathbb{B}_\delta(\bar{y})$*, we have*

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq \max_{y' \in Y \cap \mathbb{B}_{\tau(\delta)}(\bar{y})} f(x, y'). \tag{3.3.1}$$

Given $\epsilon \geq 0$ and $\bar{y} \in Y$, define the local optimal value function at $\bar{y}$ by

$$V_\epsilon(x) := V_\epsilon(x; \bar{y}) := \max_{y \in Y \cap \mathbb{B}_\epsilon(\bar{y})} f(x, y). \tag{3.3.2}$$

Note that $V_\epsilon(\bar{x}) = V_\epsilon(\bar{x}; \bar{y}) = f(\bar{x}, \bar{y})$ when $\bar{y}$ is a maximum point of $f(\bar{x}, \cdot)$ in $Y \cap \mathbb{B}_\epsilon(\bar{y})$. So $(\bar{x}, \bar{y})$ is a local minimax point if and only if there exists a function $\tau : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\tau(\delta) \to 0$ as $\delta \downarrow 0$ such that $\bar{y}$ is a maximum point of $f(\bar{x}, \cdot)$ on $Y \cap \mathbb{B}_\delta(\bar{y})$, and $\bar{x}$ is a minimum point of the local optimal value function $V_{\tau(\delta)}(x)$ on $X \cap \mathbb{B}_\delta(\bar{x})$, for any small enough δ.

Under the continuity assumption of f, Jin et al. [40, Lemma 16] showed that $(\bar{x}, \bar{y}) \in X \times Y$ is a local minimax point of problem (Min-Max) if and only if $\bar{y}$ is a local maximum point of the function $f(\bar{x}, \cdot)$ on Y, and $\bar{x}$ is a local minimum point of the function $V_\epsilon(x)$ on X for all $\epsilon \downarrow 0$. Later on, the continuity assumption on f is removed by Zhang et. al. [87, Proposition 3.9].

It has been shown in Jin et. al. [40, Remark 15] that Definition 3.3.4 remains equivalent even if we further restrict function τ in Definition 3.3.4 to be either monotonic or continuous. Thus, we assume $\tau(0) = 0$ throughout the chapter. The concept of local minimax points is a celebrated idea for characterizing the local optimality for the minimax problems. However, when we tried to characterize the local optimality, we found it important to consider the calmness property of the function τ. Thus, we introduce the following new concept for the local optimality which is the local minimax point with the function τ being calm at 0.

Definition 3.3.5 (calm local minimax point)**.** *A point* $(\bar{x}, \bar{y}) \in X \times Y$ *is a calm local minimax point of problem* (Min-Max) *if it is a local minimax point of problem* (Min-Max) *with the radius function* τ *being calm at* 0.

In Definition 3.3.4 of local minimax points, the radius R_y of the neighborhood of $\bar{y}$ for maximizing $f(x, y)$ in y is $\tau(\delta)$, i.e., the radius R_y in the local optimal value function (3.3.2) is $\tau(\delta)$. Concurrently the radius R_x of the neighborhood of $\bar{x}$ for minimizing $V_{\tau(\delta)}(x)$ is δ. Thus the ratio R_y/R_x of these two radii is $\tau(\delta)/\delta$. In Definition 3.3.5, the calmness of τ at 0 implies that for any sufficiently small δ, we have $\tau(\delta) \leq \kappa\delta$ for some $\kappa > 0$. This is equivalent to imposing a bound on the ratio R_y/R_x of the radii of the local neighborhoods where the maximization and minimization takes place respectively. Next we can give an equivalent definition of calm local minimax point, without introducing a radius function τ.

Proposition 3.3.1 (an equivalent definition of calm local minimax point)**.** *The point* $(\bar{x}, \bar{y}) \in X \times Y$ *is a calm local minimax point, if and only if there exist a* $\delta_0 > 0$ *and a* $\kappa > 0$*, such that for any* $\delta \in (0, \delta_0]$ *and any* $x \in X \cap \mathbb{B}_\delta(\bar{x})$, $y \in Y \cap \mathbb{B}_\delta(\bar{y})$*, we have*

$$f(\bar{x}, y) \le f(\bar{x}, \bar{y}) \le \max_{y' \in Y \cap \mathbb{B}_{\kappa\delta}(\bar{y})} f(x, y'). \tag{3.3.3}$$

Proof. "⇒" Let $(\bar{x}, \bar{y}) \in X \times Y$ be a calm local minimax point. Then, the relation in (3.3.1) holds with τ being calm at 0, i.e., $\tau(\delta) \le \kappa\delta$ for some $\kappa > 0$. Thus, $\max_{y' \in Y \cap \mathbb{B}_{\tau(\delta)}(\bar{y})} f(x, y') \le \max_{y' \in Y \cap \mathbb{B}_{\kappa\delta}(\bar{y})} f(x, y')$ and (3.3.3) holds for any $\delta \in (0, \delta_0]$, $x \in X \cap \mathbb{B}_\delta(\bar{x})$, $y \in Y \cap \mathbb{B}_\delta(\bar{y})$.

"⇐" This is obvious since we can let $\tau(\delta) = \kappa\delta$. □

The following example demonstrates that the concept of calm local minimax optimality is not equivalent to the concept of local minimax optimality.

Example 3.3.1. *Consider*

$$\min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} f(x, y) := -x^2 + 2xy^3 - y^6.$$

We show that $(0, 0)$ *is a local (and also global) minimax point but not a calm local minimax point.*

- *Take* $\tau(\delta) = \delta^{\frac{1}{3}}$ *and* $\delta_0 = 1$*. Then for any* $|x| \le \delta$ *and* $|y| \le \delta$ *with* $\delta \in (0, \delta_0]$ *we have*

  $$-y^6 = f(0, y) \le f(0, 0) \le \max_{y \in \mathbb{B}_{\tau(\delta)}(0)} -x^2 + 2xy^3 - y^6 = 0.$$

  Hence $(0, 0)$ *is a local minimax point.*

- *For each* x*, we have*

  $$-y^6 = f(0, y) \le f(0, 0) \le \max_{y} -x^2 + 2xy^3 - y^6 = 0,$$

  where $x^{\frac{1}{3}}$ *is the maximizer. Hence* $(0, 0)$ *is a global minimax point.*

- *By Proposition 3.3.1, one can check that* $(0, 0)$ *is not a calm local minimax* *point. Indeed, it is easy to see that $f(0,0) = 0$, and for all positive δ, κ,*

$$\max_{y \in \mathbb{B}_{\kappa\delta}(0)} f(x,y) = \begin{cases} 0, \text{ if } |x| \le \kappa^3\delta^3, \\ -x^2 + 2|x|\kappa^3\delta^3 - \kappa^6\delta^6 < 0, \text{ if } |x| > \kappa^3\delta^3. \end{cases}$$

Since $\delta^3 = o(\delta)$, for any given κ, there always exists sufficiently small δ and some x such that $\kappa^3\delta^3 < x \le \delta$. Thus, there does not exist κ in the equivalent definition in Proposition 3.3.1.

Next, we study the relation between the global minimax point and the calm local minimax point. Generally, a global minimax point may not be a local minimax point (and thus not a calm local minimax point); cf. [40, Proposition 21] for an explicit example, where the global minimax point can be neither local minimax nor a stationary point (i.e., $\nabla f(\bar{x}, \bar{y}) = 0$ in the unconstrained smooth case). Nevertheless, global minimax points can be guaranteed to be calm local minimax points if the optimal solution set of the inner maximization problem satisfies the following inner calmness condition.

Definition 3.3.6 (inner calmness [10, Definition 2.2])**.** *Consider a set-valued map $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. Given $\bar{x} \in X$ and $\bar{y} \in \Gamma(\bar{x})$, we say that the set-valued map Γ is inner calm at $(\bar{x}, \bar{y})$ w.r.t. X if there exist $\kappa > 0$ and $\delta_0 > 0$ such that*

$$\bar{y} \in \Gamma(x) + \kappa\|x - \bar{x}\|\mathbb{B} \quad \forall x \in \mathbb{B}_{\delta_0}(\bar{x}) \cap X,$$

or equivalently ([10, Lemma 2.2] or [8, Definition 2.2]), if there exists $\kappa > 0$ such that for any $x_k \to \bar{x}$ with $x_k \in X$ there exists a sequence y_k satisfying $y_k \in \Gamma(x_k)$ and for sufficiently large k, $\|y_k - \bar{y}\| \le \kappa\|x_k - \bar{x}\|$.

Using the inner calmness, the following result gives a condition under which a global minimax point is always a calm local minimax point.

Proposition 3.3.2. *Let $(\bar{x}, \bar{y})$ be a global minimax point of problem* (Min-Max)*. If the solution mapping $S(x) := \arg\max_{y \in Y} f(x,y)$ is inner calm at $(\bar{x}, \bar{y})$ w.r.t. X, then $(\bar{x}, \bar{y})$ is a calm local minimax point.*

Proof. By the inner calmness of S at $(\bar{x}, \bar{y})$, there exist $\kappa > 0$ and $\delta_0 > 0$ such that for all x in $\mathbb{B}_{\delta_0}(\bar{x}) \cap X$ there exists $\bar{y}(x) \in S(x)$ satisfying $\|\bar{y}(x) - \bar{y}\| \le \kappa\|x - \bar{x}\|$.

This implies the existence of $\bar{y}(x) \in S(x)$ in $\mathbb{B}_{\kappa\delta}(\bar{y}) \cap Y$ for any $\delta \in (0, \delta_0]$ and any x in $\mathbb{B}_\delta(\bar{x}) \cap X$. It follows that

$$\max_{y \in \mathbb{B}_{\kappa\delta}(\bar{y}) \cap Y} f(x, y) = \max_{y \in Y} f(x, y).$$

Thus, a global minimax point $(\bar{x}, \bar{y})$ is a calm local minimax point with $\tau(\delta) = \kappa\delta$. □

In [15, Proposition 6.1], Chen et al. gave conditions under which the solution mapping is Lipschitz continuous, and thus is inner calm. We use their result to give the following corollary.

Corollary 3.3.1. *Suppose that Y is convex and f satisfies the following properties:*

(i) the function f is L_f-smooth w.r.t. x, that is,

$$\|\nabla_y f(x, y) - \nabla_y f(x', y)\| \leq L_f \|x - x'\| \quad \forall x, x' \in X, y \in Y,$$

(ii) and it is gradient dominant in y for some $\alpha > 0$, i.e.,

$$L_f \|y - P_Y(y + (1/L_f)\nabla_y f(x, y))\| \geq \alpha \|y - y_p(x)\| \quad \forall x \in X, y \in Y,$$

where $P_Y(\cdot)$ is the projection operator associated with Y and $y_p(x)$ is the projection of y onto the solution set $S(x) := \arg\max_{y \in Y} f(x, y)$.

Then, the solution mapping S is inner calm w.r.t. X at any point $(\bar{x}, \bar{y}) \in \operatorname{gph} S$, and thus a global minimax point is a calm local minimax point. Particularly, by [15, Example 6.1], if Y is convex, f is L_f-smooth and and α-strongly concave, then a global minimax point is a calm local minimax point.

Note that the condition (ii) in Corollary 3.3.1 for the unconstrained case (i.e., $Y = \mathbb{R}^m$) is the error bounds of Luo and Tseng [50], and is equivalent to the well-known Polyak-Łojasiewicz condition, which is weaker than the strong concavity by [41, Theorem 2].

From the definitions, it is clear that a calm local minimax point must be a local minimax point. Next, we give conditions under which they are equivalent. To this end, we need some equivalent descriptions for the local minimax point and the calm local minimax point.

Lemma 3.3.1. *The point* $(\bar{x}, \bar{y}) \in X \times Y$ *is a local minimax point, if and only if, there exist a* $\delta_0 > 0$ *and a function* $\tau : \mathbb{R}_+ \to \mathbb{R}_+$*satisfying* $\tau(\delta) \to 0$ *as* $\delta \downarrow 0$*, such that for any* $\delta \in (0, \delta_0]$ *and any* $x \in X \cap \mathbb{B}_\delta(\bar{x}), y \in Y \cap \mathbb{B}_\delta(\bar{y})$*, we have*

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq \max_{y' \in Y \cap \mathbb{B}_{\tau(\|x-\bar{x}\|)}(\bar{y})} f(x, y').$$

Proof. Since we have assumed $\tau(0) = 0$, $\max_{y' \in Y \cap \mathbb{B}_{\tau(\|\bar{x}-\bar{x}\|)}(\bar{y})} f(x, y') = f(x, \bar{y})$.

"$\Leftarrow$" Similar to the proof for [40, Remark 15], we have the facts that $\tau(\delta) \leq \widetilde{\tau}(\delta) := \sup_{\delta' \in (0,\delta]} \tau(\delta')$ for any $\delta \in (0, \delta_0]$, $\widetilde{\tau}(\delta) \to 0$ as $\delta \downarrow 0$, and $\widetilde{\tau}$ is monotonic. Thus, without loss of generality, we assume that τ is monotonic. Since $\|x - \bar{x}\| \leq \delta$ for $x \in X \cap \mathbb{B}_\delta(\bar{x})$ and τ is monotonic, we get $\tau(\|x - \bar{x}\|) \leq \tau(\delta)$. Thus

$$\max_{y' \in Y \cap \mathbb{B}_{\tau(\|x-\bar{x}\|)}(\bar{y})} f(x, y') \leq \max_{y' \in Y \cap \mathbb{B}_{\tau(\delta)}(\bar{y})} f(x, y').$$

"$\Rightarrow$" For any given $x \in X \cap \mathbb{B}_\delta(\bar{x})$, denote $\delta' = \|x - \bar{x}\|$, then $\delta' \leq \delta \leq \delta_0$. Thus, by (3.3.1) in which δ be δ', we have

$$f(\bar{x}, \bar{y}) \leq \max_{y' \in Y \cap \mathbb{B}_{\tau(\delta')}(\bar{y})} f(x, y') = \max_{y' \in Y \cap \mathbb{B}_{\tau(\|x-\bar{x}\|)}(\bar{y})} f(x, y')$$

which finishes the proof. □

Lemma 3.3.2. *Suppose that* $f(x, \cdot)$ *is continuous for all* x *near* $\bar{x}$*. The point* $(\bar{x}, \bar{y})$ *is a calm local minimax point of problem* (Min-Max)*, if and only if, it is a local minimax point and the optimal solution mapping*

$$S_{\tau(\cdot)}(x) := \arg\max_{y \in Y \cap \mathbb{B}_{\tau(\|x-\bar{x}\|)}(\bar{y})} f(x, y)$$

is inner calm at $(\bar{x}, \bar{y})$ *w.r.t.* X*.*

Proof. "$\Rightarrow$" Since τ is calm at 0, there exists $\kappa > 0$ such that for any $x_k \to \bar{x}$, $y_k \in S_{\tau(\cdot)}(x_k)$, and sufficiently large k,

$$\|y_k - \bar{y}\| \leq \tau(\|x_k - \bar{x}\|) \leq \kappa \|x_k - \bar{x}\|.$$

By Definition 3.3.6, $S_{\tau(\cdot)}(x)$ is inner calm at $(\bar{x}, \bar{y})$ w.r.t. X.

"$\Leftarrow$" By the inner calmness of $S_{\tau(\cdot)}(x)$ at $(\bar{x}, \bar{y})$, there exist $\kappa > 0$ and $\delta_0 > 0$ such that

$$\bar{y} \in S_{\tau(\cdot)}(x) + \kappa\|x - \bar{x}\|\mathbb{B} \quad \forall x \in \mathbb{B}_{\delta_0}(\bar{x}) \cap X.$$

This means that for any $0 < \delta \leq \delta_0$ and any $x \in \mathbb{B}_\delta(\bar{x}) \cap X$, there exists $\bar{y}(x) \in S_{\tau(\cdot)}(x)$ satisfying $\|\bar{y}(x) - \bar{y}\| \leq \kappa\|x - \bar{x}\| \leq \kappa\delta$. Hence, with Lemma 3.3.1, we have

$$f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq \max_{y \in Y \cap \mathbb{B}_{\tau(\|x-\bar{x}\|)}(\bar{y})} f(x, y) = f(x, \bar{y}(x)) \leq \max_{y \in Y \cap \mathbb{B}_{\kappa\delta}(\bar{y})} f(x, y).$$

Thus, $(\bar{x}, \bar{y})$ is a calm local minimax point with $\tau(\delta) = \kappa\delta$. □

With the above equivalent description for the calm local minimax point, we can now give conditions under which the calm local minimax point is the same as the local minimax point.

Proposition 3.3.3. *Let* $(\bar{x}, \bar{y})$ *be a local minimax point of problem* (Min-Max). *Suppose that* $f(x, \cdot)$ *is continuous for all* x *near* $\bar{x}$. *Suppose that* f *is twice semidifferentiable at* $(\bar{x}, \bar{y})$ *and the separation property for the subderivative holds at* $(\bar{x}, \bar{y})$. *If for any* $h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x}, \bar{y})(h') = 0\} \setminus \{0\}$,

$$\mathrm{d}^2_{yy} f(\bar{x}, \bar{y})(h) - \mathrm{d}^2\delta_{X\times Y}((\bar{x}, \bar{y}); \mathrm{d}f(\bar{x}, \bar{y}))(0, h) < 0, \tag{3.3.4}$$

then $(\bar{x}, \bar{y})$ *is a calm local minimax point.*

Proof. Suppose that $(\bar{x}, \bar{y})$ is a local minimax point. Then there exist a $\delta_0 > 0$ and a function $\tau : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\tau(\delta) \to 0$ as $\delta \to 0$, such that for any $\delta \in (0, \delta_0]$ and any $x \in X$ satisfying $\|x - \bar{x}\| \leq \delta$, we have

$$f(\bar{x}, \bar{y}) \leq \max_{y' \in Y \cap \mathbb{B}_{\tau(\delta)}(\bar{y})} f(x, y'). \tag{3.3.5}$$

By Lemma 3.3.2, to prove that $(\bar{x}, \bar{y})$ is a calm local minimax point, we only need to show that there exists $\kappa > 0$ such that for any $x_k \to \bar{x}$ with $x_k \in X$ and any $y_k \in S_{\tau(\cdot)}(x_k)$, we have $\|y_k - \bar{y}\| \leq \kappa\|x_k - \bar{x}\|$. To the contrary, suppose that for some $x_k \to \bar{x}$ with $x_k \in X$, there exist $y_k \in S_{\tau(\cdot)}(x_k)$ such that $\|y_k - \bar{y}\|/\|x_k - \bar{x}\|$ diverges.

First, since $\|y_k - \bar{y}\| \leq \tau(\|x_k - \bar{x}\|)$ for any k , we have $y_k \to \bar{y}$. Let $h_k := (y_k - \bar{y})/\|y_k - \bar{y}\|$ for all $k \in \mathbb{N}$. Then $\|h_k\| = 1$, and thus without loss of generality, we can assume that $h_k \to h$. Set $t_k := \|y_k - \bar{y}\|$ for all $k \in \mathbb{N}$, we have $t_k \downarrow 0$ and $y_k = \bar{y} + t_k h_k$, which implies $h \in T_Y(\bar{y}) \setminus \{0\}$. Denote $u_k := (x_k - \bar{x})/\|y_k - \bar{y}\|$, then

$x_k = \bar{x} + t_k u_k$. Since $\|y_k - \bar{y}\|/\|x_k - \bar{x}\| \to \infty$, we have $\|u_k\| = \|x_k - \bar{x}\|/\|y_k - \bar{y}\| \to 0$ and thus $u_k \to 0 = u$.

By Proposition 3.2.10 (i), since $\bar{y}$ is a local maximizer of $f(\bar{x}, y)$ on Y, $\mathrm{d}_y f(\bar{x}, \bar{y})(h') \leq 0$ for any $h' \in T_Y(\bar{y})$. On the other hand, by the semidifferentiablity of f and $y_k \in S_{\tau(\cdot)}(x_k)$, we have

$$\mathrm{d}f(\bar{x}, \bar{y})(u, h) = \lim_{k\to\infty} \frac{f(x_k, y_k) - f(\bar{x}, \bar{y})}{t_k} \geq \lim_{k\to\infty} \frac{f(x_k, \bar{y}) - f(\bar{x}, \bar{y})}{t_k} = \mathrm{d}_x f(\bar{x}, \bar{y})(u).$$

Since the separation property for the subderivative holds at $(\bar{x}, \bar{y})$, we have $\mathrm{d}f(\bar{x}, \bar{y})(u, h) = \mathrm{d}_x f(\bar{x}, \bar{y})(u) + \mathrm{d}_y f(\bar{x}, \bar{y})(h) \geq \mathrm{d}_x f(\bar{x}, \bar{y})(u)$. Thus, $\mathrm{d}_y f(\bar{x}, \bar{y})(h) = 0$. Since $(\bar{x}, \bar{y})$ is a local minimax point,

$$\begin{aligned}
0 &\leq \limsup_{k\to\infty} \frac{f(\bar{x} + t_k u_k, \bar{y} + t_k h_k) - f(\bar{x}, \bar{y})}{\frac{1}{2}t_k^2} \qquad \text{by (3.3.5)}, \\
&= \limsup_{k\to\infty} \frac{f(\bar{x} + t_k u_k, \bar{y} + t_k h_k) - f(\bar{x}, \bar{y}) - t_k \mathrm{d}f(\bar{x}, \bar{y})(u_k, h_k) + t_k \mathrm{d}f(\bar{x}, \bar{y})(u_k, h_k)}{\frac{1}{2}t_k^2} \\
&= \mathrm{d}^2 f(\bar{x}, \bar{y})(0, h) + \limsup_{k\to\infty} \frac{t_k \mathrm{d}f(\bar{x}, \bar{y})(u_k, h_k)}{\frac{1}{2}t_k^2} \qquad \text{since } f \text{ is twice semidifferentiable} \\
&\leq \mathrm{d}^2 f(\bar{x}, \bar{y})(0, h) + \limsup_{\substack{t\downarrow 0, (u', h') \to (0, h) \\ \bar{x} + tu' \in X, \bar{y} + th' \in Y}} \frac{t \mathrm{d}f(\bar{x}, \bar{y})(u', h')}{\frac{1}{2}t^2} \\
&= \mathrm{d}^2 f(\bar{x}, \bar{y})(0, h) - \liminf_{\substack{t\downarrow 0, (u', h') \to (0, h) \\ \bar{x} + tu' \in X, \bar{y} + th' \in Y}} \frac{-t \mathrm{d}f(\bar{x}, \bar{y})(u', h')}{\frac{1}{2}t^2} \\
&= \mathrm{d}^2_{yy} f(\bar{x}, \bar{y})(h) - \mathrm{d}^2 \delta_{X\times Y}((\bar{x}, \bar{y}); \mathrm{d}f(\bar{x}, \bar{y}))(0, h),
\end{aligned}$$

where the last equality follows from (1.4.9), a contradiction to (3.3.4). □

To calculate the second subderivative term in (3.3.4), we need certain qualification conditions. In a more specialized case, when f is differentiable, X is convex polyhedral, and $\nabla_x f(\bar{x}, \bar{y}) = 0$, the condition (3.3.4) reduces to the second-order sufficient condition for the inner maximization problem.

Corollary 3.3.2. *Let $(\bar{x}, \bar{y})$ be a local minimax point of problem* (Min-Max)*. Suppose that $f(x, \cdot)$ is continuous for all x near $\bar{x}$. Suppose further that f is twice semidifferentiable at $(\bar{x}, \bar{y})$, the separation property holds for the subderivative of f at $(\bar{x}, \bar{y})$, and for any $h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x}, \bar{y})(h') = 0\} \setminus \{0\}$,*

$$\mathrm{d}^2_{yy} f(\bar{x}, \bar{y})(h) - \mathrm{d}^2 \delta_X(\bar{x}; \mathrm{d}_x f(\bar{x}, \bar{y}))(0) - \mathrm{d}^2 \delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x}, \bar{y}))(h) < 0, \qquad (3.3.6)$$

and either $\mathrm{d}^2\delta_X(\bar{x};\mathrm{d}_x f(\bar{x},\bar{y}))(0)$ *or* $\mathrm{d}^2\delta_Y(\bar{y};\mathrm{d}_y f(\bar{x},\bar{y}))(h)$ *is finite, then* $(\bar{x},\bar{y})$ *is a calm local minimax point. Particularly, if* f *is differentiable,* X *is convex polyhedral, and* $\nabla_x f(\bar{x},\bar{y})=0$*, then* $\mathrm{d}^2\delta_X(\bar{x};\mathrm{d}_x f(\bar{x},\bar{y}))(0)=0$ *and the condition (3.3.6) is the second-order sufficient condition for the inner maximization problem* $\max_{y\in Y} f(\bar{x},y)$ *at* $\bar{y}$*.*

Proof. Since the separation property holds for the subderivative of f at $(\bar{x},\bar{y})$, by (3.2.1) we have

$$\begin{aligned}
\mathrm{d}^2\delta_{X\times Y}((\bar{x},\bar{y});\mathrm{d}f(\bar{x},\bar{y}))(0,h) &= \liminf_{\substack{t\downarrow 0,(u',h')\to(0,h)\\ \bar{x}+tu'\in X,\bar{y}+th'\in Y}} \frac{-2\mathrm{d}_x f(\bar{x},\bar{y})(u')-2\mathrm{d}_y f(\bar{x},\bar{y})(h')}{t}\\
&\geq \liminf_{\substack{t\downarrow 0,u'\to 0\\ \bar{x}+tu'\in X}} \frac{-2\mathrm{d}_x f(\bar{x},\bar{y})(u')}{t} + \liminf_{\substack{t\downarrow 0,h'\to h\\ \bar{y}+th'\in Y}} \frac{-2\mathrm{d}_y f(\bar{x},\bar{y})(h')}{t}\\
&= \mathrm{d}^2\delta_X(\bar{x};\mathrm{d}_x f(\bar{x},\bar{y}))(0)+\mathrm{d}^2\delta_Y(\bar{y};\mathrm{d}_y f(\bar{x},\bar{y}))(h),
\end{aligned}$$

where the inequality holds since the assumption that either $\mathrm{d}^2\delta_X(\bar{x};\mathrm{d}_x f(\bar{x},\bar{y}))(0)$ or $\mathrm{d}^2\delta_Y(\bar{y};\mathrm{d}_y f(\bar{x},\bar{y}))(h)$ is finite. It follows that

$$-\mathrm{d}^2\delta_{X\times Y}((\bar{x},\bar{y});\mathrm{d}f(\bar{x},\bar{y}))(0,h)\leq -\mathrm{d}^2\delta_X(\bar{x};\mathrm{d}_x f(\bar{x},\bar{y}))(0)-\mathrm{d}^2\delta_Y(\bar{y};\mathrm{d}_y f(\bar{x},\bar{y}))(h).$$

Hence (3.3.6) implies (3.3.4). By Proposition 3.3.3, we have the desired result. Particularly, by Proposition 3.2.7, if f is differentiable, X is convex polyhedral, and $\nabla_x f(\bar{x},\bar{y})=0$, then $\mathrm{d}^2\delta_X(\bar{x};\mathrm{d}_x f(\bar{x},\bar{y}))(0)=0$. □

To end this section, we summarize relationships between Nash equilibrium, local Nash equilibrium, calm local minimax points, local minimax points, and global minimax points. Note that unlike in nonconvex optimization where global optima are always local optima (thus local optima always exist if global optima exist), without additional assumptions, a global minimax point may not be a local minimax point (and thus may not be a calm local minimax point) [39, Example 3.4].

Relations in the general case:

$$\mathcal{D}_{Nash}\subset\mathcal{D}_{localNash}\subset\mathcal{D}_{calm-local}\subset\mathcal{D}_{local},\quad \mathcal{D}_{Nash}\subset\mathcal{D}_{global}.$$

Relations for the smooth minimax problems under the second-order sufficient condition for the inner maximization problem (see Corollary 3.3.2):

$$\mathcal{D}_{Nash}\subset\mathcal{D}_{localNash}\subset\mathcal{D}_{calm-local}=\mathcal{D}_{local},\quad \mathcal{D}_{Nash}\subset\mathcal{D}_{global}.$$

Relations for the smooth nonconvex-strongly-concave case (see Corollary 3.3.1):

$$\mathcal{D}_{Nash} \subset \mathcal{D}_{global} \subset \mathcal{D}_{calm-local} = \mathcal{D}_{local}.$$

$$\mathcal{D}_{Nash} \subset \mathcal{D}_{localNash} \subset \mathcal{D}_{calm-local} = \mathcal{D}_{local}.$$

Here, the notation "$\subset$" entails that the inclusion can be strict. The notations $\mathcal{D}_{Nash}, \mathcal{D}_{localNash}, \mathcal{D}_{calm-local}, \mathcal{D}_{local}$, and $\mathcal{D}_{global}$ denote the set of Nash equilibrium, local Nash equilibrium, calm local minimax points, local minimax points, and global minimax points, respectively.

3.4 Optimality conditions for the minimax problem

In this section, we give first-order and second-order optimality conditions for the minimax problem (Min-Max).

3.4.1 First-order optimality conditions for minimax problems

First-order optimality conditions in the primal form are given as follows.

Theorem 3.4.1 (first-order optimality conditions in primary form)**.** *Let* $(\bar{x}, \bar{y}) \in X \times Y$*.*

(a) Suppose that

$$\mathrm{d}_x f(\bar{x}, \bar{y})(u) > 0 \quad \forall\, u \in T_X(\bar{x}) \setminus \{0\}, \tag{3.4.1}$$

$$\mathrm{d}_y^+ f(\bar{x}, \bar{y})(h) < 0 \quad \forall\, h \in T_Y(\bar{y}) \setminus \{0\}. \tag{3.4.2}$$

Then $(\bar{x}, \bar{y})$ *is a local Nash equilibrium and hence a calm local minimax point to problem* (Min-Max)*.*

(b) Suppose that either Y *is the whole space or* $f(\bar{x}, \cdot)$ *is semidifferentiable at* $\bar{y}$*. Suppose further that* $f(x, \cdot)$ *is continuous for any* $x \in X$ *near* $\bar{x}$*. If* $(\bar{x}, \bar{y})$ *is a calm local minimax point to the minimax problem* (Min-Max)*, then for any*

$u \in T_X(\bar{x})$, there exists $h \in T_Y(\bar{y})$ such that

$$\mathrm{d}^+ f(\bar{x}, \bar{y})(u, h) \geq 0. \tag{3.4.3}$$

Besides,

$$\mathrm{d}_y^+ f(\bar{x}, \bar{y})(h) \leq 0 \quad \forall\, h \in T_Y(\bar{y}). \tag{3.4.4}$$

(c) *Suppose that f is semidifferentiable and the separation property holds for the subderivative at $(\bar{x}, \bar{y})$. Suppose further that $f(x, \cdot)$ is continuous for any $x \in X$ near $\bar{x}$. If $(\bar{x}, \bar{y})$ is a calm local minimax point to the minimax problem* (Min-Max)*, then*

$$\mathrm{d}_x f(\bar{x}, \bar{y})(u) \geq 0 \quad \forall\, u \in T_X(\bar{x}), \tag{3.4.5}$$

$$\mathrm{d}_y f(\bar{x}, \bar{y})(h) \leq 0 \quad \forall\, h \in T_Y(\bar{y}). \tag{3.4.6}$$

Proof. (a) By Proposition 3.2.10 (ii) and the fact that $\mathrm{d}_y^+ f(\bar{x}, \bar{y})(h) = -\mathrm{d}_y(-f)(\bar{x}, \bar{y})(h)$, we have that $\bar{x}$ is a local minimizer of $f(\cdot, \bar{y})$ on X and $\bar{y}$ is a local maximizer of $f(\bar{x}, \cdot)$ on Y. Thus, $(\bar{x}, \bar{y})$ is a local Nash equilibrium to problem (Min-Max).

(b) First, since $\bar{y}$ is a local maximum point of $f(\bar{x}, \cdot)$ on Y, (3.4.4) or equivalently, $\mathrm{d}_y(-f)(\bar{x}, \bar{y})(h) \geq 0$ follows from Proposition 3.2.8 (i) (when Y is the whole space) or Proposition 3.2.10 (i) (when $f(\bar{x}, \cdot)$ is semidifferentiable at $\bar{y}$). We now prove (3.4.3). For this purpose we let $u \in T_X(\bar{x})$. Then there exist $t_k \downarrow 0, u_k \to u$ such that $x_k := \bar{x} + t_k u_k \in X$. Take $\delta_k := \|x_k - \bar{x}\|$. Then since $(\bar{x}, \bar{y})$ is a calm local minimax point, there exist $\kappa > 0$ and a sequence

$$y_k \in \operatorname*{arg\,max}_{y \in Y \cap \mathbb{B}_{\tau(\delta_k)}(\bar{y})} f(x_k, y),$$

where $\tau(\delta)$ is the function defined in the definition of the calm local minimax point, such that $\|y_k - \bar{y}\| \leq \kappa \|x_k - \bar{x}\|$ for sufficiently large k. Thus, by passing to a subsequence if necessary (without relabeling), there exists $h \in \mathbb{R}^m$ such that $h_k := (y_k - \bar{y})/t_k \to h$. By the definition of the contingent cone, we have $h \in T_Y(\bar{y})$. Hence we have

$$\begin{aligned}0 &\le \lim_{k\to\infty} \frac{f(x_k, y_k) - f(\bar{x}, \bar{y})}{t_k} \qquad \text{by (3.3.1)}\\ &\le \limsup_{\substack{t\downarrow 0\\(u',h')\to(u,h)}} \frac{f(\bar{x}+tu', \bar{y}+th') - f(\bar{x}, \bar{y})}{t}\\ &= \mathrm{d}^+ f(\bar{x}, \bar{y})(u, h).\end{aligned}$$

(c) Since f is semidifferentiable, $\mathrm{d}^+ f(\bar{x}, \bar{y})(u, h) = \mathrm{d}f(\bar{x}, \bar{y})(u, h)$ and $\mathrm{d}_y^+ f(\bar{x}, \bar{y})(h) = \mathrm{d}_y f(\bar{x}, \bar{y})(h)$. Together with the separation property, the results follow from (3.4.3) and (3.4.4). □

Remark 3.4.1. *The first-order necessary condition (3.4.5) is sharper than the one in Jiang and Chen [39, (3.2a)-(3.2b)] by the analysis in [39, Remark 3.13]. Theorem 3.4.1(c) is stronger than Zhang et al. [87, Theorem 3.12] even in the smooth case since the tangent cone is larger than the inner tangent cone. From the proof of Theorem 3.4.1(a), we can see that (3.4.2) is only used to show that $\bar{y}$ is a local maximizer of $f(\bar{x}, \cdot)$ on Y. Thus even in the smooth case, if we replace (3.4.2) by that $\bar{y}$ is a local maximizer of $f(\bar{x}, \cdot)$ on Y in Theorem 3.4.1(a), then this sufficient optimality condition is weaker than Zhang et al. [87, Theorem 3.14].*

Next, we give first-order necessary optimality conditions in the dual form for the minimax problem (Min-Max).

Theorem 3.4.2 (first-order necessary optimality conditions in the dual form)**.** *Let $Y := \{y \in \mathbb{R}^m | g(y) \in D\}$ and $(\bar{x}, \bar{y})$ be a local minimax point to the minimax problem* (Min-Max)*. Suppose that f and g are Lipschitz continuous around $(\bar{x}, \bar{y})$ and that the MSCQ holds for the system $g(y) \in D$ at $\bar{y}$. Then,*

$$0 \in -\partial_y^\circ f(\bar{x}, \bar{y}) + N_Y(\bar{y}).$$

Moreover, denote $S_\delta(\bar{x}) := \arg\max_{y\in Y\cap\mathbb{B}_\delta(\bar{y})} f(\bar{x}, y)$ for some sufficiently small $\delta > 0$. We have

$$0 \in \bigcup_{y\in S_\delta(\bar{x})} \partial_x^\circ f(\bar{x}, y) + N_X(\bar{x}).$$

Proof. Since f is Lipschitz continuous and $\bar{y}$ is local maximizer of $f(\bar{x}, y)$ on Y, we have $0 \in \widehat{\partial}_y(-f + \delta_{\mathbb{R}^n\times Y})(\bar{x}, \bar{y})$ by the Fermat's rule [73, Exercise 8.4]. By the subdifferential sum rule in [62, Theorem 2.19] and the fact that $\widehat{\partial}\phi(\bar{z}) \subseteq \partial^\circ\phi(\bar{z})$, we have $0 \in \partial_y^\circ(-f)(\bar{x},\bar{y}) + N_Y(\bar{y}) = -\partial_y^\circ f(\bar{x},\bar{y}) + N_Y(\bar{y})$.

Now, we give an upper bound for the subdifferential of the value function $V_\delta(x) = \max_{y\in Y\cap\mathbb{B}_\delta} f(x,y)$ for some sufficiently small $\delta > 0$. To apply the findings in [5], we must initially establish that requiring the MSCQ holds for the system $g(y) \in D$ at $\bar{y}$ is equivalent to requiring that the MSCQ holds for the system $g(y) \in D, \|y-\bar{y}\| \leq \delta$ at $\bar{y}$. This equivalence is evident, since the constraint $\|y-\bar{y}\| \leq \tau(\delta)$ is not active at $\bar{y}$. Then, with [5, Theorem 3.3 (i)] and the fact that $\widetilde{Y} := \{y' | y' \in Y, \|y'-\bar{y}\| \leq \delta\}$ is compact, we have V_δ is Lipschitz continuous around $\bar{x}$ and

$$\partial V_\delta(\bar{x}) \subseteq \bigcup_{y\in S_\delta(\bar{x})} \partial_x^\circ f(\bar{x},y).$$

Since $\bar{x}$ is a local minimizer of V_δ on X, by the Fermat's rule, we have $0 \in \partial V_\delta(\bar{x}) + N_X(\bar{x})$. The desired result holds. $\square$

3.4.2 Second-order optimality conditions for minimax problems

In this section, we give second-order optimality conditions for the minimax problem (Min-Max) for the general case.

Theorem 3.4.3 (Second-order optimality conditions for the constrained minimax problem)**.** *Let $(\bar{x},\bar{y}) \in X \times Y$. Suppose that f is twice semidifferentiable at $(\bar{x},\bar{y})$, the separation property for the subderivative holds at $(\bar{x},\bar{y})$.*

(a) Suppose that the first-order necessary optimality conditions

$$\mathrm{d}_x f(\bar{x},\bar{y})(u) \geq 0 \quad \textit{for all } u \in T_X(\bar{x}), \tag{3.4.7}$$
$$\mathrm{d}_y f(\bar{x},\bar{y})(h) \leq 0 \quad \textit{for all } h \in T_Y(\bar{y}), \tag{3.4.8}$$

and the second-order sufficient condition for problem $\max_{y\in Y} f(\bar{x},y)$ holds at $\bar{y}$, i.e., for any $h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x},\bar{y})(h') = 0\} \setminus \{0\}$,

$$\mathrm{d}^2_{yy} f(\bar{x},\bar{y})(h) - \mathrm{d}^2\delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x},\bar{y}))(h) < 0. \tag{3.4.9}$$

Suppose that the value $\mathrm{d}^2\delta_Y(\bar{x}; \mathrm{d}_y f(\bar{x},\bar{y}))(h)$ is finite for any $h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x},\bar{y})(h') = 0\}$. If δ_Y is twice epi-differentiable at $\bar{y}$ for $\mathrm{d}_y f(\bar{x},\bar{y})$

and for any $u \in T_X(\bar{x}) \cap \{u' | \mathrm{d}_x f(\bar{x}, \bar{y})(u') = 0\} \setminus \{0\}$*, there exists* $h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x}, \bar{y})(h') = 0\}$ *such that*

$$\mathrm{d}^2 f(\bar{x}, \bar{y})(u, h) + \mathrm{d}^2 \delta_X(\bar{x}; -\mathrm{d}_x f(\bar{x}, \bar{y}))(u) - \mathrm{d}^2 \delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x}, \bar{y}))(h) > 0, \quad (3.4.10)$$

then $(\bar{x}, \bar{y})$ *is a calm local minimax point to problem* (Min-Max) *and the following second-order growth condition holds: there exist* $\delta_0 > 0$*,* $\mu > 0$ *such that for any* $\delta \in (0, \delta_0]$*,* $x \in X \cap \mathbb{B}_\delta(\bar{x})$*, and* $y \in Y \cap \mathbb{B}_\delta(\bar{y})$*, we have*

$$f(\bar{x}, y) + \varepsilon \|y - \bar{y}\|^2 \leq f(\bar{x}, \bar{y}) \leq \max_{y' \in Y \cap \mathbb{B}_\delta(\bar{y})} f(x, y') - \mu \|x - \bar{x}\|^2.$$

(b) Let $(\bar{x}, \bar{y}) \in X \times Y$ *be a calm local minimax point to problem* (Min-Max)*. Suppose that* $f(x, \cdot)$ *is continuous for any* $x \in X$ *near* $\bar{x}$ *and that the value* $\mathrm{d}^2 \delta_X(\bar{y}; -\mathrm{d}_x f(\bar{x}, \bar{y}))(u)$ *is finite for any* $u \in T_X(\bar{x}) \cap \{u' | \mathrm{d}_x f(\bar{x}, \bar{y})(u') = 0\}$*. Then the first-order necessary optimality conditions (3.4.7)-(3.4.8) hold, the second-order necessary condition for the maximum problem* $\max_{y \in Y} f(\bar{x}, y)$ *holds at* $\bar{y}$*, i.e., for any* $h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x}, \bar{y})(h') = 0\}$*, we have*

$$\mathrm{d}^2_{yy} f(\bar{x}, \bar{y})(h) - \mathrm{d}^2 \delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x}, \bar{y}))(h) \leq 0. \quad (3.4.11)$$

We also have that for any $u \in T_X(\bar{x}) \cap \{u' | \mathrm{d}_x f(\bar{x}, \bar{y})(u') = 0\}$*, there exists* $h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x}, \bar{y})(h') = 0\}$ *such that*

$$\mathrm{d}^2 f(\bar{x}, \bar{y})(u, h) + \mathrm{d}^2 \delta_X(\bar{x}; -\mathrm{d}_x f(\bar{x}, \bar{y}))(u) - \mathrm{d}^2 \delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x}, \bar{y}))(h) \geq 0. \quad (3.4.12)$$

Proof. (a) Since f is twice semidifferentiable at $(\bar{x}, \bar{y})$, by (1.4.2) and (1.4.8), we have

$$-\mathrm{d}_y f(\bar{x}, \bar{y})(h) = \mathrm{d}_y(-f)(\bar{x}, \bar{y})(h), \quad -\mathrm{d}^2_{yy} f(\bar{x}, \bar{y})(h) = \mathrm{d}^2_{yy}(-f)(\bar{x}, \bar{y})(h).$$

By Proposition 3.2.11 (ii), the second-order sufficient condition for the maximum problem implies that $\bar{y}$ is a local maximizer of $f(\bar{x}, \cdot)$ on Y and the second-order growth condition holds. Thus, there exist $\delta_0 > 0, \varepsilon > 0$ such that for any $\delta \in (0, \delta_0], y \in Y$ satisfying $\|y - \bar{y}\| \leq \delta$, we have $f(\bar{x}, y) + \varepsilon \|y - \bar{y}\|^2 \leq f(\bar{x}, \bar{y})$. Let $\tau(\delta) := \delta$, then $\tau(\delta) \to 0$ as $\delta \downarrow 0$. Since $V_\delta(x) := V_\delta(x; \bar{y}) := \max_{y \in Y \cap \mathbb{B}_\delta(\bar{y})} f(x, y)$, we have

$$V_\delta(x) \geq f(x,y) \ \ \forall (x,y) \in \mathbb{B}_\delta(\bar{x},\bar{y}) \cap (X \times Y), \qquad \text{and } V_\delta(\bar{x}) = f(\bar{x},\bar{y}). \quad (3.4.13)$$

We break the rest proof for (a) into two steps.

Step 1: We show that for any fixed $u \in T_X(\bar{x}) \cap \{u' | \mathrm{d}_x f(\bar{x},\bar{y})(u') = 0\}$ and $\delta \in (0, \delta_0]$,

$$\begin{aligned} &\mathrm{d}^2(V_\delta + \delta_X)(\bar{x};0)(u) \\ \geq &\sup_{h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x},\bar{y})(h')=0\}} \left\{ \mathrm{d}^2 f(\bar{x},\bar{y})(u,h) + \mathrm{d}^2\delta_X(\bar{x}; -\mathrm{d}_x f(\bar{x},\bar{y})(u) - \mathrm{d}^2\delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x},\bar{y}))(h) \right\}. \end{aligned} \tag{3.4.14}$$

Since $u \in T_X(\bar{x}) \cap \{u' | \mathrm{d}_x f(\bar{x},\bar{y})(u') = 0\}$, by definition of the second subderivative in Definitions 1.4.4 and 3.2.2, there exist $t_k \downarrow 0$, $u_k \to u$ such that $\bar{x} + t_k u_k \in X$ and

$$\mathrm{d}^2(V_\delta + \delta_X)(\bar{x};0)(u) = \lim_{k\to\infty} \frac{V_\delta(\bar{x}+t_k u_k) - V_\delta(\bar{x})}{\frac{1}{2}t_k^2}. \tag{3.4.15}$$

Since δ_Y is twice epi-differentiable at $\bar{y}$ for $\mathrm{d}_y f(\bar{x},\bar{y})$, by Definition 3.2.2 for any $h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x},\bar{y})(h') = 0\}$, we can find a sequence $h_k \to h$ such that

$$-\mathrm{d}^2\delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x},\bar{y}))(h) = -\lim_{k\to\infty} \frac{\delta_Y(\bar{y}+t_k h_k) - t_k \mathrm{d}_y f(\bar{x},\bar{y})(h_k)}{\frac{1}{2}t_k^2}.$$

If $y_k := \bar{y} + t_k h_k \notin Y$ for each $k \in \mathbb{N}$, $\mathrm{d}^2\delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x},\bar{y}))(h) = \infty$ but this is impossible since we have assumed the finiteness of $\mathrm{d}^2\delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x},\bar{y}))(h)$. Hence we may assume that $y_k := \bar{y} + t_k h_k \in Y$ for each $k \in \mathbb{N}$. Then

$$-\mathrm{d}^2\delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x},\bar{y}))(h) = \lim_{k\to\infty} \frac{2\mathrm{d}_y f(\bar{x},\bar{y})(h_k)}{t_k}. \tag{3.4.16}$$

Hence we have

$$\begin{aligned}
&\mathrm{d}^2(V_\delta+\delta_X)(\bar{x};0)(u) = \liminf_{k\to\infty} \frac{V_\delta\left(\bar{x}+t_k u_k\right)-V_\delta(\bar{x})}{\frac{1}{2}t_k^2} \qquad \text{by (3.4.15)}\\
&\geq \liminf_{k\to\infty} \frac{f\left(\bar{x}+t_k u_k, \bar{y}+t_k h_k\right)-f(\bar{x},\bar{y})}{\frac{1}{2}t_k^2} \qquad \text{by (3.4.13)}\\
&= \liminf_{k\to\infty} \{ \frac{f\left(\bar{x}+t_k u_k, \bar{y}+t_k h_k\right)-f(\bar{x},\bar{y})-t_k \mathrm{d}f(\bar{x},\bar{y})(u_k,h_k)}{\frac{1}{2}t_k^2}\\
&\qquad\qquad + \frac{2\mathrm{d}_x f(\bar{x},\bar{y})(u_k)}{t_k} + \frac{2\mathrm{d}_y f(\bar{x},\bar{y})(h_k)}{t_k} \}\\
&\geq \mathrm{d}^2 f(\bar{x},\bar{y})(u,h) + \mathrm{d}^2\delta_X(\bar{x};-\mathrm{d}_x f(\bar{x},\bar{y}))(u) - \mathrm{d}^2\delta_Y(\bar{y};\mathrm{d}_y f(\bar{x},\bar{y}))(h),
\end{aligned}$$

where the last inequality follows due to (3.4.16), that $\mathrm{d}f(\bar{x},\bar{y})(u,h) = \mathrm{d}_x f(\bar{x},\bar{y})(u) + \mathrm{d}_y f(\bar{x},\bar{y})(h)$ for any $(u,h) \in \mathbb{R}^n \times \mathbb{R}^m$, the assumption that for any $h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x},\bar{y})(h') = 0\}$, $\mathrm{d}^2\delta_Y(\bar{y};\mathrm{d}_y f(\bar{x},\bar{y}))(h)$ is finite. Thus (3.4.14) holds.

Step 2: We show that for any $\delta \in (0,\delta_0]$ and $x \in X$ satisfying $\|x-\bar{x}\| \leq \delta$, we have

$$\max_{y' \in Y \cap \mathbb{B}_\delta(\bar{y})} f\left(x, y'\right) - f(\bar{x},\bar{y}) \geq \beta \|x-\bar{x}\|^2 \tag{3.4.17}$$

for some $\beta > 0$.

To the contrary, suppose that for some $\delta \in (0,\delta_0]$ and $x_k \in X$ with $\|x_k - \bar{x}\| \leq \delta$,

$$\max_{y' \in Y \cap \mathbb{B}_\delta(\bar{y})} f\left(x_k, y'\right) - f(\bar{x},\bar{y}) \leq o(t_k^2), \tag{3.4.18}$$

where $t_k := \|x_k - \bar{x}\|$. Let $u_k := (x_k - \bar{x})/\|x_k - \bar{x}\|$, we have $t_k \downarrow 0$ and $\|u_k\| = 1$. By passing to a subsequence if necessary, we may assume that $u_k \to u$ with $\|u\| = 1$. We have $u \in T_X(\bar{x}) \setminus \{0\}$. The assumed first-order condition gives us $\mathrm{d}_x f(\bar{x},\bar{y})(u) \geq 0$. If $\mathrm{d}_x f(\bar{x},\bar{y})(u) > 0$,

$$\max_{y' \in Y \cap \mathbb{B}_\delta(\bar{y})} f\left(x_k, y'\right) - f(\bar{x},\bar{y}) \geq f(x_k,\bar{y}) - f(\bar{x},\bar{y}) \geq t_k \mathrm{d}_x f(\bar{x},\bar{y})(u) + o(t_k) > o(t_k) \geq o(t_k^2),$$

which is a contradiction to (3.4.18). If $\mathrm{d}_x f(\bar{x},\bar{y})(u) = 0$, by the result we showed in Step 1, we have

$$\max_{y' \in Y \cap \mathbb{B}_\delta(\bar{y})} f\left(x_k, y'\right) - f(\bar{x},\bar{y}) = V_\delta(x_k) - V_\delta(\bar{x}) \geq \frac{1}{2} t_k^2 \theta(\bar{x},\bar{y},u,h) + o(t_k^2),$$

where

$$\theta(\bar{x},\bar{y},u,h) := \sup_{h\in T_Y(\bar{y})\cap\{h'|\mathrm{d}_y f(\bar{x},\bar{y})(h')=0\}} \{\mathrm{d}^2 f(\bar{x},\bar{y})(u,h) + \mathrm{d}^2\delta_X(\bar{x};-\mathrm{d}_x f(\bar{x},\bar{y}))(u) - \mathrm{d}^2\delta_Y(\bar{y};\mathrm{d}_y f(\bar{x},\bar{y})(h)\}.$$

It follows from (3.4.10) that $\theta(\bar{x},\bar{y},u,h) > 0$. Hence we have a contradiction to (3.4.18) and consequently (3.4.17) holds.

Combining with the fact that $\bar{y}$ is a maximizer of $f(\bar{x},\cdot)$ on $Y\cap\mathbb{B}_\delta(\bar{y})$, it follows that

$$f(\bar{x},y) \le f(\bar{x},\bar{y}) \le \max_{y'\in Y\cap\mathbb{B}_\delta(\bar{y})} f(x,y') \quad \forall (x,y)\in\mathbb{B}_\delta(\bar{x},\bar{y})\cap(X\times Y).$$

Thus, $(\bar{x},\bar{y})$ is a calm local minimax point to problem (Min-Max).

(b) First, by Theorem 3.4.1 (c) and Proposition 3.2.11 (i), we have the first- and second-order conditions for the maximization problem.

Second, since $(\bar{x},\bar{y})$ is a calm local minimax point to problem (Min-Max), by Lemma 3.3.1, there exist a $\delta_0 > 0$ and a function $\tau:\mathbb{R}_+\to\mathbb{R}_+$ which is calm at 0 satisfying $\tau(\delta)\to 0$ as $\delta\downarrow 0$, such that for any $\delta\in(0,\delta_0]$ and any $x\in X\cap\mathbb{B}_\delta(\bar{x})$, we have

$$f(\bar{x},\bar{y}) \le \max_{y'\in Y\cap\mathbb{B}_{\tau(\|x-\bar{x}\|)}(\bar{y})} f(x,y'). \tag{3.4.19}$$

For any $u\in T_X(\bar{x})\cap\{u'|\mathrm{d}_x f(\bar{x},\bar{y})(u')=0\}$, there exist $t_k\downarrow 0$, $u_k\to u$ such that $x_k := \bar{x}+t_k u_k\in X$ and

$$\mathrm{d}^2\delta_X(\bar{x};-\mathrm{d}_x f(\bar{x},\bar{y}))(u) = \lim_{k\to\infty}\frac{2\mathrm{d}_x f(\bar{x},\bar{y})(u_k)}{t_k}.$$

Since τ is calm, there exist $\kappa > 0$ and

$$y_k \in \operatorname*{arg\,max}_{y'\in Y\cap\mathbb{B}_{\tau(\|x_k-\bar{x}\|)}(\bar{y})} f(x_k,y'),$$

such that $\|y_k-\bar{y}\|\le\kappa\|x_k-\bar{x}\|$. Thus, by passing to a subsequence if necessary (without relabeling), there exists $h\in\mathbb{R}^m$ such that $h_k := (y_k-\bar{y})/t_k\to h$. By the definition of the contingent cone, we have $h\in T_Y(\bar{y})$. We now prove that

$\mathrm{d}_y f(\bar{x},\bar{y})(h)=0$. Since $y_k \in S_{\tau(\delta_k)}(x_k)$, we have

$$\mathrm{d}f(\bar{x},\bar{y})(u,h)=\lim_{k\to\infty}\frac{f(x_k,y_k)-f(\bar{x},\bar{y})}{t_k}\geq\lim_{k\to\infty}\frac{f(x_k,\bar{y})-f(\bar{x},\bar{y})}{t_k}=\mathrm{d}_x f(\bar{x},\bar{y})(u). \tag{3.4.20}$$

Since $\mathrm{d}_y f(\bar{x},\bar{y})(h')\leq 0$ for any $h'\in T_Y(\bar{y})$ by (3.4.8), we obtain by (3.4.20) that $\mathrm{d}_y f(\bar{x},\bar{y})(h)=0$. Thus,

$$\begin{aligned}
0 &\leq \limsup_{k\to\infty}\frac{f(\bar{x}+t_k u_k,\bar{y}+t_k h_k)-f(\bar{x},\bar{y})}{\frac{1}{2}t_k^2} \qquad \text{by (3.4.19)}\\
&= \limsup_{k\to\infty}\frac{f(\bar{x}+t_k u_k,\bar{y}+t_k h_k)-f(\bar{x},\bar{y})-t_k\mathrm{d}f(\bar{x},\bar{y})(u_k,h_k)}{\frac{1}{2}t_k^2}\\
&\qquad + \frac{t_k\mathrm{d}_x f(\bar{x},\bar{y})(u_k)+t_k\mathrm{d}_y f(\bar{x},\bar{y})(h_k)}{\frac{1}{2}t_k^2}\\
&= \mathrm{d}^2 f(\bar{x},\bar{y})(u,h)+\mathrm{d}^2\delta_X(\bar{x};-\mathrm{d}_x f(\bar{x},\bar{y}))(u)+\limsup_{k\to\infty}\frac{t_k\mathrm{d}_y f(\bar{x},\bar{y})(h_k)}{\frac{1}{2}t_k^2}\\
&\leq \mathrm{d}^2 f(\bar{x},\bar{y})(u,h)+\mathrm{d}^2\delta_X(\bar{x};-\mathrm{d}_x f(\bar{x},\bar{y}))(u)+\limsup_{\substack{t\downarrow 0,h'\to h\\ \bar{y}+th'\in Y}}\frac{t\mathrm{d}_y f(\bar{x},\bar{y})(h')}{\frac{1}{2}t^2}\\
&= \mathrm{d}^2 f(\bar{x},\bar{y})(u,h)+\mathrm{d}^2\delta_X(\bar{x};-\mathrm{d}_x f(\bar{x},\bar{y}))(u)-\liminf_{\substack{t\downarrow 0,h'\to h\\ \bar{y}+th'\in Y}}\frac{-t\mathrm{d}_y f(\bar{x},\bar{y})(h')}{\frac{1}{2}t^2}\\
&= \mathrm{d}^2 f(\bar{x},\bar{y})(u,h)+\mathrm{d}^2\delta_X(\bar{x};-\mathrm{d}_x f(\bar{x},\bar{y}))(u)-\mathrm{d}^2\delta_Y(\bar{y};\mathrm{d}_y f(\bar{x},\bar{y}))(h),
\end{aligned}$$

where the first equality follows from the assumption for the separation property and the second equality follows from the assumption that $\mathrm{d}^2\delta_X(\bar{x};-\mathrm{d}_x f(\bar{x},\bar{y}))(u)$ is finite for any $u\in T_X(\bar{x})\cap\{u'|\mathrm{d}_x f(\bar{x},\bar{y})(u')=0\}$. $\square$

When the separation property holds for the second subderivative, our second-order necessary optimality conditions are reduced to the following simpler but relaxed forms.

Corollary 3.4.1. *Let* $(\bar{x},\bar{y})\in X\times Y$ *be a calm local minimax point to problem* (Min-Max). *Suppose that* f *is twice semidifferentiable at* $(\bar{x},\bar{y})$*, the separation property holds for the second subderivative of* f *at* $(\bar{x},\bar{y})$*, and the value* $\mathrm{d}^2\delta_X(\bar{x};-\mathrm{d}_x f(\bar{x},\bar{y}))(u)$ *is finite for any* $u\in T_X(\bar{x})\cap\{u'|\mathrm{d}_x f(\bar{x},\bar{y})(u')=0\}$*. Then, for any* $u\in T_X(\bar{x})\cap\{u'|\mathrm{d}_x f(\bar{x},\bar{y})(u')=0\}$*, there exists* $h\in T_Y(\bar{y})\cap\{h'|\mathrm{d}_y f(\bar{x},\bar{y})(h')=0\}$ *such that*

$$\mathrm{d}^2_{xx}f(\bar{x},\bar{y})(u)+2\mathrm{d}^2_{xy}f(\bar{x},\bar{y})(u,h)+\mathrm{d}^2\delta_X(\bar{x};-\mathrm{d}_x f(\bar{x},\bar{y}))(u)\geq 0, \tag{3.4.21}$$

and for any $h \in T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x}, \bar{y})(h') = 0\}$*, we have*

$$\mathrm{d}^2_{yy} f(\bar{x}, \bar{y})(h) - \mathrm{d}^2 \delta_Y(\bar{y}; \mathrm{d}_y f(\bar{x}, \bar{y}))(h) \leq 0.$$

Proof. (3.4.11) together with (3.4.12) and (3.2.2) give us (3.4.21). □

By Proposition 3.2.11, the condition $\mathrm{d}^2_{xx} f(\bar{x}, \bar{y})(u) + \mathrm{d}^2 \delta_X(\bar{x}; -\mathrm{d}_x f(\bar{x}, \bar{y}))(u) \geq 0 \ \forall u \in T_X(\bar{x}) \cap \{u' | \mathrm{d}_x f(\bar{x}, \bar{y})(u') = 0\}$ is necessary for $\bar{x}$ to be a local minimizer for the problem $\min_{x \in X} f(x, \bar{y})$. But this corresponds to the concept of local Nash equilibrium. Since a local Nash equilibrium must be a calm local minimax point but not vice versa, the term $\mathrm{d}^2_{xy} f(\bar{x}, \bar{y})(u, h)$ can not in general be dismissed in Corollary 3.4.1.

Jiang and Chen gave necessary optimality conditions for local minimax points in [39, Theorems 3.11 and 3.17], our approach differs in the following ways:

(i) We do not require the convexity of sets X and Y. Additionally, by utilizing second subderivatives of indicator functions, we characterize optimality in all critical directions, namely $T_X(\bar{x}) \cap \{u' | \mathrm{d}_x f(\bar{x}, \bar{y})(u') = 0\}$ and $T_Y(\bar{y}) \cap \{h' | \mathrm{d}_y f(\bar{x}, \bar{y})(h') = 0\}$. However, [39, Theorems 3.11 and 3.17] only focus on a subset of critical directions.

(ii) Our optimality conditions effectively capture the nested structure of the minimax problem. Specifically, we incorporate second-order information on y in (3.4.21) (or (3.4.12)) by leveraging the property of calm local minimax points. We will provide an example in the last section to demonstrate that even when calm local minimax points are identical to local minimax points, our optimality conditions can be sharper than those in [39].

3.5 Special cases and comparisons with existing related works

In this section, we derive optimality conditions for some special cases when the functions have more properties and the constraint sets have some specific structures. To derive the corresponding optimality conditions for other cases, one can use the calculation rules presented in Proposition 3.2.7 in conjunction with the method we use in this section. We also compare our results with existing works and demonstrate that our optimality conditions for local optimality can be more appropriate for some minimax problems.

3.5.1 Set-constrained systems

In this section we consider the minimax problem

$$\min_{x \in X} \max_{y \in Y} f(x, y), \tag{3.5.1}$$

where f is twice continuously differentiable and the constraints are defined by

$$X := \{x \in \mathbb{R}^n | \phi(x) \in C\}, \quad Y := \{y \in \mathbb{R}^m | \varphi(y) \in D\},$$

where $\phi : \mathbb{R}^n \to \mathbb{R}^p$ and $\varphi : \mathbb{R}^m \to \mathbb{R}^q$ are twice continuously differentiable, $C \subseteq \mathbb{R}^p$ and $D \subseteq \mathbb{R}^q$ are convex and closed.

Define the critical cones for the minimization and the maximization problem at $(\bar{x}, \bar{y})$ respectively:

$$C_{\min}(\bar{x}, \bar{y}) := \{u \in \mathbb{R}^n | \nabla_x f\,(\bar{x}, \bar{y})^T u = 0, \nabla\phi(\bar{x})u \in T_C(\phi(\bar{x}))\},$$

$$C_{\max}(\bar{x}, \bar{y}) := \{h \in \mathbb{R}^m | \nabla_y f\,(\bar{x}, \bar{y})^T h = 0, \nabla\varphi(\bar{y})h \in T_D(\varphi(\bar{y}))\}.$$

Define the set of multipliers corresponding to the minimization and the maximization problem at $(\bar{x}, \bar{y})$ respectively:

$$\begin{aligned} \Lambda_{\min}(\bar{x}, \bar{y}) &:= \{\alpha \in N_C(\phi(\bar{x})) \mid \nabla_x f(\bar{x}, \bar{y}) + \nabla\phi(\bar{x})\alpha = 0\}, \\ \Lambda_{\max}(\bar{x}, \bar{y}) &:= \{\beta \in N_D(\varphi(\bar{y})) \mid -\nabla_y f(\bar{x}, \bar{y}) + \nabla\varphi(\bar{y})\beta = 0\}. \end{aligned}$$

Denote the Lagrangian functions to the minimization and the maximization problem by

$$L_{\min}(x, y, \alpha, \beta) := f(x, y) + \phi(x)^T\alpha - \varphi(y)^T\beta, \quad L_{\max}(y, \beta; x) := f(x, y) - \varphi(y)^T\beta.$$

Now, we can give second-order optimality conditions for the minimax problem (3.5.1).

Theorem 3.5.1. *Let $(\bar{x}, \bar{y}) \in X \times Y$. Suppose that the MSCQ holds for the system $\phi(x) \in C$ at $\bar{x}$ and $\varphi(y) \in D$ at $\bar{y}$, respectively. Suppose C is parabolically derivable* *at* $\phi(\bar{x})$ *for all vectors* $\nabla\phi(\bar{x})u$ *where* $u \in C_{\min}(\bar{x},\bar{y})$*, parabolically regular at* $\phi(\bar{x})$ *for every* $\alpha \in \Lambda_{\min}(\bar{x},\bar{y})$*; D is parabolically derivable at* $\varphi(\bar{y})$ *for all vectors* $\nabla\varphi(\bar{y})h$ *where* $h \in C_{\max}(\bar{x},\bar{y})$*, parabolically regular at* $\varphi(\bar{y})$ *for every* $\beta \in \Lambda_{\max}(\bar{x},\bar{y})$ *(e.g., when C and D are convex polyhedral sets, or second-order cones, or cones of positive semidefinite symmetric matrices).*

(a) *Suppose the second-order sufficient optimality condition for the maximization holds, i.e., for any* $h \in C_{\max}(\bar{x},\bar{y}) \setminus \{0\}$*, there exists a multiplier* $\beta \in \Lambda_{\max}(\bar{x},\bar{y})$ *such that*

$$h^T\nabla^2_{yy}L_{\max}(\bar{y},\beta;\bar{x})h + \sigma_{T^2_D(\varphi(\bar{y});\nabla\varphi(\bar{y})h)}(\beta) < 0.$$

If for any $u \in C_{\min}(\bar{x},\bar{y}) \setminus \{0\}$*, there exist* $h \in C_{\max}(\bar{x},\bar{y})$ *and a multiplier* $\alpha \in \Lambda_{\min}(\bar{x},\bar{y})$ *such that for any* $\beta \in \Lambda_{\max}(\bar{x},\bar{y})$*,*

$$\nabla^2_{(x,y)}L_{\min}(\bar{x},\bar{y},\alpha,\beta)((u,h),(u,h)) - \sigma_{T^2_C(\phi(\bar{x});\nabla\phi(\bar{x})u)}(\alpha) + \sigma_{T^2_D(\varphi(\bar{y});\nabla\varphi(\bar{y})h)}(\beta) > 0,$$

then $(\bar{x},\bar{y})$ *is a calm local minimax point to problem* (3.5.1) *with the second-order growth condition.*

(b) *Suppose that* $(\bar{x},\bar{y})$ *is a calm local minimax point to problem* (3.5.1)*. Then the following second-order necessary optimality conditions for the maximization hold: for any* $h \in C_{\max}(\bar{x},\bar{y})$*, there exists a multiplier* $\beta \in \Lambda_{\max}(\bar{x},\bar{y})$ *such that*

$$h^T\nabla^2_{yy}L_{\max}(\bar{y},\beta;\bar{x})h + \sigma_{T^2_D(\varphi(\bar{y});\nabla\varphi(\bar{y})h)}(\nabla_y f(\bar{x},\bar{y})) \leq 0,$$

and for any $u \in C_{\min}(\bar{x},\bar{y})$*, there exist* $h \in C_{\max}(\bar{x},\bar{y})$ *and a multiplier* $\alpha \in \Lambda_{\min}(\bar{x},\bar{y})$ *such that for any* $\beta \in \Lambda_{\max}(\bar{x},\bar{y})$*,*

$$\nabla^2_{(x,y)}L_{\min}(\bar{x},\bar{y},\alpha,\beta)((u,h),(u,h)) - \sigma_{T^2_C(\phi(\bar{x});\nabla\phi(\bar{x})u)}(\alpha) + \sigma_{T^2_D(\varphi(\bar{y});\nabla\varphi(\bar{y})h)}(\beta) \geq 0.$$

Proof. (i) Since f is twice continuously differentiable, it is obvious that f is twice semidifferentiable, Lipschitz continuous, and the separation property holds at any points.

(ii) We show that the value $\mathrm{d}^2\delta_X(\bar{x};-\nabla_x f(\bar{x},\bar{y}))(u)$ is finite for any $u \in T_X(\bar{x}) \cap \{\nabla_x f(\bar{x},\bar{y})\}^\perp$ and the value $\mathrm{d}^2\delta_Y(\bar{y};\nabla_y f(\bar{x},\bar{y}))(h)$ is finite for any $h \in T_Y(\bar{y}) \cap \{\nabla_y f(\bar{x},\bar{y})\}^\perp$. Besides, the nonemptyness of the multiplier sets $\Lambda_{\min}(\bar{x},\bar{y})$ and $\Lambda_{\max}(\bar{x},\bar{y})$ is equivalent to the first-order conditions (3.4.7) and (3.4.8).

By Proposition 3.2.6, $N^p_X(\bar{x}) = \widehat{N}_X(\bar{x}) = N_X(\bar{x})$ and $N^p_Y(\bar{y}) = \widehat{N}_Y(\bar{y}) = N_Y(\bar{y})$. Since $\alpha \in \Lambda_{\min}(\bar{x},\bar{y})$ and $\beta \in \Lambda_{\max}(\bar{x},\bar{y})$, we have that $-\nabla_x f(\bar{x},\bar{y}) = \nabla\phi(\bar{x})\alpha \in \nabla\phi(\bar{x})N_C(\phi(\bar{x}))$ and $\nabla_y f(\bar{x},\bar{y}) = \nabla\varphi(\bar{y})\beta \in \nabla\varphi(\bar{y})N_D(\varphi(\bar{y}))$. With [61, Proposition 4.2], this is equivalent to saying that

$$-\nabla_x f(\bar{x},\bar{y}) \in N_X(\bar{x}), \quad \nabla_y f(\bar{x},\bar{y}) \in N_Y(\bar{y}),$$

which, by (1.4.1), is equivalent to saying that

$$\begin{aligned} \nabla_x f(\bar{x},\bar{y})^T u \geq 0 \quad & \text{for all } u \in T_X(\bar{x}), \\ \nabla_y f(\bar{x},\bar{y})^T h \leq 0 \quad & \text{for all } h \in T_Y(\bar{y}). \end{aligned}$$

By Proposition 3.2.7, the value $\mathrm{d}^2\delta_X(\bar{x}; -\nabla_x f(\bar{x},\bar{y}))(u)$ is finite for any $u \in T_X(\bar{x}) \cap \{\nabla_x f(\bar{x},\bar{y})\}^\perp$ and the value $\mathrm{d}^2\delta_Y(\bar{y}; \nabla_y f(\bar{x},\bar{y}))(h)$ is finite for any $h \in T_Y(\bar{y}) \cap \{\nabla_y f(\bar{x},\bar{y})\}^\perp$.

(iii) We show that δ_Y is twice epi-differentiable at $\bar{y}$ for $\nabla_y f(\bar{x},\bar{y})$. By Proposition 3.2.6 and the MSCQ assumption, Y is parabolically regular at $\bar{y}$ for $\nabla_y f(\bar{x},\bar{y}) \in N_Y(\bar{y})$ and is parabolically derivable at $\bar{y}$ for all vectors in $T_Y(\bar{y}) \cap \{\nabla_y f(\bar{x},\bar{y})\}^\perp$. Applying Proposition 3.2.5, δ_Y is twice epi-differentiable at $\bar{y}$ for $\nabla_y f(\bar{x},\bar{y})$.

(iv) Using Proposition 3.2.7 (iii), we have

$$\mathrm{d}^2\delta_X(\bar{x}; -\nabla_x f(\bar{x},\bar{y}))(u) = \max_{\alpha\in\Lambda_{\min}(\bar{x},\bar{y})} \{\langle \alpha, \nabla^2\phi(\bar{x})(u,u)\rangle - \sigma_{T^2_C(\phi(\bar{x});\nabla\phi(\bar{x})u)}(\alpha)\},$$

$$\begin{aligned} -\mathrm{d}^2\delta_Y(\bar{y}; \nabla_y f(\bar{x},\bar{y}))(h) &= -\max_{\beta\in\Lambda_{\max}(\bar{x},\bar{y})} \{\langle \beta, \nabla^2\varphi(\bar{y})(h,h)\rangle - \sigma_{T^2_D(\varphi(\bar{y});\nabla\varphi(\bar{y})h)}(\beta)\} \\ &= \min_{\beta\in\Lambda_{\max}(\bar{x},\bar{y})} \{-\langle \beta, \nabla^2\varphi(\bar{y})(h,h)\rangle + \sigma_{T^2_D(\varphi(\bar{y});\nabla\varphi(\bar{y})h)}(\beta)\}. \end{aligned}$$

Thus,

$$\begin{aligned} &\nabla^2 f(\bar{x},\bar{y})((u,h),(u,h)) + \mathrm{d}^2\delta_X(\bar{x}; -\nabla_x f(\bar{x},\bar{y}))(u) - \mathrm{d}^2\delta_Y(\bar{y}; \nabla_y f(\bar{x},\bar{y}))(h) \\ &= \nabla^2 f(\bar{x},\bar{y})((u,h),(u,h)) + \max_{\alpha\in\Lambda_{\min}(\bar{x},\bar{y})} \{\langle \alpha, \nabla^2\phi(\bar{x})(u,u)\rangle - \sigma_{T^2_C(\phi(\bar{x});\nabla\phi(\bar{x})u)}(\alpha)\} \\ &\quad + \min_{\beta\in\Lambda_{\max}(\bar{x},\bar{y})} \{-\langle \beta, \nabla^2\varphi(\bar{y})(h,h)\rangle + \sigma_{T^2_D(\varphi(\bar{y});\nabla\varphi(\bar{y})h)}(\beta)\}. \end{aligned}$$

With Theorem 3.4.3, we have the desired optimality conditions. □

Next, we consider the special case where the constraint sets X and Y involving only equalities and inequalities. Let $C = \mathbb{R}^{p_1}_{-} \times \{0\}^{p_2}$ and $D = \mathbb{R}^{q_1}_{-} \times \{0\}^{q_2}$. Then we have

$$C_{\min}(\bar{x}, \bar{y}) = \left\{u \in \mathbb{R}^n \mid \nabla\phi_i(\bar{x})u \le 0, i \in I_\phi(\bar{x}); \nabla\phi_j(\bar{x})u = 0, j = 1, ..., p_2; \nabla_x f(\bar{x}, \bar{y})^T u = 0\right\},$$

$$C_{\max}(\bar{x}, \bar{y}) = \left\{h \in \mathbb{R}^m \mid \nabla\varphi_i(\bar{x})h \le 0, i \in I(\bar{y}); \nabla\varphi_j(\bar{y})h = 0, j = 1, ..., q_2; \nabla_y f(\bar{x}, \bar{y})^T h = 0\right\},$$

where $I_\phi(\bar{x}) := \{i = 1, ..., p_1 \mid \phi_i(\bar{x}) = 0\}$ and $I_\varphi(\bar{y}) := \{i = 1, ..., q_1 \mid \varphi_i(\bar{y}) = 0\}$,

$$\begin{aligned}
\Lambda_{\min}(\bar{x}, \bar{y}) &= \{\alpha := (\alpha_1, \alpha_2) \in \mathbb{R}^{p_1}_{+} \times \mathbb{R}^{p_2} \mid \nabla_x f(\bar{x}, \bar{y}) + \nabla\phi(\bar{x})\alpha = 0, \alpha_1 \perp \phi_{\le}(\bar{x})\}, \\
\Lambda_{\max}(\bar{x}, \bar{y}) &= \{\beta := (\beta_1, \beta_2) \in \mathbb{R}^{q_1}_{+} \times \mathbb{R}^{q_2} \mid -\nabla_y f(\bar{x}, \bar{y}) + \nabla\varphi(\bar{y})\beta = 0, \beta_1 \perp \varphi_{\le}(\bar{y})\},
\end{aligned}$$

where $\phi_{\le}(\bar{x}) := (\phi_1(\bar{x}), ..., \phi_{p_1}(\bar{x}))^T$ and $\varphi_{\le}(\bar{y}) := (\varphi_1(\bar{y}), ..., \varphi_{q_1}(\bar{y}))^T$.

Theorem 3.5.2 (inequalities and equalities systems)**.** *Let $C := \mathbb{R}^{p_1}_{-} \times \{0\}^{p_2}$, $D := \mathbb{R}^{q_1}_{-} \times \{0\}^{q_2}$ and $(\bar{x}, \bar{y}) \in X \times Y$. Suppose that the MSCQ holds for the system $\phi(x) \in C$ at $\bar{x}$ and $\varphi(y) \in D$ at $\bar{y}$, respectively (e.g., when the MFCQ holds or the functions $\phi(x)$ and $\varphi(y)$ are linear).*

(a) Suppose that for any $h \in C_{\max}(\bar{x}, \bar{y})\backslash\{0\}$, there exists a multiplier $\beta \in \Lambda_{\max}(\bar{x}, \bar{y})$ such that

$$h^T \nabla^2_{yy} L_{\max}(\bar{y}, \beta; \bar{x})h < 0,$$

and for any $u \in C_{\min}(\bar{x}, \bar{y}) \setminus \{0\}$, there exist $h \in C_{\max}(\bar{x}, \bar{y})$ and a multiplier $\alpha \in \Lambda_{\min}(\bar{x}, \bar{y})$ such that for any $\beta \in \Lambda_{\max}(\bar{x}, \bar{y})$,

$$\nabla^2_{(x,y)} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta)((u, h), (u, h)) > 0. \tag{3.5.2}$$

Then, $(\bar{x}, \bar{y})$ is a calm local minimax point to the problem (3.5.1) *with the second-order growth condition.*

(b) Suppose that $(\bar{x}, \bar{y})$ is a calm local minimax point to problem (3.5.1)*. Then for any $h \in C_{\max}(\bar{x}, \bar{y})$, there exists a multiplier $\beta \in \Lambda_{\max}(\bar{x}, \bar{y})$ such that*

$$h^T \nabla^2_{yy} L_{\max}(\bar{y}, \beta; \bar{x})h \le 0, \tag{3.5.3}$$

and for any $u \in C_{\min}(\bar{x}, \bar{y})$, there exist $h \in C_{\max}(\bar{x}, \bar{y})$ and a multiplier $\alpha \in$

$\Lambda_{\min}(\bar{x},\bar{y})$ such that for any $\beta \in \Lambda_{\max}(\bar{x},\bar{y})$,

$$\nabla^2_{(x,y)} L_{\min}(\bar{x},\bar{y},\alpha,\beta)((u,h),(u,h)) \geq 0. \tag{3.5.4}$$

Proof. Since C and D are convex polyhedral, the result follows from Proposition 3.2.7 (ii) and Theorem 3.5.1. □

To compare our results with the existing literature, which primarily focused on local minimax points, we first establish the following sufficient condition under which calm local minimax points coincide with local minimax points.

Lemma 3.5.1. *Let $(\bar{x},\bar{y})$ be a local minimax point of problem (3.5.1) with $C = \mathbb{R}^{p_1}_- \times \{0\}^{p_2}$ and $D = \mathbb{R}^{q_1}_- \times \{0\}^{q_2}$. Suppose that $\nabla_x f(\bar{x},\bar{y}) = 0$ and that the MSCQ holds for the system $\phi(x) \in C$ at $\bar{x}$ and $\varphi(y) \in D$ at $\bar{y}$, respectively. Moreover, assume that the second-order sufficient condition for the inner maximization problem holds, i.e., for any $h \in C_{\max}(\bar{x},\bar{y}) \setminus \{0\}$, there exists a multiplier $\beta \in \Lambda_{\max}(\bar{x},\bar{y})$ such that*

$$h^T \nabla^2_{yy} L_{\max}(\bar{y},\beta;\bar{x}) h < 0.$$

Then $(\bar{x},\bar{y})$ is a calm local minimax point.

Proof. By Proposition 3.2.7 (ii) and (iii), $\mathrm{d}^2\delta_X(\bar{x};\nabla_x f(\bar{x},\bar{y}))(0) = 0$ and

$$-\mathrm{d}^2\delta_Y(\bar{y};\nabla_y f(\bar{x},\bar{y}))(h) = -\max_{\beta\in\Lambda_{\max}(\bar{x},\bar{y})} \langle \beta, \nabla^2\varphi(\bar{y})(h,h)\rangle = \min_{\beta\in\Lambda_{\max}(\bar{x},\bar{y})} \langle \beta, -\nabla^2\varphi(\bar{y})(h,h)\rangle.$$

Thus, the condition (3.3.6) in Corollary 3.3.2 holds and we have the desired result.
□

We now compare our results with the one obtained by Dai and Zhang in [17, Theorems 3.2 and 3.1] for the case where the constraints for the maximization problem is independent of x.

Corollary 3.5.1. *Let $C = \mathbb{R}^{p_1}_- \times \{0\}^{p_2}$, $D = \mathbb{R}^{q_1}_- \times \{0\}^{q_2}$ and $(\bar{x},\bar{y}) \in X \times Y$. Suppose that $\nabla_x f(\bar{x},\bar{y}) = 0$ and that the MSCQ holds for the system $\phi(x) \in C$ at $\bar{x}$ and $\varphi(y) \in D$ at $\bar{y}$, respectively (e.g., when the MFCQ holds or the functions $\phi(x)$ and $\varphi(y)$ are linear).*

(i) Suppose that there exists a multiplier $\bar{\beta} \in \Lambda_{\max}(\bar{x},\bar{y})$ such that $\nabla^2_{yy} L_{\max}(\bar{y},\bar{\beta};\bar{x}) \prec 0$. Suppose further that for any $u \in C_{\min}(\bar{x},\bar{y}) \setminus \{0\}$, there exist $h \in C_{\max}(\bar{x},\bar{y})$

and a multiplier $\alpha \in \Lambda_{\min}(\bar{x}, \bar{y})$ *such that for any* $\beta \in \Lambda_{\max}(\bar{x}, \bar{y})$,

$$\nabla^2_{(x,y)} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta)((u, h), (u, h)) > 0. \tag{3.5.5}$$

Then, $(\bar{x}, \bar{y})$ *is a local minimax point to problem* (3.5.1) *with the second-order growth condition.*

(ii) *Let* $(\bar{x}, \bar{y})$ *be a local minimax point of problem* (3.5.1). *Assume that there exists a multiplier* $\beta \in \Lambda_{\max}(\bar{x}, \bar{y})$ *such that the matrix* $\nabla^2_{yy} L_{\max}(\bar{y}, \beta; \bar{x})$ *is nonsingular. Then for any* $u \in C_{\min}(\bar{x}, \bar{y})$, *there exists a multiplier* $\alpha \in \Lambda_{\min}(\bar{x}, \bar{y})$ *such that*

$$u^T \left(\sum_{i=1}^{p} \alpha_i \nabla^2 \phi_i(\bar{x}) + [\nabla^2_{xx} L_{\max} - \nabla^2_{xy} L_{\max} (\nabla^2_{yy} L_{\max})^{-1} \nabla^2_{yx} L_{\max}](\bar{y}, \beta; \bar{x}) \right) u \geq 0. \tag{3.5.6}$$

Proof. By Lemma 3.5.1, the local minimax is equivalent to the calm local minimax under the assumption in (i) or (ii).

(i) Since $\nabla^2_{yy} L_{\max}(\bar{y}, \bar{\beta}; \bar{x}) \prec 0$, Theorem 3.5.2 (a)(i) holds and we have the desired result.

(ii) Since $L_{\min}(\bar{x}, \bar{y}, \alpha, \beta) = L_{\max}(\bar{y}, \beta; \bar{x}) + \phi(\bar{x})^T \alpha$, we have that $\nabla^2_{yy} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta) = \nabla^2_{yy} L_{\max}(\bar{y}, \beta; \bar{x})$ is nonsingular. For each $u \in C_{\min}(\bar{x}, \bar{y})$, let

$$h^* := -\nabla^2_{yy} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta)^{-1} \nabla^2_{xy} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta)^T u.$$

Then we have

$$\begin{aligned}
&\sup_{h \in C_{\max}(\bar{x}, \bar{y})} \nabla^2_{(x,y)} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta)((u, h), (u, h)) \\
&\leq \sup_{h \in \mathbb{R}^m} \nabla^2_{(x,y)} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta)((u, h), (u, h)) \\
&= \sup_{h \in \mathbb{R}^m} \left\{ h^T \nabla^2_{yy} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta) h + 2u^T \nabla^2_{xy} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta) h + u^T \nabla^2_{xx} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta) u \right\} \\
&= \nabla^2_{(x,y)} L_{\min}(\bar{x}, \bar{y}, \alpha, \beta)((u, h^*), (u, h^*)) \\
&= u^T [\nabla^2_{xx} L_{\min} - \nabla^2_{xy} L_{\min} (\nabla^2_{yy} L_{\min})^{-1} \nabla^2_{yx} L_{\min}](\bar{x}, \bar{y}, \alpha, \beta) u \\
&= u^T \left(\sum_{i=1}^{p} \alpha_i \nabla^2 \phi_i(\bar{x}) + [\nabla^2_{xx} L_{\max} - \nabla^2_{xy} L_{\max} (\nabla^2_{yy} L_{\max})^{-1} \nabla^2_{yx} L_{\max}](\bar{y}, \beta; \bar{x}) \right) u.
\end{aligned}$$

Hence (3.5.6) follows from Theorem 3.5.2 (b)(ii). $\square$

From the proof of Corollary 3.5.1, we can see that (3.5.2) and (3.5.4) implies (3.5.5) and (3.5.6) respectively and hence our sufficient and necessary conditions in Theorem 3.5.2(a) and Theorem 3.5.2(b) are sharper than the one in Corollary 3.5.1(i) and Corollary 3.5.1(ii) respectively. Note that Corollary 3.5.1 obtains the same sufficient and necessary optimality conditions as the one given in [17, Theorems 3.2 and 3.1] under much weaker assumptions. In particular we do not need to assume that the Jacobian uniqueness condition holds.

3.5.2 Unconstrained case

In this section we consider the unconstrained minimax problem

$$\min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m} f(x,y). \tag{3.5.7}$$

In [40, Propositions 19 and 20], Jin et al. gave some second-order optimality conditions for local minimaxity for the unconstrained minimax problem when f is twice differentiable. For the case where f is nonsmooth, Theorem 3.4.3 gives necessary and sufficient optimality for the calm local minimaxity. An illustrative example is given as follows.

Example 3.5.1. *Consider*

$$\min_{x\in\mathbb{R}}\max_{y\in\mathbb{R}} f(x,y) := -|x|^9 + \frac{3}{5}|x|^3|y| - |y|^3.$$

We will show that $(0,0)$ *is a calm local minimax point and the necessary optimality condition in Theorem 3.4.3 holds.*

Take $\tau(\delta) = \frac{1}{\sqrt{5}}\sqrt{\delta}^3$. *Then for any* $|x| \leq \delta$ *and* $|y| \leq \delta$ *with sufficiently small* $\delta \in (0,1/2)$ *we have*

$$-|y|^3 = f(0,y) \leq f(0,0) \leq \max_{y\in[-\tau(\delta),\tau(\delta)]} -|x|^9 + \frac{3}{5}|x|^3|y| - |y|^3 = -|x|^9 + \frac{2}{5\sqrt{5}}|x|^{\frac{9}{2}}$$

where $\pm\frac{1}{\sqrt{5}}\sqrt{|x|}^3$ *is the maximizer of the above maximization problem. Since* τ *is calm at* 0*,* $(0,0)$ *is a calm local minimax point.*

Obviously, f *is not differentiable at* $(0,0)$. *We first verify that the assumptions in Theorem 3.4.3 hold at* $(0,0)$. *Then, we show that the necessary optimality conditions in Theorem 3.4.3(b) hold at* $(0,0)$.

(i) Denote $\varphi(\alpha,\beta) := -\alpha^9 + \frac{3}{5}\alpha^3\beta - \beta^3, g(x) := |x|, g(y) := |y|,$ *and* $(\bar{x},\bar{y}) = (0,0)$. *Applying Proposition 3.2.3, we know that* f *is twice semidifferentiable at* $(0,0)$ *and the separation property holds for the subderivative holds at* $(0,0)$. *Moreover, for any* $u \in \mathbb{R}$ *and* $h \in \mathbb{R}$,

$$\mathrm{d}f(0,0)(u,h) = \nabla\varphi(g(\bar{x}),g(\bar{y}))^T(g'(\bar{x};u),g'(\bar{y};h)) = 0,$$

$$\mathrm{d}_x f(0,0)(u) = \nabla_\alpha\varphi(g(\bar{x}),g(\bar{y}))g'(\bar{x};u) = 0, \tag{3.5.8}$$

$$\mathrm{d}_y f(0,0)(h) = \nabla_\beta\varphi(g(\bar{x}),g(\bar{y}))g'(\bar{y};h) = 0, \tag{3.5.9}$$

$$\begin{aligned}\mathrm{d}^2 f(0,0)(u,h) = (g'(\bar{x};u),g'(\bar{y};h))^T\nabla^2\varphi(g(\bar{x}),g(\bar{y}))(g'(\bar{x};u),g'(\bar{y};h))\\ +\nabla\varphi(g(\bar{x}),g(\bar{y}))^T(\mathrm{d}^2 g(\bar{x})(u),\mathrm{d}^2 g(\bar{y})(h)) = 0,\end{aligned}$$

$$\mathrm{d}^2_{yy} f(0,0)(h) = g'(\bar{y};h)^T\nabla^2_{\beta\beta}\varphi(g(\bar{x}),g(\bar{y}))g'(\bar{y};h) + \nabla_\beta\varphi(g(\bar{x}),g(\bar{y}))\mathrm{d}^2 g(\bar{y})(h) = 0.$$

(ii) We show that for any $u \in T_X(0)\cap\{u'|\mathrm{d}_x f(0,0)(u') = 0\}$, *the value* $\mathrm{d}^2\delta_X(0;-\mathrm{d}_x f(0,0))(u)$ *is finite. This can be seen from the fact that for any* $u \in \mathbb{R}$,

$$\mathrm{d}^2\delta_X(0;-\mathrm{d}_x f(0,0))(u) = \liminf_{t\downarrow 0,u'\to u}\frac{t\mathrm{d}_x f(0,0)(u')}{\frac{1}{2}t^2} = 0.$$

(iii) We show that the necessary optimality conditions in Theorem 3.4.3 (b) hold at $(0,0)$. *By* (3.5.8) *and* (3.5.9), *we know that* $T_X(\bar{x})\cap\{u'|\mathrm{d}_x f(\bar{x},\bar{y})(u') = 0\} = T_Y(\bar{y})\cap\{h'|\mathrm{d}_y f\,(\bar{x},\bar{y})\,h' = 0\} = \mathbb{R}$. *Besides, for any* $h \in \mathbb{R}$,

$$\mathrm{d}^2\delta_Y(0;\mathrm{d}_y f(0,0))(h) = \liminf_{t\downarrow 0,h'\to h}\frac{-t\mathrm{d}_y f(0,0)(h')}{\frac{1}{2}t^2} = 0.$$

Thus, for any $u,h \in \mathbb{R}$,

$$\mathrm{d}^2 f(0,0)(u,h) + \mathrm{d}^2\delta_X(0;-\mathrm{d}_x f(0,0))(u) - \mathrm{d}^2\delta_Y(0;\mathrm{d}_y f(0,0))(h) = 0.$$

Besides, for any $h \in \mathbb{R}$, *we have*

$$\mathrm{d}^2_{yy} f(0,0)(h) - \mathrm{d}^2\delta_Y(0;\mathrm{d}_y f(0,0))(h) \le 0.$$

Theorem 3.5.1 has the following immediate corollary. We state it here for the convenience of comparison with the existing results.

Theorem 3.5.3 (second-order conditions for the unconstrained smooth case)**.** *Sup-*

pose that f is twice differentiable at $(\bar{x}, \bar{y})$.

(a) *Suppose that $\nabla f(\bar{x}, \bar{y}) = 0$, $\nabla^2_{yy} f(\bar{x}, \bar{y}) \prec 0$, and for any $u \in \mathbb{R}^n \setminus \{0\}$, there exists $h \in \mathbb{R}^m$ such that*

$$\nabla^2 f(\bar{x}, \bar{y})((u, h), (u, h)) > 0.$$

Then $(\bar{x}, \bar{y})$ is a calm local minimax point to problem (3.5.7)*.*

(b) *Let $(\bar{x}, \bar{y})$ be a calm local minimax point to problem* (3.5.7)*. Then $\nabla f(\bar{x}, \bar{y}) = 0$, and for any $u \in \mathbb{R}^n$, there exists $h \in \mathbb{R}^m$ such that*

$$\nabla^2 f(\bar{x}, \bar{y})((u, h), (u, h)) \geq 0. \tag{3.5.10}$$

Moreover, the necessary condition $\nabla^2_{yy} f(\bar{x}, \bar{y}) \preceq 0$ holds.

We can use the above necessary condition to verify that $(0, 0)$ is not a calm local minimax point for the problem in Example 3.3.1, since $\nabla f(0, 0) = (0, 0)$, $\nabla^2_{yy} f(0, 0) = 0$ and for any $u \neq 0$ and any $h \in \mathbb{R}$, $\nabla^2 f(0, 0)((u, h), (u, h)) = -2u^2 < 0$.

By Corollary 3.3.2, when $X = \mathbb{R}^n, Y = \mathbb{R}^m$ and $(\bar{x}, \bar{y})$ is a local minimax point with $\nabla^2_{yy} f(\bar{x}, \bar{y}) \prec 0$, we have that $(\bar{x}, \bar{y})$ is a calm local minimax point. Letting $\nabla^2_{yy} f(\bar{x}, \bar{y}) \prec 0$ and $h^* := -\nabla^2_{yy} f(\bar{x}, \bar{y})^{-1} \nabla^2_{xy} f(\bar{x}, \bar{y})^T u$, the result in [40, Propositions 19 and 20] can be recovered from Theorem 3.5.3.

Corollary 3.5.2. *Let $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m$.*

(a) *Suppose that the following first-order and second-order sufficient conditions hold.*

$$\nabla_x f(\bar{x}, \bar{y}) = \nabla_y f(\bar{x}, \bar{y}) = 0, \quad \nabla^2_{yy} f(\bar{x}, \bar{y}) \prec 0,$$

$$[\nabla^2_{xx} f - \nabla^2_{xy} f (\nabla^2_{yy} f)^{-1} \nabla^2_{yx} f](\bar{x}, \bar{y}) \succ 0.$$

Then $(\bar{x}, \bar{y})$ is a (calm) local minimax point to the problem (3.5.7)*.*

(b) *Let $(\bar{x}, \bar{y})$ be a local minimax point for problem* (3.5.7)*. Suppose that $\nabla^2_{yy} f(\bar{x}, \bar{y}) \prec 0$. Then*

$$[\nabla^2_{xx} f - \nabla^2_{xy} f (\nabla^2_{yy} f)^{-1} \nabla^2_{yx} f](\bar{x}, \bar{y}) \succeq 0.$$

Proof. For each $u \in \mathbb{R}^n \setminus \{0\}$, let $h^* := -\nabla^2_{yy} f(\bar{x}, \bar{y})^{-1} \nabla^2_{xy} f(\bar{x}, \bar{y})^T u$. With Theorem 3.5.3, the rest of the proof can be given similarly to the proof for Corollary 3.5.1 (ii). □

Note that Corollary 3.5.2 (b) is only applicable when $\nabla^2_{yy} f(\bar{x}, \bar{y}) \prec 0$. This is restrictive. In contrast, Theorem 3.5.3 (b) is applicable even when $\nabla^2_{yy} f(\bar{x}, \bar{y})$ is only negative semidefinite. The following example from [39, Example 3.21] shows that there exist minimax problems for which the optimality conditions in Theorem 3.5.3(b) are applicable while the one in Corollary 3.5.2 (b) is not.

Example 3.5.2. *Consider*

$$\min_{x\in\mathbb{R}} \max_{y\in\mathbb{R}} f(x, y) := -x^4 + 4x^2y^2 - y^4.$$

We first exam that $(\bar{x}, \bar{y}) = (0, 0)$ *is a calm local minimax point, and thus is also a local minimax point. Let* $\tau(\delta) = \sqrt{2}\delta$ *and* $\delta_0 = \frac{\sqrt{2}}{2}$*. Then, for any* $\delta \in (0, \delta_0]$ *and any* $(x, y) \in \mathbb{R}^2$ *satisfying* $|x| \leq \delta$ *and* $|y| \leq \delta$*, we have*

$$-y^4 = f(0, y) \leq f(0, 0) \leq \max_{|y'|\leq\tau(\delta)} f(x, y') = 3x^4.$$

Since $\nabla^2_{yy} f(\bar{x}, \bar{y}) = 0$*, the optimality conditions in Corollary 3.5.2(b) cannot be applied. However, we can easily verify that necessary optimality conditions in Theorem 3.5.3(b) holds at* $(\bar{x}, \bar{y})$ *since*

$$\nabla^2 f(\bar{x}, \bar{y}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Besides, it is easy to verify that $(0, 0)$ *is the unique global minimax point (by the definition) and the unique local minimax point (by using the first-order condition* $\nabla f(x, y) = 0$*). However,* $(0, 0)$ *is not a local Nash equilibrium since* $f(0, 0) = 0 \geq f(x, 0) = -x^4$*.*

When reducing to the unconstrained case, the second-order necessary and sufficient optimality conditions for local minimax points in [17, Theorems 3.1 and 3.2] are the same as the result in Corollary 3.5.2. We have shown in Example 3.5.2 that our necessary optimality condition extends the result in Corollary 3.5.2 (a), and thus extends the result in [17, Theorem 3.1].

In [39], Jiang and Chen studied local minimax points and derived some necessary optimality conditions. We state their result [39, Theorem 3.17] for the unconstrained case so that we can compare it with Theorem 3.5.3 (b).

Proposition 3.5.1 ([39, Theorem 3.17]). *Let $(\bar{x},\bar{y})$ be a local minimax point for the unconstrained minimax problem* (3.5.7) *where f be twice continuously differentiable. Then it holds that $\nabla f(\bar{x},\bar{y})=0, \nabla^2_{yy} f(\bar{x},\bar{y}) \preceq 0$, and*

$$u^T \nabla^2_{xx} f(\bar{x},\bar{y}) u \ \geq 0 \text{ for all } u \in \operatorname{cl}\{\bar{u} \mid \exists \delta > 0, \bar{u} \in C_{\min}(\bar{x},y')\ \forall y' \in \mathbb{B}_\delta(\bar{y})\}, \quad (3.5.11)$$

where

$$C_{\min}(\bar{x},y') := \{u \in \mathbb{R}^n \mid \nabla_x f\,(\bar{x},y')\,u = 0\}.$$

In general we can not directly show that our condition (3.5.10) is sharper than (3.5.11). However when the set $\operatorname{cl}\{\bar{u} \mid \exists \delta > 0, \bar{u} \in C_{\min}(\bar{x},y')\ \forall y' \in \mathbb{B}_\delta(\bar{y})\}$ in (3.5.11) is equal to $\{0\}$, our condition (3.5.10) is shaper. To be more specific, consider the problem

$$\min_{x\in\mathbb{R}} \max_{y\in\mathbb{R}} -y^2 + xy + x^3 + x^4.$$

Using the first-order condition $\nabla f(\bar{x},\bar{y})=0$, there are three stationary points $(0,0)$, $(-\frac{1}{2},-\frac{1}{4})$, and $(-\frac{1}{4},-\frac{1}{8})$. Using the second-order sufficient condition in Corollary 3.5.2 (a), we can identify that $(0,0)$ and $(-\frac{1}{2},-\frac{1}{4})$ are local minimax points.

Using the second-order necessary condition in Corollary 3.5.2 (b), we can identify that $(-\frac{1}{4},-\frac{1}{8})$ is not a calm local minimax point. The problem is a smooth nonconvex-strongly-concave minimax problem satisfying the conditions required in Corollary 3.3.2, and thus local minimax points are identical to calm local minimax points.

However, for $(\bar{x},\bar{y}) = (-\frac{1}{4},-\frac{1}{8})$ and any $\delta > 0, y' \in \mathbb{B}_\delta(\bar{y})$ with $y' \neq \bar{y}$, we have $y' + 3\bar{x}^2 + 4\bar{x}^3 = y' + \frac{1}{8} \neq 0$, and thus

$$C_{\min}(\bar{x},y') = \{u \in \mathbb{R} \mid (y' + 3\bar{x}^2 + 4\bar{x}^3)u = 0\} = \{0\}.$$

Therefore,

$$\operatorname{cl}\{\bar{u} \mid \exists \delta > 0, \bar{u} \in C_{\min}(\bar{x},y')\ \forall y' \in \mathbb{B}_\delta(\bar{y})\} = \operatorname{cl}\{0\} = \{0\},$$

which means that the second-order condition for a local minimax point (3.5.11) holds. Therefore the second-order condition for a local minimax point(3.5.11) cannot rule out the possibility that $(-\frac{1}{4},-\frac{1}{8})$ is not a local minimax point.

Chapter 4

Conclusions

In this thesis, we discuss bilevel programs and minimax problems from a theoretical perspective.

In Chapter 2, we propose a second-order combined approach for solving bilevel problems, which allows us to derive new single-level formulations for bilevel problems. This approach enhances the likelihood of satisfying the partial calmness qualification condition. Moving forward, we need to study the algorithms aligned with our proposed second-order combined approach. Potential challenges may arise from two perspectives. Firstly, the value function constraint may cause difficulties. Typically, satisfying the partial calmness condition along with the usual constraint qualifications is necessary for this constraint. Secondly, challenges may arise from the second-order constraint, which transforms the problem into a nonlinear semidefinite problem.

In Chapter 3, we study a specific case of bilevel problems—minimax problems. We introduce a proper definition of local minimax points and give first-order and second-order necessary and sufficient optimality conditions. Looking ahead, several related issues require further investigation. Firstly, we need to ascertain whether the concept of local minimax points and the associated optimality conditions can be extended to minimax problems with coupling constraints. In this context, minimax problems with coupling constraints involve constraints where the maximization problem is contingent upon the minimization variable, denoted by $y \in Y(x)$ rather than $y \in Y$. Challenges may arise in the first-order and second-order sensitivity analysis of the value function, as the constraint depends on the parameter. Secondly, in our work, we use second-order tangent sets to give second-order optimality conditions for set-constrainted minimax problems, see Theorem 3.5.1. This needs the second-order tangent sets to be nonempty. However, second-order tangent sets can be empty even when the set is convex. Since the second-order tangent set and the asymptotic second-order tangent cone cannot be empty simultaneously [72, Proposition 2.1], in cases where the second-order tangent sets are empty, we may turn to the asymptotic second-order tangent cone, as discussed in [70]. However, challenges may arise in analyzing the maximal value function, given that it is typically nondifferentiable.

Appendix A

Author Contributions

In response to the external examiner's advice, this statement outlines the specific contributions made by each author in the papers included in this thesis.

Combined approach with second-order optimality conditions for bilevel programming problems:

Xiaoxiao Ma:

Completed the introduction with Dr. Wei Yao. Listed four different second-order optimality conditions for the general nonlinear optimization problem. Defined the partial calmness condition, M-stationary, S-stationary points, Clarke calmness, and MPEC LICQ. Compared equivalence between local and global solutions of the original bilevel problem. Found examples illustrating relations between different partial calmness conditions. Wrote the first version of the conclusion.

Dr. Wei Yao:

Wrote the first version of the introduction. Introduced slack variables for the lower-level program. Wrote the second-order reformulation. Proposed discussing second-order optimality conditions and relaxed critical cone indices. Found the illustrative example.

Dr. Jane Ye:

Polished the entire paper and provided detailed suggestions. Proposed citing results on Lipschitz continuity and upper estimate of Clarke subdifferential. Suggested giving stationary conditions for the general combined problem (GCP). Polished the conclusion.

Dr. Jin Zhang:

Proposed studying the second-order combined approach. Polished the entire first draft.

Calm local optimality for nonconvex-nonconcave minimax problems:

Xiaoxiao Ma:

Wrote the initial draft of the introduction. Integrated variational analysis tools. Established first- and second-order optimality conditions for nonsmooth constrained problems. Initiated defining calm local minimax points. Proved relationships between various existing minimax points. Proved first-order and second-order optimality conditions and all results/examples in Section 3.5.

Dr. Wei Yao and Dr. Jin Zhang:

Discovered an example in fair classification problem. Provided advice on minimax points and checked results. Offered many detailed suggestions.

Dr. Jane Ye:

Explored separation properties of subderivatives and second subderivatives. Provided detailed guidance and assistance in formulating the paper.

Bibliography

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

[3] A.V. Arutyunov, *Perturbations of extremum problems with constraints and necessary optimality conditions*, J. Sov. Math., 54 (1991), pp. 1342-1400.

[4] K. Bai and J.J. Ye, *Directional necessary optimality conditions for bilevel programs*, Math. Oper. Res., 47 (2022), pp. 1169-1191.

[5] K. Bai and J.J. Ye, *Directional subdifferential of the value function*, to appear in J. Optim. Theory Appl., 2024.

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

[7] R. Behling, G. Haeser, A. Ramos and D.S. Viana, *On a conjecture in second-order optimality conditions*, J. Optim. Theory Appl., 176 (2018), pp. 625-633.

[8] M. Benko, *On inner calmness*, generalized calculus, and derivatives of the normal cone mapping*, J. Nonsmooth Anal. Optim., 2 (2021), pp. 1-27.

[9] M. Benko and P. Mehlitz, *On implicit variables in optimization theory*, J. Nonsmooth Anal. Optim., 2 (2021), pp. 1-34.

[10] M. Benko, H. Gfrerer and J.V. Outrata, *Calculus for directional limiting normal cones and subdifferentials*, Set-Valued Var. Anal., 27 (2019), pp. 713-745.

[11] M. Benko, H. Gfrerer, J.J. Ye, J. Zhang and J.C. Zhou, *Second-order optimality conditions for general nonconvex optimization problems and variational analysis of disjunctive systems*, SIAM J. Optim., 33 (2023), pp. 2625-2653.

[12] J. F. Bonnans and A. Shapiro, *Pertubation Analysis of Optimization Problems*, Springer-Verlag, 2000.

[13] S. Burer and H. Dong, *Representing quadratically constrained quadratic programs as generalized copositive programs*, Oper. Res. Lett., 40 (2012), pp. 203-206.

[14] J.V. Burke, *Calmness and exact penalization*, SIAM J. Control Optim., 29 (1991), pp. 493-497.

[15] L. Chen, J. Xu and J.Z. Zhang, *Bilevel optimization without lower-level strong convexity from the hyper-objective perspective*, arXiv preprint arXiv:2301.00712v3, 2023.

[16] F.H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.

[17] Y.H. Dai and L.W. Zhang, *Optimality conditions for constrained minimax optimization*, CSIAM T. Appl. Math., 1 (2020), pp. 296-315.

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

[19] 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.

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

[21] S. Dempe and A.B. Zemkoho, *Bilevel Optimization: Advances and Next Challenges*, Springer Optim. Appl, vol. 161, 2020.

[22] S. Dempe, V. Kalashnikov, G.A. Pérez-Valdés and N. Kalashnykova, *Bilevel Programming Problems: Theory, Algorithms and Applications to Energy Systems*, Springer, Berlin, 2015.

[23] V.F. Demyanov and V.N. Malozemov, *Introduction to Minimax*, Wiley, 1974.

[24] C. Ding, D.F. Sun and J.J. Ye, *First order optimality conditions for mathematical programs with semidefinite cone complementarity constraints*, Math. Program., 147 (2014), pp. 539-579.

[25] G. Eichfelder and J. Jahn, *Set-semidefinite optimization*, J. Convex Anal., 15 (2008), pp. 767-801.

[26] T. Fiez, B. Chasnov and L. Ratliff, *Implicit learning dynamics in Stackelberg games: equilibria characterization, convergence analysis, and empirical study*, ICML (2020), pp. 3133-3144.

[27] L. Franceschi, P. Frasconi, S. Salzo, R. Grazzi and M. Pontil, *Bilevel programming for hyperparameter optimization*, ICML (2018), pp. 1563-1572.

[28] E. H. Fukuda and M. Fukushima, *A note on the squared slack variables technique for nonlinear optimization*, J. Oper. Res. Soc. Jpn., 60 (2017), pp. 262-270.

[29] J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Math. Program., 12 (1977), pp. 136-138.

[30] J. Gauvin and F. Dubeau, *Differential properties of the marginal function in mathematical programming*, Optimality and Stability in Mathematical Programming, Springer, Berlin, Heidelberg, 1982, pp. 101-119.

[31] H. Gfrerer, *On directional metric subregularity and second-order optimality conditions for a class of nonsmooth mathematical programs*, SIAM J. Optim., 23 (2013), pp. 632-665.

[32] 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.

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

[34] H. Gfrerer, J.J. Ye and J.C. Zhou, *Second-order optimality conditions for nonconvex set-constrained optimization problems*, Math. Oper. Res., 47 (2022), pp. 2344-2365.

[35] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, *Generative adversarial nets*, NeurIPS (2014), pp. 2672-2680.

[36] L. Guo, J.J. Ye and J. Zhang, *Sensitivity analysis of the maximal value function with applications in nonconvex ninimax programs*, Math. Oper. Res., 49 (2024), pp. 536-556.

[37] L. Guo, G.H. Lin, J.J. Ye and J. Zhang, *Sensitivity analysis of the value function for parametric mathematical programs with equilibrium constraints*, SIAM J. Optim., 24 (2014), pp. 1206-1237.

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

[39] J. Jiang and X.J. Chen, *Optimality conditions for nonsmooth nonconvex-nonconcave min-max problems and generative adversarial networks*, SIAM J. Math. Data Sci, 5 (2023), pp. 693-722.

[40] C. Jin, P. Netrapalli and M.I. Jordan, *What is local optimality in nonconvex-nonconcave minimax optimization?* ICML (2020), pp. 4880-4889.

[41] H. Karimi, J. Nutini and M. Schmidt, *Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition*, Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD (2016), pp. 795-811.

[42] R.Z. Ke, W. Yao, J.J. Ye and J. Zhang, *Generic property of the partial calmness condition for bilevel programming problems*, SIAM J. Optim., 32 (2022), pp. 604-634.

[43] G. Kunapuli, K. Bennett, J. Hu and J.S. Pang, *Classification model selection via bilevel programming*, Optim. Meth. Softw., 23 (2008), pp. 475-489.

[44] Y.C. Liang and J.J. Ye, *Optimality conditions and exact penalty for mathematical programs with switching constraints*, J. Optim. Theory Appl., 190 (2021), pp. 1-31.

[45] H. Li, Y. Tian, J. Zhang and A. Jadbabaie, *Complexity lower bounds for nonconvex-strongly-concave min-max optimization*, NeurIPS (2021), pp. 1792-1804.

[46] T. Lin, C. Jin and M.I. Jordan, *On gradient descent ascent for nonconvex-concave minimax problems*, ICML (2020), pp. 6083-6093.

[47] T. Lin, C. Jin and M.I. Jordan, *Near-optimal algorithms for minimax optimization*, COLT (2020), pp. 2738-2779.

[48] R.S. Liu, P. Mu, X.M. Yuan, S.Z. Zeng and J. Zhang, *A generic first-order algorithmic framework for bi-level programming beyond lower-level singleton*, ICML (2020), pp. 6305-6315.

[49] L. Luo, Y. Li and C. Chen, *Finding second-order stationary points in nonconvex-strongly-concave minimax optimization*, NeurIPS (2022), pp. 36667-36679.

[50] Z.Q. Luo and P. Tseng, *Error bounds and convergence analysis of feasible descent methods: a general approach*, Ann. Oper. Res., 46 (1993), pp. 157-178.

[51] X.X. Ma, W. Yao, J.J. Ye and J. Zhang, *Combined approach with second-order optimality conditions for bilevel programming problems*, J. Convex Anal., 30 (2023), pp. 1173-1201.

[52] X.X. Ma, W. Yao, J.J. Ye and J. Zhang, *Calm local optimality for nonconvex-nonconcave minimax problems*, arXiv preprint arXiv:2306.17443, 2023.

[53] A. Madry, A. Makelov, L. Schmidt, D. Tsipras and A. Vladu, *Towards deep learning models resistant to adversarial attacks*, ICLR (2018).

[54] P. Mehlitz, *Stationarity conditions and constraint qualifications for mathematical programs with switching constraints*, Math. Program., 181 (2020), pp. 149-186.

[55] P. Mehlitz, *On the linear independence constraint qualification in disjunctive programming*, Optimization, 69 (2020), pp. 2241-2277.

[56] P. Mehlitz and 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 Var. Anal., 30 (2022), pp. 179-205.

[57] P. Mehlitz, L.I. Minchenko and A.B. Zemkoho, *A note on partial calmness for bilevel optimization problems with linearly structured lower level*, Optim. Lett., 15 (2021), pp. 1277-1291.

[58] L. Minchenko and A. Leschov, *On strong and weak second-order necessary optimality conditions for nonlinear programming*, Optimization, 65 (2016), pp. 1693-1702.

[59] L. Minchenko and S. Stakhovski, *Parametric nonlinear programming problems under the relaxed constant rank condition*, SIAM J. Optim., 21 (2011), pp. 314-332.

[60] J. Mirrlees, *The theory of moral hazard and unobservable behaviour–part I*, Rev. Econ. Stud., 66 (1999), pp. 3-22.

[61] A. Mohammadi, B. Mordukhovich and M. Sarabi, *Parabolic regularity in geometric variational analysis*, Trans. Am. Math. Soc., 3 (2021), pp. 1711-1763.

[62] B.S. Mordukhovich, *Variational Analysis and Applications (Vol. 30)*, Cham: Springer, 2018.

[63] 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.

[64] O. Morgenstern and J.V. Neumann, *Theory of Games and Economic Behavior*, Princeton University Press, 1953.

[65] R.B. Myerson, *Game Theory*, Harvard University Press, 2013.

[66] A.S. Nemirovsky and D.B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, Wiley, 1983.

[67] M. Nouiehed, M. Sanjabi, T. Huang, J.D. Lee and M. Razaviyayn, *Solving a class of non-convex min-max games using iterative first order methods*, NeurIPS (2019), pp. 14934-14942.

[68] S. Omidshafiei, J. Pazis, C. Amato, J.P. How and J. Vian, *Deep decentralized multi-task multi-agent reinforcement learning under partial observability*, ICML (2017), pp. 2681-2690.

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

[70] W. Ouyang, J.J Ye and B.B. Zhang, *New second-order optimality conditions for directional optimality of a general set-constrained optimization problem*, arXiv preprint arXiv:2404.17696, 2024.

[71] J.P. Penot, *Optimality conditions in mathematical programming and composite optimization*, Math. Program., 67 (1994), pp. 225-245.

[72] J.P. Penot, *Second-order conditions for optimization problems with constraints*, SIAM J. Control Optim., 37 (1998), pp. 303-318.

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

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

[75] M. Sion, *On general minimax theorems*, Pac. J. Math., 8 (1958), pp. 171-176.

[76] O. Stein, *Bi-level Strategies in Semi-infinite Programming*, Springer Science and Business Media, 2013.

[77] K.K. Thekumparampil, P. Jain, P. Netrapalli and S. Oh, *Efficient algorithms for smooth minimax optimization*, NeurIPS (2019), pp. 12659-12670.

[78] M.W. Xu and J.J. Ye, *Relaxed constant positive linear dependence constraint qualification and its application to bilevel programs*, J. Global Optim., 78 (2020), pp. 181-205.

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

[80] 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.

[81] J.J. Ye and S.Y. Wu, *First order optimality conditions for generalized semi-infinite programming problems*, J. Optim. Theory Appl., 137 (2008), pp. 419-434.

[82] J.J. Ye, X.M. Yuan, S.Z. Zeng and J. Zhang, *Difference of convex algorithms for bilevel programs with applications in hyperparameter selection*, Math. Program., 198 (2023), pp. 1583-1616.

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

[84] 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.

[85] 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.

[86] 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.

[87] G. Zhang, P. Poupart and Y. Yu, *Optimality and stability in non-convex smooth games*, J. Mach. Learn. Res., 23 (2022), pp. 1-71.

[88] J.Y. Zhang, M.Y. Hong and S.Z. Zhang, *On lower iteration complexity bounds for the convex concave saddle point problems*, Math. Program., 194 (2022), pp. 901-935.