

Optimality Conditions for Constrained Minimax Optimization

Yu-Hong Dai^{*†} and Liwei Zhang[‡]

Abstract

Minimax optimization problems arises from both modern machine learning including generative adversarial networks, adversarial training and multi-agent reinforcement learning, as well as from tradition research areas such as saddle point problems, numerical partial differential equations and optimality conditions of equality constrained optimization. For the unconstrained continuous nonconvex-nonconcave situation, Jin, Netrapalli and Jordan (2019) carefully considered the very basic question: what is a proper definition of local optima of a minimax optimization problem, and proposed a proper definition of local optimality called local minimax. We shall extend the definition of local minimax point to constrained nonconvex-nonconcave minimax optimization problems. By analyzing Jacobian uniqueness conditions for the lower-level maximization problem and the strong regularity of Karush-Kuhn-Tucker conditions of the maximization problem, we provide both necessary optimality conditions and sufficient optimality conditions for the local minimax points of constrained minimax optimization problems.

Key words: constrained minimax optimization, value function, Jacobian uniqueness conditions, strong regularity, necessary optimality conditions, sufficient optimality conditions.

AMS subject classification: 90C30

1 Introduction

Minimax optimization problems arises from both modern machine learning including generative adversarial networks, adversarial training and multi-agent reinforcement learning, as well as from tradition research areas such as saddle point problems, numerical partial differential equations and optimality conditions of equality constrained optimization. Let m, n, m_1, m_2, n_1 and n_2 be positive integers, $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{m_1}$, $g : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{m_2}$, $H : \mathbb{R}^n \rightarrow \mathbb{R}^{n_1}$ and $G : \mathbb{R}^n \rightarrow \mathbb{R}^{n_2}$ be given functions. We are interested in the constrained minimax optimization problem of the form

$$\min_{x \in \Phi} \max_{y \in Y(x)} f(x, y), \quad (1.1)$$

^{*}LSEC, ICMSEC, AMSS, Chinese Academy of Sciences, Beijing 100190, China. *Email:* dyh@lsec.cc.ac.cn. This author was supported by the Natural Science Foundation of China (No. 11991020, 11631013, 11971372 and 11991021) and Beijing Academy of Artificial Intelligence (BAAI).

[†]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.

[‡]Corresponding author. School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China. *Email:* lwzhang@dlut.edu.cn. This author was supported by the Natural Science Foundation of China (No. 11971089 and 11731013).

where $f : \mathfrak{X}^n \times \mathfrak{X}^m \rightarrow \mathfrak{K}$, $\Phi \subset \mathfrak{X}^n$ is a feasible set of decision variable x defined by

$$\Phi = \{x \in \mathfrak{X}^n : H(x) = 0, G(x) \leq 0\} \quad (1.2)$$

and $Y : \mathfrak{X}^n \rightrightarrows \mathfrak{X}^m$ is a set-valued mapping defined by

$$Y(x) = \{y \in \mathfrak{X}^m : h(x, y) = 0, g(x, y) \leq 0\}. \quad (1.3)$$

For the unconstrained continuous nonconvex-nonconcave situation, Jin, Netrapalli and Jordan (2019) [5] carefully considered the very basic question: what is a proper definition of local optima of a minimax optimization problem, and proposed a proper definition of local optimality called local minimax. We shall extend this definition of local minimax point for the constrained minimax optimization problem (1.1).

Definition 1.1 *A point $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ is said to be a local minimax point of Problem (1.1) if there exists $\delta_0 > 0$ and a function $\eta : (0, \delta_0] \rightarrow \mathfrak{K}_+$ satisfying $\eta(\delta) \rightarrow 0$ as $\delta \rightarrow 0$, such that for any $\delta \in (0, \delta_0]$ and any $(x, y) \in [\mathbf{B}_\delta(x^*) \cap \Phi] \times [Y(x^*) \cap \mathbf{B}_\delta(y^*)]$, we have*

$$f(x^*, y) \leq f(x^*, y^*) \leq \max \{f(x, z) : z \in Y(x) \cap \mathbf{B}_{\eta(\delta)}(y^*)\}. \quad (1.4)$$

The minimax optimization problem is essentially a bi-level programming problem and the local minimax point is closely related to the so-called pessimistic solution of bi-level programming problem, see [7]. There have been many results about optimality conditions for bi-level programming. Dempe (1992)[6] demonstrated necessary optimality conditions and the sufficient optimality conditions for the bi-level programming when the lower level problem is a convex optimization problem satisfying the Mangasarian-Fromovitz constraint qualification, the second-order sufficient optimality condition and the constant rank constraint qualification. Falk (1995)[4] discussed the optimality conditions when the lower level has a local unique solution and the upper level problem is unconstrained. Ye and Zhu (1995,1997)[14] established necessary optimality conditions for bi-level programming based on the generalized gradient of value function. Dempe et. al (2007)[8] derived necessary optimality conditions for bi-level programming when the solution set of the lower level problem satisfies some calmness property. Dempe and Zemkoho (2013)[9] also developed necessary optimality conditions based on the value function reformulation of bi-level programming and the assumption that the value function is locally convex. Dempe et. al (2014)[10] derived a new type upper subdifferential necessary optimality conditions for the pessimistic version of bi-level programming problem. Even recently, Mehlitz and Zemkoho (2019)[11] studied sufficient optimality conditions for bi-level programming.

Although many results about optimality conditions for bi-level programming are available, they are established based on different solution notations. For example, many works involve value function of the lower level problem, which is usually defined as the global optimal value. This restricts the application of the theoretical results. In this paper, we shall discuss optimality conditions for constrained minimax optimization whose solution is specified as the local minimax point given in Definition 1.1.

2 Differential of the Value Function

2.1 Under Jacobian uniqueness

Let $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point and f, h, g be twice continuously differentiable around (x^*, y^*) . For a point $x \in \mathfrak{X}^n$ around x^* , we use (P_x) to denote the following problem

$$\begin{aligned} \max_{z \in \mathfrak{X}^m} \quad & f(x, z) \\ \text{s.t.} \quad & h(x, z) = 0, \\ & g(x, z) \leq 0. \end{aligned} \tag{2.1}$$

The Lagrangian of Problem (P_x) is defined by

$$\mathcal{L}(x; z, \mu, \lambda) = f(x, z) + \mu^T h(x, z) - \lambda^T g(x, z).$$

Definition 2.1 Let $(\mu^*, \lambda^*) \in \mathfrak{X}^{m_1} \times \mathfrak{X}^{m_2}$ be a point. We say that Jacobian uniqueness conditions of Problem (P_{x^*}) are satisfied at (y^*, μ^*, λ^*) if

(a) The point (y^*, μ^*, λ^*) is a Karush-Kuhn-Tucker point of Problem (P_{x^*}) ; namely,

$$\begin{aligned} \nabla_y \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) &= 0, \\ h(x^*, y^*) &= 0, \\ 0 \leq \lambda^* \perp g(x^*, y^*) &\leq 0. \end{aligned}$$

(b) The linear independence constraint qualification holds at y^* ; namely, the set of vectors

$$\{\nabla_y h_1(x^*, y^*), \dots, \nabla_y h_{m_1}(x^*, y^*)\} \cup \{\nabla_y g_i(x^*, y^*) : i \in I_{x^*}(y^*)\}$$

are linearly independent, where $I_{x^*}(y^*) = \{i : g_i(x^*, y^*) = 0, i = 1, \dots, m_2\}$.

(c) The strict complementarity condition holds at y^* for λ^* ; namely,

$$\lambda_i^* - g_i(x^*, y^*) > 0, \quad i = 1, \dots, m_2.$$

(d) The second-order sufficient optimality condition holds at (y^*, μ^*, λ^*) ,

$$\langle \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) d_y, d_y \rangle < 0 \quad \forall d_y \in C_{x^*}(y^*),$$

where $C_{x^*}(y^*)$ is the critical cone of Problem (P_{x^*}) at y^* ,

$$C_{x^*}(y^*) = \left\{ d_y \in \mathfrak{X}^m : \mathcal{J}_y h(x^*, y^*) d_y = 0; \nabla_y g_i(x^*, y^*) d_y \leq 0, i \in I_{x^*}(y^*); \nabla_y f(x^*, y^*) d_y \leq 0 \right\}.$$

Lemma 2.1 Let $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point around which f, h, g are twice continuously differentiable. Let $(\mu^*, \lambda^*) \in \mathfrak{X}^{m_1} \times \mathfrak{X}^{m_2}$ such that Jacobian uniqueness conditions of Problem (P_{x^*}) are satisfied at (x^*, μ^*, λ^*) . Then there exist $\delta_0 > 0$ and $\varepsilon_0 > 0$, and a twice continuously differentiable mapping $(y, \mu, \lambda) : \mathbf{B}_{\delta_0}(x^*) \rightarrow \mathbf{B}_{\varepsilon_0}(y^*) \times \mathbf{B}_{\varepsilon_0}(\mu^*) \times \mathbf{B}_{\varepsilon_0}(\lambda^*)$ such that Jacobian uniqueness conditions of Problem (P_x) are satisfied at $(y(x), \mu(x), \lambda(x))$ when $x \in \mathbf{B}_{\delta_0}(x^*)$.

By introducing a set of auxiliary variables w_1, \dots, w_{m_2} , we consider the following equality constrained optimization problem (\mathbb{P}_x).

$$\begin{aligned} \max \quad & f(x, z) \\ \text{s.t.} \quad & h(x, z) = 0, \\ & g(x, z) + w \circ w = 0, \\ & z \in \mathfrak{X}^m, w \in \mathfrak{X}^{m_2}, \end{aligned} \tag{2.2}$$

where \circ is the Hadamard product.

Remark 2.1 Let $w^* \in \mathfrak{X}^{m_2}$ with $w_i^* = \sqrt{-g_i(x^*, y^*)}$ for $i = 1, \dots, m_2$. With the help of Proposition 3.2 in [1], we may prove that the point $(y^*, w^*, \mu^*, \lambda^*)$ satisfies Karush-Kuhn-Tucker conditions, the linear independence constraint qualification as well as the second-order sufficient optimality condition for Problem (2.2) hold at $(y^*, w^*, \mu^*, \lambda^*)$ if Jacobian uniqueness conditions of Problem (\mathbb{P}_{x^*}) are satisfied at (x^*, μ^*, λ^*) .

Define the optimal value function

$$\varphi(x) = f(x, y(x)), \quad x \in \mathbf{B}_{\delta_0}(x^*), \tag{2.3}$$

where $y(x)$ is defined by Lemma 2.1.

Proposition 2.1 If the assumptions of Lemma 2.1 are satisfied and φ is defined by (2.3), then

$$\nabla_x \phi(x) = \nabla_x \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \tag{2.4}$$

and

$$\begin{aligned} \nabla^2 \phi(x) = \quad & \nabla_{xx}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ & - \begin{bmatrix} \nabla_{yx}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ 0 \\ \mathcal{J}_x h(x, y(x)) \\ \mathcal{J}_x g(x, y(x)) \end{bmatrix}^T K(x)^{-1} \begin{bmatrix} \nabla_{yx}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ 0 \\ \mathcal{J}_x h(x, y(x)) \\ \mathcal{J}_x g(x, y(x)) \end{bmatrix}, \end{aligned} \tag{2.5}$$

where

$$K(x) = \begin{bmatrix} \nabla_{yy}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) & 0 & \mathcal{J}_y h(x, y(x))^T & \mathcal{J}_y g(x, y(x))^T \\ 0 & -2\text{Diag}(\lambda(x)) & 0 & 2\text{Diag}(\sqrt{-g(x, y(x))}) \\ \mathcal{J}_y h(x, y(x)) & 0 & 0 & 0 \\ \mathcal{J}_y g(x, y(x)) & 2\text{Diag}(\sqrt{-g(x, y(x))}) & 0 & 0 \end{bmatrix}. \tag{2.6}$$

Proof. Consider Problem (2.2) and define its Lagrange function as

$$\mathbb{L}(x; z, w, \mu, \lambda) = f(x, z) + \mu^T h(x, z) - \lambda^T (g(x, z) + w \circ w) = \mathcal{L}(x; z, \mu, \lambda) - \lambda^T w \circ w.$$

From Remark 2.1, we can know that, when $x \in \mathbf{B}_{\delta_0}(x^*)$, for $w(x) = \sqrt{-g(x, y(x))}$, the point given by $(y(x), w(x), \mu(x), \lambda(x))$ satisfies Kurash-Kuhn-Tucker conditions for Problem (\mathbb{P}_x), the linear

independence constraint qualification holds at $(y(x), w(x))$ and the second-order sufficient optimality condition holds at $(y(x), w(x), \mu(x), \lambda(x))$. Kurash-Kuhn-Tucker conditions for Problem (\mathbb{P}_x) at $(y(x), w(x), \mu(x), \lambda(x))$ can be expressed as

$$T(x(x), y(x), w(x), \mu(x), \lambda(x)) = 0, \quad (2.7)$$

where

$$T(x, y, w, \mu, \lambda) = \begin{bmatrix} \nabla_y \mathbb{L}(x; y, w, \mu, \lambda) \\ \nabla_w \mathbb{L}(x; y, w, \mu, \lambda) \\ h(x, y) \\ g(x, y) + w \circ w \end{bmatrix}.$$

Differentiating both sides of (2.7) with respect to x yields

$$\mathcal{J}_{(y,w,\mu,\lambda)} T(x(x), y(x), w(x), \mu(x), \lambda(x)) \begin{bmatrix} \mathcal{J}y(x) \\ \mathcal{J}w(x) \\ \mathcal{J}\mu(x) \\ \mathcal{J}\lambda(x) \end{bmatrix} + \mathcal{J}_x T(x(x), y(x), w(x), \mu(x), \lambda(x)) = 0. \quad (2.8)$$

Noting that

$$\mathcal{J}_{(y,w,\mu,\lambda)} T(x(x), y(x), w(x), \mu(x), \lambda(x)) = K(x)$$

and

$$\mathcal{J}_x T(x(x), y(x), w(x), \mu(x), \lambda(x)) = \begin{bmatrix} \nabla_{x,y}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ \mathbf{0} \\ \mathcal{J}_x h(x, y(x)) \\ \mathcal{J}_x g(x, y(x)) \end{bmatrix},$$

we obtain (2.8) that

$$\begin{bmatrix} \mathcal{J}y(x) \\ \mathcal{J}w(x) \\ \mathcal{J}\mu(x) \\ \mathcal{J}\lambda(x) \end{bmatrix} = -K(x)^{-1} \begin{bmatrix} \nabla_{x,y}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ \mathbf{0} \\ \mathcal{J}_x h(x, y(x)) \\ \mathcal{J}_x g(x, y(x)) \end{bmatrix}. \quad (2.9)$$

Noting that

$$h(x, y(x)) = 0, \quad g(x, y(x)) + w(x) \circ w(x) = 0,$$

we have

$$\begin{aligned} \varphi(x) &= f(x, y(x)) \\ &= f(x, y(x)) + \mu(x)^T h(x, y(x)) - \lambda(x)^T [g(x, y(x)) + w(x) \circ w(x)] \\ &= \mathbb{L}(x; y(x), w(x), \mu(x), \lambda(x)). \end{aligned}$$

Thus we get

$$\begin{aligned} \nabla \varphi(x) &= \nabla_x \mathbb{L}(x; y(x), w(x), \mu(x), \lambda(x)) + \mathcal{J}y(x)^T \nabla_y \mathbb{L}(x; y(x), w(x), \mu(x), \lambda(x)) \\ &\quad + \mathcal{J}\mu(x)^T h(x, y(x)) + \mathcal{J}\lambda(x)^T [g(x, y(x)) + w(x) \circ w(x)] \\ &= \nabla_x \mathbb{L}(x; y(x), w(x), \mu(x), \lambda(x)) \\ &= \nabla_x \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \end{aligned}$$

and

$$\begin{aligned}
\nabla^2 \varphi(x) &= \nabla_{xx}^2 \mathbb{L}(x; y(x), w(x), \mu(x), \lambda(x)) + \mathcal{J}_{y,w,\mu,\lambda} \nabla_x \mathbb{L}(x; y(x), w(x), \mu(x), \lambda(x)) \mathcal{J} \begin{bmatrix} y(x) \\ w(x) \\ \mu(x) \\ \lambda(x) \end{bmatrix} \\
&= \nabla_{xx}^2 \mathbb{L}(x; y(x), w(x), \mu(x), \lambda(x)) \\
&\quad + \begin{bmatrix} \nabla_{xy}^2 \mathbb{L}(x; y(x), \mu(x), \lambda(x)) & 0 & \mathcal{J}_x h(x, y(x))^T & \mathcal{J}_x g(x, y(x)) \end{bmatrix} \begin{bmatrix} \mathcal{J} y(x) \\ \mathcal{J} w(x) \\ \mathcal{J} \mu(x) \\ \mathcal{J} \lambda(x) \end{bmatrix} \\
&= \nabla_{xx}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\
&\quad + \begin{bmatrix} \nabla_{xy}^2 \mathbb{L}(x; y(x), \mu(x), \lambda(x)) & 0 & \mathcal{J}_x h(x, y(x))^T & \mathcal{J}_x g(x, y(x)) \end{bmatrix} \begin{bmatrix} \mathcal{J} y(x) \\ \mathcal{J} w(x) \\ \mathcal{J} \mu(x) \\ \mathcal{J} \lambda(x) \end{bmatrix}.
\end{aligned}$$

Combing this with (2.9), we obtain (2.4) and (2.5). \square

2.2 Strong regularity of Karush-Kuhn-Tucker system

We use $\Lambda_{x^*}(y^*)$ to denote the set of all $(\mu^*, \lambda^*) \in \mathfrak{X}^{m_1} \times \mathfrak{X}^{m_2}$ satisfying Karush-Kuhn-Tucker conditions at y^* for Problem (P_{x^*}) .

Definition 2.2 Let $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point at which $\Lambda_{x^*}(y^*) \neq \emptyset$. We say that the strong second-order sufficient optimality condition holds at y^* for Problem (P_{x^*}) if

$$\sup_{(\mu, \lambda) \in \Lambda_{x^*}(y^*)} \langle \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu, \lambda) d_y, d_y \rangle < 0 \quad \forall d_y \in \text{aff } C_{x^*}(y^*) \setminus \{0\},$$

where $C_{x^*}(y^*)$ is the critical cone of Problem (P_{x^*}) at y^* .

Definition 2.3 Let $(x, y) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point. We say that **Assumption A** holds at $y \in Y(x)$ for Problem (P_x) if $\Lambda_x(y) \neq \emptyset$, the linear independence constraint qualification and the strong second-order sufficient optimality condition hold at y .

Lemma 2.2 Let $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point around which f, h, g are twice continuously differentiable. Suppose that **Assumption A** holds at y^* for Problem (P_{x^*}) . Then there exist $\delta_0 > 0$ and $\varepsilon_0 > 0$, and a locally Lipschitz continuous mapping $(y, \mu, \lambda) : \mathbf{B}_{\delta_0}(x^*) \rightarrow \mathbf{B}_{\varepsilon_0}(y^*) \times \mathbf{B}_{\varepsilon_0}(\mu^*) \times \mathbf{B}_{\varepsilon_0}(\lambda^*)$ satisfying $(y(x^*), \mu(x^*), \lambda(x^*)) = (y^*, \mu^*, \lambda^*)$ and

$$\begin{aligned}
\nabla_y \mathcal{L}(x; y(x), \mu(x), \lambda(x)) &= 0, \\
h(x, y(x)) &= 0, \\
g(x, y(x)) - \Pi_{\mathfrak{X}^{m_2}}(\lambda(x) + g(x, y(x))) &= 0
\end{aligned} \tag{2.10}$$

for $x \in \mathbf{B}_{\delta_0}(x^*)$. Moreover, **Assumption A** holds at $y(x)$ for Problem (P_x) when $x \in \mathbf{B}_{\delta_0}(x^*)$.

Proof. The first part of this lemma is from Robinson (1980) [13]. For the second part, Karush-Kuhn-Tucker conditions of Problem (P_x) at $(y(x), \mu(x), \lambda(x))$ are from (2.10) and the linear independence constraint qualification of $y(x)$ for Problem (P_x) comes from the continuity of $\mathcal{J}_y h$ and $\mathcal{J}_y g$ and the linear independence constraint qualification for (P_{x^*}) at y^* . We only need to prove the strong second-order sufficient optimality condition for Problem (P_x) at $(y(x), \mu(x), \lambda(x))$. Let $\alpha_{x^*} = \{i : \lambda_i^* > 0, i = 1, \dots, m_2\}$. Then the affine space of the critical cone $C_{x^*}(y^*)$ under **Assumption A** is expressed as

$$\text{aff } C_{x^*}(y^*) = \{d_y \in \mathfrak{R}^m : \mathcal{J}_y h(x^*, y^*) d_y = 0; \nabla_y g_i(x^*, y^*)^T d_y = 0, i \in \alpha_{x^*}\}.$$

Noting that for $\alpha_x = \{i : \lambda_i(x) > 0, i = 1, \dots, m_2\}$, we have

$$\text{aff } C_x(y(x)) = \{d_y \in \mathfrak{R}^m : \mathcal{J}_y h(x, y(x)) d_y = 0; \nabla_y g_i(x, y(x))^T d_y = 0, i \in \alpha_x\}.$$

Since $\lambda(x)$ is continuous at x^* and $\lambda_i^* > 0$ for α_{x^*} , we have $\alpha_x \supseteq \alpha_{x^*}$, and in turn,

$$\text{aff } C_x(y(x)) \subseteq \Gamma^*(x) = \{d_y \in \mathfrak{R}^m : \mathcal{J}_y h(x, y(x)) d_y = 0; \nabla_y g_i(x, y(x))^T d_y = 0, i \in \alpha_{x^*}\}. \quad (2.11)$$

It follows from the strong second-order sufficient condition for Problem (P_{x^*}) at y^* that the matrix $\nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*)$ is negatively definite on $\Gamma^*(x^*)$. Thus we have for small $\delta_0 > 0$ and $x \in \mathbf{B}_{\delta_0}(x^*)$ that $\nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*)$ is negatively definite on $\Gamma^*(x)$, which with (2.11) implies that $\nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*)$ is negatively definite on $\text{aff } C_x(y(x))$. Finally, from the continuity of the matrix $\nabla_{yy}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x))$ with respect to x around x^* , we obtain that $\nabla_{yy}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x))$ is negatively definite over $\text{aff } C_x(y(x))$, indicating the strong second-order sufficient optimality condition holds at $y(x)$ for (P_x) when x is around x^* . \square

Remark 2.2 Let **Assumption A** be satisfied. Then from Lemma 2.2, for $x \in \mathbf{B}_{\delta_0}(x^*)$, $y(x)$ is a unique local maximizer of Problem (P_x) .

For a linear operator $W : \mathfrak{R}^{m_2} \rightarrow \mathfrak{R}^{m_2}$, define

$$\mathcal{A}(x, W) = \begin{bmatrix} \nabla_{yy}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) & \mathcal{J}_y h(x, y(x))^T & \mathcal{J}_y g(x, y(x))^T \\ \mathcal{J}_y h(x, y(x)) & 0 & 0 \\ (I - W) \mathcal{J}_y g(x, y(x)) & 0 & W \end{bmatrix}, \quad (2.12)$$

a set-valued mapping $\mathbb{A}_B : \mathfrak{R}^n \rightrightarrows \mathfrak{R}^{m_2 \times m_2}$ by

$$\mathbb{A}_B(x) = \{\mathcal{A}(x, W) : W \in \partial_B \Pi_{\mathfrak{R}_-^{m_2}}(\lambda(x) + g(x, y(x)))\}, \quad (2.13)$$

and a set-valued mapping $\mathbb{A}_C : \mathfrak{R}^n \rightrightarrows \mathfrak{R}^{m_2 \times m_2}$ by

$$\mathbb{A}_C(x) = \{\mathcal{A}(x, W) : W \in \partial \Pi_{\mathfrak{R}_-^{m_2}}(\lambda(x) + g(x, y(x)))\}. \quad (2.14)$$

Proposition 2.2 Let $(x^*, y^*) \in \mathfrak{R}^n \times \mathfrak{R}^m$ be a point around which f, h, g are twice continuously differentiable. Suppose that **Assumption A** holds at y^* for Problem (P_{x^*}) . Then every element in $\mathbb{A}_C(x^*)$ is nonsingular.

Proof. Let $V \in \mathbb{A}_B(x^*)$. Then there exists an element $W \in \partial_B \Pi_{\mathfrak{X}^{m_2}}(\lambda^* + g(x^*, y^*))$ such that

$$V = \begin{bmatrix} \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) & \mathcal{J}_y h(x^*, y^*)^T & \mathcal{J}_y g(x^*, y^*)^T \\ \mathcal{J}_y h(x^*, y^*) & 0 & 0 \\ (I - W)\mathcal{J}_y g(x^*, y^*) & 0 & W \end{bmatrix}.$$

Define

$$\alpha = \{i : g_i(x^*, y^*) = 0, \lambda_i > 0\}, \beta = \{i : g_i(x^*, y^*) = 0, \lambda_i = 0\}, \gamma = \{i : g_i(x^*, y^*) < 0, \lambda_i = 0\}.$$

Hence W can be expressed as

$$W = \text{Diag}(w_1, \dots, w_{m_2})$$

with

$$w_i \begin{cases} = 0 & i \in \alpha, \\ \in [0, 1] & i \in \beta, \\ = 1 & i \in \gamma. \end{cases}$$

Then V may be expressed as

$$V = \begin{bmatrix} \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) & \mathcal{J}_y h(x^*, y^*)^T & \mathcal{J}_y g_\alpha(x^*, y^*)^T & \mathcal{J}_y g_\beta(x^*, y^*)^T & \mathcal{J}_y g_\gamma(x^*, y^*)^T \\ \mathcal{J}_y h(x^*, y^*) & 0 & 0 & 0 & 0 \\ \mathcal{J}_y g_\alpha(x^*, y^*) & 0 & 0 & W_\beta & 0 \\ (I_\beta - W_\beta)\mathcal{J}_y g_\beta(x^*, y^*) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{|\gamma|} \end{bmatrix}.$$

For $\xi_1 \in \mathfrak{R}^m$, $\xi_2 \in \mathfrak{R}^{m_1}$, $\xi_3 \in \mathfrak{R}^{|\alpha|}$, $\xi_4 \in \mathfrak{R}^{|\beta|}$ and $\xi_5 \in \mathfrak{R}^{|\gamma|}$, consider

$$V \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{bmatrix} = 0.$$

or equivalently

$$\begin{aligned} \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) \xi_1 + \mathcal{J}_y h(x^*, y^*)^T \xi_2 + \mathcal{J}_y g_\alpha(x^*, y^*)^T \xi_3 \\ + \mathcal{J}_y g_\beta(x^*, y^*)^T \xi_4 + \mathcal{J}_y g_\gamma(x^*, y^*)^T \xi_5 &= 0, \\ \mathcal{J}_y h(x^*, y^*) \xi_1 &= 0, \\ \mathcal{J}_y g_\alpha(x^*, y^*) \xi_1 &= 0, \\ (I_{|\beta|} - W_\beta)\mathcal{J}_y g_\beta(x^*, y^*) \xi_1 + W_\beta \xi_4 &= 0, \\ \xi_5 &= 0, \end{aligned}$$

which implies $\xi_5 = 0$ and

$$\begin{aligned} \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) \xi_1 + \mathcal{J}_y h(x^*, y^*)^T \xi_2 + \mathcal{J}_y g_\alpha(x^*, y^*)^T \xi_3 + \mathcal{J}_y g_\beta(x^*, y^*)^T \xi_4 &= 0, \\ \mathcal{J}_y h(x^*, y^*) \xi_1 &= 0, \\ \mathcal{J}_y g_\alpha(x^*, y^*) \xi_1 &= 0, \\ (I_{|\beta|} - W_\beta)\mathcal{J}_y g_\beta(x^*, y^*) \xi_1 + W_\beta \xi_4 &= 0. \end{aligned} \tag{2.15}$$

Without loss of generality, we assume that $w_i \in (0, 1]$ so that W_β is nonsingular. From the fourth row of (2.15), we have

$$\xi_4 = -W_\beta^{-1}(I_{|\beta|} - W_\beta)\mathcal{J}_y g_\beta(x^*, y^*)\xi_1. \quad (2.16)$$

From the second row and the third row, we have

$$\mathcal{J}_y h(x^*, y^*)\xi_1 = 0, \mathcal{J}_y g_\alpha(x^*, y^*)\xi_1 = 0,$$

which implies $\xi_1 \in \text{aff } C_{x^*}(y^*)$. Substituting ξ_4 of (2.16) in (2.15) and premultiplying ξ_1^T to both sides of the first row of (2.15), we obtain

$$\xi_1^T \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) \xi_1 - \xi_1^T \mathcal{J}_y g_\beta(x^*, y^*)^T W_\beta^{-1} (I_{|\beta|} - W_\beta) \mathcal{J}_y g_\beta(x^*, y^*) \xi_1 = 0. \quad (2.17)$$

Since $TW_\beta^{-1}(I_{|\beta|} - W_\beta)$ is diagonal with $W_{ii} = (1 - w_i)/w_i > 0$, we have that this matrix is positively definite, and

$$-\xi_1^T \mathcal{J}_y g_\beta(x^*, y^*)^T W_\beta^{-1} (I_{|\beta|} - W_\beta) \mathcal{J}_y g_\beta(x^*, y^*) \xi_1 \leq 0.$$

Therefore we obtain $\xi_1 = 0$ from $\xi_1 \in \text{aff } C_{x^*}(y^*)$ and the strong second-order sufficient optimality condition. Obviously we have $\xi_4 = 0$. Substitute $\xi_1 = 0$ and $\xi_4 = 0$ to the first row of (2.15) and using linear independence constraint qualification, we get $\xi_2 = 0$ and $\xi_3 = 0$. Therefore, matrix V is nonsingular. The proof is complete. \square

Corollary 2.1 *Let $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point around which f, h, g are twice continuously differentiable. Suppose that **Assumption A** holds at y^* for Problem (P_{x^*}) . Then every element in $\mathbb{A}_B(x^*)$ is nonsingular.*

Proof. Let $V \in \mathbb{A}_B(x^*)$. Then there exists an element $W \in \partial_B \Pi_{\mathfrak{X}_-^{m_2}}(\lambda^* + g(x^*, y^*))$ such that

$$V = \begin{bmatrix} \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) & \mathcal{J}_y h(x^*, y^*)^T & \mathcal{J}_y g(x^*, y^*)^T \\ \mathcal{J}_y h(x^*, y^*) & 0 & 0 \\ (I - W)\mathcal{J}_y g(x^*, y^*) & 0 & W \end{bmatrix}.$$

Define

$$\alpha = \{i : g_i(x^*, y^*) = 0, \lambda_i > 0\}, \beta = \{i : g_i(x^*, y^*) = 0, \lambda_i = 0\}, \gamma = \{i : g_i(x^*, y^*) < 0, \lambda_i = 0\}.$$

There exists a partition of β , say (β_1, β_2) , namely $\beta_1 \cup \beta_2 = \beta$ and $\beta_1 \cap \beta_2 = \emptyset$, such that

$$W = \text{Diag}(w_1, \dots, w_{m_2})$$

with

$$w_i = \begin{cases} 0 & i \in \tilde{\alpha}, \\ 1 & i \in \tilde{\gamma}, \end{cases}$$

where $\tilde{\alpha} = \alpha \cup \beta_1$ and $\tilde{\gamma} = \beta_2 \cup \gamma$. Then V may be expressed as

$$V = \begin{bmatrix} \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) & \mathcal{J}_y h(x^*, y^*)^T & \mathcal{J}_y g_{\tilde{\alpha}}(x^*, y^*)^T & \mathcal{J}_y g_{\tilde{\gamma}}(x^*, y^*)^T \\ \mathcal{J}_y h(x^*, y^*) & 0 & 0 & 0 \\ \mathcal{J}_y g_{\tilde{\alpha}}(x^*, y^*) & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{|\tilde{\gamma}|} \end{bmatrix}.$$

The nonsingularity of V can be proved in a similar way as in the proof of Proposition 2.2. \square

Proposition 2.3 Let $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point around which f, h, g are twice continuously differentiable. Suppose that **Assumption A** holds at y^* for Problem (P_{x^*}) . Let $\delta_0 > 0$ be given in Lemma 2.2. Then the set-valued mapping $\mathbb{A}_B(x)$ ($\mathbb{A}_C(x)$) is upper semicontinuous at x^* , and for small $\delta \in (0, \delta_0)$, every element in $\mathbb{A}_B(x)$ ($\mathbb{A}_C(x)$) is nonsingular when $x \in \mathbf{B}(x^*, \delta)$.

Proposition 2.4 Let $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point around which f, h, g are twice continuously differentiable. Suppose that **Assumption A** holds at y^* for Problem (P_{x^*}) . Let $\delta_0 > 0$, $\varepsilon_0 > 0$ and $(y(\cdot), \mu(\cdot), \lambda(\cdot))$ be given in Lemma 2.2. Then for $x \in \mathbf{B}_{\delta_0}(x^*)$,

(i) The directional derivative of $(y(\cdot), \mu(\cdot), \lambda(\cdot))$ at x satisfies

$$\begin{pmatrix} y'(x; d_x) \\ \mu'(x; d_x) \\ \lambda'(x; d_x) \end{pmatrix} \in \left\{ -\mathcal{A}(x, W)^{-1} \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) d_x \\ \mathcal{J}_x h(x, y(x)) d_x \\ (I - W) \mathcal{J}_x g(x, y(x)) d_x \end{pmatrix} : W \in \partial_B \Pi_{\mathfrak{X}_-^{m_2}}(\lambda(x) + g(x, y(x))) \right\}. \quad (2.18)$$

(ii) The B -subdifferential of $(y(\cdot), \mu(\cdot), \lambda(\cdot))$ at x satisfies

$$\partial_B \begin{pmatrix} y \\ \mu \\ \lambda \end{pmatrix} (x) \in \left\{ -\mathcal{A}(x, W)^{-1} \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ \mathcal{J}_x h(x, y(x)) \\ (I - W) \mathcal{J}_x g(x, y(x)) \end{pmatrix} : W \in \partial_B \Pi_{\mathfrak{X}_-^{m_2}}(\lambda(x) + g(x, y(x))) \right\}. \quad (2.19)$$

(iii) Clarke generalized Jacobian of $(y(\cdot), \mu(\cdot), \lambda(\cdot))$ at x satisfies

$$\partial \begin{pmatrix} y \\ \mu \\ \lambda \end{pmatrix} (x) \in \left\{ \mathcal{A}(x, W)^{-1} \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ \mathcal{J}_x h(x, y(x)) \\ (I - W) \mathcal{J}_x g(x, y(x)) \end{pmatrix} : W \in \partial \Pi_{\mathfrak{X}_-^{m_2}}(\lambda(x) + g(x, y(x))) \right\}. \quad (2.20)$$

Proof. It follows from (2.10), for $d_x \in \mathfrak{X}^n$ that

$$\begin{aligned} & \nabla_{yx}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) d_x + \nabla_{yy}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) y'(x; d_x) \\ & \quad + \mathcal{J}_y h(x, y(x)) \mu'(x; d_x) - \mathcal{J}_y g(x, y(x))^T \lambda'(x; d_x) = 0, \\ & \quad \mathcal{J}_x h(x, y(x)) d_x + \mathcal{J}_y h(x, y(x)) y'(x; d_x) = 0, \\ & \mathcal{J}_x g(x, y(x)) d_x + \mathcal{J}_y g(x, y(x)) y'(x; d_x) - \Pi'_{\mathfrak{X}_-^{m_2}}(\lambda(x) + g(x, y(x)); \lambda'(x; d_x) \\ & \quad + \mathcal{J}_x g(x, y(x)) d_x + \mathcal{J}_y g(x, y(x)) y'(x; d_x)) = 0. \end{aligned} \quad (2.21)$$

Since $\Pi_{\mathfrak{X}_-^{m_2}}$ is semismooth everywhere, we have from [12] that there exists an matrix $\widetilde{W} \in \partial_B \Pi_{\mathfrak{X}_-^{m_2}}(\lambda(x) + g(x, y(x)))$ such that

$$\begin{aligned} & \Pi'_{\mathfrak{X}_-^{m_2}}(\lambda(x) + g(x, y(x)); \lambda'(x; d_x) + \mathcal{J}_x g(x, y(x)) d_x + \mathcal{J}_y g(x, y(x)) y'(x; d_x)) \\ & \quad = \widetilde{W} [\lambda'(x; d_x) + \mathcal{J}_x g(x, y(x)) d_x + \mathcal{J}_y g(x, y(x)) y'(x; d_x)]. \end{aligned} \quad (2.22)$$

Substituting (2.22) into (2.21), we may rewrite (2.21) as

$$\mathcal{A}(x, \widetilde{W}) \begin{pmatrix} y'(x; d_x) \\ \mu'(x; d_x) \\ \lambda'(x; d_x) \end{pmatrix} = - \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) d_x \\ \mathcal{J}_x h(x, y(x)) d_x \\ (I - \widetilde{W}) \mathcal{J}_x g(x, y(x)) d_x \end{pmatrix},$$

which implies (2.18).

We use $(y(x); \mu(x); \lambda(x))$ to denote $(y(x)^T, \mu(x)^T, \lambda(x)^T)^T$. Let $\mathcal{D}_1(\delta_0)$ be the set of all differentiable points of $(y(x); \mu(x); \lambda(x))$ in $\mathbf{B}_{\delta_0}(x^*)$ and $\mathcal{D}_2(\delta_0)$ be the set of all differentiable points such that $\lambda(x) - g(x, y(x)) > 0$ in $\mathbf{B}_{\delta_0}(x^*)$ (namely; the points at which $\Pi_{\mathfrak{X}_-^{m_2}}$ is differentiable), and define $\mathcal{D}(\delta_0) = \mathcal{D}_1(\delta_0) \cap \mathcal{D}_2(\delta_0)$. For every $V \in \partial_B(y(x); \mu(x); \lambda(x))$, there exists a sequence $x^k \rightarrow x$ with $x^k \in \mathcal{D}(\delta_0)$ such that

$$\mathcal{J}_x(y(x^k); \mu(x^k); \lambda(x^k)) \rightarrow V.$$

It follows from (2.10) and $x^k \in \mathcal{D}(\delta_0)$ that

$$\begin{aligned} & \nabla_{yx}^2 \mathcal{L}(x^k; y(x^k), \mu(x^k), \lambda(x^k)) + \nabla_{yy}^2 \mathcal{L}(x^k; y(x^k), \mu(x^k), \lambda(x^k)) \mathcal{J}y(x^k) \\ & \quad + \mathcal{J}_y h(x^k, y(x^k)) \mathcal{J}\mu(x^k) - \mathcal{J}_y g(x^k, y(x^k))^T \mathcal{J}\lambda(x^k) = 0, \\ & \quad \mathcal{J}_x h(x^k, y(x^k)) + \mathcal{J}_y h(x^k, y(x^k)) \mathcal{J}y(x^k) = 0, \\ & \mathcal{J}_x g(x^k, y(x^k)) + \mathcal{J}_y g(x^k, y(x^k)) \mathcal{J}y(x^k) - \mathcal{J}\Pi_{\mathfrak{X}_-^{m_2}}(\lambda(x^k) + g(x^k, y(x^k))) (\mathcal{J}\lambda(x^k) \\ & \quad + \mathcal{J}_x g(x^k, y(x^k)) + \mathcal{J}_y g(x^k, y(x^k)) \mathcal{J}y(x^k)) = 0 \end{aligned}$$

or, equivalently,

$$\mathcal{J} \begin{pmatrix} y \\ \mu \\ \lambda \end{pmatrix} (x^k) = -\mathcal{A}(x^k, W^k)^{-1} \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}(x^k; y(x^k), \mu(x^k), \lambda(x^k)) \\ \mathcal{J}_x h(x^k, y(x^k)) \\ (I - W^k) \mathcal{J}_x g(x^k, y(x^k)) \end{pmatrix} \quad (2.23)$$

with

$$W^k = \mathcal{J}\Pi_{\mathfrak{X}_-^{m_2}}(\lambda(x^k) + g(x^k, y(x^k))).$$

Let $W = \lim_{k \rightarrow \infty} W^k$ (or assume that W is an limit operator of W^k). Then $W \in \partial_B \Pi_{\mathfrak{X}_-^{m_2}}(\lambda(x) + g(x, y(x)))$. Taking the limit in both sides of (2.23) as $k \rightarrow \infty$, we obtain the result in (2.19). From the definition of Clarke generalized Jacobian, we obtain (2.20) from (2.19). \square

Define

$$H(x, W) = \mathcal{A}(x, W)^{-1} \begin{pmatrix} \nabla_{yx}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ \mathcal{J}_x h(x, y(x)) \\ (I - W) \mathcal{J}_x g(x, y(x)) \end{pmatrix}. \quad (2.24)$$

Then we obtain the directional derivative, B-subdifferential and Clarke generalized subdifferential of ϕ at x by the following corollary.

Corollary 2.2 *Let $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point around which f, h, g are twice continuously differentiable. Suppose that **Assumption A** holds at y^* for Problem (P_{x^*}) . Let $\delta_0 > 0$, $\varepsilon_0 > 0$ and $(y(\cdot), \mu(\cdot), \lambda(\cdot))$ be given in Lemma 2.2. Then φ is locally Lipschitz continuous in $\mathbf{B}_{\delta_0}(x^*)$ and for $x \in \mathbf{B}_{\delta_0}(x^*)$,*

(i) *The directional derivative of φ at x satisfies*

$$\begin{aligned} \varphi'(x; d_x) & \in \nabla_x \mathcal{L}(x; y(x), \mu(x), \lambda(x)) d_x \\ & - \left\{ \nabla_{y,\mu,\lambda} \mathcal{L}(x; y(x), \mu(x), \lambda(x))^T H(x, W) d_x : W \in \partial_B \Pi_{\mathfrak{X}_-^{m_2}}(\lambda(x) + g(x, y(x))) \right\}. \end{aligned} \quad (2.25)$$

(ii) *The B-subdifferential of φ at x satisfies*

$$\begin{aligned} \partial_B \varphi(x) & \in \nabla_x \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ & - \left\{ H(x, W)^T : W \in \partial_B \Pi_{\mathfrak{X}_-^{m_2}}(\lambda(x) + g(x, y(x))) \right\} \nabla_{y,\mu,\lambda} \mathcal{L}(x; y(x), \mu(x), \lambda(x)). \end{aligned} \quad (2.26)$$

(ii) The Clarke generalized subdifferential of φ at x satisfies

$$\begin{aligned} \partial\varphi(x) \in & \nabla_x \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ & - \left\{ H(x, W)^T : W \in \partial \Pi_{\mathfrak{R}^{m_2}}(\lambda(x) + g(x, y(x))) \right\} \nabla_{y, \mu, \lambda} \mathcal{L}(x; y(x), \mu(x), \lambda(x)). \end{aligned} \quad (2.27)$$

3 Optimality Conditions

Suppose that $\varphi(x)$ is defined by (2.3). Then the constrained minimax problem (1.1) is locally reduced to

$$\begin{aligned} \min \quad & \varphi(x) = f(x, y(x)) \\ \text{s.t.} \quad & x \in \Phi \cap \mathbf{B}_{\delta_0}(x^*), \end{aligned} \quad (3.1)$$

where $y(x)$ is a local minimizer of (P_x) around y^* and Φ is defined by (1.2).

For $x^* \in \Phi$, the Mangasarian-Fromovitz constraint qualification is said to hold at x^* the constraint set Φ if

- (a) The set of vectors $\nabla H_j(x^*), j = 1, \dots, n_1$ are linearly independent;
- (b) There exists a vector $\bar{d} \in \mathfrak{R}^n$ such that

$$\nabla H_j(x^*)^T \bar{d} = 0, j = 1, \dots, n_1, \nabla G_i(x^*)^T \bar{d} < 0, i \in I(x^*),$$

where $I(x^*) = \{i : G_i(x^*) = 0, i = 1, \dots, n_2\}$.

Define the critical cone of Problem (3.1) at x^* is defined by

$$C(x^*) = \{d_x \in \mathfrak{R}^n : \mathcal{J}H(x^*)d_x = 0; \nabla G_i(x^*)^T d_x \leq 0, i \in I(x^*); \varphi'(x^*; d_x) \leq 0\}. \quad (3.2)$$

We now derive necessary optimality conditions and second-order sufficient optimality conditions for Problem (1.1) under Jacobian uniqueness conditions for (P_{x^*}) . In this case, the critical cone $C(x^*)$ can be expressed as

$$C(x^*) = \{d_x \in \mathfrak{R}^n : \mathcal{J}H(x^*)d_x = 0; \nabla G_i(x^*)^T d_x \leq 0, i \in I(x^*); \nabla_x \mathcal{L}(x^*; y^*, \mu^*, \lambda^*)^T d_x \leq 0\}. \quad (3.3)$$

Theorem 3.1 (Necessary Optimality Conditions) *Let $(x^*, y^*) \in \mathfrak{R}^n \times \mathfrak{R}^m$ be a point around which f, h, g are twice continuously differentiable and H, G are twice continuously differentiable around x^* . Let (x^*, y^*) be a local minimax point of Problem (1.1). Assume that the linear independence constraint qualification holds at y^* for constraint set $Y(x^*)$. Then there exists a unique vector $(\mu^*, \lambda^*) \in \mathfrak{R}^{m_1} \times \mathfrak{R}^{m_2}$ such that*

$$\begin{aligned} \nabla_y \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) &= 0, \\ h(x^*, y^*) &= 0, \\ 0 &\geq \lambda^* \perp g(x^*, y^*) \leq 0. \end{aligned} \quad (3.4)$$

For any $d_y \in C_{x^*}(y^*)$, we have that

$$\langle \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) d_y, d_y \rangle \leq 0. \quad (3.5)$$

Assuming Problem (P_{x^*}) satisfies Jacobian uniqueness conditions at (y^*, μ^*, λ^*) and the Mangasarian-Fromovitz constraint qualification holds at x^* for the constraint set Φ , there exists $(u^*, v^*) \in \mathfrak{R}^{n_1} \times \mathfrak{R}^{n_2}$ such that

$$\begin{aligned} \nabla_x \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) + \mathcal{J}H(x^*)^T u^* + \mathcal{J}G(x^*)^T v^* &= 0, \\ H(x^*) &= 0, \\ 0 \leq v^* \perp G(x^*) \leq 0. \end{aligned} \quad (3.6)$$

The set of all (u^*, v^*) satisfying (3.6), denoted by $\Lambda(x^*)$, is nonempty compact convex set. Furthermore, for every $d_x \in C(x^*)$ where $C(x^*)$ is defined by (3.3),

$$\begin{aligned} \max_{(u,v) \in \Lambda(x^*)} \left\{ \left\langle \left[\sum_{j=1}^{n_1} u_j \nabla_{xx}^2 H_j(x^*) + \sum_{i=1}^{n_2} v_i \nabla_{xx}^2 G_i(x^*) \right] d_x, d_x \right\rangle \right. \\ \left. + \left\langle \left[\nabla_{xx}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) - N(x^*)^T K(x^*)^{-1} N(x^*) \right] d_x, d_x \right\rangle \geq 0, \right. \end{aligned} \quad (3.7)$$

where $K(x)$ is defined by (2.6) and $N(x)$ is defined by

$$N(x) = \begin{bmatrix} \nabla_{x,y}^2 \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ 0 \\ \mathcal{J}_x h(x, y(x)) \\ \mathcal{J}_x g(x, y(x)) \end{bmatrix}. \quad (3.8)$$

Proof. Since y^* is a local minimizer of (P_{x^*}) and the linear independence constraint qualification holds at y^* for the constraint set

$$Y(x^*) = \{y \in \mathfrak{R}^m : h(x^*, y) = 0, g(x^*, y) \leq 0\},$$

we may obtain the first-order and second-order necessary optimality conditions (3.4) and (3.5) from [3]. Noting that x^* is a local minimizer of the following problem

$$\begin{aligned} \min \quad & \varphi(x) = f(x, y(x)) \\ \text{s.t.} \quad & x \in \Phi \cap \mathbf{B}_{\delta_0}(x^*). \end{aligned}$$

The Lagrange function of the above problem is

$$L(x, u, v) = \varphi(x) + u^T H(x) + v^T G(x).$$

It follows from [2] that there exist u^* and v^* such that

$$\begin{aligned} \nabla_x L(x^*, u^*, v^*) &= 0, \\ H(x^*) &= 0, \\ 0 \leq u^* \perp G(x^*) \leq 0. \end{aligned} \quad (3.9)$$

Since the Mangasarian-Fromovitz constraint qualification holds at x^* for the constraint set Φ , the set of all vectors (u^*, v^*) satisfying (3.9) is a nonempty compact convex set. From the formula for $\nabla \varphi(x)$ in (2.4), we obtain (3.6) from (3.9) and $\Lambda(x^*)$ is nonempty compact convex. It also follows from [2] that the second-order necessary optimality conditions at x^* can be expressed as

$$\max_{(u,v) \in \Lambda(x^*)} \left\{ \left\langle \left[\nabla^2 \varphi(x^*) + \sum_{j=1}^{n_1} u_j \nabla_{xx}^2 H_j(x^*) + \sum_{i=1}^{n_2} v_i \nabla_{xx}^2 G_i(x^*) \right] d_x, d_x \right\rangle \right\} \geq 0, \quad \forall d_x \in C(x^*). \quad (3.10)$$

From the expression of $\nabla^2 \varphi(x)$ in (2.5), we obtain (3.7) from (3.10). The proof is complete. \square

Theorem 3.2 (Second-order Sufficient Optimality Conditions) Let $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point around which f, h, g are twice continuously differentiable and H, G are twice continuously differentiable around x^* . Assume that $x^* \in \Phi$ and $y^* \in Y(x^*)$. Let $(\mu^*, \lambda^*) \in \mathfrak{X}^{m_1} \times \mathfrak{X}^{m_2}$. Suppose that Problem (P_{x^*}) satisfies Jacobian uniqueness conditions at (y^*, μ^*, λ^*) , $\Lambda(x^*) \neq \emptyset$, and for every $d_x \in C(x^*) \setminus \emptyset$ (where $C(x^*)$ is defined by (3.3)),

$$\sup_{(u,v) \in \Lambda(x^*)} \left\{ \left\langle \left[\sum_{j=1}^{n_1} u_j \nabla_{xx}^2 H_j(x^*) + \sum_{i=1}^{n_2} v_i \nabla_{xx}^2 G_i(x^*) \right] d_x, d_x \right\rangle + \left\langle \left[\nabla_{xx}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) - N(x^*)^T K(x^*)^{-1} N(x^*) \right] d_x, d_x \right\rangle \right\} > 0, \quad (3.11)$$

where $K(x)$ is defined by (2.6) and $N(x)$ is defined by (3.8). Then there exist $\delta_1 \in (0, \delta_0)$, $\varepsilon_1 \in (0, \varepsilon_0)$ (where δ_0 and ε_0 are given by Lemma (2.1)) and $\gamma_1 > 0, \gamma_2 > 0$ such that for $x \in \mathbf{B}_{\delta_1}(x^*) \cap \Phi$ and $y \in \mathbf{B}_{\varepsilon_1}(y^*) \cap Y(x^*)$,

$$f(x^*, y) + \gamma_1 \|y - y^*\|^2/2 \leq f(x^*, y^*) \leq \sup_{z \in Y(x) \cap \mathbf{B}_{\varepsilon_0}(y^*)} f(x, z) - \gamma_2 \|x - x^*\|^2/2, \quad (3.12)$$

which implies that (x^*, y^*) a local minimax point of Problem (1.1).

Proof. As Jacobian uniqueness conditions hold at y^* for Problem (P_{x^*}) , we know that the local second-order descent condition holds for Problem (P_{x^*}) at y^* . Thus there exist $\gamma_1 > 0$ and $\varepsilon_1 \in (0, \varepsilon_0)$ such that

$$f(x^*, y) + \gamma_1 \|y - y^*\|^2/2 \leq f(x^*, y^*), \quad y \in \mathbf{B}_{\varepsilon_1}(y^*) \cap Y(x^*).$$

From the formula for $\nabla \varphi(x)$ in (2.4) and the formula for $\nabla^2 \varphi(x)$ in (2.5), we have from the definition (3.3) that

$$C(x^*) = \{d_x \in \mathfrak{X}^n : \mathcal{J}H(x^*)d_x = 0; \nabla G_i(x^*)^T d_x \leq 0, i \in I(x^*); \nabla \varphi(x^*)^T d_x \leq 0\},$$

which is just the critical cone of Problem (3.1), and (3.11) is the second-order sufficient optimality condition for Problem (3.1). Then the second-order growth condition of Problem (3.1) holds at x^* from [2]; namely, there exist $\gamma_2 > 0$ and $\delta_1 \in (0, \delta_0)$ such that

$$\varphi(x^*) + \gamma_2 \|x - x^*\|^2/2 \leq \varphi(x), \quad x \in \mathbf{B}_{\delta_1}(x^*) \cap \Phi,$$

which combining the expression

$$\varphi(x) = \sup_{z \in Y(x) \cap \mathbf{B}_{\varepsilon_0}(y^*)} f(x, z)$$

yields

$$f(x^*, y^*) \leq \sup_{z \in Y(x) \cap \mathbf{B}_{\varepsilon_0}(y^*)} f(x, z) - \gamma_2 \|x - x^*\|^2/2.$$

Therefore the inequalities in (3.12) are demonstrated. \square

In the following, we derive necessary optimality conditions for Problem (1.1) under **Assumption A** for (P_{x^*}) . Define the outer approximation of $\partial \varphi(x)$ by

$$\begin{aligned} \widetilde{\partial} \varphi(x) = & \nabla_x \mathcal{L}(x; y(x), \mu(x), \lambda(x)) \\ & - \left\{ H(x, W)^T \nabla_{y, \mu, \lambda} \mathcal{L}(x; y(x), \mu(x), \lambda(x)) : W \in \partial \Pi_{\mathfrak{X}^{m_2}}(\lambda(x) + g(x, y(x))) \right\}. \end{aligned} \quad (3.13)$$

Theorem 3.3 (First-order Necessary Optimality Conditions under **Assumption A**) Let $(x^*, y^*) \in \mathfrak{X}^n \times \mathfrak{X}^m$ be a point around which f, h, g are twice continuously differentiable and H, G are continuously differentiable around x^* . Let (x^*, y^*) be a local minimax point of Problem (1.1). Assume that linear independence constraint qualification holds at y^* for constraint set $Y(x^*)$. Then there exists a unique vector $(\mu^*, \lambda^*) \in \mathfrak{X}^{m_1} \times \mathfrak{X}^{m_2}$ such that

$$\begin{aligned} \nabla_y \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) &= 0, \\ h(x^*, y^*) &= 0, \\ 0 &\geq \lambda^* \perp g(x^*, y^*) \leq 0. \end{aligned} \quad (3.14)$$

For any $d_y \in C_{x^*}(y^*)$, we have that

$$\langle \nabla_{yy}^2 \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) d_y, d_y \rangle \leq 0. \quad (3.15)$$

Suppose that Problem (P_{x^*}) satisfies **Assumption A** at (y^*, μ^*, λ^*) and the Mangasarian-Fromovitz constraint qualification holds at x^* for the constraint set Φ . Then there exists $(u^*, v^*) \in \mathfrak{X}^{n_1} \times \mathfrak{X}^{n_2}$ and $W \in \partial \Pi_{\mathfrak{X}^{m_2}}(\lambda^* + g(x^*, y^*))$ such that

$$\begin{aligned} \nabla_x \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) - H(x^*, W)^T \nabla_{y, \mu, \lambda} \mathcal{L}(x^*; y^*, \mu^*, \lambda^*) + \mathcal{J}H(x^*)^T u^* + \mathcal{J}G(x^*)^T v^* &= 0, \\ H(x^*) &= 0, \\ 0 &\leq v^* \perp G(x^*) \leq 0, \end{aligned} \quad (3.16)$$

where $H(x, W)$ is defined by (2.24). The set of all (u^*, v^*) satisfying (3.16), denoted by $\Lambda(x^*)$, is a nonempty compact convex set.

Proof. Properties (3.14) and (3.15) are obvious from Theorem 3.1. Now we prove property (3.16). It follows from Corollary 2.2 that φ is locally Lipschitz continuous, also directionally differentiable in $\mathbf{B}(x^*, \delta_0)$. Thus we can easily get that $0 \in \mathfrak{X}^n$ is an optimal solution to the following problem

$$\begin{aligned} \min_{d_x} \quad & \varphi'(x^*; d_x) \\ \text{s.t.} \quad & d \in T_\Phi(x^*), \end{aligned} \quad (3.17)$$

where $T_\Phi(x^*)$ is the tangent cone of Φ at x^* . Since the Mangasarian-Fromovitz constraint qualification holds at x^* for the constraint set Φ , we have from [2] that

$$T_\Phi(x^*) = \{d_x \in \mathfrak{X}^n : \mathcal{J}H(x^*)d_x = 0, \nabla G_i(x^*)^T d_x \leq 0, i \in I(x^*)\}. \quad (3.18)$$

Since $\partial\varphi(x^*) \subset \widetilde{\partial}\varphi(x^*)$, from (2.25) and (2.27), we obtain

$$\varphi'(x^*; d_x) \leq \max\{v^T d_x : v \in \widetilde{\varphi}(x^*)\} = \delta^*(d_x | \widetilde{\varphi}(x^*)). \quad (3.19)$$

It follows from (3.18) and (3.19) and $0 \in \mathfrak{X}^n$ is the minimizer of Problem (3.17), we have that $0 \in \mathfrak{X}^n$ is the optimal solution to the following convex problem

$$\begin{aligned} \min_{d_x} \quad & \delta^*(d_x | \widetilde{\partial}\varphi(x^*)) \\ \text{s.t.} \quad & \nabla H_j(x^*)^T d_x = 0, j = 1, \dots, n_1, \\ & \nabla G_i(x^*)^T d_x = 0, i \in I(x^*). \end{aligned} \quad (3.20)$$

Noting that the Mangasarian-Fromovitz constraint qualification holds at x^* for the constraint set Φ , we have that Slater condition holds for convex optimization problem (3.20). Then, from the optimality conditions for convex programming, we have that there exist u^* and $v_i^*, i \in I(x^*)$ such that

$$0 \in \widetilde{\partial}\varphi(x^*) + \mathcal{J}H(x^*)^T u^* + \sum_{i \in I(x^*)} v_i^* \nabla G_i(x^*). \quad (3.21)$$

Therefore there exists $W \in \partial\Pi_{\mathfrak{X}_-^{m_2}}(\lambda^* + g(x^*, y^*))$ such that (3.16) is satisfied. Now we prove by contradiction that $\Lambda(x^*)$ is compact. Assuming that the set is unbounded, there there exist a sequence $W^k \in \partial\Pi_{\mathfrak{X}_-^{m_2}}(\lambda^* + g(x^*, y^*))$, u^k and v^k such that (W^k, u^k, v^k) satisfies (3.16) and $\|(u^k, v^k)\| \rightarrow \infty$. Let $(\bar{u}^k, \bar{v}^k) = (u^k, v^k)/\|(u^k, v^k)\|$ and without loss of generality assume that $(\bar{u}^k, \bar{v}^k) \rightarrow (\bar{u}, \bar{v})$. Thus we have

$$\begin{aligned} \nabla_x \mathcal{L}(x^*; y^*, \mu^*, \lambda^*)/\|(u^k, v^k)\| + \mathcal{J}H(x^*)^T \bar{u}^k + \mathcal{J}G(x^*)^T \bar{v}^k &= 0, \\ H(x^*) &= 0, \\ 0 \leq \bar{v}^k \perp G(x^*) &\leq 0. \end{aligned} \quad (3.22)$$

Taking the limit of (3.22) as $k \rightarrow \infty$, we obtain

$$\begin{aligned} \mathcal{J}H(x^*)^T \bar{u} + \mathcal{J}G(x^*)^T \bar{v} &= 0, \\ H(x^*) &= 0, \\ 0 \leq \bar{v} \perp G(x^*) &\leq 0. \end{aligned} \quad (3.23)$$

The system (3.23) implies $u = 0$ and $v = 0$ from the Mangasarian-Fromovitz constraint qualification, this contradicts with $\|(\bar{u}, \bar{v})\| = 1$. The proof is complete. \square

4 Some Concluding Remarks

In this paper, we have successfully extended the definition of local minimax point from unconstrained minimax optimization problems. to constrained minimax optimization problems. By analyzing Jacobian uniqueness conditions for the lower-level maximization problem and the strong regularity of Karush-Kuhn-Tucker conditions of the maximization problem, we provided both necessary optimality conditions and sufficient optimality conditions for the local minimax points of constrained minimax optimization problems, see Theorems 3.1-3.3.

As the current study is theoretical, we are looking for more applications of the constrained minimax optimization problems in modern machine learning and also traditional research areas in future. Furthermore, it remains under investigation how to design numerical algorithms which can converge to a local minimax point of constrained minimax optimization problems.

References

- [1] Bertsekas, D. P., Constrained Optimization and Lagrange Multiplier Methods, Academic Press, New York, 1982.
- [2] Bonnans J. F. and Shapiro A., Perturbation Analysis of Optimization Problems, Springer-Verlag, New York, 2000.

- [3] Nocedal J. and Wright S. J., *Numerical Optimization*, Springer Press, 1999.
- [4] Falk J. E. and Liu J., *On bilevel programming, Part I: nonlinear cases*, *Mathematical Programming*, 70(1995), pp. 47-72.
- [5] Jin C., Netrapalli P. and Jordan M. I., *What is local optimality in nonconvex-nonconcave minimax optimization?* arXiv:1902.00618v2 [cs.LG] 3 Jun 2019.
- [6] Dempe S., *A necessary and a sufficient optimality condition for bilevel programming problems*, *Optimization*, 1992, Vol. 25, pp. 341-354.
- [7] Dempe S., *Foundations of Bilevel Programming*, Kluwer, Dordrecht, 2002.
- [8] Dempe S., Dutta J. and Mordukhovich B. S., *New necessary optimality conditions in optimistic bilevel programming*, *Optimization*, 56:5-6(2007), pp. 577-604.
- [9] Dempe S. and Zemkoho A. B., *The bilevel programming problem: reformulations, constraint qualifications and optimality conditions*, *Math. Program., Ser. A*, 138(2013), pp. 447-473.
- [10] Dempe S., Mordukhovich B. S. and Zemkoho A. B., *Necessary optimality conditions in pessimistic bilevel programming*, *Optimization*, 63:4(2014), pp. 505-533.
- [11] Mehlitz P. and Zemkoho A. B., *Sufficient optimality conditions in bilevel programming*, arXiv:1911.01647v1 [math.OC] 5 Nov 2019.
- [12] Qi L., *Convergence analysis of some algorithms for solving nonsmooth equations*, *Mathematics of Operations Research*, 18:1(1993), pp. 227-244.
- [13] Robinson, S. M., *Strongly regular generalized equations*, *Mathematics of Operations Research*. 5(1980), pp. 43-62.
- [14] Ye J. J. and Zhu D. L., *Optimality conditions for bilevel programming problems*, *Optimization*, 33(1995), pp. 9-27. with correction in *Optimization*, 39(1997), pp. 361-366.