arXiv:2303.01474v1 [math.OC] 2 Mar 2023

最大值函数的敏感性分析及其在非凸极小极大规划中的应用

郭磊*、Jane J. Ye†、张进 ‡

2023年3月3日

摘要

本文对最大值函数进行了敏感性分析，该函数是参数化最大化问题的最优值函数。我们的目标是研究最大值函数的各种次微分。通过巧妙运用一些敏感性分析技术，我们获得了最大值函数的Fr´echet次微分、极限次微分和水平次微分的上界估计。所得上界估计仅取决于所有解的并集，而不取决于其凸包，也不取决于解集中唯一的某个解。最后，我们将所得结果应用于为非凸极小极大问题建立一些新的必要最优性条件。在非凸-凹情形下，我们的Wolfe 对偶方法优于一阶方法，因为其必要条件更为严格，且约束资格条件更弱。

关键词：敏感性分析、值函数、Wolfe 对偶、非凸极小极大问题、非利普希茨连续性、必要最优性条件。

MSC2020 分类：90C30、90C31、90C47。

1 引言

参数优化问题的值函数敏感性分析在许多领域都非常有用，例如双层优化、极小极大规划、随机规划、鲁棒优化、动态规划以及经济学中的比较静态分析。

本文考虑以下参数化最大化问题

$$(P_x) \qquad \max_y \ f(x,y) \ \text{ s.t. } \ y \in \mathcal{F}(x) := \{y \in \mathbb{R}^m : g(x,y) \le 0\},$$

*华东理工大学商学院，中国上海 200237。电子邮箱：lguo@ecust.edu.cn。

†维多利亚大学数学与统计学系，加拿大不列颠哥伦比亚省维多利亚市 V8W 2Y2。电子邮箱：janeye@uvic.ca。

‡通讯作者。数学系，南方科技大学国际数学中心，南方科技大学。国家应用数学中心（深圳），鹏城实验室，中国广东深圳。电子邮箱：zhangj9@sustech.edu.cn。

其中 $f : \mathrm{IR}^{n+m} \to \mathrm{IR}$ $g : \mathrm{IR}^{n+m} \to \mathrm{IR}^q$。除非另有说明，我们假设$f(x, y)$ $g(x, y)$是连续可微函数。将该最大化问题的值函数或最大值函数定义为

$$V(x) := \sup\{f(x, y) : y \in \mathcal{F}(x)\},$$

其中，按约定，若 $\mathcal{F}(x) = \emptyset$，则 $V(x) = -\infty$；而最优解映射定义为

$$\mathcal{S}(x) := \{y \in \mathcal{F}(x) : f(x, y) = V(x)\}.$$

最大值函数 $V(x)$ 显然与相应的最小值函数有关，后者定义为

$$v(x) := \inf_y\{-f(x, y) : y \in \mathcal{F}(x)\}$$

按以下方式

$$V(x) = -v(x). \tag{1.1}$$

本文的首要目标是研究最大值函数的各种次微分。即使所有涉及的函数都是光滑的，优化问题的最优值函数本质上仍然是非光滑的。文献中对值函数的广义微分性质进行了深入研究。一类研究关注值函数的方向导数（例如，Bonnans 和 Shapiro [3], Fiacco 和 Ishizuka [7], Gauvin 和 Dubeau [10], Janin [20]）。另一类研究考察值函数的次微分（例如，Clarke [4], Gauvin [9], Guo 等人 [17], Lucet 和 Ye [26, 27], Mordukhovich 等人 [31], Thibault [45], Ye和Zhang [51]）。本文主要关注后一类研究，部分原因是我们将运用所得结果，为极小极大问题建立必要最优性条件。

最小值函数的极限次微分的上估计相对容易获得（例如，Clarke [4], Guo 等人[17], Lucet 和 Ye [26, 27]）。利用关系式 (1.1)，人们希望通过相应最小值函数的上估计，获得最大值函数的极限次微分的上估计。如果最大值函数 $V(x)$ 是局部利普希茨连续的，这一点便可以做到。事实上，众所周知，如果一个函数是局部利普希茨连续的，那么它的 Clarke 次微分是齐次的，即

$$\partial^c V(x) = \partial^c(-v)(x) = -\partial^c v(x),$$

而极限次微分则不是齐次的，即

$$\partial V(x) = \partial(-v)(x) \neq -\partial v(x),$$

其中，∂ 和 ∂^c 分别表示极限次微分算子和Clarke次微分算子。因此，若要通过研究最大值函数 $V(x)$的极限次微分

来研究最小值函数 $v(x)$的极限次微分，可以先确保该值函数是局部利普希茨连续的，然后使用以下上估计

$$\partial V(x) \subseteq \partial^c V(x) = \partial^c(-v)(x) = -\partial^c v(x) = -\operatorname{co} \partial v(x), \tag{1.2}$$

其中，co C 表示集合 C的凸包。一般而言，除非函数 $V(x)$是 Clarke 正则的，否则（1.2）中的包含关系可能是严格的（例如，约束集是一个固定紧集时，如 Danskin 定理中的情形，或者值函数是光滑的情形）。

一般而言，极限次微分 $\partial v(x)$的上估计涉及解集 $\mathcal{S}(x)$中的所有解（例如，参见第 3 节中的（3.2）），因此， $\partial v(x)$中的任意向量都可以仅用 $\mathcal{S}(x)$中的一个解来表示。然而，根据（1.2），极限次微分 $\partial V(x)$的上估计将涉及 $\mathcal{S}(x)$中所有解的并集的凸包，因此， $\partial V(x)$中的任意向量都取决于所有解。因此，在实践中使用极限次微分会很复杂，甚至难以实现。本文旨在研究能否利用解集中的所有解，或仅用一个解，来获得极限次微分 $\partial V(x)$的估计。我们首先考虑 $V(x)$是局部利普希茨连续的情形。此时，（1.2）中的关系成立。我们将给出一些充分条件，以便推导出 $\partial v(x)$的易于计算的凸上界估计。在这种情况下，凸包运算将是多余的，并且可以得到 $\partial V(x)$的易于计算的上估计。其次，我们考虑值函数不一定局部利普希茨连续，但仍希望用解集中的一个解来表示最大函数的极限次微分中的任意向量的情形。假设函数 $-f(x,y)$ $g(x,y)$关于 y是凸的，我们采用 Wolfe 对偶（Wolfe [46]），使最大值函数 $V(x)$等于问题（P_x）的 Wolfe 对偶的最小值函数（见（3.10））。由于最大化问题的 Wolfe 对偶是一个最小化问题，我们无需采用 Clarke次微分来获得 $\partial V(x)$的上估计。因此，我们成功地得到了最大值函数 $V(x)$的极限次微分或水平次微分的上估计，其中使用了解集中的所有解；并且可以仅用解集中的一个解来表示最大值函数的极限次微分或水平次微分中的任意向量。

极小极大问题是数学、经济学和计算机科学中的一个重要研究焦点（例如，冯·诺伊曼 [32], 冯·诺伊曼和摩根斯坦 [33], Sion [43]）。现有的大多数极小极大问题研究成果都聚焦于凸凹情形，并研究如下形式的光滑极小极大问题

$$(P_{\text{minimax}}) \qquad \min_{x\in X} \max_{y\in Y} f(x,y), \tag{1.3}$$

其中，函数 $f(\cdot,y)$ 对每个 $y \in Y$ 都是凸函数，而 $f(x,\cdot)$ 对每个 $x \in X$ 都是凹函数，且 X、Y 是闭凸集。在这种情形下，著名的

凸凹情形极小极大定理（Sion [43,推论 3.3]）指出，如果 X 或 Y 是紧集，那么

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

这一性质意味着，两个参与者（决策为 x 的参与者和决策为 y 的参与者）之间的博弈是同时进行的，因此可以使用纳什均衡来定义最优性，这正是同时行动博弈中广为人知的最优性概念。若且唯若，我们称（$\bar{x}$，$\bar{y}$）$\in X\times Y$ 是极小极大问题（P_{minimax}）在纳什（均衡）意义下的最优解

$$f(\bar{x},y)\le f(\bar{x},\bar{y})\le f(x,\bar{y}),\quad \forall (x,y)\in X\times Y.$$

在凸凹情形下，纳什均衡意义下的一阶必要且充分驻点条件如下

$$0\in \nabla_x f(\bar{x},\bar{y})+\mathcal{N}_X(\bar{x}),\qquad 0\in -\nabla_y f(\bar{x},\bar{y})+\mathcal{N}_Y(\bar{y}), \tag{1.4}$$

其中，$\mathcal{N}_C(c)$ 表示集合 C 在点 c 处的法锥，其含义依据凸分析确定。

然而，实践中的许多极小极大问题并不属于凸凹情形。特别是，近年来出现的许多机器学习任务只能表述为（P_{minimax}），其中 f 在 x 中是非凸的。示例包括生成对抗网络（GAN）（Goodfellow 等人 [14]）、分布鲁棒学习、具有不可分解损失的学习，以及多分布鲁棒学习（参见 Rafique 等人 [37]及其中的参考文献）。现在考虑一般极小极大问题（1.3），其中函数 $f(x,y)$ 是连续可微的，且 X、Y 是闭集。在这一一般情形下，“min”和“max”的顺序无法交换，因此该博弈是序贯行动博弈。对于序贯行动博弈，使用 Stackelberg 均衡的概念更为合理。若 $\bar{y}$ 是内层最大化问题 $\max_{y\in Y} f(\bar{x},y)$ 的全局最优解，且 $\bar{x}$ 是问题的局部最优解，则我们称（$\bar{x}$，$\bar{y}$）且 $\bar{x}\in X$、$\bar{y}\in Y$ 是（P_{minimax}）在 Stackelberg（均衡）意义下的局部最优解

$$(P_V)\qquad \min_{x\in X} V(x),$$

其中 $V(x)=\sup_{y\in Y} f(x,y)$。注意，在 Stackelberg 意义下的最优性概念中，$\bar{y}$ 必须是内层最大化问题的全局解。本文将研究的正是这种最优性。近年来，研究者探讨了内层最大化问题的局部解为 $\bar{y}$ 的极小极大问题的局部最优性概念；例如参见 [6, 21, 24]。

当解 $\bar{x}$ 处满足受限上紧性条件（定义 3.2）时（例如，当 Y 是紧集时），值函数在 $\bar{x}$ 附近 Lipschitz连续，并且

$$\partial V(\bar{x})\subseteq \operatorname{co}\left\{\nabla_x f(\bar{x},y') : y'\in \mathcal{S}(\bar{x})\right\}.$$

因此，如果 $\bar{x}$ 是问题（P_V）的局部最优解，那么我们有

$$0 \in \partial V(\bar{x}) + \mathcal{N}_X(\bar{x}), \tag{1.5}$$

其中， $\mathcal{N}_C(c)$表示集合 C 在点 c处的极限法锥，由此我们得到

$$0 \in \operatorname{co}\left\{\nabla_x f(\bar{x}, y') : y' \in \mathcal{S}(\bar{x})\right\} + \mathcal{N}_X(\bar{x}). \tag{1.6}$$

此外，根据Carathéodory定理，存在

$$\lambda_i \geq 0, y_i \in \mathcal{S}(\bar{x})\ i = 1, \ldots, n+1 \text{ with } \sum_{i=1}^{n+1} \lambda_i = 1,$$

使得

$$0 \in \sum_{i=1}^{n+1} \lambda_i \nabla_x f(\bar{x}, y_i) + \mathcal{N}_X(\bar{x}), \quad 0 \in -\nabla_y f(\bar{x}, y_i) + \mathcal{N}_Y(y_i), \quad i = 1, \ldots, n+1. \tag{1.7}$$

然而，这类必要最优性条件并不十分实用。它可用于验证某个可行解并非局部最优解。但即使这种验证也并不容易，因为它涉及整个解集 $\mathcal{S}(\bar{x})$。

由于极小极大问题是双层规划或半无限规划的特殊情形，因此可以采用一阶方法，即用其一阶最优性条件替代内层最大化问题。对于 $f(x, y)$关于变量 y为凹且 Y 为凸集的情形，根据一阶方法，(P_{minimax})等价于

$$\min_{x \in X, y \in Y} \ f(x, y) \quad \text{s.t. } 0 \in \nabla_y f(x, y) + \mathcal{N}_Y(y).$$

设$(\bar{x}, \bar{y})$是(P_{minimax})在Stackelberg意义下的一个局部最优解。若 $\bar{y}$ 位于集合 Y的内部，则$(\bar{x}, \bar{y})$是如下问题的一个局部最优解

$$\min_{x \in X, y \in \mathbb{R}^m} \ f(x, y) \ \text{ s.t. } \nabla_y f(x, y) = 0.$$

进一步假设 f 二阶连续可微，且等式约束系统 $\nabla_y f(x, y)$在$(\bar{x}, \bar{y})$处满足约束品性条件。 $= 0.$ 于是，根据一阶必要最优性条件，存在一个乘子 $\bar{u}$，使得

$$\begin{aligned} 0 \in \nabla_x f(\bar{x}, \bar{y}) + \nabla^2_{xy} f(\bar{x}, \bar{y})\bar{u} + \mathcal{N}_X(\bar{x}), \\ \nabla^2_{yy} f(\bar{x}, \bar{y})\bar{u} = 0, \ \nabla_y f(\bar{x}, \bar{y}) = 0. \end{aligned} \tag{1.8}$$

比较 Stackelberg 意义下最优性的两个驻点系统 (1.7) 和 (1.8) 可见，一般而言，系统 (1.7) 需要解集 $\mathcal{S}(\bar{x})$中的所有解，而系统 (1.8) 只需一个解，但要求内层最大化问题具有凹性，且涉及二阶导数。与系统 (1.7) 不同，(1.8) 是关于变量 x、 y、 u的一个可追踪变分不等式系统。

本文研究极小极大问题 $\min_{x \in X} \max_{y \in \mathcal{F}(x)} f(x, y)$。对于这一一般极小极大问题，我们得到了以下两个新结果。

(i) 考虑内层最大化问题为凹的非凸-凹情形。在这种情况下，可以采用带均衡约束的数学规划（MPEC）方法。但若使用 Karush-Kuhn-Tucker（KKT）条件替代内层最大化问题，并对原变量和乘子进行最小化，则重构问题的局部最优解并不对应原极小极大问题的局部最优解（参见 Dempe 和 Dutta [5] 关于双层规划的讨论）。此外，相应 MPEC 的约束资格条件通常非常强（例如参见 Gfrerer 和 Ye [12]）。本文采用 Wolfe 对偶方法证明，在合理的约束资格条件下，按 Stackelberg 意义定义的极小极大问题的局部最优解满足强（S-）驻点条件；众所周知，这是数学规划与均衡约束的最强驻点条件。因此，相较于 MPEC 方法，我们的新方法能够在较弱的约束资格条件下导出更精细的最优性条件（参见定理 4.3、4.4 和 4.5）。

(ii) 对于一般的非凸-非凹情形，在解映射 $\mathcal{S}(x)$ 内半连续或值函数 $V(x)$ 凹这两种情况中的任意一种下，我们证明，最优性条件仅取决于解集 $\mathcal{S}(\bar{x})$ 中的一个解（见定理 4.7 和 4.8）。应用于 $\mathcal{F}(x) = Y$ 的情形时，我们得到：在解映射 $\mathcal{S}(x)$ 内半连续，或 $f(x,y)$ 联合凹且 Y 凸这两种情况中的任意一种下，以 Stackelberg 表述的最优性条件与以 Nash 表述的最优性条件一致。也就是说，系统 (1.4) 成立，其中 $\mathcal{N}_X(\bar{x})$ 是集合 X 在 $\bar{x}$ 处的极限法锥。

本文其余部分安排如下。第 2 节介绍本文所需的一些预备知识和初步结果。第 3 节给出最大值函数的 Fr´echet 次微分、极限次微分和水平次微分的上界估计。第 4 节建立非凸极小极大问题的必要最优性条件。我们还将新的最优性条件应用于 GAN 的一个特殊情形，以说明我们方法的优势。第 5 节总结全文。

2 预备知识与预备结果

2.1 基本符号

对于任意 $x \in \mathrm{IR}^n$，$\|x\|$ 表示其在 IR^n中的欧几里得范数。对于 $x \in \mathrm{IR}^n$ 和 $\delta > 0$，我们用 $\mathcal{B}_\delta(x) := \{y : \|y - x\| < \delta\}$ 表示以 x 为中心、半径为 δ 的开球，并用 $\mathbb{B}$ 表示闭单位球。子集 $\Omega \subseteq \mathrm{IR}^n$ 的边界和内部分别记为 bdΩ 和 intΩ。对于向量值函数 $\phi = (\phi_1, \ldots, \phi_m) : \mathrm{IR}^n \to \mathrm{IR}^m$ 及点 $x \in \mathrm{IR}^n$，我们用 $\mathcal{I}_\phi(x) := \{i : \phi_i(x) = 0\}$ 表示在 x 处的活跃指标集，并用 $\nabla\phi(x) \in \mathbb{R}^{n\times m}$ 表示 ϕ 在 x处的转置雅可比矩阵。对于集值映射 $\Phi : \mathrm{IR}^n \rightrightarrows \mathrm{IR}^m$，我们用其图表示为

gph$\Phi := \{(x,y): y \in \Phi(x)\}$。给定集合 $\Omega \subseteq \mathrm{IR}^n$ 和点 $x \in \mathrm{IR}^n$， x到 Ω 的距离记为 dist$(x, \Omega) := \inf\{\|y-x\|: y \in \Omega\}$.。对于两个向量 $a, b \in \mathrm{IR}^n$， a 垂直于 b 的关系记为 $a \perp b$。对于序列 $\{a^k\}_{k=1}^{\infty}$ 和点 a， $\{a^k\}_{k=1}^{\infty}$ 当 $k \to \infty$ 时收敛到 a 的条件记为 $a^k \to a$。

2.2 变分分析

本小节介绍一些变分分析的背景材料；更多细节参见 Clarke[4], Mordukhovich [30], Rockafellar 和 Wets [41] 。

对于函数 $\varphi : \mathrm{IR}^n \to [-\infty, \infty]$ 以及点 $x^* \in \mathrm{IR}^n$，其中 $\varphi(x^*)$ 为有限值，函数 φ 在 x^* 处的正则（或 Fr´echet）次微分定义为

$$\widehat{\partial}\varphi(x^*) := \left\{v : \varphi(x) \geq \varphi(x^*) + v^\top(x-x^*) + o(\|x-x^*\|)\right\},$$

φ 在 x^* 处的极限（或 Mordukhovich）次微分定义为

$$\partial\varphi(x^*) := \left\{v : \exists x^k \to_\varphi x^*, \exists v^k \in \widehat{\partial}\varphi(x^k) \quad \text{s.t. } v^k \to v\right\},$$

φ在 x^* 处的水平（或奇异 Mordukhovich）次微分定义为

$$\partial^\infty\varphi(x^*) := \left\{v : \exists x^k \to_\varphi x^*, v^k \in \widehat{\partial}\varphi(x^k) \text{ and } t_k \downarrow 0, \quad \text{s.t. } t_k v^k \to v\right\},$$

其中， $o(\cdot)$ 表示当 $\alpha \downarrow 0$ 时 $o(\alpha)/\alpha \to 0$，而 $x^k \to_\varphi x^*$ 表示 $x^k \to x^*$ 且 $\varphi(x^k) \to \varphi(x^*)$。众所周知，根据 Rockafellar 和 Wets [41, 定理 9.13]，下半连续函数 φ 在 x 处是 Lipschitz连续的，当且仅当 $\partial^\infty\varphi(x) = \{0\}$。如果 φ 在 x^* 处是 Lipschitz连续的，则其在 x^* 处的 Clarke次微分为

$$\partial^c\varphi(x^*) = \mathrm{co}\,\partial\varphi(x^*).$$

集合 Ω 在 $x^* \in \Omega$ 处的正则法锥是一个闭锥，定义如下:

$$\widehat{\mathcal{N}}_\Omega(x^*) := \{\xi : \xi^T(x-x^*) \leq o(\|x-x^*\|) \text{ for each } x \in \Omega\},$$

Ω 在 $x^* \in \Omega$ 处的极限法锥是一个闭锥，定义如下:

$$\mathcal{N}_\Omega(x^*) := \{\xi : \xi = \lim_{k\to\infty} \xi^k \text{ with } \xi^k \in \widehat{\mathcal{N}}_\Omega(x^k) \text{ and } x^k \to_\Omega x^*\},$$

其中， $x^k \to_\Omega x^*$ 表示 $x^k \to x^*$，且对所有 k，均有 $x^k \in \Omega$。

2.3 约束资格条件

在本小节中，我们引入一种统一的约束资格条件（CQ），用于后续分析。首先，我们回顾以下参数非线性规划的若干约束资格条件及其性质

$$(\mathcal{P}_p) \qquad \begin{array}{ll} \max\limits_x & f(p,x) \\ \text{s.t.} & x \in \mathcal{X}(p) := \{x : g(p,x) \leq 0, h(p,x) = 0\}, \end{array} \tag{2.1}$$

其中， $f: \mathrm{IR}^m \times \mathrm{IR}^n \to \mathrm{IR}$ $g: \mathrm{IR}^m \times \mathrm{IR}^n \to \mathrm{IR}^p$ 和 $h: \mathrm{IR}^m \times \mathrm{IR}^n \to \mathrm{IR}^q$ 均为连续可微函数。

约束资格条件是施加于约束系统上的条件，旨在确保局部极小点满足 Karush-Kuhn-Tucker（KKT）条件。当约束函数均为仿射函数时，最优解满足 KKT 条件无需额外条件；而对于非线性约束系统，则需要满足某种约束资格条件。最弱的约束资格条件是所谓的 Guignard 约束资格条件，但难以验证。可验证的约束资格条件包括众所周知的线性无关约束资格条件（LICQ）、Mangasarian-Fromovitz（MFCQ）和 Slater 条件。文献中还提出了一些较弱且可验证的约束资格条件，例如常秩约束资格条件（CRCQ）、松弛常秩约束资格条件（RCRCQ）、松弛常正线性相关条件（RCPLD）以及拟正规性（Andreani 等人 [1], Bertsekas 和 Ozdaglar [2], Minchenko 和 Stakhovski [29]）。上述所有约束资格条件都强于误差界性质／度量次正则性／平稳性；例如参见 Ye 和 Zhou [52, 定理 7。4]。

由于参数非线性规划 (2.1) 含有参数，因此约束资格条件依赖于该参数。现在我们回顾上述约束资格条件的参数化版本。

定义 2.1（LICQ、MFCQ、CRCQ、RCRCQ、RCPLD、准范数性）
（i）若偏梯度向量族在 $\bar{x} \in \mathcal{X}(\bar{p})$ 处

$$\{\nabla_x g_i(\bar{p}, \bar{x}) : i \in \mathcal{I}_g(\bar{p}, \bar{x})\} \cup \{\nabla_x h_j(\bar{p}, \bar{x}) : j = 1, \ldots, q\}, \tag{2.2}$$

其中 $\mathcal{I}_g(\bar{p}, \bar{x}) := \{i : g_i(\bar{p}, \bar{x}) = 0\}$，线性无关，则称 LICQ 在 $\mathcal{I}_g(\bar{p})$ 处成立。
（ii）若向量族 (2.2) 正线性无关，则称 MFCQ 在 $\bar{x} \in \mathcal{X}(\bar{p})$ 处成立。也就是说，

$$\nabla_x g(\bar{p}, \bar{x})\lambda^g + \nabla_x h(\bar{p}, \bar{x})\lambda^h = 0,\ 0 \le \lambda^g \perp -g(\bar{p}, \bar{x}) \ge 0 \Longrightarrow (\lambda^g, \lambda^h) = 0.$$

(iii) 如果存在 $\delta > 0$，使得对于所有指标集 $\mathcal{I}_1 \subseteq \mathcal{I}_g(\bar{p}, \bar{x})$ 和 $\mathcal{I}_2 \subseteq \{1, \ldots, q\}$，偏梯度向量组

$$\{\nabla_x g_i(p, x), \nabla_x h_j(p, x) : i \in \mathcal{I}_1, j \in \mathcal{I}_2\}$$

在所有 $(p, x) \in \mathcal{B}_\delta(\bar{p}, \bar{x})$ 上具有相同的秩。
(iv) 如果存在 $\delta > 0$，使得对于所有 $\mathcal{I} \subseteq \mathcal{I}_g(\bar{p}, \bar{x})$，偏梯度向量组 $\{\nabla_x g_i(p, x)\ \nabla_x h_j(p, x) : i \in \mathcal{I}, j = 1, \ldots, q\}$ 在所有 $(p, x) \in \mathcal{B}_\delta(\bar{p}, \bar{x})$ 上具有相同的秩，则称 RCRCQ 在 $\bar{x} \in \mathcal{X}(\bar{p})$ 处成立。
(v) 设 $\mathcal{J} \subseteq \{1, \ldots, q\}$ 使得 $\{\nabla_x h_j(\bar{p}, \bar{x}) : j \in \mathcal{J}\}$ 是 span $\{\nabla_x h_j(\bar{p}, \bar{x}) : j = 1, \ldots, q\}$ 的一个基。若存在 $\delta > 0$ 使得，则称 RCPLD（常秩正线性相关性条件）在 $\bar{x} \in \mathcal{X}(\bar{p})$ 处成立

- $\{\nabla_x h_j(p, x) : j = 1, \ldots, q\}$ 对于每个 $(p, x) \in \mathcal{B}_\delta(\bar{p}, \bar{x})$，秩相同；

• 对于每个 $\mathcal{I} \subseteq \mathcal{I}_g(\bar{p}, \bar{x})$，如果 $\{\nabla_x g_i(\bar{p}, \bar{x}):\ i \in \mathcal{I}\} \cup \{\nabla_x h_j(\bar{p}, \bar{x}):\ j \in \mathcal{J}\}$ 正线性相关，那么对于所有 $(p, x) \in \mathcal{B}_\delta(\bar{p}, \bar{x})$，$\{\nabla_x g_i(p, x)\ \nabla_x h_j(p, x):\ i \in \mathcal{I}, j \in \mathcal{J}\}$ 线性相关。

(vi) 如果不存在非零向量 (λ, μ) 使得，我们就称拟正规性在 $\bar{x} \in \mathcal{X}(\bar{p})$ 处成立

- $\nabla_x g(\bar{p}, \bar{x})\lambda + \nabla_x h(\bar{p}, \bar{x})\mu = 0,\ 0 \le \lambda \perp -g(\bar{p}, \bar{x}) \ge 0$;
- 存在一个收敛到 $(\bar{p}, \bar{x})$ 的序列 $\{(p^k, x^k)\}$，使得对所有 k,

$$g_i(p^k, x^k) > 0 \text{ if } \lambda_i > 0,\ \mu_i h_i(p^k, x^k) > 0 \text{ if } \mu_i \neq 0.$$

定义 2.2（Robinson 稳定性）（Gfrerer 和 Mordukhovich *[11,* 定义 1.*1]*）如果存在正标量 κ, δ 使得，我们就称参数可行映射 $\mathcal{X}(p)$ 在 $(\bar{p}, \bar{x}) \in \text{gph}\mathcal{X}$ 处满足 Robinson 稳定性（RS条件）

$$\text{dist}(x, \mathcal{X}(p)) \le \kappa \left[\sum_{i=1}^{p} \max\{g_i(p, x), 0\} + \sum_{i=1}^{q} |h_i(p, x)|\right], \quad \forall (p, x) \in \mathcal{B}_\delta(\bar{p}, \bar{x}).$$

如果在上述不等式中省略参数 p 的所有依赖关系（即 $g(p, x)$ 和 $h(p, x)$ 都与 p 无关），那么 Robinson 稳定性就变为局部误差界性质。因此，RS 意味着：对于所有充分接近 $\bar{p}$ 的参数，误差界性质都以相同的常数 κ 成立。尽管 RS 一词最早由 Gfrerer 和 Mordukhovich [11]，提出，但这一性质最早由 Robinson [38] 研究。文献中还使用了其他术语来称呼这一性质；例如，在 Minchenko 和 Stakhovski [29] 以及其他一些出版物中，该性质被称为 R-正则性。对于解映射 $\mathcal{S}(x)$ 满足 RS 的情形，Ye 和 Zhu [53] 以及 Ye 等人 [54] 分别将该性质称为一致弱尖锐极小值和一致参数误差界。

对于参数非线性规划（2.1），CRCQ、RCRCQ、RCPLD 和拟正规性都弱于 MFCQ，而 MFCQ 又弱于 LICQ。通过将 Andreani等人 [1，在非参数情形下的定理 1] 证明调整到参数情形，可以很容易验证，对于参数规划，RCRCQ 蕴含 RCPLD。此外，RCPLD 和 $\mathcal{X}(p)$ 的内半连续性可以推出 RS（Mehlitz 和 Monchenko [28，定理3.3]）。

此外，某些约束资格条件（例如 LICQ）具有局部持久性，即如果它们在某一点成立，那么在该点附近的所有可行点处也都成立。但有些约束资格条件并不具有局部持久性，例如 Guignard 约束资格条件。对于参数非线性规划 (2.1)，如果某个约束资格条件在某个邻域内任意满足 (x, p) 的 $x \in \mathcal{X}(p)$ 处成立，其中该邻域包含 $(\bar{x}, \bar{p})$，则称该约束资格条件在 $\bar{x} \in \mathcal{X}(\bar{p})$ 处具有局部持久性。

接下来，我们通过扩展 Guo 等人 [16, 定义 3.3] 中的乘子稳定性，并将其与

局部持久性相结合，为参数非线性规划引入一类统一的约束资格条件。为方便表述，对于给定点 $\beta, x \in \mathrm{IR}^n$ 和$p \in \mathrm{IR}^m$，我们定义乘子集

$$\mathcal{M}(\beta,p,x) := \left\{ (\lambda,\mu): \begin{array}{l} \beta - \nabla_x g(p,x)\lambda - \nabla_x h(p,x)\mu = 0 \\ 0 \le \lambda \perp -g(p,x) \ge 0,\ h(p,x) = 0 \end{array} \right\}. \tag{2.3}$$

注意，当 $\beta = \nabla_x f(p,x)$时，乘子集是参数优化问题(2.1)在 $x \in \mathcal{X}(p)$处的所有KKT乘子构成的集合。

定义2.3（稳定参数约束资格条件）如果对参数优化问题(2.1)的约束区域施加的某个条件在 $\bar{x} \in \mathcal{X}(\bar{p})$处局部持久，并且在$\bar{x} \in \mathcal{X}(\bar{p})$处满足以下条件，则称该条件在$\bar{x} \in \mathcal{X}(\bar{p})$处是稳定参数约束资格条件；并且

(i) 如果 $\bar{x} \in \mathcal{X}(\bar{p})$是问题$(\mathcal{P}_{\bar{p}})$的局部极小点，那么 $\mathcal{M}(\nabla_x f(\bar{p},\bar{x})\ \bar{p},\bar{x}) \neq \emptyset$;

(ii) 对于任意给定的向量序列 $\{\beta^k\}_{k=1}^{\infty} \subseteq \mathrm{IR}^n$、 $\{x^k\}_{k=1}^{\infty} \subseteq \mathrm{IR}^n$和 $\{p^k\}_{k=1}^{\infty} \subseteq \mathrm{IR}^m$，若它们满足 $\beta^k \to \bar{\beta}$、 $x^k \to \bar{x}$ ，且当 $k \to \infty$ 时 $p^k \to \bar{p}$ ，并且对所有k都有 $\mathcal{M}(\beta^k,p^k,x^k) \neq \emptyset$ ，则存在一个乘子序列 $\{(\lambda^k,\mu^k) \in \mathcal{M}(\beta^k,p^k,x^k)\}_{k=1}^{\infty}$ ，它在某个子序列上收敛到某个$(\bar{\lambda},\bar{\mu})$，使得$(\bar{\lambda},\bar{\mu}) \in \mathcal{M}(\bar{\beta},\bar{p},\bar{x})$。

不难看出，定义2.1中所定义的 LICQ 和 MFCQ 都是稳定参数化约束资格条件。下述结果表明，定义2.1和定义2.2中的所有条件，按照定义2.3的意义，都是稳定参数化约束资格条件。

命题 2.1 设 $\bar{x} \in \mathcal{X}(\bar{p})$。那么，CRCQ、RCRCQ、RCPLD、拟正规性和 RS条件在 $\bar{x} \in \mathcal{X}(\bar{p})$处都是稳定参数化约束资格条件。

证明。由定义可立即得出 CRCQ、RCRCQ 和 RS条件的局部持久性。RCPLD 和拟正规性的局部持久性，则分别可通过将 Andreani 等人 [1, 定理 4] 以及 Ozdaglar 和 Bertsekas [36, 引理 2] 在非参数情形下的证明改用于参数情形而得出。

由 Andreani 等人 [1, 推论 1] 可知，RCPLD 是一种约束资格条件。因此，更强的 CRCQ 和 RCRCQ 也都是约束资格条件。根据 Ozdaglar 和 Bertsekas [36, 命题 1], , 拟正规性是一种约束资格条件。显然，RS条件蕴含系统 $\{x:\ g(\bar{p},x) \le 0, h(\bar{p},x) = 0\}$ 的局部误差界。因此，问题 (2.1) 在 $p = \bar{p}$ 下满足 Clarke 镇定性。故 RS条件是一种约束资格条件（Clarke [4, 命题 6.4.4]）。

采用 Andreani 等人定理 2 的证明技巧 [1], , 我们可以证明，若满足常秩正线性相关性条件，则定义2.3(ii)成立。因此，更强的常秩约束资格条件和 RCRCQ 也具有这一性质。接下来我们证明拟正规性蕴含定义2.3(ii)。对于给定的点 $\beta, x \in \mathrm{IR}^n$ 和

$p \in \mathrm{IR}^m$，令

$$\mathcal{M}^e(\beta,p,x) := \left\{ (\lambda,\mu): \begin{array}{l} \beta - \nabla_x g(p,x)\lambda - \nabla_x h(p,x)\mu = 0 \\ 0 \le \lambda \perp -g(p,x) \ge 0,\ h(p,x) = 0 \\ \text{there exists a sequence } \{(p^l,x^l)\} \text{ converging to } (p,x) \\ \text{such that for all } l, \\ g_i(p^l,x^l) > 0 \text{ if } \lambda_i > 0,\ \mu_i h_i(p^l,x^l) > 0 \text{ if } \mu_i \ne 0 \end{array} \right\} \quad (2.4)$$

为增强乘子集。当 $\beta = \nabla_x f(p,x)$ 时，增强乘子集是参数优化问题 (2.1) 在 $x \in \mathcal{X}(p)$ 处的所有增强KKT乘子组成的集合。令 $\beta^k \to \bar{\beta}$, $x^k \to \bar{x}$ 和 $p^k \to \bar{p}$ 为 $k \to \infty$，且对所有 k，$\mathcal{M}(\beta^k,p^k,x^k) \ne \emptyset$ 。根据 Giorgi等人 [13, 定理4, 6], 任意 KKT 点都是增强KKT点，这意味着对于每个 $(\lambda^k,\mu^k) \in \mathcal{M}(\beta^k,p^k,x^k)$，都存在 $(\tilde{\lambda}^k,\tilde{\mu}^k) \in \mathcal{M}^e(\beta^k,p^k,x^k) \subseteq \mathcal{M}(\beta^k,p^k,x^k)$。接着，运用 Ye和Zhang [51, 定理 3], 中的证明技巧，可以证明，若在 $\bar{x} \in \mathcal{X}(\bar{p})$ 处满足拟正规性，则序列$\{(\tilde{\lambda}^k,\tilde{\mu}^k)\}$ 有界。取一个子序列，我们$\bar{\ }$ $\bar{\ }$ $\bar{\ }$得到 $(\lambda,\bar{\mu})$，使得 $(\lambda,\bar{\mu}) \in \mathcal{M}(\beta,\bar{p},\bar{x})$。最后，我们证明 RS条件蕴含定义2.3中的条件 (ii)。由 Gfrerer 和 Mordukhovich [11, 定理 3] 可知，若 RS条件在 $(\bar{p},\bar{x})$ 处成立，则在 $(\bar{p},\bar{x})$ 处满足所谓的有界乘子性质，也就是说，存在 $\kappa > 0$ ，使得对所有充分大的 k,

$$\mathcal{M}(\beta^k,p^k,x^k) \cap \kappa\|\beta^k\|\mathbb{B} \ne \emptyset.$$

令 $(\lambda^k,\mu^k) \in \mathcal{M}(\beta^k,p^k,x^k) \cap \kappa\|\beta^k\|\mathbb{B}$。由于 $\beta^k \to \bar{\beta}$ as $k \to \infty$，因此 $\{\beta^k\}$ 有界。于是，$\{(\lambda^k,\mu^k)\}$ 有界，且存在收敛子序列。由于 $(\lambda^k,\mu^k) \in \mathcal{M}(\beta^k,p^k,x^k)$，因此

$$\beta^k - \nabla_x g(p^k,x^k)\lambda^k - \nabla_x h(p^k,x^k)\mu^k = 0,$$
$$h(p^k,x^k) = 0,\ 0 \le \lambda^k \perp -g(p^k,x^k) \ge 0.$$

沿着 (λ^k,μ^k) 收敛到 $(\bar{\lambda},\bar{\mu})$ 的子序列取极限，可得$\bar{\ }$ $\bar{\ }$$(\lambda,\bar{\mu}) \in \mathcal{M}(\beta,\bar{p},\bar{x})$。证明完毕。

■

本文还需要使用以下针对非参数非线性规划的约束资格条件概念。

定义2.4（稳定约束资格条件）令 $\bar{x} \in \mathcal{X} := \{x:\ g(x) \le 0, h(x) = 0\}$。若定义2.3中的所有条件均成立，且省略所有对参数p 的依赖，则称稳定约束资格条件在 $\bar{x}$ 处成立。

值得注意的是，文献中针对非参数非线性规划的大多数约束资格条件，即 LICQ、MFCQ、CRCQ、RCRCQ、RCPLD 和拟正规性，按照定义2.4的含义，都是稳定约束资格条件。

3 最大值函数的次微分

本节研究最大值函数 $V(x)$ 的 Fr´echet、极限和水平子微分的上界估计，其中该函数具有单个最优解，或其值为所有最优解的并集。我们将分析分为两种情况：一种是 $V(x)$ 局部利普希茨连续；另一种是 $V(x)$ 可能不是局部利普希茨连续。对于第一种情况，我们将利用最优解映射的内半连续性，这是一项相对较强的条件；对于第二种情况，则将采用 Wolfe 对偶。

给定标量 $r \in \{0,1\}$，最大化问题（$P_{\bar{x}}$）的广义拉格朗日函数定义为 $\mathcal{L}^r(\bar{x},y,\lambda) := rf(\bar{x},y) - g(\bar{x},y)^\top\lambda$.为简便起见，当$r=1$ 时，我们省略上标 r。我们将问题（$P_{\bar{x}}$）在 $y \in \mathcal{S}(\bar{x})$ 处的广义乘子集记为

$$\Sigma^r(\bar{x},y) := \{\lambda : \nabla_y \mathcal{L}^r(\bar{x},y,\lambda) = 0,\ 0 \le -g(\bar{x},y) \perp \lambda \ge 0\},$$

当 $r=1$时，我们省略上标 r。

3.1 V 局部利普希茨连续的情形

如引言中所述，当最大值函数局部利普希茨连续时，根据（1.2），最大值函数的次微分可估计为

$$\partial V(\bar{x}) \subseteq \partial^c V(\bar{x}) \subseteq -\mathrm{co}\,\partial v(\bar{x}). \tag{3.1}$$

一般而言， $\partial v(\bar{x})$ 的上估计具有以下形式（例如，Lucet 和 Ye [27,定理4.4]）：

$$\partial v(\bar{x}) \ \subseteq \bigcup_{y\in\mathcal{S}(\bar{x})} \{-\nabla_x \mathcal{L}(\bar{x},y,\lambda) : \lambda \in \Sigma(\bar{x},y)\}. \tag{3.2}$$

因此，根据（3.1）—（3.2），只要 V 在 $\bar{x}$ 处是 Lipschitz连续的，就有

$$\partial V(\bar{x}) \subseteq \mathrm{co} \bigcup_{y\in\mathcal{S}(\bar{x})} \{\nabla_x \mathcal{L}(\bar{x},y,\lambda) : \lambda \in \Sigma(\bar{x},y)\}. \tag{3.3}$$

在本小节中，我们研究确保上估计（3.2）能够采取以下更简单形式的充分条件

$$\partial v(\bar{x}) \subseteq \{-\nabla_x \mathcal{L}(\bar{x},\bar{y},\lambda) : \lambda \in \Sigma(\bar{x},\bar{y})\}, \tag{3.4}$$

对于给定的 $\bar{y} \in \mathcal{S}$（$\bar{x}$）。由于函数 $\nabla_x \mathcal{L}$（x， y， λ）关于 λ 是线性的，且乘子集 Σ（$\bar{x}$，$\bar{y}$）具有凸性，（3.4）中的上估计是凸的，因此（3.3）中的凸包运算是多余的。

本小节的第一个结果依赖于内半连续性，它是单值映射连续性的集合值推广。

定义 3.1（内半连续性）（Mordukhovich *[30, 定义 1.63]*）给定 $\bar{y} \in \mathcal{S}(\bar{x})$，若对任意序列 $x^k \to \bar{x}$，都存在一个收敛到 $\bar{y}$ 的序列 $y^k \in \mathcal{S}(x^k)$，则称解映射 $\mathcal{S}(x)$ 在 $(\bar{x}, \bar{y})$ 处内半连续。

根据该定义，很容易看出，若解映射 $\mathcal{S}(x)$ 在 $\bar{x}$ 处下半连续，则对于每个 $\bar{y} \in \mathcal{S}(\bar{x})$，它在 $(\bar{x}, \bar{y})$ 处内半连续。此外，根据 Mehlitz 和 Minchenko [28, 引理2]，$\mathcal{S}(x)$ 在 $(\bar{x}, \bar{y})$ 处的内半连续性可由 $(\bar{x}, \bar{y})$ 附近存在一致弱尖锐极小值来保证（定义和充分条件见 [48, 53]）。

现在我们证明，在内半连续性成立的条件下，可以利用任意给定的解 $\bar{y} \in \mathcal{S}(\bar{x})$ 得到 $\partial V(\bar{x})$ 的上估计。

定理 3.1 设 $\bar{y} \in \mathcal{S}(\bar{x})$。假设解映射 $\mathcal{S}(x)$ 在 $(\bar{x}, \bar{y})$ 处内半连续。再假设系统 $g(x, y) \leq 0$ 在 $(\bar{x}, \bar{y})$ 处满足局部误差界。那么 $V(x) = -v(x)$ 在 $\bar{x}$ 处连续，且

$$\partial v(\bar{x}) \subseteq \{-\nabla_x \mathcal{L}(\bar{x}, \bar{y}, \lambda) : \lambda \in \Sigma(\bar{x}, \bar{y})\}, \tag{3.5}$$

$$\partial^\infty v(\bar{x}) \subseteq \{\nabla_x g(\bar{x}, \bar{y})\lambda : \lambda \in \Sigma^0(\bar{x}, \bar{y})\}. \tag{3.6}$$

进一步假设

$$\{\nabla_x g(\bar{x}, \bar{y})\lambda : \lambda \in \Sigma^0(\bar{x}, \bar{y})\} = \{0\}. \tag{3.7}$$

那么，V在 $\bar{x}$ 处是 Lipschitz连续的，并且 $\partial V(\bar{x}) \subseteq \partial^c V(\bar{x}) \subseteq \{\nabla_x \mathcal{L}(\bar{x}, \bar{y}, \lambda) : \lambda \in \Sigma(\bar{x}, \bar{y})\}$。

证明。在 $(\bar{x}, \bar{y})$ 处的内半连续性条件下，不难验证最小值函数 $v(x)$ 在 $\bar{x}$ 处连续。令 $\theta(x, y) := -f(x, y) + I_{\mathrm{gph}\mathcal{F}}(x, y)$，其中 $I_{\mathrm{gph}\mathcal{F}}$ 是示性函数。那么 $v(x) = \inf_y \theta(x, y)$。由于 $\mathcal{S}(x)$ 在 $(\bar{x}, \bar{y})$ 处内半连续，根据 Mordukhovich 定理 1.108] 以及极限次微分的和法则，我们有

$$\begin{aligned}
\partial v(\bar{x}) &\subseteq \{x^* : (x^*, 0) \in -\nabla f(\bar{x}, \bar{y}) + \mathcal{N}_{\mathrm{gph}\mathcal{F}}(\bar{x}, \bar{y})\}, \\
\partial^\infty v(\bar{x}) &\subseteq \{x^* : (x^*, 0) \in \mathcal{N}_{\mathrm{gph}\mathcal{F}}(\bar{x}, \bar{y})\}.
\end{aligned}$$

因此，在系统 $g(x, y) \leq 0$ 在 $(\bar{x}, \bar{y})$ 处满足局部误差界条件的情况下，根据 $\mathcal{N}_{\mathrm{gph}\mathcal{F}}(\bar{x}, \bar{y})$ 的显式表达式（例如，Henrion 等人 [18, 定理4.1] 或 Gfrerer 和 Ye[12, 命题 4]），所需的包含关系 (3.5)-(3.6) 立即成立。

由于 (3.6) 和 (3.7) 蕴含 $\partial^\infty v(\bar{x}) = \{0\}$，因此在条件 (3.7) 下，$v(x)$（以及 $V(x)$）在 $\bar{x}$ 处是 Lipschitz连续的（Rockafellar 和 Wets [41, 定理 9.13]）。因此

$$\partial V(\bar{x}) \subseteq \partial^c V(\bar{x}) \subseteq -\mathrm{co}\, \partial v(\bar{x}).$$

结合条件 (3.5)，即可得出所需结果。 ■

回顾一下，对于系统 g，当且仅当满足以下条件时，MFCQ 在 $(\bar{x},\bar{y})$ 处成立：$x\ y\ \leq 0$ 在 $(\bar{x},\bar{y})$ 处

$$\nabla g(\bar{x},\bar{y})\lambda = 0, \quad 0 \leq -g(\bar{x},\bar{y}) \perp \lambda \geq 0 \Longrightarrow \lambda = 0, \tag{3.8}$$

当且仅当系统 $g(\bar{x},y)\ \leq 0$ 在 $\bar{y}$ 处满足 MFCQ

$$\nabla_y g(\bar{x},\bar{y})\lambda = 0, \quad 0 \leq -g(\bar{x},\bar{y}) \perp \lambda \geq 0 \Longrightarrow \lambda = 0. \tag{3.9}$$

蕴含关系 (3.9) 显然强于关系 (3.8)。因此，在条件 (3.9) 下，$g(x,y)\ \leq 0$ 在 $(\bar{x},\bar{y})$ 处满足局部误差界，而且 (3.7) 成立。以下推论可直接得出。

推论3.1 设 $\bar{y} \in \mathcal{S}(\bar{x})$。假设 $\mathcal{S}(x)$ 在 $(\bar{x},\bar{y})$ 处内半连续，且系统 $g(\bar{x},y)\ \leq 0$ 在 $\bar{y}$ 处满足 MFCQ。那么，$V(x)$ 在 $\bar{x}$ 处 Lipschitz连续，且 $\partial V(x) \subseteq \{\nabla_x \mathcal{L}(\bar{x},\bar{y},\lambda) : \lambda \in \Sigma(\bar{x},\bar{y})\}$。

根据 Robinson [39, 定理3，2], 当 MFCQ 和二阶充分条件成立时，最优解映射是内半连续的。因此，以下结果可立即得出。

推论3.2 令 $\bar{y} \in \mathcal{S}(\bar{x})$。假设对于任意靠近 $\bar{x}$ 的 x，$-f(x,\cdot)$ $g(x,\cdot)$ 都是凸函数。设系统 $g(\bar{x},y)\ \leq 0$ 在 $\bar{y}$ 处满足 MFCQ，并且对于任意 $\lambda \in \Sigma(\bar{x},\bar{y})$，二阶充分条件在 $(\bar{x},\lambda)$ 处成立：

$$d^\top \nabla^2_{xx}\mathcal{L}(\bar{x},\bar{y},\lambda)d < 0, \quad \forall d \in \mathcal{C}(\bar{x},\bar{y})\backslash\{0\},$$

其中

$$\mathcal{C}(\bar{x},\bar{y}) = \left\{ d : \begin{array}{l} \nabla_y g_i(\bar{x},\bar{y})^\top d = 0\ i \in \mathcal{I}_g(\bar{x},\bar{y}) \text{ and } \lambda_i > 0 \\ \nabla_y g_i(\bar{x},\bar{y})^\top d \leq 0\ i \in \mathcal{I}_g(\bar{x},\bar{y}) \text{ and } \lambda_i = 0 \end{array} \right\}.$$

Th且 $V(x)$ 在 $\bar{x}$ 处是 Lipschitz连续的，并且 $\partial V(x) \subseteq \{\nabla_x \mathcal{L}(\bar{x},\bar{y},\lambda) : \lambda \in \Sigma(\bar{x},\bar{y})\}$.

证明。根据所涉及函数的凸性假设可知，全局最优解与局部最优解相同。然后根据 Robinson [39,定理3.2], ，解映射 $\mathcal{S}$ 是连续的，因此在 $(\bar{x},\bar{y})$ 处内半连续。由此即可得到所需结果，该结果来自推论3.1,

■

现在我们确定另一种情形：$\partial V(\bar{x})$ 的上估计仅涉及一个任意指定的解 $\bar{y}$。

定理 3.2 假设 $\bar{y} \in \mathcal{S}(\bar{x})$，且 $f(x,y)$ 联合凹，$g(x,y)$ 关于 (x,y) 联合拟凸。设存在一个开集 $\mathcal{O} \ni \bar{x}$，使得 $\mathcal{F}(x)$ 非空，且目标函数 $f(x,y)$ 在 $\mathcal{F}(x)$ 上有上界。假设系统 $g(x,y)\ \leq 0$ 在 $(\bar{x},\bar{y})$ 处满足约束资格条件。则 $V(x)$ 是凹函数，在 $\bar{x}$ 处 Lipschitz连续，且

$$\partial V(\bar{x}) \subseteq \{\nabla_x \mathcal{L}(\bar{x},\bar{y},\lambda) : \lambda \in \Sigma(\bar{x},\bar{y})\}.$$

证明。由于 $-f(x,y)$ 联合凸，且 $g(x,y)$ 关于 (x,y) 联合拟凸，因此 $v(x)$ 是凸函数（Fiacco and Kyparisis [8, 命题 2.1]）。于是，根据 Ye and Wu [50,命题 4.1]，我们有

$$\partial v(\bar{x}) = \{-\nabla_x \mathcal{L}(\bar{x}, \bar{y}, \lambda) : \lambda \in \Sigma(\bar{x}, \bar{y})\}.$$

在这些假设下，不难验证，$V(x)$ 在 $\bar{x}$ 附近是有限的。这一点结合 $v(x)$ 的凸性，意味着 $v(x)$ 在 $\bar{x}$ 处必为 Lipschitz连续（参见 Rockafellar[40, 定理 10.4]）。于是，根据（3.1）以及拉格朗日乘子集的凸性，我们有

$$\partial V(\bar{x}) \subseteq -\operatorname{co} \partial v(\bar{x}) = \{\nabla_x \mathcal{L}(\bar{x}, \bar{y}, \lambda) : \lambda \in \Sigma(\bar{x}, \bar{y})\}.$$

■

3.2 V 可能不是局部 Lipschitz连续的情形

为了获得最小值函数次微分的易于计算的凸上界估计，在前一小节中，除了隐含假设最大值函数是局部 Lipschitz连续的以外，我们还假设了一些相对较强的条件，即最优解映射是内半连续的，或者该问题是联合凸问题。在本小节中，我们不采用这些条件，而是直接研究最大值函数的次微分。

下面的例子说明，第 3.1 小节中的假设不成立，但本小节所需的所有条件均成立。

例3.1 考虑以下简单例子:

$$\max_y \ y(x+1) \quad \text{s.t.} \ -5 \le x + y \le 0.$$

通过简单计算可得，最优解 $\mathcal{S}(x)$ 为

$$\mathcal{S}(x) = \begin{cases} \{-x\} & \text{if } x > -1, \\ [-4, 1] & \text{if } x = -1, \\ \{-x-5\} & \text{if } x < -1. \end{cases}$$

容易验证，对于任何满足 $\bar{x} = -1, \bar{y} \in (-4,1)$ 的 $(\bar{x}, \bar{y})$，$\mathcal{S}(x)$ 都不是内半连续的，而且该问题不是联合凸的。

为给出本小节的主要结果，我们假设 $-f(x,y)$，$g_i(x,y)$ $i = 1, \ldots, q$ 在 (x,y) 上二阶连续可微，并且对参考点 $\bar{x}$ 附近的所有 x，在 y 上均为凸函数。现在，我们引入 Wolfe [46] 提出的问题 $(P_{\bar{x}})$ 的对偶问题，如下所示:

$$(D_{\bar{x}}) \qquad \begin{aligned} \min_{y,\lambda} \quad & \mathcal{L}(\bar{x}, y, \lambda) \\ \text{s.t.} \quad & \nabla_y \mathcal{L}(\bar{x}, y, \lambda) = 0, \ \lambda \ge 0. \end{aligned}$$

我们用 $\Omega(\bar{x})$ 表示可行域，并且

$$V_D(\bar{x}) := \inf_{y,\lambda}\{\mathcal{L}(\bar{x}, y, \lambda) : (y, \lambda) \in \Omega(\bar{x})\} \tag{3.10}$$

问题（$D_{\bar{x}}$）的最优值函数。

引理 3.1（Wolfe *[46, 定理 1]*）假设 $f(x, y)$ 关于变量 y 是凹的，并且对任意给定的 x，$g_i(x, y)$ $i = 1$，……，q 关于变量 y 是凸的。则弱对偶性成立，即有 $V(x) \leq V_D(x)$。

对于给定的标量 $r \in \{0, 1\}$ 以及 $(y, \lambda) \in \Omega(\bar{x})$，我们将问题（$D_{\bar{x}}$）在 (y, λ) 处的广义乘子集合记为：

$$\Xi^r(\bar{x}, y, \lambda) := \left\{u : \nabla^2_{yy}\mathcal{L}(\bar{x}, y, \lambda)u = 0,\ 0 \leq -\nabla_y g(\bar{x}, y)^\top u - rg(\bar{x}, y) \perp \lambda \geq 0\right\}.$$

我们对集合 $\Xi^r(\bar{x}, y, \lambda)$ 作一些说明。首先注意到，我们总有 $0 \in \Xi^r(\bar{x}, y, \lambda)$。对于问题（$D_{\bar{x}}$），若 MFCQ 在 $(y, \lambda) \in \Omega(\bar{x})$ 处成立，则 $\Xi^0(\bar{x}, y, \lambda) = \{0\}$.。此外，$\Xi^0(\bar{x}, y, \lambda) \subseteq \Xi^1(\bar{x}, y, \lambda)$，因为 $g(\bar{x}, y) \leq 0$。

利用弱对偶性，我们可以将最大化问题的值函数作为最小化问题的值函数来研究。因此，在下述定理中，无需对所有最优解取凸包，即可得到最大值函数次微分的上界估计。

对于最小化问题，Clarke [4, 假设 6.5.1] 中首次引入、并在 Guo 等人 [17, 定义 3.8], 中命名的所谓受限下确界紧性条件，用于确保值函数的下半连续性。对于最大化问题，我们提出如下受限上紧性条件。

定义 3.2（受限上紧性）我们称受限上紧性条件在 $\bar{x}$ 处成立，如果 $V(\bar{x})$ 有限，且存在正标量 ϵ、δ 以及一个紧集 Ω，使得对于所有满足 $V(x) > V(\bar{x}) - \epsilon$ 的 $x \in \mathcal{B}_\delta(\bar{x})$，总有 $\mathcal{S}(x) \cap \Omega \neq \emptyset$。

容易看出，受限上紧性条件弱于上确界紧性条件，即存在 $\delta > 0$、$\alpha < V(\bar{x})$ 和一个紧集 Ω，使得

$$\{y \in \mathcal{F}(x) : f(x, y) \geq \alpha, x \in \mathcal{B}_\delta(\bar{x}) \subseteq \Omega.$$

此外，容易看出，上确界紧性条件弱于可行解映射 $\mathcal{F}$ 的一致紧性，即对于 $\bar{x}$ 的某个邻域 V，$\bigcup_{x\in V}\mathcal{F}(x)$ 的闭包是紧的。

定理 3.3 设 $\bar{y} \in \mathcal{S}(\bar{x})$ 且 $\lambda \in \Sigma(\bar{x}, \bar{y})$。假设在 $(\bar{x}, \bar{y}, \lambda)$ 处，对于 (x, y, λ) 中如下约束系统，某个约束资格条件成立：

$$\nabla_y \mathcal{L}(x, y, \lambda) = 0,\ \lambda \geq 0. \tag{3.11}$$

则有

$$\widehat{\partial}V(\bar{x}) \subseteq \left\{\nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) + \nabla^2_{xy}\mathcal{L}(\bar{x},\bar{y},\bar{\lambda})u : u \in \Xi^1(\bar{x},\bar{y},\bar{\lambda})\right\}. \tag{3.12}$$

假设在 $\bar{x}$ 处满足受限上紧性，对于 $(P_{\bar{x}})^{\tilde{}}$ 在所有 $y \in \mathcal{S}(\bar{x})$ 处均满足稳定参数约束资格条件，并且对于约束系统 (3.11)，在所有 $(\bar{x},\tilde{y},\lambda)$ 处均满足稳定约束资格条件，其中$\tilde{y} \in \mathcal{S}(\bar{x})$ 且 $\lambda \in \Sigma(\bar{x},\tilde{y})$。那么我们有

$$\partial V(\bar{x}) \subseteq \bigcup_{y\in\mathcal{S}(\bar{x})} \left\{\nabla_x \mathcal{L}(\bar{x},y,\lambda) + \nabla^2_{xy}\mathcal{L}(\bar{x},y,\lambda)u : u \in \Xi^1(\bar{x},y,\lambda), \lambda \in \Sigma(\bar{x},y)\right\}, \tag{3.13}$$

$$\partial^\infty V(\bar{x}) \subseteq \bigcup_{y\in\mathcal{S}(\bar{x})} \left\{\nabla^2_{xy}\mathcal{L}(\bar{x},y,\lambda)u : u \in \Xi^0(\bar{x},y,\lambda), \lambda \in \Sigma(\bar{x},y)\right\}. \tag{3.14}$$

$\widehat{}$证明。(i) 设 $\xi \in \partial V(\bar{x})$。根据正则次微分的定义，对于任意 $\epsilon > 0$，存在 $\delta_\epsilon > 0$ 使得

$$V(x) - V(\bar{x}) - \xi^\top(x-\bar{x}) \geq -\epsilon\|x-\bar{x}\|, \quad \forall x \in \mathcal{B}_{\delta_\epsilon}(\bar{x}).$$

根据引理 3.1 中的弱对偶性，我们有

$$V_D(x) - \xi^\top(x-\bar{x}) + \epsilon\|x-\bar{x}\| \geq f(\bar{x},\bar{y}), \quad \forall x \in \mathcal{B}_{\delta_\epsilon}(\bar{x}).$$

因此，根据 $V_D(x)$ 的定义可得

$$\mathcal{L}(x,y,\lambda) - \xi^\top(x-\bar{x}) + \epsilon\|x-\bar{x}\| \geq f(\bar{x},\bar{y}), \quad \forall x \in \mathcal{B}_{\delta_\epsilon}(\bar{x}),\ \forall (y,\lambda) \in \Omega(x). \tag{3.15}$$

$^-$ $^-$ $^-$由于根据 $\lambda \in \Sigma(\bar{x},\bar{y})$，$(\bar{y},\lambda) \in \Omega(\bar{x})$，关系式 (3.15) 表明 $(\bar{x},\bar{y},\lambda)$ 是问题的一个局部极小点

$$\begin{aligned} \min_{x,y,\lambda} \quad & \mathcal{L}(x,y,\lambda) - \xi^\top(x-\bar{x}) + \epsilon\|x-\bar{x}\| \\ \text{s.t.} \quad & \nabla_y \mathcal{L}(x,y,\lambda) = 0,\ \lambda \geq 0. \end{aligned} \tag{3.16}$$

$^-$ $^-$由于在（$\bar{x},\bar{y},\lambda$）处满足约束资格条件，因此（$\bar{x},\bar{y},\lambda$）是问题 (3.16) 的 KKT 点，即存在一个乘子 u，使得

$$\begin{aligned} &\xi \in \nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) + \epsilon\mathbb{B} + \nabla^2_{xy}\mathcal{L}(\bar{x},\bar{y},\bar{\lambda})u, \\ &\nabla^2_{yy}\mathcal{L}(\bar{x},\bar{y},\bar{\lambda})u = 0,\ 0 \leq -\nabla_y g(\bar{x},\bar{y})^\top u - g(\bar{x},\bar{y}) \perp \bar{\lambda} \geq 0. \end{aligned}$$

公式 (3.12) 由 $\Xi^1(\bar{x},\bar{y},\bar{\lambda})$ 的定义以及 ϵ 的任意性得出。

(ii) 设 $\xi \in \partial V(\bar{x})$。根据极限次微分的定义，存在序列 $x^k \to \bar{x}$ with $V(x^k) \to V(\bar{x})$ 和 $\xi^k \in \widehat{\partial}V(x^k)$，使得 $\xi^k \to \xi$。由于 $V(x^k) \to V(\bar{x})$，根据受限上紧性假设，存在有界序列 $\{y^k \in \mathcal{S}(x^k)\}$。不失一般性，假设 $y^k \to \tilde{y}$ as $k \to \infty$。因此，根据 $f(x,y)$ 的连续性，可得

$$V(\bar{x}) = \lim_{k\to\infty} V(x^k) = \lim_{k\to\infty} f(x^k,y^k) = f(\bar{x},\tilde{y}).$$

因此，我们有 $\tilde{y} \in \mathcal{S}(\bar{x})$。

由于问题（P_{x^k}）在 k 充分大时于 $y^k \in \mathcal{S}(x^k)$ 处满足稳定参数化约束资格条件，而问题（$P_{\bar{x}}$）在 $\tilde{y} \in \mathcal{S}(\bar{x})$ 处满足稳定参数化约束资格条件，因此，对所有 k 充分大的情形，都存在 KKT 乘子 λ^k，使得

$$\nabla_y f(x^k, y^k) - \nabla_y g(x^k, y^k)\lambda^k = 0,\ 0 \le -g(x^k, y^k) \perp \lambda^k \ge 0.$$

这意味着

$$\lambda^k \in \mathcal{M}(\nabla_y f(x^k, y^k), x^k, y^k),$$

其中，$\mathcal{M}(\cdot)$ 是如（2.3）中定义的乘子集。根据稳定参数化约束资格条件的定义，我们可以找到一个子序列，使得 $\{\lambda^k\}$ 收敛到某个 $\tilde{\lambda}$，且满足 $\lambda^k \in \Sigma(x^k, y^k)$ 和 $\tilde{\lambda} \in \Sigma(\bar{x}, \tilde{y})$。不失一般性，我们假设整个序列 $\{(y^k, \lambda^k)\}\tilde{}$ 在 $k \to \infty$ 时收敛到（$\tilde{y}, \lambda$）。

由于约束系统（3.11）在（$\bar{x}, \tilde{y}, \tilde{\lambda}$）处满足稳定约束资格条件，且当 $k \to \infty$ 时（x^k, y^k, λ^k）$\to\tilde{}$（$\bar{x}, \tilde{y}, \lambda$），根据局部持久性可知，对所有 k 充分大的情形，该约束系统（3.11）在（x^k, y^k, λ^k）处满足此约束资格条件。因此，由结果（i）可知，将（$\bar{x}, \bar{y}, \bar{\lambda}$）替换为（$x^k, y^k, \lambda^k$）后，（3.12）仍成立。于是，根据（3.12），关系 $\xi^k \in \widehat{\partial} V(x^k)$ 意味着存在 u^k，使得

(a) $\xi^k = \nabla_x \mathcal{L}(x^k, y^k, \lambda^k) + \nabla^2_{xy} \mathcal{L}(x^k, y^k, \lambda^k) u^k$,

(b) $\nabla^2_{yy} \mathcal{L}(x^k, y^k, \lambda^k) u^k = 0$,

（c） $0 \le -\nabla_y g(x^k, y^k)^\top u^k - g(x^k, y^k) \perp \lambda^k \ge 0$。

由函数 $\nabla_x \mathcal{L}(x, y, \lambda)$ 和 $g(x, y)$ 的连续性可知，当 $k \to \infty$ 时，

$$\xi^k - \nabla_x \mathcal{L}(x^k, y^k, \lambda^k) \to \xi - \nabla_x \mathcal{L}(\bar{x}, \tilde{y}, \tilde{\lambda}), \tag{3.17}$$

$$g(x^k, y^k) \to g(\bar{x}, \bar{y}). \tag{3.18}$$

回顾一下，约束系统 (3.11) 在 $(\bar{x}, \tilde{y}, \lambda)$ 处满足稳定约束资格条件。于是，由稳定约束资格条件的条件 (ii)，根据关系式 (a)-(c) 和 (3.17)-(3.18)，可知存在 u，使得

- $\xi = \nabla_x \mathcal{L}(\bar{x}, \tilde{y}, \tilde{\lambda}) + \nabla^2_{xy} \mathcal{L}(\bar{x}, \tilde{y}, \tilde{\lambda}) u$,
- $\nabla^2_{yy} \mathcal{L}(\bar{x}, \tilde{y}, \tilde{\lambda}) u = 0,\ 0 \le -\nabla_y g(\bar{x}, \tilde{y})^\top u - g(\bar{x}, \tilde{y}) \perp \tilde{\lambda} \ge 0.$

于是，由上述结果以及$\Xi^1(\bar{x}, \tilde{y}, \tilde{\lambda})$的定义，立即可得公式 (3.13)。

(iii) 令 $\xi \in \partial^\infty V(\bar{x})$。根据地平线次微分的定义，存在$x^k \to \bar{x}$，使得 $V(x^k) \to V(\bar{x})$，$t_k \to 0^+$，且 $\xi^k \in \widehat{\partial} V(x^k)$ 满足 $t_k \xi^k \to \xi$。与结果 (ii) 的证明类似，由关系 $\xi^k \in \widehat{\partial} V(x^k)$ 可知，对所有充分大的 k，存在 u_k，使得结果 (ii) 证明中的 (a)–(c) 成立。将 (a)–(c) 乘以 t_k，得到

(a1) $t_k\xi^k = t_k\nabla_x\mathcal{L}(x^k, y^k, \lambda^k) + \nabla^2_{xy}\mathcal{L}(x^k, y^k, \lambda^k)(t_k u^k)$,

(b1) $\nabla^2_{yy}\mathcal{L}(x^k, y^k, \lambda^k)(t_k u^k) = 0$, (c1) $0 \le -\nabla_y g(x^k, y^k)^\top(t_k u^k) - g(x^k, y^k)t_k \perp \lambda^k \ge 0$.

由函数 $\nabla_x\mathcal{L}(x, y, \lambda)$ 和 $g(x, y)$ 的连续性，以及 $t_k\xi^k \to \xi$ 和 $t_k \to 0^+$ 这一事实，可知当 $k \to \infty$ 时，

$$t_k\xi^k - t_k\nabla_x\mathcal{L}(x^k, y^k, \lambda^k) \to \xi,\ g(x^k, y^k)t_k \to 0. \tag{3.19}$$

˜注意到，在 $(\bar{x}, \tilde{y}, \lambda)$ 处，约束系统 (3.11) 满足稳定约束资格条件。于是，根据稳定约束资格条件的条件 (ii)，利用关系式 (a1)-(c1) 和 (3.19)，可知存在 u ，使得

- $\xi = \nabla^2_{xy}\mathcal{L}(\bar{x}, \tilde{y}, \tilde{\lambda})u$,
- $\nabla^2_{yy}\mathcal{L}(\bar{x}, \tilde{y}, \tilde{\lambda})u = 0,\ 0 \le -\nabla_y g(\bar{x}, \tilde{y})^\top u \perp \tilde{\lambda} \ge 0.$

由上述结果及 $\Xi^0(\bar{x}, \tilde{y}, \tilde{\lambda})$ 的定义，立即可得公式 (3.14)。

■

根据 Clarke [4, 定理 6.5 的推论1，$\bar{x}$ 处满足受限上紧性，且所有 $y \in \mathcal{S}(\bar{x})$ 处均满足 MFCQ，因此值函数 $V(x)$ 在 $\bar{x}$ 处是 Lipschitz 连续。以下推论给出了值函数满足利普希茨连续性的另一充分条件。

推论 3.3 设在 $\bar{x}$ 处满足受限上紧性，并且对于问题 $(P_{\bar{x}})$，在所有 $y \in \mathcal{S}(\bar{x})$ 处均满足稳定参数化约束资格条件；对于约束系统 (3.11)，在所有 $(\bar{x}, \tilde{y}, \lambda)$ 处均满足稳定约束资格条件，其中 $\tilde{y} \in \mathcal{S}(\bar{x})$ 且 $\lambda \in \Sigma(\bar{x}, \bar{y})$。此外，假设值函数 $V(x)$ 在 $\bar{x}$ 处连续。若

$$\bigcup_{y\in\mathcal{S}(\bar{x})} \left\{\nabla^2_{xy}\mathcal{L}(\bar{x}, y, \lambda)u : u \in \Xi^0(\bar{x}, y, \lambda), \lambda \in \Sigma(\bar{x}, y)\right\} = \{0\}, \tag{3.20}$$

则 $V(x)$ 在 $\bar{x}$ 处是 Lipschitz连续的，且上估计 (3.13) 成立。在上述假设之外，若

$$\bigcup_{y\in\mathcal{S}(\bar{x})} \left\{\nabla_x\mathcal{L}(\bar{x}, y, \lambda) + \nabla^2_{xy}\mathcal{L}(\bar{x}, y, \lambda)u : u \in \Xi^1(\bar{x}, y, \lambda), \lambda \in \Sigma(\bar{x}, y)\right\} = \{\zeta\},$$

则 $V(x)$ 严格可微，且 $\nabla V(\bar{x}) = \zeta$。

证明。由于 $V(x)$ 在 $\bar{x}$ 处连续，因此有 $0 \in \partial^\infty V(\bar{x})$（Rockafellar 和 Wets [41,定理 8.9]）。由定理 3.3 和条件 (3.20) 可得 $\partial^\infty V(\bar{x}) \subseteq \{0\}$。因此，

$\partial^\infty V(\bar{x}) = \{0\}$，因此 $V(x)$ 在 $\bar{x}$ 处是 Lipschitz连续的（Rockafellar 和 Wets[41, 定理 9.13]）。证明的其余部分可由定理 3.3 立即推出。 ■

推论 3.3 中需要值函数 $V(x)$ 的连续性。下面的命题给出了保证值函数连续性的一个非常弱的充分条件。事实上，该条件甚至比 $\bar{x}$ 处的 RS条件更弱；而后者又比某个解 $\bar{y} \in \mathcal{F}(\bar{x})$ 处的 MFCQ 更弱。Gauvin 和 Dubeau 在 [10, 定理 3.3 和 5.1] 中证明，后一条件与可行映射 $\mathcal{F}(x)$ 在 $\bar{x}$ 附近的一致紧性相结合，足以保证连续性。

命题 3.1 考虑参数优化问题 (P_x)。假设在 $\bar{x}$ 处满足受限上紧性，并且存在 $\bar{y} \in S(\bar{x})$，使得 $\lim_{x\to\bar{x}} d(\bar{y}, \mathcal{F}(x)) = 0$。那么 $V(x)$ 在 $\bar{x}$ 处连续。

证明。在 $\bar{x}$ 处满足受限上紧性的条件下，容易验证 $V(x)$ 在 $\bar{x}$ 处上半连续；类似论证见 Guo 等人[17, 定理3.9]的证明。现在我们证明 $V(x)$ 在 $\bar{x}$ 处的下半连续性。设 $\{x_k\}$ 是一个序列，使得 $x_k \to \bar{x}$ 且 $V(x_k) \to \liminf_{x\to\bar{x}} V(x)$。那么 $\lim_{k\to\infty} \mathrm{dist}(\bar{y}, \mathcal{F}(x_k)) = 0$。对每个 k，取 $y_k \in \mathcal{F}(x_k)$ 使得 $\|y_k - \bar{y}_k\| < \mathrm{dist}(\bar{y}, \mathcal{F}(x_k)) + 1/k$。由此可得 $(x_k;\ y_k) \to (\bar{x};\ \bar{y})$，因此

$$V(\bar{x}) = f(\bar{x};\bar{y}) = \lim_{k\to\infty} f(x_k;y_k) \le \lim_{k\to\infty} V(x_k) = \liminf_{x\to\bar{x}} V(x)$$

这表明 V 在 $\bar{x}$ 处下半连续。证明完毕。 ■

在定理 3.3 中，上估计取决于具有 $r = \{0, 1\}$ 的集合 $\Xi^r(\bar{x}, y, \lambda)$。现在我们探讨使 $\Xi^0(\bar{x}, y, \lambda) = \Xi^1(\bar{x}, y, \lambda) = \{0\}$ 成立的充分条件。设 $y \in \mathcal{S}(\bar{x})$ 且 $\lambda \in \Sigma(\bar{x}, y)$。回顾一下，问题 $(P_{\bar{x}})$ 的临界锥 $\mathcal{C}(\bar{x}, y)$ 可写为

$$\left\{ d:\ \begin{array}{l} \nabla_y g_i(\bar{x},y)^\top d = 0,\ \text{if } i \in \mathcal{I}_g(\bar{x},y), \lambda_i > 0 \\ \nabla_y g_i(\bar{x},y)^\top d \le 0,\ \text{if } i \in \mathcal{I}_g(\bar{x},y), \lambda_i = 0 \end{array} \right\},$$

并且 t 若二阶充分条件（二阶充分条件）在问题 $(P_{\bar{x}})$ 的 (y, λ) 处成立，则

$$\langle d, \nabla^2_{yy}\mathcal{L}(\bar{x}, y, \lambda)d\rangle < 0, \quad \forall d \in \mathcal{C}(\bar{x}, y)\backslash\{0\}.$$

命题 3.2 设 $y \in \mathcal{S}(\bar{x})$ 且 $\lambda \in \Sigma(\bar{x}, y)$。若二阶充分条件在问题 $(P_{\bar{x}})$ 的 (y, λ) 处成立，则 $\Xi^0(\bar{x}, y, \lambda) = \Xi^1(\bar{x}, y, \lambda) = \{0\}$。

证明。反之，假设 $0 \ne u \in \Xi^1(\bar{x}, y, \lambda)$。那么我们有

$$\nabla^2_{yy}\mathcal{L}(\bar{x}, y, \lambda)u = 0, \tag{3.21}$$

$$0 \le -\nabla_y g(\bar{x}, y)^\top u - g(\bar{x}, y) \perp \lambda \ge 0. \tag{3.22}$$

由 (3.22) 可知，对于 $i = 1,\dots, m$,

$$\lambda_i \ge 0, -\nabla_y g_i(\bar{x}, y)^\top u + g_i(\bar{x}, y) \ge 0, \lambda_i(\nabla_y g_i(\bar{x}, y)^\top u - g_i(\bar{x}, y)) = 0.$$

若 $i \in \mathcal{I}_g(\bar{x}, y)$ 且 $\lambda_i > 0$，则容易得到 $\nabla_y g_i(\bar{x}, y)^\top u = 0$。若 $i \in \mathcal{I}_g(\bar{x}, y)$ 且 $\lambda_i = 0$，则容易得到 $\nabla_y g_i(\bar{x}, y)^\top u \leq 0$。因此，$u \in \mathcal{C}(\bar{x}, y)$。于是，SOSC 成立意味着

$$\langle u, \nabla^2_{yy} \mathcal{L}(\bar{x}, y, \lambda) u \rangle < 0,$$

这与 (3.21) 矛盾。因此，$\Xi^1(\bar{x}, y, \lambda) = \{0\}$。由于 $\Xi^0(\bar{x}, y, \lambda) \subseteq \Xi^1(\bar{x}, y, \lambda)$，容易得到 $\Xi^0(\bar{x}, y, \lambda) = \{0\}$。 ■

现在我们说明，若 $f(x, y)$ 和 $g(x, y)$ 关于 x 和 y 可分离，或者二阶充分条件成立，则上界估计 (3.13)-(3.14) 会简单得多。特别地，无需涉及二阶导数。

推论 3.4 设 $\bar{y} \in \mathcal{S}(\bar{x})$ 和 $\lambda \in \Sigma(\bar{x}, \bar{y})$。假设 $f(x, y)$ 和 $g(x, y)$ 关于 x 和 y 可分离，且约束系统 (3.11) 在 $(\bar{x}, \bar{y}, \lambda)$ 处满足约束资格条件，或者问题 $(P_{\bar{x}})$ 在 $(\bar{y}, \lambda)$ 处满足二阶充分条件。则有

$$\widehat{\partial} V(\bar{x}) \subseteq \left\{ \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\lambda}) \right\}.$$

证明。若 $f(x, y)$ 和 $g(x, y)$ 关于 x 和 y 是可分离的，则对所有 u，有 $\nabla^2_{xy} \mathcal{L}(\bar{x}, y, \lambda) u = 0$。若问题 $(P_{\bar{x}})$ 在 $(\bar{y}, \bar{\lambda})$ 处满足二阶充分条件，则由命题 3.2，$\Xi^0(\bar{x}, \bar{y}, \bar{\lambda}) = \Xi^1(\bar{x}, \bar{y}, \bar{\lambda}) = \{0\}$。。这意味着问题 $(D_{\bar{x}})$ 在 $(\bar{y}, \bar{\lambda}) \in \Omega(\bar{x})$ 处满足 MFCQ，因而系统 (3.11) 在 $(\bar{x}, \tilde{y}, \lambda)$ 处满足 MFCQ。所需的包含关系立即由定理 3.3 的 (3.12) 式得出。

■

推论 3.5 假设对于所有 $y \in \mathcal{S}(\bar{x})$，$(P_{\bar{x}})$ 满足受限上紧性，且 $(P_{\bar{x}})$ 满足稳定参数约束资格条件。此外，进一步假设以下条件之一成立：$f(x, y)$ 和 $g(x, y)$ 关于 x 和 y 可分离，且约束系统 (3.11) 在所有 $(\bar{x}, \tilde{y}, \lambda)$ 处满足稳定约束资格条件；或者 $(P_{\bar{x}})$ 在所有 $(\tilde{y}, \lambda)$ 处满足二阶充分条件，其中 $\tilde{y} \in \mathcal{S}(\bar{x})$ 且 $\lambda \in \Sigma(\bar{x}, \tilde{y})$。那么

$$\partial V(\bar{x}) \subseteq \bigcup_{y \in \mathcal{S}(\bar{x})} \left\{ \nabla_x \mathcal{L}(\bar{x}, y, \lambda) : \lambda \in \Sigma(\bar{x}, y) \right\}, \ \partial^\infty V(\bar{x}) \subseteq \{0\}.$$

此外，若 $\mathcal{F}(x)$ 在某个 $\bar{y} \in \mathcal{S}(\bar{x})$ 处满足 RS条件，则 $V(x)$ 在 $\bar{x}$ 处是 Lipschitz连续的。

证明。第一部分的结果可通过与推论 3.4 的证明中类似的论证，由 (3.13)-(3.14) 立即得出。假设在某个 $\bar{y} \in \mathcal{S}(\bar{x})$ 处，$\mathcal{F}(x)$ 满足 RS条件。根据命题 3.1，值函数 $V(x)$ 在 $\bar{x}$ 处连续。因此，$0 \in \partial^\infty V(\bar{x})$（Rockafellar 和 Wets [41, 定理 8.9]）。于是可得 $\partial^\infty V(\bar{x}) = \{0\}$，从而 $V(x)$ 在 $\bar{x}$ 处 Lipschitz连续（Rockafellar 和 Wets [41, 定理 9.13]）。

■

值得注意的是，我们上述结果可能适用于某些最优解处 MFCQ 不成立的情形，因此 Gauvin 和 Dubeau [10, 定理 5.3] 可能不适用。

此外，我们得到了极限次微分的上界估计，并可由此得到 Clarke 次微分的上界估计。但 Gauvin 和 Dubeau [10, 定理 5.3]仅给出了 Clarke 次微分的上界估计，无法据此得到极限次微分的上界估计。

在本节的最后一部分，我们讨论扰动为规范形式时，最大值函数的次微分与乘子集之间的关系。考虑以下非参数最大化问题

$$\max_{y} \quad c(y) \quad \text{s.t.} \quad d(y) \leq 0, \tag{3.23}$$

其中 $c: \mathbb{R}^m \to \mathbb{R}, d: \mathbb{R}^m \to \mathbb{R}^p$ 是二阶连续可微函数。我们假设最优解存在。在经济学中，最大化问题 (3.23) 可用于描述预算约束下的生产最大化问题；例如参见 I ntriligator [19, 第 149] 页。令 $U(x) := \max\{c(y): d(y) - x \leq 0\}$ 为最大值函数。在经济理论中，众所周知，如果最大值函数连续可微且乘子唯一，那么最大值函数的梯度就等于该乘子。由于最大值函数的梯度表示约束右端项变化时最大值的变化率，因此乘子可解释为影子价格。以下推论将这一结果推广到更一般的情形，即最大值函数可能并不光滑，且乘子集可能不是单元素集。

推论3.6 考虑问题 (3.23)。设 $\bar{y}$ 是问题 (3.23) 的一个最优解，且$\Sigma(\bar{y})$ 是在 $\bar{y}$处相应的乘子集。那么，最大值函数 $U(x)$ 在 $\bar{x} = 0$ 处Lipschitz连续，并且

$$\partial U(0) \subseteq \Sigma(\bar{y})$$

满足以下假设之一:

1) 在 $\bar{x} = 0$ 处具有受限上紧性，且问题 (3.23) 的所有解处均满足稳定约束资格条件（例如，LICQ、MFCQ、CRCQ、RCRCQ、RCPLD，或准-˜正规性成立），并且系统的所有 ($\tilde{y}$, λ) 处均满足稳定约束资格条件

$$\nabla c(y) - \nabla d(y)\lambda = 0, \ \lambda \geq 0,$$

˜其中 $\tilde{y}$ 是一个最优解，且 $\lambda \in \Sigma(\tilde{y})$。

2) 函数 $c(y)$ 是强凹的，且 $d(y)$ 是凸的。问题 (3.23) 在唯一最优解 $\bar{y}$处满足稳定约束资格条件。

3) 函数 $c(y)$ 是强凹的，而 $d(y)$ 是线性的。

证明。1) 由于 $f(x,y) := c(y)$ 和 $g(x,y) := d(y) - x$ 在变量 x 和 y 上是可分离的，因此结论由推论 3.5 得出。

2) 由于 $c(y)$ 是强凹函数，且 $d(y)$ 是凸函数，因此受限上紧性成立，二阶充分条件也成立。此外，$\bar{y}$ 是唯一解。由推论 3.5 可得该结果。

3) 由于 $d(y)$ 是线性函数，结果 2) 所要求的稳定约束资格条件自动成立。因此，所需结果可由结果 2) 得出。 ■

4 极小极大问题的必要最优性条件

本节中，我们针对以下形式的非凸极小极大问题建立最优性条件：

$$(P_{\text{minimax}}) \qquad \min_{x\in X} \max_{y\in\mathcal{F}(x)} f(x,y), \quad \mathcal{F}(x) := \{y : g(x,y) \le 0\}, \tag{4.1}$$

其中 $f : \mathrm{IR}^{n+m} \to \mathrm{IR}$ $g : \mathrm{IR}^{n+m} \to \mathrm{IR}^q$，且 X 是 IR^n 的闭子集。除非另有说明，我们假设 $f(x,y)$ $g(x,y)$ 是连续可微函数。

4.1 基于 MPEC 方法的非凸-凹极小极大问题最优性条件

本小节回顾求解非凸-凹极小极大问题的 MPEC 方法（例如，参见 Stein [44]）。假设 $f(x,y)$ $g(x,y)$ 二阶连续可微，且对于所有 $x \in X$，$-f(x,y)$ $g(x,y)$ 关于变量 y 是凸的。设 $X := \{x : h(x) \le 0\}$，其中函数 $h : \mathrm{IR}^n \to \mathrm{IR}^r$ 连续可微。将内层最大化问题替换为其 KKT 条件，可得到以下 MPEC重构：

$$\begin{aligned} (\mathrm{P}_{\mathrm{MPEC}}) \qquad \min_{x,y,\lambda} \quad & f(x,y) \\ \text{s.t.} \quad & x \in X,\ \nabla_y \mathcal{L}(x,y,\lambda) = 0, \\ & 0 \le -g(x,y) \perp \lambda \ge 0. \end{aligned}$$

我们注意到，($\mathrm{P}_{\mathrm{MPEC}}$) 的目标函数与变量 λ 无关。因此，在非常宽松的条件下，容易推出 ($\mathrm{P}_{\mathrm{minimax}}$) 与 ($\mathrm{P}_{\mathrm{MPEC}}$) 在全局极小解意义下等价（Dempe 和 Dutta [5]），因为 ($\mathrm{P}_{\mathrm{MPEC}}$) 的任一全局极小解都对应于 ($\mathrm{P}_{\mathrm{minimax}}$) 的全局极小解。然而，由于引入了额外乘子作为隐式变量，($\mathrm{P}_{\mathrm{MPEC}}$) 的局部极小解未必对应于原极小极大问题的局部极小解。运用 Dempe 和 Dutta [5]，中定理 2.1 和 3.2 的证明技巧，我们可以得到以下结果。

定理 4.1（等价定理）设 $(\bar{x},\bar{y})$ 是 $(\mathrm{P}_{\mathrm{minimax}})^-$ $^-$的局部最优解，且约束资格条件在 $\bar{y} \in \mathcal{F}(\bar{x})$ 处成立。那么对于所有 $\lambda \in \Sigma(\bar{x},\bar{y})$，点 $(\bar{x},\bar{y},\lambda)$ 是 ($\mathrm{P}_{\mathrm{MPEC}}$) 的局部最优解。另一方面，设稳定参数约束资格条件在 $(\bar{x},\bar{y})$ 处成立，并且

¯ ¯$(\bar{x},\bar{y},\lambda)$ 是 $(\mathrm{P}_{\mathrm{MPEC}})$ 的局部最优解，对所有 $\lambda\in\Sigma(\bar{x},\bar{y})$ 均成立。那么 $(\bar{x},\bar{y})$ 是 $(\mathrm{P}_{\mathrm{minimax}})$ 的局部最优解。

利用文献中针对数学规划与均衡约束的驻点概念（例如 [42, 49]），可以推导出 $(\mathrm{P}_{\mathrm{MPEC}})$ 的驻点条件。对于 $(\mathrm{P}_{\mathrm{MPEC}})$ 的任意可行点 $(\bar{x},\bar{y},\lambda)$，定义以下指标集。

$$\begin{aligned}
\mathcal{I}_{0+} &= \mathcal{I}_{0+}(\bar{x},\bar{y},\bar{\lambda}) := \{i : -g_i(\bar{x},\bar{y}) = 0, \bar{\lambda}_i > 0\},\\
\mathcal{I}_{+0} &= \mathcal{I}_{+0}(\bar{x},\bar{y},\bar{\lambda}) := \{i : -g_i(\bar{x},\bar{y}) > 0, \bar{\lambda}_i = 0\},\\
\mathcal{I}_{00} &= \mathcal{I}_{00}(\bar{x},\bar{y},\bar{\lambda}) := \{i : -g_i(\bar{x},\bar{y}) = 0, \bar{\lambda}_i = 0\}.
\end{aligned}$$

¯定义 4.1（驻点条件）若存在乘子 $(\bar{u},\bar{\alpha},\beta)\in \mathrm{IR}^m\times\mathrm{IR}^q\times\mathrm{IR}^q$ 使得

$$\begin{aligned}
&0\in\nabla_x f(\bar{x},\bar{y})+\nabla_x g(\bar{x},\bar{y})\bar{\alpha}+\nabla^2_{xy}\mathcal{L}(\bar{x},\bar{y},\bar{\lambda})\bar{u}+\mathcal{N}_X(\bar{x}),\\
&\nabla_y f(\bar{x},\bar{y})+\nabla_y g(\bar{x},\bar{y})\bar{\alpha}+\nabla^2_{yy}\mathcal{L}(\bar{x},\bar{y},\bar{\lambda})\bar{u}=0,\\
&-\nabla_y g(\bar{x},\bar{y})^\top\bar{u}-\bar{\beta}=0,\\
&\nabla_y\mathcal{L}(\bar{x},\bar{y},\bar{\lambda})=0,\ 0\le -g(\bar{x},\bar{y})\perp\bar{\lambda}\ge 0,\\
&\bar{\alpha}_i=0\quad i\in\mathcal{I}_{+0},\ \bar{\beta}_i=0\quad i\in\mathcal{I}_{0+}.
\end{aligned}\tag{4.2}$$

¯若 (4.2) 成立且

$$\bar{\alpha}_i\bar{\beta}_i\ge 0\ \forall i\in\mathcal{I}_{00}.$$

¯若 (4.2) 成立且

$$\text{either } (\bar{\alpha}_i,\bar{\beta}_i)>0 \text{ or } \bar{\alpha}_i\bar{\beta}_i=0\ \forall i\in\mathcal{I}_{00}.$$

¯若满足（4.2），则称（$\bar{x},\bar{y},\lambda$）为 S-驻点，且

$$(\bar{\alpha}_i,\bar{\beta}_i)\ge 0\ \forall i\in\mathcal{I}_{00}.\tag{4.3}$$

容易看出，S-驻性是定义 4.1 中最强的驻点条件。由于在 MPEC-LICQ 条件下，$(\mathrm{P}_{\mathrm{MPEC}})$ 的局部极小解可以是 S- 驻点（ Scheel 和 Scholtes [42, 定理 2]），由定理 4.1 可立即得到以下结果。

¯定理 4.2 设（$\bar{x},\bar{y}$）是 $(\mathrm{P}_{\mathrm{minimax}})$ 的局部最优解，且 $\lambda\in\Sigma(\bar{x},\bar{y})$。若¯ ¯在（$\bar{x},\bar{y},\lambda$）处满足 MPEC-LICQ，即对于以下系统，在（$\bar{x},\bar{y},\lambda$）处满足 LICQ:

$$h(x)\le 0,\ \nabla_y\mathcal{L}(x,y,\lambda)=0, g_i(x,y)=0\ i\in\mathcal{I}_{0+}\cup\mathcal{I}_{00}, \lambda_i=0\ i\in\mathcal{I}_{+0}\cup\mathcal{I}_{00},\tag{4.4}$$

¯那么，点 $(\bar{x},\bar{y},\lambda)$ 是 $(\mathrm{P}_{\mathrm{MPEC}})$ 的 S-驻点。

我们指出，MPEC-LICQ 是一个非常严格的条件。若乘子集 $\Sigma(\bar{x},\bar{y})$ 不是单点集，且约束函数 $g(x,y) = g(y)$ 与 x 无关，则该条件必然不成立；更多讨论见 Gfrerer 和 Ye [12]。事实上，当 $\Sigma(\bar{x},\bar{y})$ 不是单点集时，梯度族 $\{\nabla g_i(\bar{y}) : i \in \mathcal{I}_g(\bar{y})\}$ 线性相关，导致系统 (4.4) 中各函数的梯度族线性相关。因此，系统 (4.4) 不满足 LICQ。若我们希望找到较弱的 M-驻点或 C-驻点，则可以将 MPEC-LICQ 弱化为较弱的数学规划问题专用约束规范（例如，Ye [49]）。我们省略对这些较弱驻点条件的详细讨论，因为我们将在第4.2小节中，将最强的 S-驻点条件与通过最大值函数的次微分推导出的条件进行比较。

4.2 Wolf对偶方法下非凸-凹极小极大问题的最优性条件

在本小节中，我们应用第3.2小节的结果，推导非凸-凹极小极大问题的最优性条件。这里所说的非凸-凹极小极大问题是指，在问题 (4.1) 中，对于所有 $x \in X$， $f(x,y)$ 是凹函数，且 $g(x,y)$ 关于变量 y 是凸函数。

当 $\bar{x}$ 是问题 (P_V) 的局部最优解，且为集合 X 的内点时，由费马法则可知 $0 \in \partial V(\bar{x})$。因此，最大值函数的利普希茨连续性并非必要条件，且在相当弱的假设下，以下最优性条件成立。

定理 4.3 假设对于所有 $x \in X$， $f(x,y)$ 是凹函数，且 $g(x,y)$ 关于变量 y 是凸函数，并且 $f(x,y)$ $g(x,y)$ 二阶连续可微。设 $\bar{x} \in \mathrm{int} X$ 是问题 (P_V) 的局部极小值，且 $\bar{y} \in \mathcal{S}(\bar{x})$。假设问题 $(P_{\bar{x}})$ 在 $\bar{y}$ 处满足 KKT条件，其乘子为 λ，并且约束系统的 CQ 在 $(\bar{x},\bar{y},\lambda)$ 处成立：

$$\nabla_y \mathcal{L}(x,y,\lambda) = 0,\ \lambda \geq 0. \tag{4.5}$$

于是，存在一个向量 $\bar{u}$，使得下列系统成立：

$$\begin{aligned} &\nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) + \nabla^2_{xy}\mathcal{L}(\bar{x},\bar{y},\bar{\lambda})\bar{u} = 0, \\ &\nabla^2_{yy}\mathcal{L}(\bar{x},\bar{y},\bar{\lambda})\bar{u} = 0,\ \nabla_y \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) = 0, \\ &0 \leq -\nabla_y g(\bar{x},\bar{y})^\top \bar{u} - g(\bar{x},\bar{y}) \perp \bar{\lambda} \geq 0,\ 0 \leq -g(\bar{x},\bar{y}) \perp \bar{\lambda} \geq 0. \end{aligned} \tag{4.6}$$

证明。根据费马法则以及 $\bar{x}$ 的最优性，可得 $0 \in \partial V(\bar{x})$，因为 $\bar{x}$ 是 X 的一个内点。于是立即由定理 3.3 得到结论。 ■

当 $\bar{x}$ 是集合 X 的边界点时，我们确实需要最大值函数在 $\bar{x}$ 处具有利普希茨连续性。在这种情况下，在一些能够保证值函数具有局部利普希茨连续性的额外假设下，最优性条件成立。

定理 4.4 假设对于所有 $x \in X$， $f(x,y)$ 是凹的，且 $g(x,y)$ 关于变量 y 是凸的，并且 $f(x,y)$ $g(x,y)$ 是二阶连续可微的。令 $\bar{x} \in \mathrm{bdy} X$ 为问题（P_V）的一个局部解。假设对于内层问题，在 $\bar{x}$ 处满足受限 sup-紧性，并且以下条件之一成立：

(a) 对于约束系统 (4.5)，MFCQ 在所有 $y \in \mathcal{S}(\bar{x})$ 处均成立；当 $\tilde{y} \in \mathcal{S}(\bar{x})$ 且 $\lambda \in \Sigma(\bar{x},\tilde{y})$ 时，稳定约束资格条件在所有 $(\bar{x},\tilde{y},\lambda)$ 处均成立；

(b) RS条件在某个 $y \in \mathcal{S}(\bar{x})$ 处成立，稳定参数约束资格条件在所有 $y \in \mathcal{S}(\bar{x})$ 处成立；当 $\tilde{y} \in \mathcal{S}(\bar{x})$ 且 $\lambda \in \Sigma(\bar{x},\tilde{y})$ 时，稳定约束资格条件在所有 $(\bar{x},\tilde{y},\lambda)$ 处成立。假设

$$\bigcup_{y\in\mathcal{S}(\bar{x}),\lambda\in\Sigma(\bar{x},y)} \left\{ \nabla^2_{xy}\mathcal{L}(\bar{x},y,\lambda)u : \begin{array}{l} \nabla^2_{yy}\mathcal{L}(\bar{x},y,\lambda)u = 0, \\ 0 \le -\nabla_y g(\bar{x},y)^\top u \perp \lambda \ge 0 \end{array} \right\} = \{0\}. \tag{4.7}$$

那么，存在一点 $\bar{y} \in \mathcal{S}(\bar{x})$、 $\lambda \in \Sigma(\bar{x},\bar{y})$ 以及一个乘子 $\bar{u}$ ，使得以下系统成立。

$$\begin{aligned} &0 \in \nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) + \nabla^2_{xy}\mathcal{L}(\bar{x},\bar{y},\bar{\lambda})\bar{u} + \mathcal{N}_X(\bar{x}), \\ &\nabla_y \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) = 0,\ \nabla^2_{yy}\mathcal{L}(\bar{x},\bar{y},\bar{\lambda})\bar{u} = 0, \\ &0 \le -\nabla_y g(\bar{x},\bar{y})^\top \bar{u} - g(\bar{x},\bar{y}) \perp \bar{\lambda} \ge 0,\ 0 \le -g(\bar{x},\bar{y}) \perp \bar{\lambda} \ge 0. \end{aligned} \tag{4.8}$$

证明。由于 $\bar{x} \in \arg\min_{x\in X} V(x)$，根据费马法则， $0 \in \partial(V + I_X)(\bar{x})$，其中 $I_C(c)$ 表示集合 C 在 c 处的示性函数。

(a) 根据 Clarke [4, 定理 6.5 的推论1，2] 在受限上紧性和 MFCQ 条件下， $V(x) = -v(x)$ 在 $\bar{x}$ 处是 Lipschitz连续的。因此，根据 Rockafellar 和 Wets[41, 推论 10.10]， $0 \in \partial V(\bar{x}) + \mathcal{N}_X(\bar{x})$。注意到 MFCQ 是一种稳定参数约束资格条件，所需结论由定理 3.3 立即得出。

(b) 根据推论 3.3， $V(x)$ 在 $\bar{x}$ 处是 Lipschitz连续的，进而根据 Rockafellar 和 Wets[41, 推论 10.10]， $0 \in \partial V(\bar{x}) + \mathcal{N}_X(\bar{x})$。因此，所需结论立即由定理 3.3 得出。

■

我们对通过 MPEC 方法（定理 4.2）和 Wolfe 对偶方法（定理 4.3 和 4.4）推导出的驻点条件作一些说明。首先，给出如下蕴含关系。

定理 4.5 设 $\bar{x} \in X$。假设存在 $\bar{y} \in \mathcal{S}(\bar{x})$、 $\lambda \in \Sigma(\bar{x},\bar{y})$ 和乘子 $\bar{u}$，使得（4.8）成立。那么，$(\bar{x},\bar{y},\lambda)$ 是 ($\mathrm{P}_{\mathrm{MPEC}}$) 的 S-驻点。

证明。假设条件 (4.8) 成立。令 $\bar{\alpha} = -\bar{\lambda}$ 和 $\bar{\beta} = -\nabla_y g(\bar{x},\bar{y})^\top \bar{u}$。根据这些定义和 (4.8)，(4.2) 的第三行和第四行成立。我们还注意到

$$\begin{aligned} &\nabla_x f(\bar{x},\bar{y}) + \nabla_x g(\bar{x},\bar{y})\bar{\alpha} = \nabla_x f(\bar{x},\bar{y}) - \nabla_x g(\bar{x},\bar{y})\bar{\lambda} = \nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}), \\ &\nabla_y f(\bar{x},\bar{y}) + \nabla_y g(\bar{x},\bar{y})\bar{\alpha} = \nabla_y f(\bar{x},\bar{y}) - \nabla_y g(\bar{x},\bar{y})\bar{\lambda} = \nabla_y \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) = 0. \end{aligned}$$

因此，由这两个条件和条件 (4.8) 可知，(4.2) 的第一行和第二¯行成立。根据互补条件 $0 \leq -g(\bar{x},\bar{y}) \perp \lambda \geq 0$，可知 $\bar{\alpha}_i = -\lambda_i = 0$ 当且仅当 $g_i(\bar{x},\bar{y}) < 0$。根据条件 (4.8) 最后一行中的两个互补条件，可知 $-\nabla_y g_i(\bar{x},\bar{y})^\top \bar{u} - g_i(\bar{x},\bar{y}) = 0$ 并且 $g_i(\bar{x},\bar{y}) = 0$ 当且仅当 $\bar{\lambda}_i < 0$,$-\nabla_y g_i(\bar{x},\bar{y})^\top \bar{u} \geq 0$ 当且仅当 $g_i(\bar{x},\bar{y}) = \bar{\lambda}_i = 0$。由此可知，$\bar{\beta}_i = -\nabla_y g_i(\bar{x},\bar{y})^\top \bar{u} = 0$ 当且仅当 $\bar{\lambda}_i < 0$，且 $\bar{\beta}_i = -\nabla_y g_i(\bar{x},\bar{y})^\top \bar{u} \geq 0$ 当且仅当 $g_i(\bar{x},\bar{y}) = \bar{\lambda}_i = 0$。因此，条件 (4.3) 和条件 (4.2) 的最后一行成立。

■

定理 4.5 表明，利用 Wolfe 对偶方法得到的驻点条件比利用 MPEC 方法得到的驻点条件更强。现在我们比较所需的充分条件。正如定理 4.2 后所述，定理 4.2 所要求的 MPEC-LICQ 是一个非常严格的条件，而定理 4.3 和 4.4 所要求的条件则是非线性规划问题中常用的条件。

根据推论 3.4 和 3.5，我们得到以下结果。

定理 4.6 设对于所有$x \in X$， $f(x,y)$ 关于变量 y 是凹函数，且 $g(x,y)$ 关于变量 y 是凸函数； $f(x,y)$ $g(x,y)$ 二阶连续可微。设 $\bar{x}$ 是问题 (P_V) 的一个局部解。

¯(i) 假设 $\bar{x} \in \text{int}X$, $\bar{y} \in \mathcal{S}(\bar{x})$ 和 $\lambda \in \Sigma(\bar{x},\bar{y})$。设 $f(x,y)$ 与_$g(x,y)$ 关于 x 和 y 可分离，并且约束¯ ¯系统 (4.5) 在 $(\bar{x},\bar{y},\lambda)$ 处满足约束资格条件，或者问题 $(P_{\bar{x}})$ 在 $(\bar{y},\lambda)$ 处满足二阶充分条件。则存在 λ 使得以下系统成立：

$$0 = \nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}),\ \nabla_y \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) = 0,\ 0 \leq -g(\bar{x},\bar{y}) \perp \bar{\lambda} \geq 0.$$

(ii) 假设 $\bar{x} \in \text{bdy}X$。对于内层问题，假设在 $\bar{x}$ 处满足受限上确界紧性，在某个 $y \in \mathcal{S}(\bar{x})$ 处满足 RS条件，并且在所有 $y \in \mathcal{S}(\bar{x})$ 处均满足稳定参数约束资格条件。若 $f(x,y)$ 和 $g(x,y)$ 关于 x 和 y 和 ˜ 可分离，且在约束系统 (4.5) 的所有 $(\bar{x},\tilde{y},\lambda)$ 处满足稳定约束资格条件，其中 $\tilde{y} \in \mathcal{S}(\bar{x})$ 且 ˜ ˜ ˜ $\lambda \in \Sigma(\bar{x},\tilde{y})$；或者在所有 $(\tilde{y},\lambda)$ 处满足二阶充分条件，其中 $\tilde{y} \in \mathcal{S}(\bar{x})$ 且 $\lambda \in \Sigma(\bar{x},\tilde{y})$，那么存在 ¯ $\bar{y} \in \mathcal{S}(\bar{x})$ 和 λ，使得

$$0 \in \nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) + \mathcal{N}_X(\bar{x}),\ \nabla_y \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) = 0,\ 0 \leq -g(\bar{x},\bar{y}) \perp \bar{\lambda} \geq 0. \tag{4.9}$$

^证明。结论 (i) 由推论 3.4 以及以下事实立即得出： $0 \in \partial V(\bar{x})$。假设 $\bar{x} \in \text{bdy}X$。由 $\bar{x}$ 的最优性可知， $0 \in \partial(V + I_X)(\bar{x})$。接着由推论 3.5 可得， V 在 $\bar{x}$ 处是 Lipschitz 连续的。因此， $0 \in \partial V(\bar{x}) + \mathcal{N}_X(\bar{x})$。结论 (ii) 由推论 3.5 得出。

■

4.3 非凸非凹极小极大问题的最优性条件

本节讨论一般的非凸—非凹情形。设 $(\bar{x},\bar{y})$ 是 Stackelberg 意义下一般非凸—非凹极小极大问题的一个局部解。若值函数 $V(x)$ 是 Lipschitz连续的，则

$$0 \in \operatorname{co} \bigcup_{y\in\mathcal{S}(\bar{x})} \left\{\nabla_x \mathcal{L}(\bar{x},y,\bar{\lambda}) : \bar{\lambda}\in\Sigma(\bar{x},y)\right\} + \mathcal{N}_X(\bar{x}). \tag{4.10}$$

此外，根据 Carathéodory定理，集合 $\mathcal{S}(\bar{x})$ 的凸包可表示为 $\mathcal{S}(\bar{x})$ 中至多 $n+1$ 个点的凸包。这正是 Ye 和 Zhu [53, 定理4.1] 中提出的必要最优性条件类型。请注意，很容易看出，[53,定理4.1] 的表述遗漏了受限上紧性条件。

本小节将推导最优性条件，使得 (4.10) 被替换为

$$0 \in \left\{\nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) : \bar{\lambda}\in\Sigma(\bar{x},\bar{y})\right\} + \mathcal{N}_X(\bar{x}),$$

其中，$\bar{y}$ 是从 $\mathcal{S}(\bar{x})$ 中选取的一个给定解。

第一个结果基于解映射 $\mathcal{S}(x)$ 的内半连续性。

定理4.7 设 $(\bar{x},\bar{y})$ 是极小极大问题 (P_{minimax}) 的一个局部解。假设解映射 $\mathcal{S}(x)$ 在 $(\bar{x},\bar{y})$ 处内半连续，系统 $g(x,y)\le 0$ 在 $(\bar{x},\bar{y})$ 处满足局部误差界，并且

$$\{\nabla_x g(\bar{x},\bar{y})\lambda : \lambda\in\Sigma^0(\bar{x},\bar{y})\} = \{0\}. \tag{4.11}$$

¯然后存在 λ 使得系统（4.9）成立。

证明. 在上述假设下，根据定理 3.1可知，$V(x)$在 $\bar{x}$处是Lipschitz连续的，并且

$$\partial V(\bar{x}) \subseteq \{\nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{\lambda}) : \bar{\lambda}\in\Sigma(\bar{x},\bar{y})\}$$

将上述包含关系与费马法则（1.5）相结合即可完成证明。 ■

以下推论可由推论3.1和费马法则（1.5）立即得到。

推论4.1 设（$\bar{x}$，$\bar{y}$）是极小极大问题（P_{minimax}）的一个局部解。假设解映射 $\mathcal{S}(x)$在（$\bar{x}$，$\bar{y}$）处内半连续，并且系统$g(\bar{x}$，$y)\le 0$ 在 $y=\bar{y}$处满足MFCQ。则存在 λ 使得系统（4.9）成立。

第二个结果基于值函数 $V(x)$的凹性。

定理 4.8 设 $(\bar{x},\bar{y})$ 是极小极大问题 (P_{minimax}) 的一个局部解。假设$f(x,y)$ 关于 (x,y) 是联合凹的，且 $g(x,y)$ 关于 (x,y) 是联合拟凸的，并且存在一个开集$\mathcal{O}\ni\bar{x}$，使得 $\mathcal{F}(x)$ 非空，且目标函数 $f(x,y)$ 在$\mathcal{F}(x)$ 上有上界。进一步假设 $g(x,y)$ 在 $(\bar{x},\bar{y})$ 处满足一个约束资格条件 ≤ 0。则存在 λ，使得系统 (4.9) 成立。

证明。由定理 3.2， $V(x)$ 是凹的，并且在 $\bar{x}$ 处Lipschitz连续。此外

$$\partial V(\bar{x}) \subseteq \{\nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{\lambda}) : \bar{\lambda} \in \Sigma(\bar{x}, \bar{y})\}.$$

因此，将上述包含关系与费马法则 (1.5) 结合即可完成证明。 ■

4.4 生成对抗网络中的应用

本节将把我们新的最优性条件应用于一种称为生成对抗网络（GAN）的对抗建模框架，该框架最早由 Goodfellow 等人提出。[14]生成网络生成新的样本，而判别网络则试图区分这些样本与真实数据。设 $G(z;x)$ 是由 x 参数化的生成器，$D(s;y)$ 是样本 s 来自数据而非生成器在数据上的分布的概率，其中 y 是判别器的参数。给定 x，判别网络经过训练，以最大化为训练样本和来自生成网络的样本分配正确标签的概率。因此，与 GAN 对应的极小极大问题可以写为

$$\min_x \max_y f(x,y) := \mathbb{E}_{s\sim p_{data}(s)}\left[\log D(s;y)\right] + \mathbb{E}_{z\sim p_{latent}(z)}\left[\log\left(1 - D(G(z;x);y)\right)\right], \quad (4.12)$$

其中， p_{data} 是数据分布， p_{latent} 是潜在分布。

考虑这样一种情形：生成器是采用恒等激活函数的单层网络，而判别器也是采用逻辑 sigmoid 激活函数的单层网络。也就是说，

$$G(z;x) = x + z, \quad D(s;y) = \frac{1}{1 + e^{y^T(s-\bar{s})}}, \quad (4.13)$$

其中， $x \in \mathbb{R}^n$ 是生成器的参数， $y \in \mathbb{R}^n$ 是判别器的参数， $\bar{s} = \mathbb{E}_{s\sim p_{data}}[s]$ 是数据集上的平均向量，参见 Grimmer 等人 [15] 中简单的 CIFAR 生成对抗网络。

为简单起见，分别从 p_{data} 生成一个样本 s ，并从 p_{latent} 生成一个样本 z ，我们有

$$\begin{aligned} f(x,y) =& \log\left(\frac{1}{1+e^{y^T(s-\bar{s})}}\right) + \log\left(1 - \frac{1}{1+e^{y^T(x+z-\bar{s})}}\right) \\ =& y^T(x+z-\bar{s}) - \log\left(1+e^{y^T(s-\bar{s})}\right) - \log\left(1+e^{y^T(x+z-\bar{s})}\right). \end{aligned}$$

注意到

$$\nabla_x f(x,y) = \frac{1}{1+e^{y^T(x+z-\bar{s})}} y, \nabla_y f(x,y) = \frac{1}{1+e^{y^T(x+z-\bar{s})}}(x+z-\bar{s}) - \frac{e^{y^T(s-\bar{s})}}{1+e^{y^T(s-\bar{s})}}(s-\bar{s}).$$

$$\nabla^2_{xx} f(x,y) = -\frac{e^{y^T(x+z-\bar{s})}}{\left[1+e^{y^T(x+z-\bar{s})}\right]^2} yy^T \preceq 0,$$

$$\nabla^2_{yy} f(x,y) = -\frac{e^{y^T(x+z-\bar{s})}}{\left[1+e^{y^T(x+z-\bar{s})}\right]^2}(x+z-\bar{s})(x+z-\bar{s})^T - \frac{e^{y^T(s-\bar{s})}}{\left[1+e^{y^T(s-\bar{s})}\right]^2}(s-\bar{s})(s-\bar{s})^T \preceq 0.$$

因此，$\min_x \max_y f(x,y)$ 是一个非凸-凹极小极大问题。定义 $\bar{x} = s - z$。我们有 $\nabla_y f(\bar{x},y) = \frac{1-e^{y^T(s-\bar{s})}}{1+e^{y^T(s-\bar{s})}}(s-\bar{s})$，因此 $\nabla_y f(\bar{x},y) = 0$ 当且仅当 $y^T(s-\bar{s}) = 0$。由 $f(x,y)$ 关于 y对每个 x的凹性可得，$\mathcal{S}(\bar{x}) = \left\{y \in \mathbb{R}^n : \ y^T(s-\bar{s}) = 0\right\}$，进而 $\max_y f(\bar{x},y) = -2\log 2$。另一方面，注意到 $f(x,0) = -2\log 2$，且 $\nabla_y f(x,0) = \frac{1}{2}(x+z-s)$。因此，对于所有 $x \neq \bar{x}$， $\nabla_y f(x,0) \neq 0$，所以

$$\max_y f(x,y) > f(x,0) = -2\log 2 = \max_y f(\bar{x},y) \text{ for all } x \neq \bar{x}. \tag{4.14}$$

Thus a所有 $\min_x \max_y f(x,y)$ 的最优解均由 $\{(\bar{x},\bar{y}) : \ \bar{y} \in \mathcal{S}(\bar{x})$ 给出 $\}$.

我们声称，对于上述例子的所有最优解，定理 4.3 中的最优性条件都成立；但通常采用的、基于纳什均衡的一阶条件

$$\nabla_x f(\bar{x},\bar{y}) = 0, \quad \nabla_y f(\bar{x},\bar{y}) = 0 \tag{4.15}$$

仅适用于其中一个解，即最优解 $(\bar{x},0)$。回顾一下，驻点条件 (4.15) 被广泛用于刻画 Jin 等人 [24] 定义的（局部）纳什均衡和局部极小极大点，并用于分析各种极小极大问题的梯度下降上升算法的收敛性，参见例如 Ozdaglar 和 Bertsekas [36], 中的定义 2.1 和定理 3.4、Lu 等人 [25], 中的命题 3.1 以及定理 1 和 2、Lin 等人 [22], 中的定义 4.10 以及命题 4.11 和 4.12、Lin 等人 [23], 中的定义 5 和定理 12、Yang 等人 [47], 中的定义 1 和定理 3.2、Ostrovskii 等人 [35] 中的定义 2.1、备注 2.2 和定理 4.1。接下来证明我们的结论。事实上，一方面， $\nabla_x f(\bar{x},\bar{y}) = 0$ 意味着 $\bar{y} = 0$。这说明，满足 $0 \neq \bar{y} \in \mathcal{S}(\bar{x})$ 的最优解 $(\bar{x},\bar{y})$ 无法用通常的一阶条件 (4.15) 来刻画。另一方面，很容易验证定理 4.3 中的必要条件均成立，即对于任意 $\bar{y} \in \mathcal{S}(\bar{x})$，都存在 $\bar{u} = -\bar{y}$ 使得

$$\nabla_x f(\bar{x},\bar{y}) + \nabla^2_{xy} f(\bar{x},\bar{y})\bar{u} = 0, \quad \nabla^2_{yy} f(\bar{x},\bar{y})\bar{u} = 0, \quad \nabla_y f(\bar{x},\bar{y}) = 0. \tag{4.16}$$

此外，可以在生成对抗网络 (4.12) 中添加一些约束函数，使得带均衡约束的数学规划-线性无关约束资格条件不成立，而定理 4.3 所要求的条件仍然成立。例如，取 $g(x,y) = \left(\frac{1}{2}\|y\|^2 - \frac{1}{2}, y_1 - 1\right)^T$, $s - \bar{s} = (0,1)^T$，以及 $\bar{y} = (1,0)^T \in \mathbb{R}^2$，则 $\bar{x} = s - z, \bar{y} = (1,0)^T$仍然是最优解，因为 $y = \bar{y}$ 和 $y = 0$ 均满足约束 $g(x,y) \leq 0$，对所有x 均成立，并且不等式 (4.14) 也成立。由于

$$\nabla_y f(x,y) - \nabla_y g(x,y)^T\lambda = \frac{1}{1+e^{y^T(x+z-\bar{s})}}(x+z-\bar{s}) - \frac{e^{y^T(s-\bar{s})}}{1+e^{y^T(s-\bar{s})}}(s-\bar{s}) - \lambda_1 y - \lambda_2(1,0)^T,$$

我们有

$$\nabla_y f(\bar{x},y) - \nabla_y g(\bar{x},y)^T\lambda = -\lambda_1 y - \lambda_2(1,0)^T.$$

因此，很容易得到 $\Sigma(\bar{x},\bar{y}) = \{(0,0)\}$。由于 $\mathcal{I}_{00} = \{1,2\}$，向量 $\nabla_{(x,y,\lambda)} g_1(\bar{x},\bar{y}) = (0,0,1,0,0,0)^T$和 $\nabla_{(x,y,\lambda)} g_2(\bar{x},\bar{y}) = (0,0,1,0,0,0)^T$线性相关，定理 4.2 中定义的 MPEC-LICQ 不成立。但对于约束系统 (4.5)，在 $(\bar{x},\bar{y},0)$ 处 LICQ 成立，因为矩阵

$$(\nabla_{x,y,\lambda}\nabla_y \mathcal{L}(x,y,\lambda), \nabla_{x,y,\lambda}\lambda) = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$

$\bar{}$ $\bar{}$具有满列秩。由于乘子 $\lambda_1 = 0$、$\lambda_2 = 0$，约束情形下的必要最优性条件与 (4.16)中的无约束情形相同。这表明，通过 Wolfe 对偶方法得到的定理 4.5 中的驻点条件比通过 MPEC 方法得到的驻点条件更为严格，而定理 4.3 和 4.4 所要求的条件通常弱于 MPEC-LICQ。

5 总结

本文针对最大值函数的 Fr´échet、极限和水平次微分，分别在解集的所有解的并集或仅一个解的情形下，推导了上界估计。基于所得结果，我们为非凸（非）凹极小极大问题建立了必要最优性条件。

致谢

作者感谢责任编辑和两位审稿人提出的宝贵意见和建设性建议。特别感谢其中一位审稿人建议采用一个弱于 RS条件的值函数连续性的充分条件；我们在论文的早期版本中使用了该条件。

参考文献

[1] R. Andreani、G. Haeser、M.L. Schuverdt 和 P.J. Silva，《放宽的常数正线性相关约束资格条件及其应用》，《数学规划》，135（2002年），255–273。[2] D.P. Bertsekas 和 A.E. Ozdaglar，《约束优化中的伪正规性与拉格朗日乘子理论》，《优化理论与应用杂志》，114（2002年），287–343。[3] J.F. Bonnans 和 A. Shapiro，《优化问题的扰动分析》，施普林格，纽约，2000。[4] F.H. Clarke，《优化与非光滑分析》，SIAM，1990。

[5] S. Dempe 和 J. Dutta，《双层规划是带均衡约束的数学规划的特例吗？》，《数学规划》，131（2012年），37–48。[6] Y-H. Dai 和 L. Zhang，《约束极小极大优化的最优性条件》，《中国工业与应用数学学会学报》，1（2020年），296-315。[7] A.V. Fiacco 和 Y. Ishizuka，《非线性规划的灵敏度与稳定性分析》，《运筹学年刊》，27（1990年），215–235。[8] A.V. Fiacco 和 J. Kyparisis，《参数非线性规划中最优值函数的凸性与凹性》，《优化理论与应用杂志》，48（1986年），95–126。[9] J. Gauvin，《数学规划中边际函数的广义梯度》，《数学运筹学》，4（1979年），458–463。[10] J. Gauvin 和 F. Dubeau，《数学规划中边际函数的微分性质》，《数学规划中的最优性与稳定性》。施普林格，柏林、海德堡，1982年，101–119。[11] H. Gfrerer 和 B. Mordukhovich，《通过变分分析研究参数约束系统的 Robinson 稳定性》，SIAM 优化杂志，27（2017年），438–465。[12] H. Gfrerer 和 J.J. Ye，《通过变分分析为带均衡约束的数学规划建立新的约束资格条件》，SIAM 优化杂志，27（2017年），842–865。[13] G. Giorgi、B. Jim´enez 和 V. Novo，《多目标优化中的近似Karush-Kuhn-Tucker条件》，《优化理论与应用杂志》，171（2016年），70–89。[14] I. Goodfellow、J. Pouget-Abadie、M. Mirza 等，《生成对抗网络》，《神经信息处理系统进展》，27（2014年），2672-2680。[15] B. Grimmer、H. Lu、P. Worah 等，《通过连续时间系统研究非凸-非凹极小极大优化的极限行为》，国际算法学习理论会议，PMLR，167（2022年），465-487。[16] L. Guo、G.H. Lin 和 J.J. Ye，《带几何约束的参数数学规划的稳定性分析及其应用》，SIAM 优化杂志，22（2012年），1151–1176。[17] L. Guo、G.H. Lin、J.J. Ye 和 J. Zhang，《带均衡约束的参数数学规划值函数的灵敏度分析》，SIAM 优化杂志，24（2014年），1206–1237。[18] R. Henrion、A. Jourani 和 J.V. Outrata，《一类多值函数的平稳性》，SIAM 优化杂志，13（2002年），603–618。[19] M.D. Intriligator，《数学优化与经济理论》，第2版。SIAM，费城，2002。[20] R. Janin，《非线性规划中边际函数的方向导数》，载于《灵敏度、稳定性与参数分析》（第110–126页）。施普林格，柏林、海德堡，1984。

[21] J. Jiang 和 X.J. Chen，非光滑非凸-非凹极小极大问题及生成对抗网络的最优性条件，arXiv2203.10914v1。[22] T. Lin、C. Jin 和 M.I. Jordan，关于非凸-凹极小极大问题的梯度下降上升，国际机器学习会议，PMLR，119（2020），6083-6093。[23] T. Lin、C. Jin 和 M.I. Jordan，极小极大优化的近乎最优算法，学习理论会议，PMLR，125（2020），2738-2779。[24] C. Jin、P. Netrapalli 和 M.I. Jordan，非凸-非凹极小极大优化中的局部最优性是什么，2019年国际机器学习会议。[25] S. Lu、I. Tsaknakis、M. Hong 和 Y. Chen，单边非凸极小极大问题的混合分块逐次逼近：算法与应用，《IEEE信号处理汇刊》，68（2020），3676-3691。[26] Y. Lucet 和 J.J. Ye，具有变分不等式约束的优化问题的值函数敏感性分析，《SIAM控制与优化杂志》，40（2001），699–723。[27] Y. Lucet 和 J.J. Ye，勘误：具有变分不等式约束的优化问题的值函数敏感性分析，《SIAM控制与优化杂志》，41（2002），1315–1319。[28] P. Mehlitz 和 L. Minchenko，在放宽的常数正线性相关约束资格条件下集值映射的R-正则性及其在参数优化和双层优化中的应用，《集值与变分分析》，30（2022），179-205。[29] L. Minchenko 和 S. Stakhovski，放宽的常数秩条件下的参数非线性规划问题，《SIAM优化杂志》，21（2011），314–332。[30] B.S. Mordukhovich，《变分分析与广义微分 I：基础理论，II：应用》，数学科学基础丛书 330，施普林格，柏林，2006年。[31] B.S. Mordukhovich、N.M. Nam 和 N.D. Yen，参数数学规划中边际函数的次梯度，《数学规划》，116（2009），369–396。[32] J. 冯·诺伊曼，关于博弈论，《数学年刊》，100（1928），295–320。[33] J. 冯·诺伊曼和 O. 摩根斯坦，《博弈论与经济行为》，普林斯顿大学出版社，2007年。[34] M. Nouiehed、M. Sanjabi、T. Huang、J.D. Lee 和 M. Razaviyayn，使用迭代一阶方法求解一类非凸极小极大博弈，《神经信息处理系统进展》，32（2019）。[35] D.M. Ostrovskii、A. Lowy 和 M. Razaviyayn，在非凸-凹光滑极小极大问题中高效搜索一阶纳什均衡，《SIAM优化杂志》，31（2021），2508-2538。

[36] A.E. Ozdaglar 和 D.P. Bertsekas，约束优化中伪正规性与拟正则性之间的关系，《优化方法与软件》，19（2004），493–506。[37] H. Rafique、M. Liu、Q. Lin 和 T. Yang，弱凸极小极大优化：可证明的算法及其在机器学习中的应用，arXiv预印本 arXiv:1810.02060v4，2021年。[38] S.M. Robinson，不等式系统的稳定性理论，第二部分：可微非线性系统，《SIAM数值分析杂志》，13（1976），497–513。[39] S.M. Robinson，广义方程及其解，第二部分：非线性规划中的应用，《数学规划研究》，19（1982），200–221。[40] R.T. Rockafellar，《凸分析》，普林斯顿大学出版社，1970年。[41] R.T. Rockafellar 和 R.J.-B. Wets，《变分分析》，施普林格，1998年。[42] H. Scheel 和 S. Scholtes，带互补约束的数学规划：驻点性、最优性与灵敏度，《数学运筹学》，25（2000），1-22。[43] M. Sion，关于一般极小极大定理，《太平洋数学杂志》，8（1958），171-176。[44] O. Stein，《半无限规划中的双层策略》，克吕维尔学术出版社，波士顿，2003年。[45] L. Thibault，关于最优值函数的次微分，《SIAM控制与优化杂志》29（1991），1019–1036。[46] P. Wolfe，非线性规划的对偶定理，《应用数学季刊》，19（1961），239–244。[47] J. Yang、N. Kiyavash 和 N. He，一类非凸-非凹极小极大问题的全局收敛与方差缩减，《神经信息处理系统进展》，33（2020），1153-1165。[48] J.J. Ye，新的统一参数误差界，《优化理论与应用杂志》，98（1998），197-219。[49] J.J. Ye，带均衡约束的数学规划的必要且充分最优性条件，《数学分析与应用杂志》，307（2005），350-369。[50] J.J. Ye 和 S.Y. Wu，广义半无限规划问题的一阶最优性条件，《优化理论与应用杂志》，137（2008），419-434。[51] J.J. Ye 和 J. Zhang，强化的 Karush-Kuhn-Tucker 条件与更弱的约束资格条件，《数学规划》，139（2013），353–381。[52] J.J. Ye 和 J.C. Zhou，二阶锥互补问题误差界性质的可验证充分条件，《数学规划》，171（2018），361–395。[53] J.J. Ye 和 D.L. Zhu，双层规划问题的最优性条件，《优化》，33（1995），9–27。

[54] J.J. Ye、Q.J. Zhu 和 D.L. Zhu，《广义双层规划问题的精确罚与必要最优性条件》，《SIAM 优化杂志》7（1997年），481–507。[55] J. Zhang、M. Hong 和 S. Zhang，《关于凸凹鞍点问题的较低的迭代复杂度界》，《数学规划》194（2022年），901–935。