

稀疏加组稀疏优化问题的一阶和二阶方向稳定点研究

吴青青, 彭定涛*, 苏妍妍

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2023年10月20日; 录用日期: 2023年12月21日; 发布日期: 2023年12月29日

摘要

在本文中, 我们考虑一类非凸非光滑的无约束稀疏加组稀疏优化问题, 其损失函数是二阶连续可微函数(可能非凸), 惩罚项是稀疏惩罚与组稀疏惩罚的组合, 其稀疏惩罚是 l_1 范数, 组稀疏惩罚是折叠凹惩罚函数。目前, 计算这类带有凸加非凸惩罚优化问题的方向稳定点的研究较少, 但利用方向导数定义的方向稳定点比次微分所定义的稳定点(critical点、lifted稳定点等)能更好的刻画解的局部最优性质。因此, 本文主要通过方向稳定点来刻画模型的最优性条件。首先, 本文引入了一阶、二阶方向稳定点的概念, 探讨了它们与问题局部解的关系。其次, 给出了一阶、二阶方向导数的具体表达式, 这为进一步分析和求解此类问题提供了理论基础。

关键词

稀疏加组稀疏优化问题, 非凸惩罚, 方向稳定点, 局部最优性质

On the First-Order and Second-Order Directional Stationary Points of Sparse Plus Group Sparse Optimization Problems

Qingqing Wu, Dingtao Peng*, Yanyan Su

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Oct. 20th, 2023; accepted: Dec. 21st, 2023; published: Dec. 29th, 2023

Abstract

In this paper, we consider a class of nonconvex, nonsmooth and unconstrained sparse plus group
*通讯作者。

文章引用: 吴青青, 彭定涛, 苏妍妍. 稀疏加组稀疏优化问题的一阶和二阶方向稳定点研究[J]. 运筹与模糊学, 2023, 13(6): 7464-7476. DOI: 10.12677/orf.2023.136734

sparse optimization problems, in which the loss function is a twice continuously differentiable function (possibly nonconvex), and the penalty term is a combination of sparse penalty and group sparse penalty, where the sparse penalty is a ℓ_1 norm, and the group sparse penalty is a general folded concave penalty function. At present, there are few studies on the calculation of directional stationary points for this type of optimization problems with convex plus nonconvex penalties, and directional stationary points characterized by directional derivatives can better show the local optimal properties of solutions than other stationary points defined by subdifferentials (e.g. critical points, lifted stationary points, etc.). Therefore, in this paper, the optimality conditions of the model are characterized by means of directional stationary points. Firstly, we introduce the concepts of first-order and second-order directional stationary points, and discuss their relations with local solutions of the problem. Secondly, the concrete expressions of the first and second order directional derivatives are given, which provide a theoretical basis for further analyzing and solving this kind of problems.

Keywords

Sparse Plus Group Sparse Optimization Problem, Nonconvex Penalty, Directional Stationary Points, Local Optimality

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来, 稀疏加组稀疏优化问题得到了广泛的研究, 包括图像恢复、变量选择、机器学习等众多领域[1] [2] [3] [4]。一般的稀疏加组稀疏优化问题[3]如下:

$$\min_{\mathbf{x} \in \mathbb{R}^n} F(\mathbf{x}) = f(\mathbf{x}) + \lambda_1 \|\mathbf{x}\|_0 + \lambda_2 \sum_{j=1}^J \#(\|\mathbf{x}_{(j)}\|_2), \quad (1)$$

其中, 损失函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 二阶连续可微(可能非凸), 本文用 $\|\cdot\|$ 表示 ℓ_2 范数。变量 $\mathbf{x} = (\mathbf{x}_{(1)}^T, \mathbf{x}_{(2)}^T, \dots, \mathbf{x}_{(J)}^T)^T \in \mathbb{R}^n$ 具有 J 个分组结构, 第 j 组变量 $\mathbf{x}_{(j)} \in \mathbb{R}^{n^j}$, $\sum_{j=1}^J n^j = n$, $\|\mathbf{x}\|_0$ 表示 \mathbf{x} 中非零分量的个数, $\|\mathbf{x}_{(j)}\|_2$ 表示 \mathbf{x} 的第 j 组变量 $\mathbf{x}_{(j)}$ 的 ℓ_2 范数, $\sum_{j=1}^J \#(\|\mathbf{x}_{(j)}\|_2)$ 表示 \mathbf{x} 中非零组的个数, 正则化参数 $\lambda_1, \lambda_2 > 0$ 。

稀疏优化问题的主要目的是寻求欠定线性方程组的稀疏解, 从而提高模型的计算效率和可解释性。稀疏性通常用 ℓ_0 范数刻画, 即 $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{x}\|_0 \ll n$ 。 ℓ_0 范数可以直接惩罚非零元素的数量, 有助于找到仅有少数重要特征或特征组合的解, 去除无用或冗余的特征, 从而降低数据表示的复杂度, 提高预测识别的准确性。2006年 Yuan 和 Lin [4]首次提出将组稀疏性作为先验信息, 组稀疏性考虑了特征之间的相关性, 是把向量的分量进行分组再考察各组是否整体为零, 体现为向量中非零分量或零分量集中出现在某些区域。因此, 综合考虑变量的稀疏和组稀疏结构可以获得更具实际意义的结果。

然而, 问题(1)的正则项是非凸、非光滑、非 Lipschitz 甚至不连续的 NP 难问题[5] [6]。Donoho 等人 [7]提出将 ℓ_0 范数凸松弛为 ℓ_1 范数。Fan 和 Li [6] [8]提出了一种折叠凹惩罚函数, 并证明了该非凸非光滑优化模型的解具有无偏性、oracle 等性质。后来, 学者们给出了一系列折叠凹惩罚函数, 包括 ℓ_p ($0 < p < 1$) 范数、SCAD 罚(Smoothly Clipped Absolute Deviation Penalty) [6]、MCP (Minimax Concave Penalty) [9]、Capped- ℓ_1 罚(Continuous Exact ℓ_0 Penalty) [10]等, 非凸非光滑松弛问题已经成为近年来研究热点之一[5]

[11][12][13][14][15]。常用的折叠凹惩罚函数有以下几类：

1) Capped- ℓ_1 惩罚函数为

$$\phi^{Capped-\ell_1}(t) := \lambda \min\left\{\frac{t}{\alpha}, 1\right\} = \begin{cases} \frac{\lambda t}{\alpha}, & 0 \leq t \leq \alpha, \\ \lambda, & \alpha < t. \end{cases} \quad (\lambda > 0, \alpha > 0)$$

2) MCP 函数为

$$\phi^{MCP}(t) := \lambda \int_0^t \left(1 - \frac{\tau}{\alpha\lambda}\right)_+ d\tau = \begin{cases} \lambda t - \frac{t^2}{2\alpha}, & 0 \leq t \leq \alpha\lambda, \\ \frac{1}{2}\alpha\lambda^2, & \alpha\lambda < t. \end{cases} \quad (\lambda > 0, \alpha > 1)$$

3) SCAD 惩罚函数为

$$\phi^{SCAD}(t) := \lambda \int_0^t \min\left\{1, \frac{\left(\alpha - \frac{\tau}{\lambda}\right)_+}{\alpha - 1}\right\} d\tau = \begin{cases} \lambda t, & 0 \leq t \leq \lambda, \\ \frac{1}{\alpha - 1} \left(-\frac{\lambda^2}{2} + \alpha\lambda t - \frac{1}{2}t^2\right), & \lambda < t \leq \alpha\lambda, \\ \frac{1}{2}(\alpha + 1)\lambda^2, & \alpha\lambda < t. \end{cases} \quad (\lambda > 0, \alpha > 2)$$

其中 $\beta_+ := \begin{cases} 0, & \beta \leq 0, \\ \beta, & \beta > 0. \end{cases}$

对于问题(1)，本文考虑如下非凸松弛模型：

$$\min_{\mathbf{x} \in \mathbb{R}^n} F(\mathbf{x}) = f(\mathbf{x}) + \lambda_1 \|\mathbf{x}\| + \lambda_2 \sum_{j=1}^J \phi(\|\mathbf{x}_{(j)}\|), \quad (2)$$

其中，正则化参数 $\lambda_1, \lambda_2 > 0$ ， $\|\mathbf{x}\| := \sum_{i=1}^n |x_i|$ ， $\phi: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ 为一般折叠凹惩罚函数形式，只需满足以下性质：

- 1) ϕ 在 $[0, \infty)$ 上是局部 Lipschitz 的，单调不减， $\phi(0) = 0$ 且当 $t > 0$ 时， $\phi(t) > 0$ ；
- 2) $\phi'(0) := \phi'(0^+) > 0$ 。易见，Capped- ℓ_1 、MCP、SCAD 等惩罚函数均满足上述条件。

对于无约束非凸非光滑问题，Ahn 和 Pang [16][17] 通过方向导数定义了一阶、二阶 \mathbf{d} -稳定点，Cui [18] 指出通过方向导数定义的 \mathbf{d} -稳定点可以推导出 lifted 稳定点、critical 稳定点等，进而得出非光滑问题的一阶、二阶 \mathbf{d} -稳定点强于次微分所定义的稳定点。然而，目前的研究表明计算非光滑问题的一阶、二阶 \mathbf{d} -稳定点是困难的。2020 年 Peng 和 Chen [19] 给出了无约束组稀疏优化问题的一阶、二阶 \mathbf{d} -稳定点的有效刻画以及与该问题局部最优解的关系，提出了光滑化方法，并证明了光滑化问题与松弛问题二阶 \mathbf{d} -稳定点的一致性。在此基础上，Su [20] 对稀疏组稀疏优化问题的连续松弛模型的一阶 \mathbf{d} -稳定点进行了刻画，并给出了光滑化问题和松弛问题一阶 \mathbf{d} -稳定点的一致性。不同于上述研究，本文考虑的稀疏惩罚是凸函数加组稀疏惩罚是一类折叠凹惩罚函数的组合问题，研究此类问题的一阶、二阶 \mathbf{d} -稳定点的有效刻画以及局部最优性质，并给出了一阶、二阶方向导数的具体表达式。

本文的结构如下：在第二节，我们介绍了一阶 \mathbf{d} -稳定点及其局部最优性质，并给出松弛问题(2)的一阶 \mathbf{d} -稳定点的特征和相关结论；在第三节，我们给出了二阶 \mathbf{d} -稳定点的概念和性质，并给出了问题(2)的二阶 \mathbf{d} -稳定点的具体表达式；最后一节是一个简单的总结。

符号标记：对于 $n \in \mathbb{N}^+$ ， $[n] := \{1, 2, \dots, n\}$ ， $J(\mathbf{x}) := \{j: \mathbf{x}_{(j)} \neq \mathbf{0}, j \in [J]\}$ 表示向量 \mathbf{x} 的非零组的下标。

$I_j(\mathbf{x}) := \{i: x_{(j)i} \neq 0, \forall j \in [J], i \in [n^j]\}$ 表示向量 \mathbf{x} 的第 j 组中非零元素的下标。设 A, B 为两个集合, $A \setminus B := \{i: i \in A \text{ 且 } i \notin B\}$ 。 $D(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n: I_j(\mathbf{y}) = I_j(\mathbf{x}), \forall j \in [J]\}$ 表示与 \mathbf{x} 有相同的非零元素下标的向量全体。

2. 一阶 d-稳定点

2.1. 一阶 d-稳定点的局部最优性

本节引入问题(2)的几个重要概念: 凸差函数、方向导数以及方向稳定点[16] [19] [21]。

为方便分析, 本节假定惩罚函数可以改写为凸差(Difference-of-Convex)形式, 即设惩罚函数具有下述形式。

假设 2.1 [16] [19] (凸差函数) 设折叠凹惩罚函数 $\phi: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ 具有凸差形式: $\phi(t) := g(t) - h(t)$, 其中 $h(t) := \max_{1 \leq \nu \leq \bar{\nu}} h_\nu(t)$, ν 和 $\bar{\nu} \in \mathbb{N}^+$, g 和 h_ν 在 $t \in (0, \infty)$ 上都是可微凸函数, 且满足:

$$g'(0) := g'(0^+), h'_\nu(0) := h'_\nu(0^+), \forall 1 \leq \nu \leq \bar{\nu}.$$

容易验证, 前述几类折叠凹惩罚函数均具有凸差形式:

1) Capped- ℓ_1 罚函数:

$$\begin{aligned} \phi^{\text{Capped-}\ell_1}(t) &= g^{\text{Capped-}\ell_1}(t) - h^{\text{Capped-}\ell_1}(t) \\ g^{\text{Capped-}\ell_1}(t) &= \frac{\lambda t}{\alpha}, h^{\text{Capped-}\ell_1}(t) = \max\left\{0, \frac{\lambda t}{\alpha} - \lambda\right\}, (\alpha > 0, \lambda > 0) \end{aligned}$$

2) MCP 函数:

$$\begin{aligned} \phi^{\text{MCP}}(t) &= g^{\text{MCP}}(t) - h^{\text{MCP}}(t) \\ g^{\text{MCP}}(t) = \lambda t, h^{\text{MCP}}(t) &= \begin{cases} \frac{t^2}{2\alpha}, & 0 \leq t \leq \alpha\lambda, \\ \lambda t - \frac{1}{2}\alpha\lambda^2, & \alpha\lambda < t, \end{cases} (\alpha > 1, \lambda > 0) \end{aligned}$$

3) SCAD 罚函数:

$$\begin{aligned} \phi^{\text{SCAD}}(t) &= g^{\text{SCAD}}(t) - h^{\text{SCAD}}(t) \\ g^{\text{SCAD}}(t) = \lambda t, h^{\text{SCAD}}(t) &= \begin{cases} 0, & 0 \leq t \leq \lambda, \\ \frac{(t-\lambda)^2}{2(\alpha-1)}, & \lambda < t \leq \alpha\lambda, (\alpha > 2, \lambda > 0) \\ \lambda t - \frac{1}{2}(\alpha+1)\lambda^2, & \alpha\lambda < t. \end{cases} \end{aligned}$$

在假设 2.1 之下, 问题(2)可以写成如下形式:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} F(\mathbf{x}) &= f(\mathbf{x}) + \lambda_1 \|\mathbf{x}\| + \lambda_2 \sum_{j=1}^J \left[g(\|\mathbf{x}_{(j)}\|) - \max_{1 \leq \nu \leq \bar{\nu}} h_\nu(\|\mathbf{x}_{(j)}\|) \right] \\ &= f(\mathbf{x}) + \lambda_1 \sum_{j=1}^J \sum_{i=1}^{n^j} |x_{(j)i}| + \lambda_2 \sum_{j=1}^J \left[g(\|\mathbf{x}_{(j)}\|) - \max_{1 \leq \nu \leq \bar{\nu}} h_\nu(\|\mathbf{x}_{(j)}\|) \right]. \end{aligned} \tag{3}$$

当 $\bar{\nu} = 1$ 时, 问题(3)可退化为

$$\min_{\mathbf{x} \in \mathbb{R}^n} F(\mathbf{x}) = f(\mathbf{x}) + \lambda_1 \sum_{j=1}^J \sum_{i=1}^{n^j} |x_{(j)i}| + \lambda_2 \sum_{j=1}^J \left[g(\|\mathbf{x}_{(j)}\|) - h(\|\mathbf{x}_{(j)}\|) \right].$$

为了方便表述, 令 $u(\mathbf{x}) = \|\mathbf{x}\|$, $w(\mathbf{x}) = \|\mathbf{x}\|$, 则问题(3)可以改写为:

$$\min_{\mathbf{x} \in \mathbb{R}^n} F(\mathbf{x}) = f(\mathbf{x}) + \lambda_1 \sum_{j=1}^J \sum_{i=1}^{n^j} u(x_{(j)i}) + \lambda_2 \sum_{j=1}^J \left[g \circ w(\mathbf{x}_{(j)}) - \max_{l \leq v \leq \bar{v}} h_v \circ w(\mathbf{x}_{(j)}) \right]. \quad (4)$$

d -稳定点是描述方向导数的必要条件, 接下来我们介绍一阶 d -稳定点的定义以及局部最优性质。

定义 2.1 [19] [21] (一阶 d -稳定点) 设 $Q: \mathbb{R}^n \rightarrow \mathbb{R}$ 在给定 $\mathbf{x}^* \in \mathbb{R}^n$ 处方向可微, 若 Q 的方向导数 Q' 满足

$$Q'(\mathbf{x}^*; \mathbf{x} - \mathbf{x}^*) = \lim_{t \downarrow 0} \frac{Q(\mathbf{x}^* + t(\mathbf{x} - \mathbf{x}^*)) - Q(\mathbf{x}^*)}{t} \geq 0, \forall \mathbf{x} \in \mathbb{R}^n,$$

则称 \mathbf{x}^* 为 Q 的一阶(方向) d -稳定点。值得注意的是, 若 Q 在 \mathbf{x}^* 处可微, 则有

$$Q'(\mathbf{x}^*; \mathbf{x} - \mathbf{x}^*) = \langle \nabla Q(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle = 0.$$

引理 2.1 [19] (一阶 d -稳定点的局部最优性质) 已知 $F: \mathbb{R}^n \rightarrow \mathbb{R}$ 在 $\mathbf{x} \in \mathbb{R}^n$ 上是局部 Lipschitz 连续且方向可微的, 则以下结论成立:

- 1) 如果 \mathbf{x}^* 是 F 的局部极小点, 则 \mathbf{x}^* 也是 F 的一阶 d -稳定点;
- 2) 如果 F 在 \mathbf{x}^* 满足一阶增长性条件, 即存在 \mathbf{x}^* 的一个领域 Θ 和 $\delta > 0$, 使得

$$F(\mathbf{x}) \geq F(\mathbf{x}^*) + \delta \|\mathbf{x} - \mathbf{x}^*\|, \forall \mathbf{x} \in \Theta,$$

等价于在 \mathbf{x}^* 满足

$$F'(\mathbf{x}^*; \mathbf{x} - \mathbf{x}^*) > 0, \forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{x}^*\}.$$

Peng 和 Chen 在[19]中给出了引理 2.1 的证明。根据定义 2.1, 对任意 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $u(\mathbf{x})$, $w(\mathbf{x})$ 的一阶方向导数如下:

$$u'(\mathbf{x}; \mathbf{y}) = \begin{cases} \|\mathbf{y}\|, & \|\mathbf{x}\| = 0, \\ \text{sgn}(\mathbf{x})^\top \mathbf{y}, & \|\mathbf{x}\| \neq 0. \end{cases} \quad w'(\mathbf{x}; \mathbf{y}) = \begin{cases} \|\mathbf{y}\|, & \|\mathbf{x}\| = 0, \\ \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|}, & \|\mathbf{x}\| \neq 0. \end{cases} \quad (5)$$

其中 $\text{sgn}(a) := \begin{cases} 1, & a > 0, \\ 0, & a = 0, \\ -1, & a < 0. \end{cases}$ 进一步, $\forall j \in [J], i \in [n^j]$, 有

$$u'(x_{(j)i}; y_{(j)i}) = \begin{cases} |y_{(j)i}|, & |x_{(j)i}| = 0, \\ \text{sgn}(x_{(j)i}) y_{(j)i}, & |x_{(j)i}| \neq 0, \end{cases} \quad w'(\mathbf{x}_{(j)}; \mathbf{y}_{(j)}) = \begin{cases} \|\mathbf{y}_{(j)}\|, & \|\mathbf{x}_{(j)}\| = 0, \\ \frac{\langle \mathbf{x}_{(j)}, \mathbf{y}_{(j)} \rangle}{\|\mathbf{x}_{(j)}\|}, & \|\mathbf{x}_{(j)}\| \neq 0. \end{cases} \quad (6)$$

实际上, 在一维情况下 ℓ_1 范数和 ℓ_2 范数是等价的。

2.2. 问题(3)的一阶 d -稳定点

本节给出问题(3)的一阶 d -稳定点的具体表达式以及特征刻画。

定理 2.1 在假设 2.1 成立的条件下, 问题(3)在点 \mathbf{x}^* 处的方向导数具体表达式如下:

$$\begin{aligned}
 F'(\mathbf{x}^*; \mathbf{y}) &= \langle \nabla f(\mathbf{x}^*), \mathbf{y} \rangle + \lambda_1 \sum_{j \in [J] \setminus J(\mathbf{x}^*)} \sum_{i \in [n^j]} |y_{(j)i}| + \lambda_1 \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in [n^j] \setminus I_j(\mathbf{x}^*)} |y_{(j)i}| \\
 &\quad + \lambda_1 \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in I_j(\mathbf{x}^*)} \operatorname{sgn}(x_{(j)i}) y_{(j)i} + \lambda_2 \sum_{j \in [J] \setminus J(\mathbf{x}^*)} \left[g' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \|\mathbf{y}_{(j)}\| \\
 &\quad + \lambda_2 \sum_{j \in J(\mathbf{x}^*)} \left[g' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \frac{\langle \mathbf{x}_{(j)}^*, \mathbf{y}_{(j)} \rangle}{\|\mathbf{x}_{(j)}^*\|}, \quad \forall \mathbf{y} \in \mathbb{R}^n.
 \end{aligned} \tag{7}$$

证明. 首先, 注意到凸函数 $h_\nu: \mathbb{R}^+ \rightarrow \mathbb{R}^+$, $w: \mathbb{R}^{n^j} \rightarrow \mathbb{R}^+$ 和复合函数 $h_\nu \circ w: \mathbb{R}^{n^j} \rightarrow \mathbb{R}^+$ 均是方向可微的。 $\forall j \in [J]$, 有

$$(h \circ w)'(\mathbf{x}_{(j)}^*; \mathbf{x}_{(j)} - \mathbf{x}_{(j)}^*) = h'(\|\mathbf{x}_{(j)}^*\|; \mathbf{x}_{(j)} - \mathbf{x}_{(j)}^*) = \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu(\|\mathbf{x}_{(j)}^*\|) w'(\mathbf{x}_{(j)}^*; \mathbf{x}_{(j)} - \mathbf{x}_{(j)}^*).$$

同理, 函数 g 也可得出上述结论。根据定义 2.1 和(6), 可得

$$\begin{aligned}
 F'(\mathbf{x}^*; \mathbf{x} - \mathbf{x}^*) &= \langle \nabla f(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle + \lambda_1 \sum_{j=1}^J \sum_{i=1}^{n^j} u'(\mathbf{x}_{(j)i}^*; x_{(j)i} - x_{(j)i}^*) \\
 &\quad + \lambda_2 \sum_{j=1}^J \left[g' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] w'(\mathbf{x}_{(j)}^*; \mathbf{x}_{(j)} - \mathbf{x}_{(j)}^*) \\
 &= \langle \nabla f(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle + \lambda_1 \sum_{j \in [J] \setminus J(\mathbf{x}^*)} \sum_{i \in [n^j]} u'(\mathbf{x}_{(j)i}^*; x_{(j)i} - x_{(j)i}^*) \\
 &\quad + \lambda_1 \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in [n^j] \setminus I_j(\mathbf{x}^*)} u'(\mathbf{x}_{(j)i}^*; x_{(j)i} - x_{(j)i}^*) + \lambda_1 \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in I_j(\mathbf{x}^*)} u'(\mathbf{x}_{(j)i}^*; x_{(j)i} - x_{(j)i}^*) \\
 &\quad + \lambda_2 \sum_{j \in [J] \setminus J(\mathbf{x}^*)} \left[g' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] w'(\mathbf{x}_{(j)}^*; \mathbf{x}_{(j)} - \mathbf{x}_{(j)}^*) \\
 &\quad + \lambda_2 \sum_{j \in J(\mathbf{x}^*)} \left[g' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] w'(\mathbf{x}_{(j)}^*; \mathbf{x}_{(j)} - \mathbf{x}_{(j)}^*).
 \end{aligned}$$

$\forall \mathbf{y} = \mathbf{x} - \mathbf{x}^* \in \mathbb{R}^n$ 有

$$\begin{aligned}
 F'(\mathbf{x}^*; \mathbf{y}) &= \langle \nabla f(\mathbf{x}^*), \mathbf{y} \rangle + \lambda_1 \sum_{j \in [J] \setminus J(\mathbf{x}^*)} \sum_{i \in [n^j]} |y_{(j)i}| + \lambda_1 \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in [n^j] \setminus I_j(\mathbf{x}^*)} |y_{(j)i}| \\
 &\quad + \lambda_1 \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in I_j(\mathbf{x}^*)} \operatorname{sgn}(x_{(j)i}) y_{(j)i} + \lambda_2 \sum_{j \in [J] \setminus J(\mathbf{x}^*)} \left[g' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \|\mathbf{y}_{(j)}\| \\
 &\quad + \lambda_2 \sum_{j \in J(\mathbf{x}^*)} \left[g' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \frac{\langle \mathbf{x}_{(j)}^*, \mathbf{y}_{(j)} \rangle}{\|\mathbf{x}_{(j)}^*\|}.
 \end{aligned}$$

故得结论。 □

定理 2.2 在假设 2.1 成立的条件下, 设 $\mathbf{x}^* \in \mathbb{R}^n$ 是问题(3)的一个一阶 \mathbf{d} -稳定点, 则下述结论成立:

1) $\forall j \in J(\mathbf{x}^*), i \in I_j(\mathbf{x}^*),$

$$\left| \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} \right| = \left| \lambda_1 \operatorname{sgn}(x_{(j)i}^*) + \lambda_2 \left[g' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \frac{x_{(j)i}^*}{\|\mathbf{x}_{(j)}^*\|} \right|,$$

特别地, 当 $\bar{\nu} = 1$ 时,

$$\left| \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} \right| = \left| \lambda_1 \operatorname{sgn}(x_{(j)i}^*) + \lambda_2 \left[\mathbf{g}' \circ w(\mathbf{x}_{(j)}^*) - h' \circ w(\mathbf{x}_{(j)}^*) \right] \frac{x_{(j)i}^*}{\|\mathbf{x}_{(j)}^*\|} \right|.$$

2) $\forall j \in J(\mathbf{x}^*)$ 若满足 $[n^j] \setminus I_j(\mathbf{x}^*) \neq \emptyset$ (即 $|I_j(\mathbf{x}^*)| < n^j$), 则对任意 $i \in [n^j] \setminus I_j(\mathbf{x}^*)$, 有

$$\left| \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} \right| \leq \lambda_1.$$

3) $\forall j \in [J] \setminus J(\mathbf{x}^*), i \in [n^j]$, 有

$$\left| \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} \right| \leq \lambda_1 n^j (J - |J(\mathbf{x}^*)|) + \lambda_2 \sqrt{n} \sum_{j \in [J] \setminus J(\mathbf{x}^*)} \left[\mathbf{g}' \circ w(\mathbf{0}) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{0}) \right].$$

特别地, 当 $\bar{\nu} = 1$ 时,

$$\left| \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} \right| \leq \lambda_1 n^j (J - |J(\mathbf{x}^*)|) + \lambda_2 \sqrt{n} \sum_{j \in [J] \setminus J(\mathbf{x}^*)} \left[\mathbf{g}' \circ w(\mathbf{0}) - h' \circ w(\mathbf{0}) \right].$$

证明. 1) 根据定理 2.2, $\forall \mathbf{y} = \mathbf{x} - \mathbf{x}^* \in \mathbb{R}^n$, 有

$$\begin{aligned} 0 \leq & \sum_{j \in [J]} \sum_{i \in [n^j]} \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} y_{(j)i} + \lambda_1 \sum_{j \in [J] \setminus J(\mathbf{x}^*)} \sum_{i \in [n^j]} |y_{(j)i}| + \lambda_1 \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in [n^j] \setminus I_j(\mathbf{x}^*)} |y_{(j)i}| \\ & + \lambda_1 \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in I_j(\mathbf{x}^*)} \operatorname{sign}(x_{(j)i}^*) y_{(j)i} + \lambda_2 \sum_{j \in [J] \setminus J(\mathbf{x}^*)} \left[\mathbf{g}' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \|y_{(j)}\| \\ & + \lambda_2 \sum_{j \in J(\mathbf{x}^*)} \left[\mathbf{g}' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \frac{\langle \mathbf{x}_{(j)}^*, y_{(j)} \rangle}{\|\mathbf{x}_{(j)}^*\|}. \end{aligned} \quad (8)$$

特别地, 任取 $\mathbf{y} \in D(\mathbf{x}^*)$, 得

$$0 \leq \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in I_j(\mathbf{x}^*)} \left\{ \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} + \lambda_1 \operatorname{sign}(x_{(j)i}^*) + \lambda_2 \left[\mathbf{g}' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \frac{x_{(j)i}^*}{\|\mathbf{x}_{(j)}^*\|} \right\} \cdot y_{(j)i}, \quad (9)$$

由 $\mathbf{y} \in D(\mathbf{x}^*)$ 的任意性, 对任意 $j \in J(\mathbf{x}^*)$, $i \in I_j(\mathbf{x}^*)$, 有

$$0 = \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} + \lambda_1 \operatorname{sign}(x_{(j)i}^*) + \lambda_2 \left[\mathbf{g}' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \frac{x_{(j)i}^*}{\|\mathbf{x}_{(j)}^*\|}, \quad (10)$$

即

$$\left| \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} \right| = \left| \lambda_1 \operatorname{sign}(x_{(j)i}^*) + \lambda_2 \left[\mathbf{g}' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \frac{x_{(j)i}^*}{\|\mathbf{x}_{(j)}^*\|} \right|.$$

当 $\bar{\nu} = 1$ 时,

$$\left| \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} \right| = \left| \lambda_1 \operatorname{sign}(x_{(j)i}^*) + \lambda_2 \left[\mathbf{g}' \circ w(\mathbf{x}_{(j)}^*) - h' \circ w(\mathbf{x}_{(j)}^*) \right] \frac{x_{(j)i}^*}{\|\mathbf{x}_{(j)}^*\|} \right|.$$

2) 根据已知条件, 我们可取任意满足(8)的 $\hat{j} \in J(\mathbf{x}^*)$, $\mathbf{x}_{(\hat{j})}^* = (x_{(\hat{j})1}^*, x_{(\hat{j})2}^*, \dots, x_{(\hat{j})n^{\hat{j}}}^*)^\top \in \mathbb{R}^{n^{\hat{j}}}$ 和任意取

$\hat{i} \in [n^j] \setminus I_i(\mathbf{x}^*)$, 定义 \mathbf{y}^+ , \mathbf{y}^- 如下:

$$y_{(j)\hat{i}}^+ = \begin{cases} 1, & j = \hat{j}, i = \hat{i}, \\ 0, & \text{其他}, \end{cases} \quad y_{(j)\hat{i}}^- = \begin{cases} -1, & j = \hat{j}, i = \hat{i}, \\ 0, & \text{其他}. \end{cases}$$

则 $\langle \mathbf{x}_{(j)}^*, \mathbf{y}_{(j)}^\pm \rangle = 0$, $\|\mathbf{y}_{(j)}^\pm\| = 1$, 且 $\bar{y}_{(\bar{j})i}^\pm = 0$, $\forall j \neq \bar{j}, i \in [n^j]$ 。根据(6)可得

$$u'(\mathbf{x}_{(j)\hat{i}}^*; \mathbf{y}_{(j)\hat{i}}^\pm) = u'(\mathbf{0}; \mathbf{y}_{(j)\hat{i}}^\pm) = |\mathbf{y}_{(j)\hat{i}}^\pm| = 1, \quad w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}^\pm) = \frac{\langle \mathbf{x}_{(j)}^*, \mathbf{y}_{(j)}^\pm \rangle}{\|\mathbf{x}_{(j)}^*\|} = 0.$$

进而, 由(8)可得

$$\begin{aligned} 0 &\leq \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in [n^j] \setminus I_j(\mathbf{x}^*)} [\nabla f(\mathbf{x}^*)]_{(j)i} \mathbf{y}_{(j)i}^\pm + \lambda_1 \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in [n^j] \setminus I_j(\mathbf{x}^*)} |\mathbf{y}_{(j)i}^\pm| \\ &\quad + \lambda_2 \sum_{j \in J(\mathbf{x}^*)} \left[\mathbf{g}' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(j)}^*) \right] \frac{\langle \mathbf{x}_{(j)}^*, \mathbf{y}_{(j)}^\pm \rangle}{\|\mathbf{x}_{(j)}^*\|} \\ &= \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in [n^j] \setminus I_j(\mathbf{x}^*)} \pm [\nabla f(\mathbf{x}^*)]_{(j)i} + \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in [n^j] \setminus I_j(\mathbf{x}^*)} \lambda_1, \end{aligned}$$

即 $\pm [\nabla f(\mathbf{x}^*)]_{(j)i} \leq \lambda_1$, 从而 $|\pm [\nabla f(\mathbf{x}^*)]_{(j)i}| \leq \lambda_1$ 。

3) 与 2) 的证明类似, 任取 $\bar{j} \in [J] \setminus J(\mathbf{x}^*)$, $\bar{i} \in [n^{\bar{j}}]$, 可知 $\mathbf{x}_{(\bar{j})\bar{i}}^* = \mathbf{0}$, 定义 \mathbf{y}^+ , \mathbf{y}^- 如下:

$$\bar{y}_{(j)\bar{i}}^+ = \begin{cases} 1, & j = \bar{j}, i = \bar{i}, \\ 0, & \text{其他}, \end{cases} \quad \bar{y}_{(j)\bar{i}}^- = \begin{cases} -1, & j = \bar{j}, i = \bar{i}, \\ 0, & \text{其他}, \end{cases}$$

则 $\langle \mathbf{x}_{(\bar{j})}^*, \bar{\mathbf{y}}_{(\bar{j})}^\pm \rangle = 0$, 且 $\bar{y}_{(\bar{j})i}^\pm = 0$, $\forall j \neq \bar{j}, i \in [n^j]$ 。由(8)得

$$\begin{aligned} 0 &\leq \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \sum_{\bar{i} \in [n^{\bar{j}}]} [\nabla f(\mathbf{x}^*)]_{(\bar{j})\bar{i}} \bar{y}_{(\bar{j})\bar{i}}^\pm + \lambda_1 \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \sum_{\bar{i} \in [n^{\bar{j}}]} |\bar{y}_{(\bar{j})\bar{i}}^\pm| \\ &\quad + \lambda_2 \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \left[\mathbf{g}' \circ w(\mathbf{x}_{(\bar{j})}^*) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{x}_{(\bar{j})}^*) \right] \|\bar{\mathbf{y}}_{(\bar{j})}^\pm\| \\ &= \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \sum_{\bar{i} \in [n^{\bar{j}}]} \pm [\nabla f(\mathbf{x}^*)]_{(\bar{j})\bar{i}} + \lambda_1 n^{\bar{j}} (J - |J(\mathbf{x}^*)|) + \lambda_2 \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \left[\mathbf{g}' \circ w(\mathbf{0}) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{0}) \right] \|\bar{\mathbf{y}}_{(\bar{j})}^\pm\| \\ &\leq \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \sum_{\bar{i} \in [n^{\bar{j}}]} \pm [\nabla f(\mathbf{x}^*)]_{(\bar{j})\bar{i}} + \lambda_1 n^{\bar{j}} (J - |J(\mathbf{x}^*)|) + \lambda_2 \sqrt{n} \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \left[\mathbf{g}' \circ w(\mathbf{0}) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{0}) \right], \end{aligned}$$

即

$$\begin{aligned} \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \sum_{\bar{i} \in [n^{\bar{j}}]} \pm [\nabla f(\mathbf{x}^*)]_{(\bar{j})\bar{i}} &\leq \lambda_1 n^{\bar{j}} (J - |J(\mathbf{x}^*)|) + \lambda_2 \sqrt{n} \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \left[\mathbf{g}' \circ w(\mathbf{0}) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{0}) \right], \\ \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \sum_{\bar{i} \in [n^{\bar{j}}]} \left| [\nabla f(\mathbf{x}^*)]_{(\bar{j})\bar{i}} \right| &\leq \lambda_1 n^{\bar{j}} (J - |J(\mathbf{x}^*)|) + \lambda_2 \sqrt{n} \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \left[\mathbf{g}' \circ w(\mathbf{0}) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{0}) \right], \end{aligned}$$

从而

$$\left| \left[\nabla f(\mathbf{x}^*) \right]_{(j)\bar{i}} \right| \leq \lambda_1 n^{\bar{j}} \left(J - |J(\mathbf{x}^*)| \right) + \lambda_2 \sqrt{n} \sum_{\bar{j} \in [J] \setminus J(\mathbf{x}^*)} \left[g' \circ w(\mathbf{0}) - \max_{1 \leq \nu \leq \bar{\nu}} h'_\nu \circ w(\mathbf{0}) \right]. \quad \square$$

3. 二阶 d-稳定点

3.1. 二阶 d-稳定点的局部最优性

首先, 我们介绍二阶 d-稳定点的定义[19] [21], 本文采用沿一个方向的二阶方向导数来定义问题(2)的二阶 d-稳定点, 这样定义的导数简单易理解.

定义 3.1 [19] (二阶方向导数) 设 $Q: \mathbb{R}^n \rightarrow \mathbb{R}$ 是局部 Lipschitz 连续且方向可微的, 对任意 $\mathbf{x}^*, \mathbf{y} \in \mathbb{R}^n$, 若以下极限存在:

$$\lim_{z \rightarrow \mathbf{y}, t \downarrow 0} \frac{Q(\mathbf{x}^* + tz) - Q(\mathbf{x}^*) - tQ'(\mathbf{x}^*; z)}{\frac{1}{2}t^2}, \quad (11)$$

则称 Q 在 \mathbf{x}^* 处沿方向 \mathbf{y} 二阶方向可导, 其极限称为 Q 在 \mathbf{x}^* 处沿方向 \mathbf{y} 二阶方向导数, 记为 $Q^{(2)}(\mathbf{x}^*; \mathbf{y})$. 若 $\forall \mathbf{y} \in \mathbb{R}^n$, 上述极限均存在, 则 Q 在点 \mathbf{x}^* 二阶方向可微.

事实上, 极限(11)存在表明对 $\forall \varepsilon > 0, \exists \delta > 0$, 使得 $\forall \mathbf{x} \in \mathbb{R}^n$, 有

$$\left\{ \begin{array}{l} \|\mathbf{x} - \mathbf{x}^*\| \leq \delta, 0 < t \leq \delta \\ \left\| \frac{\mathbf{x} - \mathbf{x}^*}{t} - \mathbf{y} \right\| \leq \delta \end{array} \right. \Rightarrow \left| \frac{Q(\mathbf{x}) - Q(\mathbf{x}^*) - tQ'(\mathbf{x}^*; \frac{\mathbf{x} - \mathbf{x}^*}{t})}{\frac{1}{2}t^2} - Q^{(2)}(\mathbf{x}^*; \mathbf{y}) \right| \leq \varepsilon$$

成立. 显然, 若极限(11)存在, 则有以下式成立

$$Q^{(2)}(\mathbf{x}^*; \mathbf{y}) = \lim_{t \downarrow 0} \frac{Q(\mathbf{x}^* + t\mathbf{y}) - Q(\mathbf{x}^*) - tQ'(\mathbf{x}^*; \mathbf{y})}{\frac{1}{2}t^2}.$$

假设 Q 在点 \mathbf{x}^* 二阶可微, 则 Hessian 阵 $\nabla^2 Q(\mathbf{x}^*)$ 存在, 且 $\forall \mathbf{y} \in \mathbb{R}^n$, 有

$$Q^{(2)}(\mathbf{x}^*; \mathbf{y}) = \langle \nabla^2 Q(\mathbf{x}^*) \mathbf{y}, \mathbf{y} \rangle.$$

定义 3.2 [19] (二阶 d-稳定点) 设 $Q: \mathbb{R}^n \rightarrow \mathbb{R}$ 在 $\mathbf{x}^* \in \mathbb{R}^n$ 处二阶方向可微, 若 \mathbf{x}^* 是 Q 的一阶 d-稳定点, 且

$$\forall \mathbf{y} \in \mathbb{R}^n, Q'(\mathbf{x}^*; \mathbf{y}) = 0 \Rightarrow Q^{(2)}(\mathbf{x}^*; \mathbf{y}) \geq 0,$$

则称 \mathbf{x}^* 是 Q 的二阶 d-稳定点.

文献[19] [21]给出了如下 F 二阶方向可微时的局部最优性质.

引理 3.1 [19] (二阶 d-稳定点的局部最优性质) 已知 $F: \mathbb{R}^n \rightarrow \mathbb{R}$ 在 $\mathbf{x}^* \in \mathbb{R}^n$ 处二阶方向可微, 则以下结论成立:

- 1) 若 \mathbf{x}^* 是 F 的局部极小点, 则 \mathbf{x}^* 是 F 的二阶 d-稳定点;
- 2) 若 \mathbf{x}^* 是 F 的强局部极小点(即: 存在 \mathbf{x}^* 的邻域 Θ 和标量 $\delta > 0$, 使得 $\forall \mathbf{x} \in \Theta$, 都有 $F(\mathbf{x}) \geq F(\mathbf{x}^*) + \delta \|\mathbf{x} - \mathbf{x}^*\|^2$), 当且仅当 \mathbf{x}^* 是 F 的一阶 d-稳定点, 且 $\forall \mathbf{y} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, 有

$$F'(\mathbf{x}^*; \mathbf{y}) = 0 \Rightarrow F^{(2)}(\mathbf{x}^*; \mathbf{y}) > 0.$$

根据定义 3.1 可计算得出在点 \mathbf{x} 处的 ℓ_1 范数和 ℓ_2 范数的二阶方向导数如下:

$$u^{(2)}(\mathbf{x}; \mathbf{y}) = 0, \quad w^{(2)}(\mathbf{x}; \mathbf{y}) = \begin{cases} 0, & \|\mathbf{x}\|_2 = 0, \\ \frac{(\|\mathbf{x}\| \|\mathbf{y}\|)^2 - |\langle \mathbf{x}, \mathbf{y} \rangle|^2}{\|\mathbf{x}\|^3}, & \|\mathbf{x}\|_2 \neq 0. \end{cases} \quad (12)$$

对于任意 $j \in [J]$, 有

$$w^{(2)}(\mathbf{x}_{(j)}; \mathbf{y}_{(j)}) = \begin{cases} 0, & \|\mathbf{x}_{(j)}\|_2 = 0, \\ \frac{(\|\mathbf{x}_{(j)}\| \|\mathbf{y}_{(j)}\|)^2 - |\langle \mathbf{x}_{(j)}, \mathbf{y}_{(j)} \rangle|^2}{\|\mathbf{x}_{(j)}\|^3}, & \|\mathbf{x}_{(j)}\|_2 \neq 0. \end{cases} \quad (13)$$

为了求出复合函数的二阶方向导数, 文献[22, Proposition 3.2]和[19, Lemma 3.2]给出了如下引理。

引理 3.2 [19] 设 $\rho: \mathbb{R}^m \rightarrow \mathbb{R}$ 是在 $\Phi(\mathbf{x}) \in \mathbb{R}^m$ 上局部 Lipschitz 连续, 且 $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 在 $\mathbf{x} \in \mathbb{R}^n$ 上局部 Lipschitz 连续。若复合函数 $G = \rho \circ \Phi: \mathbb{R}^n \rightarrow \mathbb{R} \in \mathbb{R}^n$ 满足以下条件中的任意一条, 则 G 在点 \mathbf{x} 处二阶方向可微:

- 1) ρ 在 $\Phi(\mathbf{x})$ 处半光滑可微(即: ρ 在 $\Phi(\mathbf{x})$ 附近方向可微, $\nabla \rho$ 在 $\Phi(\mathbf{x})$ 处半光滑), 且 Φ 在点 \mathbf{x} 处二阶方向可微;
- 2) ρ 在 $\Phi(\mathbf{x})$ 处二阶方向可微, 且 Φ 在点 \mathbf{x} 附近分段仿射(线性);
- 3) ρ 在 $\Phi(\mathbf{x})$ 附近分段仿射(线性), 且 Φ 在点 \mathbf{x} 处二阶方向可微。

此外, $\forall \mathbf{y} \in \mathbb{R}^n$, 有

$$\text{当(1)成立, } G^{(2)}(\mathbf{x}; \mathbf{y}) = \Phi'(\mathbf{x}; \mathbf{y})^\top (\nabla \rho)'(\Phi(\mathbf{x}); \Phi'(\mathbf{x}; \mathbf{y})) + \nabla \rho(\Phi(\mathbf{x}))^\top \Phi^{(2)}(\mathbf{x}; \mathbf{y}), \quad (14)$$

$$\text{当(2)成立, } G^{(2)}(\mathbf{x}; \mathbf{y}) = \rho^{(2)}(\Phi(\mathbf{x}); \Phi'(\mathbf{x}; \mathbf{y})), \quad (15)$$

$$\text{当(3)成立, } G^{(2)}(\mathbf{x}; \mathbf{y}) = \rho'(\Phi(\mathbf{x}); \Phi^{(2)}(\mathbf{x}; \mathbf{y})). \quad (16)$$

3.2. 问题(2)的二阶 d-稳定点

为了研究问题(2)的二阶 d-稳定点, 给出如下假设。

假设 3.1 [19] 惩罚函数 $\phi: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ 具有凸差形式, $\phi(t) := g(t) - h(t)$, 其中 g 是 $t \in [0, \infty)$ 上的仿射(线性)函数, 且 $g'(0) := g'(0^+)$, h 是 $t \in (0, \infty)$ 上的凸函数且半光滑可微, $h'(0) := h'(0^+)$ 。

容易验证, 前述的折叠凸惩罚函数中 MCP、SCAD 满足假设 3.1 的条件。

定理 3.1 若假设 3.1 成立, $\forall \mathbf{y} \in \mathbb{R}^n$, 则问题(2)在点 \mathbf{x}^* 处的二阶方向导数有以下表达式:

$$F^{(2)}(\mathbf{x}^*; \mathbf{y}) = \langle \nabla^2 f(\mathbf{x}^*) \mathbf{y}, \mathbf{y} \rangle + \sum_{j=1}^J \left[g'(\|\mathbf{x}_{(j)}^*\|) - h'(\|\mathbf{x}_{(j)}^*\|) \right] w^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) - \sum_{j=1}^J w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) H'(\|\mathbf{x}_{(j)}^*\|; w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)})), \quad (17)$$

其中 $w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)})$ 和 $w^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)})$ 可分别由(6)和(13)得到, 对任意 $t \in [0, \infty)$, $H(t) := h'(t)$ 。

证明. 因为 f 是二阶连续可微的, 因此 $f^{(2)}(\mathbf{x}^*; \mathbf{y}) = \langle \nabla^2 f(\mathbf{x}^*) \mathbf{y}, \mathbf{y} \rangle$ 。由于 ℓ_1 范数的二阶方向导数为零, 式中凸差部分的二阶方向导数可由[19, Lemma 3.5]中的证明得到。□

下面给出问题(2)在假设 3.1 条件下的二阶 d-稳定点的特征。

推论 3.1 若假设 3.1 成立, 设 $\mathbf{x}^* \in \mathbb{R}^n$ 是问题(2)的二阶 \mathbf{d} -稳定点, 则 $\forall \mathbf{y} \in D(\mathbf{x}^*)$, 有

$$\langle \nabla^2 f(\mathbf{x}^*) \mathbf{y}, \mathbf{y} \rangle + \sum_{j \in J(\mathbf{x}^*)} \left[(g \circ w)^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) - (h \circ w)^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) \right] \geq 0, \tag{18}$$

其中 $\forall j \in J(\mathbf{x}^*)$,

$$w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) = \frac{\langle \mathbf{x}_{(j)}^*, \mathbf{y}_{(j)} \rangle}{\|\mathbf{x}_{(j)}^*\|},$$

$$w^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) = \frac{\left(\|\mathbf{x}_{(j)}^*\| \|\mathbf{y}_{(j)}\| \right)^2 - \left| \langle \mathbf{x}_{(j)}^*, \mathbf{y}_{(j)} \rangle \right|^2}{\|\mathbf{x}_{(j)}^*\|^3},$$

$$(g \circ w)^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) = g'(\|\mathbf{x}_{(j)}^*\|) w^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}),$$

$$(h \circ w)^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) = w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) H'(\|\mathbf{x}_{(j)}^*\|; w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)})) + h'(\|\mathbf{x}_{(j)}^*\|) w^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}).$$

证明. 因为 \mathbf{x}^* 是问题(2)的二阶 \mathbf{d} -稳定点, 则 \mathbf{x}^* 一定是问题(2)的一阶 \mathbf{d} -稳定点, 根据定理 2.3 证明中的(9), 对任意的 $\mathbf{y} \in D(\mathbf{x}^*)$, 有

$$0 \leq \sum_{j \in J(\mathbf{x}^*)} \sum_{i \in I_j(\mathbf{x}^*)} \left\{ \left[\nabla f(\mathbf{x}^*) \right]_{(j)i} + \lambda_1 \operatorname{sgn}(x_{(j)i}^*) + \lambda_2 \left[g' \circ w(\mathbf{x}_{(j)}^*) - \max_{1 \leq v \leq \bar{d}} h'_v \circ w(\mathbf{x}_{(j)}^*) \right] \frac{x_{(j)i}^*}{\|\mathbf{x}_{(j)}^*\|} \right\} \cdot y_{(j)i}.$$

从而

$$\begin{aligned} & \left[\nabla f(\mathbf{x}^*) \right]_{(j)} + \lambda_1 \operatorname{sgn}(x_{(j)}^*) + \lambda_2 \left[g' \circ w(\mathbf{x}_{(j)}^*) - h' \circ w(\mathbf{x}_{(j)}^*) \right] \frac{x_{(j)}^*}{\|\mathbf{x}_{(j)}^*\|} = 0, \\ \Rightarrow & \sum_{j \in J(\mathbf{x}^*)} \left\langle \left[\nabla f(\mathbf{x}^*) \right]_{(j)} + \lambda_1 \operatorname{sgn}(x_{(j)}^*) + \lambda_2 \left[g' \circ w(\mathbf{x}_{(j)}^*) - h' \circ w(\mathbf{x}_{(j)}^*) \right] \frac{x_{(j)}^*}{\|\mathbf{x}_{(j)}^*\|}, \mathbf{y}_{(j)} \right\rangle = 0, \\ \Rightarrow & \sum_{j=1}^J \left\langle \left[\nabla f(\mathbf{x}^*) \right]_{(j)} + \lambda_1 \operatorname{sgn}(x_{(j)}^*) + \lambda_2 \left[g' \circ w(\mathbf{x}_{(j)}^*) - h' \circ w(\mathbf{x}_{(j)}^*) \right] \frac{x_{(j)}^*}{\|\mathbf{x}_{(j)}^*\|}, \mathbf{y}_{(j)} \right\rangle = 0, \end{aligned}$$

因此

$$0 = F'(\mathbf{x}^*; \mathbf{y}).$$

因为 \mathbf{x}^* 是问题(2)的二阶 \mathbf{d} -稳定点, 故有 $F^{(2)}(\mathbf{x}^*; \mathbf{y}) \geq 0$, 由定义 3.2 和定理 3.1, 得

$$\begin{aligned} 0 \leq & \langle \nabla^2 f(\mathbf{x}^*) \mathbf{y}, \mathbf{y} \rangle + \sum_{j=1}^J \left[g'(\|\mathbf{x}_{(j)}^*\|) - h'(\|\mathbf{x}_{(j)}^*\|) \right] w^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) \\ & - \sum_{j=1}^J w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) H'(\|\mathbf{x}_{(j)}^*\|; w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)})). \end{aligned}$$

根据(6)和(13), $\forall j \notin J(\mathbf{x}^*)$, 有 $w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) = w^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) = 0$, 进而 $\forall j \in J(\mathbf{x}^*)$, 有

$$w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) = \frac{\langle \mathbf{x}_{(j)}^*, \mathbf{y}_{(j)} \rangle}{\|\mathbf{x}_{(j)}^*\|},$$

$$w^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) = \frac{(\|\mathbf{x}_{(j)}^*\| \|\mathbf{y}_{(j)}\|)^2 - |\langle \mathbf{x}_{(j)}^*, \mathbf{y}_{(j)} \rangle|^2}{\|\mathbf{x}_{(j)}^*\|^3},$$

$$(g \circ w)^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) = g'(\|\mathbf{x}_{(j)}^*\|) w^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}),$$

$$(h \circ w)^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) = w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) H'(\|\mathbf{x}_{(j)}^*\|; w'(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)})) + h'(\|\mathbf{x}_{(j)}^*\|) w^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}).$$

故

$$F^{(2)}(\mathbf{x}^*; \mathbf{y}) = \langle \nabla^2 f(\mathbf{x}^*) \mathbf{y}, \mathbf{y} \rangle + \sum_{j \in J(\mathbf{x}^*)} [(g \circ w)^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)}) - (h \circ w)^{(2)}(\mathbf{x}_{(j)}^*; \mathbf{y}_{(j)})] \geq 0. \quad \square$$

4. 总结

本文研究了带有 ℓ_1 稀疏惩罚和非凸折叠凹组稀疏惩罚的稀疏加组稀疏优化问题(2)，分别对问题(2)的一阶、二阶 \mathbf{d} -稳定点进行了刻画和分析，并分析了问题(2)中 \mathbf{d} -稳定点和局部最优解的关系。为进一步分析和求解此类问题提供了一定理论基础。后续我们的工作将围绕此类问题的其他稳定点的具体刻画以及约束优化问题的稳定点做进一步的分析。

基金项目

国家自然科学基金项目(11861020, 12261020)、贵州省科技计划项目(ZK [2021] 009, [2018] 5781)、贵州省高层次留学人才创新创业择优资助重点项目([2018] 03)和贵州省教育厅青年科技人才成长项目([2018] 121)。

参考文献

- [1] Bech, A. and Hallak, N. (2019) Optimization Problems Involving Group Sparsity Terms. *Mathematical Programming*, **178**, 39-67. <https://doi.org/10.1007/s10107-018-1277-1>
- [2] Hu, Y., Li, C. and Meng, K. (2017) Group Sparse Optimization via $L_{p,q}$ Regularization. *Journal of Machine Learning Research*, **18**, 1-52.
- [3] Li, W., Bian, W. and Toh, K.-C. (2022) Difference-of-Convex Algorithms for a Class of Sparse Group L_0 Regularized Optimization Problems. *SIAM Journal on Optimization*, **32**, 1614-1641. <https://doi.org/10.1137/21M1443455>
- [4] Yuan, M. and Lin, Y. (2006) Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **68**, 49-67. <https://doi.org/10.1111/j.1467-9868.2005.00532.x>
- [5] Bian, W. and Chen, X. (2017) Optimality and Complexity for Constrained Optimization Problems with Nonconvex Regularization. *Mathematics of Operations Research*, **42**, 1063-1084. <https://doi.org/10.1287/moor.2016.0837>
- [6] Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- [7] Donoho, D.L. (2006) Compressed Sensing. *IEEE Transactions on Information Theory*, **52**, 1289-1306. <https://doi.org/10.1109/TIT.2006.871582>
- [8] Fan, J. and Li, R. (2006) Statistical Challenges with High Dimensionality: Feature Selection in Knowledge Discovery. *Proceedings of the International Congress of Mathematicians*, **3**, 595-622. <https://doi.org/10.4171/022-3/31>
- [9] Zhang, C.H. (2010) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *The Annals of Statistics*, **38**, 894-942. <https://doi.org/10.1214/09-AOS729>

-
- [10] Ong, C.S. and An, L.T.H. (2013) Learning Sparse Classifiers with Difference of Convex Functions Algorithms. *Optimization Methods and Software*, **28**, 830-854. <https://doi.org/10.1080/10556788.2011.652630>
- [11] Bian, W. and Chen, X. (2020) A Smoothing Proximal Gradient Algorithm for Nonsmooth Convex Regression with Cardinality Penalty. *SIAM Journal on Numerical Analysis*, **58**, 858-883. <https://doi.org/10.1137/18M1186009>
- [12] Chen, X., Pan, L.L. and Xiu, N. (2022) Solution Sets of Three Sparse Optimization Problems for Multivariate Regression. *Journal of Global Optimization*, **87**, 347-371. <https://doi.org/10.1007/s10898-021-01124-w>
- [13] Pan, L. and Chen, X. (2021) Group Sparse Optimization for Images Recovery Using Capped Folded Concave Functions. *SIAM Journal on Imaging Sciences*, **14**, 1-25. <https://doi.org/10.1137/19M1304799>
- [14] Soubies, E., Blanc-Féraud, L. and Aubert, G. (2017) A Unified View of Exact Continuous Penalties for L_2 - L_0 Minimization. *SIAM Journal on Optimization*, **27**, 2034-2060. <https://doi.org/10.1137/16M1059333>
- [15] Zhang, Y., Zhang, N. and Sun, D. (2020) An Efficient Hessian Based Algorithm for Solving Large-Scale Sparse Group Lasso Problems. *Mathematical Programming*, **179**, 223-263. <https://doi.org/10.1007/s10107-018-1329-6>
- [16] Ahn, M., Pang, J.-S. and Xin, J. (2017) Difference-of-Convex Learning: Directional Stationarity, Optimality, and Sparsity. *SIAM Journal on Optimization*, **27**, 1637-1665. <https://doi.org/10.1137/16M1084754>
- [17] Pang, J.S., Razaviyayn, M. and Alvarado, A. (2017) Computing B-Stationary Points of Nonsmooth DC Programs. *Mathematics of Operations Research*, **42**, 95-118. <https://doi.org/10.1287/moor.2016.0795>
- [18] Cui, Y., Pang, J. and Sen, B. (2018) Composite Difference-Max Programs for Modern Statistical Estimation Problems. *SIAM Journal on Optimization*, **28**, 3344-3374. <https://doi.org/10.1137/18M117337X>
- [19] Peng, D. and Chen, X. (2020) Computation of Second-Order Directional Stationary Points for Group Sparse Optimization. *Optimization Methods and Software*, **35**, 348-376. <https://doi.org/10.1080/10556788.2019.1684492>
- [20] 苏妍妍, 彭定涛. 稀疏加组稀疏优化问题的方向稳定点及其光滑化方法[J]. 应用数学进展, 2022, 11(10): 7464-7477.
- [21] Rockafellar, R.T. and Wets, R.J.B. (2009) Variational Analysis. Springer Science & Business Media, Berlin.
- [22] Chang, T.H., Hong, M. and Pang, J.-S. (2017) Local Minimizers and Second-Order Conditions in Composite Piecewise Programming via Directional Derivatives.