

一类带MCP函数正则化问题的稳定性研究

赵立斌¹, 李明华²

¹重庆交通大学数学与统计学院, 重庆

²重庆文理学院数学与大数据学院, 重庆

收稿日期: 2024年3月2日; 录用日期: 2024年3月26日; 发布日期: 2024年4月30日

摘要

本文研究一类带极大极小凹惩罚(MCP)函数正则化问题, 该模型由二次可微的损失函数和MCP函数组成。首先我们研究了MCP函数的邻近次微分和极限次微分, 然后利用该次微分表达式和图像导数的运算法则, 得到带MCP正则化问题目标函数次微分图像导数。最后, 利用该图像导数表达式分别建立了该正则化问题次微分强度量次正则的充分条件和充要条件。

关键词

MCP函数, 次微分, 图像导数, 强度量次正则

Study on the Strong Metric Subregularity for a Class of Regularization Problems with the MCP Penalty Functions

Libin Zhao¹, Minghua Li²

¹School of Mathematics and Statistics, Chongqing Jiaotong University, Chongqing

²School of Mathematics and Big Data, Chongqing University of Arts and Sciences, Chongqing

Received: Mar. 2nd, 2024; accepted: Mar. 26th, 2024; published: Apr. 30, 2024

Abstract

In this paper, we study a class of regularization problems with the minimax concave penalty (MCP) function, which consists of a twice differentiable loss function and the MCP penalty function. First, we study the prox-regular subdifferential and limiting subdifferential of the MCP penalty function. Next, we obtain the graphical derivative of regularization problems with the MCP penalty function by using the subdifferential expression. Finally, we use the graphical derivative to establish a suf-

efficient condition, a sufficient and necessary condition for the strong metric subregularity of sub-differential of the regularization problems, respectively.

Keywords

MCP Function, Subdifferential, Graphical Derivative, Strong Metric Subregularity

Copyright © 2024 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

考虑如下无约束优化问题:

$$\min_{x \in \mathbb{R}^n} F(x) := f(x) + g(x), \quad (1)$$

其中 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是二次可微函数, $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 是一个可分离变量的非凸函数。

MCP 函数是一种重要的可分离变量的函数, 它有两个标量 $a > 2$ 和 $\lambda > 0$, 并以原点为唯一的零点:

$$g(x) = \sum_{i=1}^n g_{a,\lambda}(x_i) \quad (2)$$

其中

$$g_{a,\lambda}(x_i) := \begin{cases} 2\lambda|x_i| - \frac{x_i^2}{a}, & |x_i| \leq a\lambda, \\ a\lambda^2, & |x_i| \geq a\lambda. \end{cases} \quad (3)$$

在机器学习和统计学中, 为了防止模型过拟合和提高模型的泛化能力, 需要通过引入凹惩罚函数来对模型进行正则化。变量选择在高维数据分析中起着非常重要的作用, 在对模型正则化过程中可以通过对参数的惩罚来实现变量选择的效果, 从而在一定程度上实现变量选择的功能, 然而许多方法容易忽略在变量选择过程中的随机误差。2001年, Fan 和 Li [1]为了解决这个问题, 提出了一种新的非凸惩罚函数, 即光滑裁剪绝对偏差(smoothly clipped absolute deviation, SCAD)罚函数。伴随着 SCAD 罚函数的出现, oracle 性质成为了评价变量选择方法的标准。2010年, Zhang [2]提出了新的非凸 MCP 函数来进行变量选择, 并证明了 MCP 函数具有 oracle 性质。

在正则化模型中, 通过引入 MCP 函数可以对模型的复杂度进行控制, 防止模型过拟合, 提高模型的泛化能力。在高维数据的情况下, MCP 函数可以帮助减少模型的自由度, 提高模型的稳定性和可解释性。MCP 函数的参数 λ 和 a 可以根据具体问题进行调节, 以满足不同的需求, 这使得 MCP 函数可以适用于各种不同的优化问题。

目前为止, 由于 MCP 函数良好的 oracle 性质和它可以同时实现变量选择与参数估计, 带 MCP 函数正则化问题被广泛应用于机器学习、统计学习、信号处理、图像处理等领域, 详见参考文献[3] [4] [5] [6] [7]。为了计算带 MCP 函数正则化问题, Zhang [2]采用了惩罚线性无偏选择(PLUS)算法, 在此之后 Shi 等[8]开发了一种有效的交替方向乘子法(ADMM)求解带 MCP 函数的正则化问题。

为了研究模型(1)相关算法的收敛速度, 我们需要研究该模型的稳定性, 如强度量次正和二阶增长条件, 他们已经被用来建立优化算法的收敛速度, 详见文献[9] [10] [11]。为了刻画强度量次正则[12] [13] [14]

和二阶增长条件[13] [15], 许多学者研究了图像导数, 它是变分分析中的一类重要工具。Chieu 等人[13] 利用次微分的图形导数研究了有限维中二阶增长条件和次微分的强度量次正则之间的关系, Dontchev 等人[12]给出了集值映射强度量次正则的图像导数准则, Bello-Cruz 等人[15]利用图像导数刻画了 ℓ_1 -正则化优化问题的二阶增长条件。到目前为止, 大多数文献都是对带 MCP 函数正则化问题的应用与相关算法的研究, 还没有文献对模型(1)次微分的强度量次正则进行研究, 而强度量次正则模型在模型相关算法的收敛性分析中起着非常重要的作用。

受文献[12] [13] [15]的启发, 本文研究了模型(1)次微分的强度量次正则。本文的结构如下: 第二节, 我们介绍了一些基本的定义和引理, 并利用次微分的定义得到了 MCP 函数次微分的表达式。第三节, 利用次微分的运算法则和集值映射的图像导数工具得到了 MCP 函数次微分的图像导数表达式, 最后利用该表达式分别建立了模型(1)次微分强度量次正则的充分条件和充要条件。

2. 预备知识

本文中, \mathbb{R} 表示实数集, \mathbb{R}^n 是 n 维欧氏空间, $\|\cdot\|$ 和 $\langle \cdot, \cdot \rangle$ 分别表示 \mathbb{R}^n 中的范数和内积。 $\text{sign}: \mathbb{R} \rightarrow \mathbb{R}$ 表示符号函数。函数 $h: \mathbb{R}^n \rightarrow \mathbb{R}$ 的定义域记作 $\text{dom } h := \{x \in \mathbb{R}^n : h(x) \neq \emptyset\}$ 。 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是一个二次连续可微函数, $\nabla^2 f$ 表示函数 f 的 Hessian 矩阵, $(\nabla^2 f(\cdot))_{i,j}$ 表示 Hessian 矩阵的第 i 行第 j 列元素, $(\nabla^2 f(\cdot))_{i,\cdot}$ 表示 Hessian 矩阵的第 i 行所有元素。对于矩阵 $A \in \mathbb{R}^{m \times n}$, 定义 $\ker A := \{w \in \mathbb{R}^n \mid Aw = 0\}$ 。对集值映射 $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, 定义 F 的图为 $\text{gph} F = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid y \in F(x)\}$; 定义集值映射 F 的逆映射 $F^{-1}: \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ 为: $F^{-1}(y) = \{x \mid y \in F(x)\}$ 。

2.1. 基础知识

定义 1 [16] 设 $x \in \mathbb{R}^n$ 。对于函数 $h: \mathbb{R}^n \rightarrow \mathbb{R}$ 和向量 $v \in \mathbb{R}^n$, 如果存在 $\rho > 0$ 和 x 的一个邻域 V , 使得

$$h(y) \geq h(x) + \langle v, y - x \rangle - \rho \|y - x\|^2 \quad \forall y \in V,$$

则称 v 是函数 h 在 x 处的邻近次梯度。所有这样的 v 构成的集合称为 h 在 x 处的邻近次微分, 记作 $\partial_\rho h(x)$ 。

定义 2 [16] 函数 $h: \mathbb{R}^n \rightarrow \mathbb{R}$ 在 $x \in \mathbb{R}^n$ 处的极限次微分定义如下:

$$\partial h(x) := \{v \in \mathbb{R}^n : x_k \xrightarrow{h} x, v_k \rightarrow v, v_k \in \partial_\rho h(x_k)\},$$

其中 $x_k \xrightarrow{h} x$ 意味着 $x_k \rightarrow x$, $h(x_k) \rightarrow h(x)$ 。显然 $\partial_\rho h(x) \subseteq \partial h(x)$ 。

在本文中我们使用的都是极限次微分, 简称为次微分。

定义 3 [12] 设 Ω 是 \mathbb{R}^n 中的非空子集, $\bar{x} \in \Omega$ 。集合 Ω 在 \bar{x} 处的切锥定义为:

$$T_\Omega(\bar{x}) := \{v \in \mathbb{R}^n \mid \exists t_k \downarrow 0, v_k \rightarrow v \text{ 满足 } \bar{x} + t_k v_k \in \Omega, \forall k \in \mathbb{N}\}.$$

定义 4 [12] 对于集值映射 $H: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 和点 $(x, y) \in \text{gph} H$ 。 H 在 (x, y) 处的图像导数是集值映射 $DH(x|y): \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, 其图像是 $\text{gph} H$ 在 (x, y) 处的切锥 $T_{\text{gph} H}$:

$$v \in DH(x|y)(u) \Leftrightarrow (u, v) \in T_{\text{gph} H}(x, y).$$

或者说, $v \in DH(x|y)(u)$ 当且仅当存在序列 $u_k \rightarrow u$, $v_k \rightarrow v$ 和 $t_k \rightarrow 0$ 使得对于所有 k ,

$$y + t_k v_k \in H(x + t_k u_k).$$

定义 5 [12] 映射 $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 。对于点 $(\bar{x}, \bar{y}) \in \text{gph} F$, 如果存在 $k > 0$ 和 \bar{x} 的一个邻域 V , 使得

$$d(x; F^{-1}(\bar{y})) \leq kd(\bar{y}; F(x) \cap V) \quad \forall x \in V,$$

则称映射 F 在 \bar{x} 处对于 \bar{y} 是度量次正则的。如果 \bar{x} 是 $H^{-1}(\bar{y})$ 的孤立点, 则称 F 在 \bar{x} 处对于 \bar{y} 是强度度量次正则的。

引理 1 [12] 集值映射 $H: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 在 (\bar{x}, \bar{y}) 是局部闭的, H 在 (\bar{x}, \bar{y}) 是强度度量次正则的当且仅当

$$DH(\bar{x} | \bar{y})^{-1}(0) = \{0\}. \quad (4)$$

引理 2 [12] 对于函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 在 x 处可微, 集值映射 $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 和任意的 $y \in F(x)$, 有

$$D(f + F)(x | f(x) + y) = \nabla f(x) + DF(x | y).$$

2.2. MCP 函数的次微分表达式

命题 1 MCP 函数的次微分表达式:

$$\partial g(x) = \left\{ s \in \mathbb{R}^n \left| \begin{array}{ll} s_j \in [-2\lambda, 2\lambda], & x_j = 0 \\ s_j = 2\lambda \text{sign}(x_j) - \frac{2x_j}{a}, & 0 < |x_j| \leq a\lambda \\ s_j = 0, & |x_j| \geq a\lambda \end{array} \right. \right\}. \quad (5)$$

证明 因为函数 $g(x)$ 是连续的可分离变量的函数, 因此对于任意的 $\bar{x} \in \text{dom } g$, 有

$$\partial g(\bar{x}) = \partial g_{a,\lambda}(\bar{x}_1) \times \cdots \times \partial g_{a,\lambda}(\bar{x}_n).$$

因为单变量函数 $g_{a,\lambda}(x_j)$ 除在原点外是连续可微的, 因此, 当 $x_j \neq 0$ 时,

$$\nabla g_{a,\lambda}(x_j) = \left\{ s_j \in \mathbb{R} \left| \begin{array}{ll} s_j = 2\lambda \text{sign}(x_j) - \frac{2x_j}{a}, & 0 < |x_j| \leq a\lambda \\ s_j = 0, & |x_j| \geq a\lambda \end{array} \right. \right\}.$$

当 $x_j = 0$ 时, 我们需要先求得 $x_j = 0$ 时的邻近次微分。设 $y \in (-a\lambda, a\lambda)$, 则 $g_{a,\lambda}(y) = 2\lambda|y| - \frac{y^2}{a}$ 。设 $\rho \geq \frac{1}{a}$, 则

$$2\lambda|y| - \frac{y^2}{a} + \rho y^2 \geq \langle s_j, y \rangle = s_j y.$$

如果 $y > 0$, 我们有 $s_j \leq 2\lambda + \rho y - \frac{y}{a}$ 。因此 $s_j \leq 2\lambda$, 即 $\forall y \in (0, a\lambda)$ 和 $\rho \geq \frac{1}{a}$

$$g_{a,\lambda}(y) \geq g_{a,\lambda}(0) + \langle s_j, y \rangle - \rho y^2.$$

如果 $y = 0$, 那么对于任意的 $s_j \in \mathbb{R}^n$ 都成立。

如果 $y < 0$, 我们有 $s_j \geq -2\lambda + \rho y - \frac{y}{a}$ 。因此 $s_j \geq 2\lambda$, 即 $\forall y \in (-a\lambda, 0)$ 和 $\rho \geq \frac{1}{a}$

$$g_{a,\lambda}(y) \geq g_{a,\lambda}(0) + \langle s_j, y \rangle - \rho y^2.$$

综合上述三种情况, 我们可以得到当 $x_j = 0$ 时, $\partial_\rho g_{a,\lambda}(x_j) = \{s_j \in \mathbb{R} | s_j \in [-2\lambda, 2\lambda]\}$ 。再根据极限次微分的定义, 我们可以得到 $\partial g_{a,\lambda}(x_j) = \{s_j \in \mathbb{R} | s_j \in [-2\lambda, 2\lambda], x_j = 0\}$ 。因此, MCP 函数的次微分表达式为:

$$\partial g(x) = \left\{ s \in \mathbb{R}^n \begin{cases} s_j \in [-2\lambda, 2\lambda], & x_j = 0 \\ s_j = 2\lambda \text{sign}(x_j) - \frac{2x_j}{a}, & 0 < |x_j| \leq a\lambda \\ s_j = 0, & |x_j| \geq a\lambda \end{cases} \right\}.$$

3. 带 MCP 函数正则化问题次微分的强度量次正则

在本节中, 我们研究问题(1)次微分的强度量次正则性。为了使用引理 1 来刻画 ∂F 的强度量次正则性, 我们需要先计算 ∂g 的图像导数。

3.1. ∂g 的图像导数

为了研究 ∂g 的图像导数, 设 $x^* \in \text{dom } g$, $\bar{s} \in \partial g(x^*)$, 定义:

$$I := \{j \in \{1, \dots, n\} \mid x_j^* = 0\},$$

$$J := \{j \in I \mid |\bar{s}_j| < 2\lambda\},$$

$$K := \{j \in I \mid |\bar{s}_j| = 2\lambda\},$$

$$M := \{j \in \{1, \dots, n\} \mid 0 < |x_j^*| < a\lambda\},$$

$$O := \{j \in \{1, \dots, n\} \mid x_j^* = a\lambda\},$$

$$P := \{j \in \{1, \dots, n\} \mid x_j^* = -a\lambda\},$$

$$H(x^*) := \{u \in \mathbb{R}^n \mid \{u_j = 0, j \in J \text{ 或 } u_j \bar{s}_j \geq 0, j \in K\}\}$$

命题 2 $D(\partial g)(x^* | \bar{s})(u)$ 是非空的当且仅当 $u \in H(x^*)$ 。此外, 对任意的 $u \in H(x^*)$ 有

$$D(\partial g)(x^* | \bar{s})(u) = \left\{ v \in \mathbb{R}^n \begin{cases} \bar{s}_j v_j \leq 0, u_j = 0 \text{ 或 } v_j = \frac{-2u_j}{a}, & j \in K \\ v_j = -\frac{2u_j}{a}, & j \in M \\ v_j = 0, & j \in N \\ u_j \leq 0, v_j = -\frac{2u_j}{a} \text{ 或 } u_j \geq 0, v_j = 0, & j \in O \\ u_j \leq 0, v_j = 0 \text{ 或 } u_j \geq 0, v_j = -\frac{2u_j}{a}, & j \in P \end{cases} \right\}. \quad (6)$$

证明 取 $v \in D(\partial g)(x^* | \bar{s})(u)$, 根据图像导数的定义, 存在序列 $t^k \downarrow 0$ 和 $(u^k, v^k) \rightarrow (u, v)$ 使得 $\bar{s} + t^k v^k \in \partial g(x^* + t^k u^k)$ 。下面考虑 j 的六种情形。

情形 1.1 $j \in J$, 即 $x_j^* = 0$ 且 $|\bar{s}_j| < 2\lambda$ 。

若 $\bar{s}_j = 0$ 且 $u_j^k \neq 0$, 则 $(x^* + t^k u^k)_j \neq 0$, $(\bar{s} + t^k v^k)_j = 2\lambda \text{sign}(x^* + t^k u^k)_j - \frac{2(x^* + t^k u^k)_j}{a}$ 。此时存在足够大的常数 L 和 $\varepsilon > 0$, 使得当 $k > L$ 时, $(\bar{s} + t^k v^k)_j \cap B_\varepsilon(2\lambda) = \emptyset$, 因此这种情况是不满足的。若 $\bar{s}_j = 0$ 且 $u_j^k = 0$, 则 $(x^* + t^k u^k)_j = 0$, $(\bar{s} + t^k v^k)_j \in [-a\lambda, a\lambda]$ 。当 $k \rightarrow \infty$, $u_j = 0$, $v_j \in \mathbb{R}$ 。

若 $\bar{s}_j \neq 0$ 且 $u_j^k \neq 0$, 则 $(x^* + t^k u^k)_j \neq 0$, $(\bar{s} + t^k v^k)_j = 2\lambda \text{sign}(x^* + t^k u^k)_j - \frac{2(x^* + t^k u^k)_j}{a}$ 。当 $k \rightarrow \infty$ 时,

$(\bar{s} + t^k v^k)_j \in [-a\lambda, a\lambda] \neq 2\lambda \text{sign}(x^* + t^k u^k)_j - \frac{2(x^* + t^k u^k)_j}{a}$, 因此这种情况是不满足。若 $\bar{s}_j = 0$ 且 $u_j^k = 0$, 则 $(x^* + t^k u^k)_j = 0$, $(\bar{s} + t^k v^k)_j \in [-a\lambda, a\lambda]$ 。当 $k \rightarrow \infty$, $u_j = 0$, $v_j \in \mathbb{R}$ 。

情形 1.2 $j \in K$ 即 $x_j^* = 0$ 且 $|\bar{s}_j| = 2\lambda$ 。

如果有一个序列 $(x^*, \bar{s})_j + t^k (u^k, v^k)_j$, 使得 $\left| (\bar{s} + t^k v^k)_j \right| < 2\lambda = |\bar{s}_j|$, 我们可以得到 $\bar{s}_j v_j^k < 0$, 当 $k \rightarrow \infty$, 有 $\bar{s}_j v_j \leq 0$ 。对于 $u_j \neq 0$, 有 $\left| (x^* + t^k u^k)_j \right| > 0$ 。因为 $x_j^* = 0$, 当 $k \rightarrow \infty$, 有

$$\text{sign}(x_j^* + t^k u_j^k) = \text{sign}(u_j^k) = \text{sign}(\bar{s}_j + t^k v_j^k) = \text{sign}(\bar{s}_j),$$

即 $\bar{s}_j u_j \geq 0$ 。

若 $u_j > 0$, 有 $(x^* + t^k u^k)_j > 0$, 则 $(\bar{s} + t^k v^k)_j = 2\lambda - \frac{2t^k u_j^k}{a}$, 此时, $v_j^k = -\frac{2u_j^k}{a}$, 这意味着 $v_j = -\frac{2u_j}{a}$ 。

若 $u_j < 0$, 有 $(x^* + t^k u^k)_j < 0$, 则 $(\bar{s} + t^k v^k)_j = -2\lambda - \frac{2t^k u_j^k}{a}$, 此时, $v_j^k = -\frac{2u_j^k}{a}$, 这意味着 $v_j = -\frac{2u_j}{a}$ 。

因此, 当 $u_j \neq 0$ 时, 有 $\bar{s}_j v_j \leq 0$, $\bar{s}_j u_j \geq 0$ 和 $v_j = -\frac{2u_j}{a}$ 。

对于 $u_j = 0$, 我们可以分为三类情况讨论:

1) $(x^* + t^k u^k)_j > 0$, 此时 $v_j = -\frac{2u_j}{a} = 0$ 。

2) $(x^* + t^k u^k)_j < 0$, 此时 $v_j = -\frac{2u_j}{a} = 0$ 。

3) $(x^* + t^k u^k)_j = 0$, 此时 $\left| (\bar{s} + t^k v^k)_j \right| \leq 2\lambda$ 。那么有 $\bar{s}_j v_j^k \leq 0$, 当 $k \rightarrow \infty$, 有 $\bar{s}_j v_j \leq 0$ 。

因此, 当 $u_j = 0$ 时, 有 $\bar{s}_j v_j \leq 0$ 和 $\bar{s}_j u_j = 0$ 。

情形 1.3 $j \in M$, 即 $0 < |x_j^*| < a\lambda$ 。

此时 $|\bar{s}_j| = 2\lambda \text{sign}(x_j^*) - \frac{2x_j^*}{a}$ 。对于足够大的 k , $0 < \left| (x^* + t^k u^k)_j \right| < a\lambda$, 由(5)我们可以得到

$$(\bar{s} + t^k v^k)_j = 2\lambda \text{sign}(x^* + t^k u^k)_j - \frac{2(x^* + t^k u^k)_j}{a}.$$

当 $0 < x_j^* < a\lambda$ 时, 有 $v_j^k = -\frac{2u_j^k}{a}$; 当 $-a\lambda < x_j^* < 0$ 时, 有 $v_j^k = -\frac{2u_j^k}{a}$ 。这意味着 $v_j = -\frac{2u_j}{a}$ 。

因此对任意的 $j \in M$, $v_j = -\frac{2u_j}{a}$ 。

情形 1.4 $j \in N$, 即 $|x_j^*| > a\lambda$ 。

对于足够大的 k , $(x^* + t^k u^k)_j > a\lambda$, 由(5)知此时

$$|\bar{s}_j| = 0 = \left| (\bar{s} + t^k v^k)_j \right|.$$

因此 $v_j^k = 0$, 这意味着 $v_j = 0$ 。

情形 1.5 $j \in O$, 即 $x_j^* = a\lambda$ 。

由(5)知此时 $\bar{s}_j = 2\lambda - \frac{2a\lambda}{a} = 0$ 。若 $u_j^k \leq 0$, 此时 $(x^* + t^k u^k)_j \in (0, a\lambda]$, 由(5)可得 $(\bar{s} + t^k v^k)_j = -\frac{2(t^k u^k)_j}{a}$,

因此 $v_j^k = -\frac{2u_j^k}{a}$, 这意味着 $v_j = -\frac{2u_j}{a}$; 若 $u_j^k \geq 0$, 此时 $(x^* + t^k u^k)_j \in [a\lambda, \infty)$, 由(5)可得 $(\bar{s} + t^k v^k)_j = 0$, 因此 $v_j^k = 0$, 这意味着 $v_j = 0$ 。

情形 1.6 $j \in P$, 即 $x_j^* = -a\lambda$ 。

由(5)式知此时 $\bar{s}_j = -2\lambda - \frac{-2a\lambda}{a} = 0$ 。若 $u_j^k \geq 0$, 此时 $(x^* + t^k u^k)_j \in [-a\lambda, 0)$, 由(5)可得

$(\bar{s} + t^k v^k)_j = -\frac{2(t^k u^k)_j}{a}$, 因此 $v_j^k = -\frac{2u_j^k}{a}$, 这意味着 $v_j = -\frac{2u_j}{a}$; 若 $u_j^k \leq 0$, 此时 $(x^* + t^k u^k)_j \in (\infty, -a\lambda]$, 由(5)可得 $(\bar{s} + t^k v^k)_j = 0$, 因此 $v_j^k = 0$, 这意味着 $v_j = 0$ 。

综合以上六个情形, 我们得出 $u \in H(x^*)$, 同时也验证了(6)中的包含“ \subseteq ”。为了证明逆向包“ \supseteq ”, 取 $u \in H(x^*)$ 和任意的 v 属于(6)的右半部分, 对于任意的 $t^k \downarrow 0$ 。我们证明了 $\bar{s} + t^k v \in \partial g(x^* + t^k u)$, 从而验证了 $v \in D(\partial g)(x^* | \bar{s})(u)$ 。对于任意的 $t \in \mathbb{R}$, 我们定义集值映射:

$$\partial g_{a,\lambda}(t) = \begin{cases} 0, & |t| \geq a\lambda \\ 2\lambda \text{sign}(t) - \frac{2t}{a}, & 0 < |t| \leq a\lambda. \\ [-\lambda, \lambda], & t = 0 \end{cases}$$

与证明“ \subseteq ”一样, 我们考虑 j 的六种情形。

情形 2.1 $j \in J$, 即 $x_j^* = 0$ 且 $|\bar{s}_j| < 2\lambda$ 。

因为 $u \in H(x^*)$, 所以有 $(x^* + t^k u)_j = 0$, 当 k 足够大时 $(\bar{s} + t^k v)_j \in [-2\lambda, 2\lambda]$ 。因此

$$(\bar{s} + t^k v)_j \in \partial g_{a,\lambda}(x^* + t^k u)_j.$$

情形 2.2 $j \in K$, 即 $x_j^* = 0$ 且 $|\bar{s}_j| = 2\lambda$ 。

若 $u_j = 0$, 我们有 $(x^* + t^k u)_j = 0$; 并且对于足够大的 k , 因为 $\bar{s}_j v_j \leq 0$, 所以 $|\bar{s} + t^k v)_j| \leq |\bar{s}_j| = 2\lambda$ 。

若 $u_j \neq 0$, 我们有 $v_j = -\frac{2u_j}{a}$ 。当 k 足够大时, 因为 $\bar{s}_j u_j \geq 0$, 所以有

$$\begin{aligned} (\bar{s} + t^k v)_j &= 2\lambda - \frac{2u_j}{a} \in [-2\lambda, 2\lambda], u_j > 0 \\ \text{或} (\bar{s} + t^k v)_j &= -2\lambda - \frac{2u_j}{a} \in [-2\lambda, 2\lambda], u_j < 0. \end{aligned}$$

综合上述两种情况, 我们有 $(\bar{s} + t^k v)_j \in \partial g_{a,\lambda}(x^* + t^k u)_j$ 。

情形 2.3 $j \in M$, 即 $0 < |x_j^*| < a\lambda$ 。

如果 $0 < x_j^* < a\lambda$, 由(5)知 $\bar{s}_j = 2\lambda - \frac{2x_j^*}{a}$ 。对于足够大的 k , 我们有 $(x^* + t^k u)_j \in (0, a\lambda)$, 因为 $v_j = -\frac{2u_j}{a}$, 所以

$$(\bar{s} + t^k v)_j = 2\lambda - \frac{2(x^* + t^k u)_j}{a} \in \partial g_{a,\lambda}(x^* + t^k u)_j.$$

如果 $-a\lambda < x_j^* < 0$, 由(5)知 $\bar{s}_j = -2\lambda - \frac{2x_j^*}{a}$ 。对于足够大的 k , 我们有 $(x^* + t^k u)_j \in (-a\lambda, 0)$, 因为 $v_j = -\frac{2u_j}{a}$, 所以

$$(\bar{s} + t^k v)_j = -2\lambda - \frac{2(x^* + t^k u)_j}{a} \in \partial g_{a,\lambda}(x^* + t^k u)_j.$$

情形 2.4 $j \in N$, 即 $|x_j^*| > a\lambda$ 。

对于足够大的 k , 我们有 $(x^* + t^k u)_j \in (a\lambda, \infty)$ 。因为 $v_j = 0$, 所以

$$\bar{s}_j = (\bar{s} + t^k v)_j = 0 \in \partial g_{a,\lambda}(x^* + t^k u)_j.$$

情形 2.5 $j \in O$, 即 $x_j^* = a\lambda$ 。

由(4)式知此时 $\bar{s}_j = 2\lambda - \frac{2a\lambda}{a} = 0$ 。若 $u_j \leq 0$, 当 k 足够大时, $0 < (x^* + t^k u)_j \leq a\lambda$, $v_j = -\frac{2u_j}{a}$, 因此

$$(\bar{s} + t^k v)_j = 0 - \frac{2t^k u_j}{a} = 2\lambda - \frac{2(a\lambda + t^k u_j)}{a}.$$

若 $u_j \geq 0$, 当 k 足够大时, $(x^* + t^k u)_j \geq a\lambda$, 此时 $v_j = 0$, 因此

$$(\bar{s} + t^k v)_j = \bar{s}_j = 0.$$

综合上述两种情况, 我们有 $(\bar{s} + t^k v)_j \in \partial g_{a,\lambda}(x^* + t^k u)_j$ 。

情形 2.6 $j \in P$, 即 $x_j^* = -a\lambda$ 。

由(5)式知此时 $\bar{s}_j = -2\lambda - \frac{2(-a\lambda)}{a} = 0$ 。若 $u_j \geq 0$, 当 k 足够大时, $-a\lambda \leq (x^* + t^k u)_j < 0$, $v_j = -\frac{2u_j}{a}$, 因此

$$(\bar{s} + t^k v)_j = 0 - \frac{2t^k u_j}{a} = -2\lambda - \frac{2(-a\lambda + t^k u_j)}{a}.$$

若 $u_j \leq 0$, 当 k 足够大时, $(x^* + t^k u)_j \leq a\lambda$, 此时 $v_j = 0$, 因此

$$(\bar{s} + t^k v)_j = \bar{s}_j = 0.$$

综合上述两种情况, 我们有 $(\bar{s} + t^k v)_j \in \partial g_{a,\lambda}(x^* + t^k u)_j$ 。

综上所述, 当 $j \in J \cup K \cup M \cup N \cup O \cup P$ 时, 总有 $(\bar{s} + t^k v)_j \in \partial g(x^* + t^k u)_j$, 故

$$v \in D(\partial g)(x^* | \bar{s})(u).$$

3.2. ∂F 的强度量次正则

为了研究 ∂F 的强度量次正则, 设 $x^* \in \text{dom}F$, $0 \in \partial F(x^*)$, f 在 x^* 处是二次可微的。定义:

$$K_1 := \left\{ j \in \{1, \dots, n\} \mid x_j^* = 0, \left| (\nabla f(x^*))_j \right| = 2\lambda \right\},$$

$$M_1 := \{ j \in \{1, \dots, n\} \mid 0 < |x_j^*| < a\lambda \},$$

$$N_1 := \{ j \in \{1, \dots, n\} \mid |x_j^*| > a\lambda \},$$

$$P_1 := \{ j \in \{1, \dots, n\} \mid |x_j^*| = a\lambda \},$$

$$U := \left\{ u \in R^E \mid u_j (\nabla f(x^*))_j \leq 0, j \in K_1 \right\},$$

$$H_E^1(x^*) := \begin{cases} (\nabla^2 f(x^*))_{i,j}, & i \in K_1 \cup N_1 \cup P_1, j \in E \\ \left(\nabla^2 f(x^*) - \frac{2}{a}I\right)_{i,j}, & i \in M_1, j \in E \end{cases},$$

$$H_E^2(x^*) := \begin{cases} (\nabla^2 f(x^*))_{i,j}, & i \in N_1, j \in E \\ \left(\nabla^2 f(x^*) - \frac{2}{a}I\right)_{i,j}, & i \in K_1 \cup M_1 \cup P_1, j \in E \end{cases},$$

其中 $E = \{j \mid j \in K_1 \cup M_1 \cup N_1 \cup P_1\}$ 。

定理 1 如果 $\{\ker H_E^1(x^*) \cup \ker H_E^2(x^*)\} \cap U = \{0\}$, 则 ∂F 在 $(x^*, 0)$ 处是强度量次正则的。

证明 因为 $F = f + g$, f 在 x^* 处是可微的, 所以有 $\partial F(x^*) = \nabla f(x^*) + \partial g(x^*)$ 。由引理 2 可知:

$$D(\partial F)(x^* | 0)(u) = \nabla^2 f(x^*)u + D(\partial g)(x^* | -\nabla f(x^*))(u).$$

设 $\{\ker H_E^1(x^*) \cup \ker H_E^2(x^*)\} \cap U = \{0\}$, $0 \in D(\partial F)(x^* | 0)(u)$, 有

$$-\nabla^2 f(x^*)u \in D(\partial g)(x^* | -\nabla f(x^*))(u).$$

由命题 2, 当 $j \notin E$ 时, $u_j = 0$; 当 $j \in K_1$ 时, 有

$$\left(\left(\nabla^2 f(x^*) - \frac{2}{a}I\right)u\right)_j = 0, u_j (\nabla f(x^*))_j \leq 0, (\nabla^2 f(x^*)u)_j (\nabla f(x^*))_j \leq 0,$$

或者

$$u_j = 0, u_j (\nabla f(x^*))_j \leq 0, (\nabla^2 f(x^*)u)_j (\nabla f(x^*))_j \leq 0.$$

当 $j \in M_1$ 时, 有 $\left(\left(\nabla^2 f(x^*) - \frac{2}{a}I\right)u\right)_j = 0$; 当 $j \in N_1$ 时, 有 $(\nabla^2 f(x^*)u)_j = 0$; 当 $j \in P_1$ 时, 有

$$(\nabla^2 f(x^*)u)_j = 0 \text{ 或 } \left(\left(\nabla^2 f(x^*) - \frac{2}{a}I\right) \cdot u\right)_j = 0.$$

因为 $\{\ker H_E^1(x^*) \cup \ker H_E^2(x^*)\} \cap U = \{0\}$, 所以有 $u = 0$, 这意味着

$$D(\partial F)(x^* | 0)^{-1}(0) = \{0\}.$$

再由引理 1 知 ∂F 在 $(x^*, 0)$ 处是强度量次正则的。

为了建立 ∂F 强度量次正则的充要条件, 我们引出如下记号:

$$J_1 := \left\{j \in \{1, \dots, n\} \mid x_j^* = 0 \text{ 和 } (\nabla f(x^*))_j \in (-2\lambda, 2\lambda)\right\},$$

$$U := \left\{u \in \mathbb{R}^n \mid u_j = 0, j \in J_1 \text{ 和 } u_j (\nabla f(x^*))_j \leq 0, j \in K_1\right\},$$

$$H_E^3(x^*) := \begin{cases} (\nabla^2 f(x^*))_{j,j}, & j \in K_1 \cup N_1 \cup P_1 \\ \left(\nabla^2 f(x^*) - \frac{2}{a}I\right)_{j,j}, & j \in M_1 \end{cases},$$

$$H_E^4(x^*) := \begin{cases} (\nabla^2 f(x^*))_{j,j}, & j \in N_1 \\ \left(\nabla^2 f(x^*) - \frac{2}{a}I\right)_{j,j}, & j \in K_1 \cup M_1 \cup P_1 \end{cases}.$$

定理 2 ∂F 在 $(x^*, 0)$ 处是强度量次正则的当且仅当

$$\begin{cases} \{\ker H_E^3(x^*) \cup \ker H_E^4(x^*)\} \cap \mathcal{U} = \{0\} \\ \left(\nabla^2 f(x^*)u \right)_j \left(\nabla f(x^*) \right)_j \leq 0, & j \in K_1 \end{cases}$$

证明 因为 $F = f + g$, f 在 x^* 处是可微的, 所以有 $\partial F(x^*) = \nabla f(x^*) + \partial g(x^*)$ 。由引理 2 可知:

$$D(\partial F)(x^* | 0)(u) = \nabla^2 f(x^*)u + D(\partial g)(x^* | -\nabla f(x^*))(u). \quad (7)$$

由引理 1 知 ∂F 在 $(x^*, 0)$ 是强度量次正则的当且仅当 $D(\partial F)(x^* | 0)^{-1}(0) = \{0\}$, 即

$$0 \in D(\partial F)(x^* | 0)(u) \Rightarrow u = 0. \quad (8)$$

由命题 2 和(7), 我们有(8)成立当且仅当

$$\begin{cases} u_j \left(\nabla f(x^*) \right)_j \leq 0, \left(\nabla^2 f(x^*)u \right)_j \left(\nabla f(x^*) \right)_j \leq 0, \left(\left(\nabla^2 f(x^*) - \frac{2}{a}I \right)u \right)_j = 0 \text{ 或 } u_j = 0, & j \in K_1 \\ \left(\left(\nabla^2 f(x^*) - \frac{2}{a}I \right)u \right)_j = 0, & j \in M_1 \\ \left(\nabla^2 f(x^*)u \right)_j = 0, & j \in N_1 \\ \left(\nabla^2 f(x^*)u \right)_j = 0 \text{ 或 } \left(\left(\nabla^2 f(x^*) - \frac{2}{a}I \right)u \right)_j = 0, & j \in P_1 \\ u_j = 0, & \text{if } j \in J_1 \end{cases} \Rightarrow u = 0$$

注意到

$$\begin{cases} u_j \left(\nabla f(x^*) \right)_j \leq 0, \left(\left(\nabla^2 f(x^*) - \frac{2}{a}I \right)u \right)_j = 0 \text{ 或 } u_j = 0, & j \in K_1 \\ \left(\left(\nabla^2 f(x^*) - \frac{2}{a}I \right)u \right)_j = 0, & j \in M_1 \\ \left(\nabla^2 f(x^*)u \right)_j = 0, & j \in N_1 \\ \left(\nabla^2 f(x^*)u \right)_j = 0 \text{ 或 } \left(\left(\nabla^2 f(x^*) - \frac{2}{a}I \right)u \right)_j = 0, & j \in P_1 \\ u_j = 0, & j \in J_1 \end{cases} \Rightarrow u = 0.$$

当且仅当 $\{\ker H_E^3(x^*) \cup \ker H_E^4(x^*)\} \cap \mathcal{U} = \{0\}$ 。

4. 结束语

在本文中, 我们分别建立了带 MCP 函数正则化问题次微分强度量次正则的充分条件和充要条件。首先, 我们根据次微分的定义得到了 MCP 函数的次微分的表达式, 并在此基础上求得了 MCP 函数次微分的图像导数。最后, 通过求和准则将 MCP 函数次微分的图像导数与正则化问题(1)次微分的图像导数建立联系, 得到了正则化问题(1)次微分强度量次正则的充分条件和充要条件。后续我们将研究正则化问题(1)的二阶增长条件, 并用强度量次正则或二阶增长条件研究该模型相关算法的收敛性。

基金项目

本文在重庆市自然科学基金(CSTB2022NSCQ-MSX0409, CSTB2022NSCQMSX0406, CQ-NCAM-2021-02)和重庆市教育委员会科技项目(KJZD-M202201303, KJQN202201349)资助下完成。

参考文献

- [1] Fan, J.Q. 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>
- [2] Zhang, C. (2012) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *The Annals of Statistics*, **38**, 894-942. <https://doi.org/10.1214/09-AOS729>
- [3] Andriyana, Y., Fitriani, R., Tantular, B., et al. (2023) Modeling the Cigarette Consumption of Poor Households Using Penalized Zero-Inflated Negative Binomial Regression with Minimax Concave Penalty. *Mathematics*, **11**, 3192. <https://doi.org/10.3390/math11143192>
- [4] Kumar, P.P., Vamshi, R.H. and Sekhar, C.S. (2022) Iteratively Reweighted Minimax-Concave Penalty Minimization for Accurate Low-Rank Plus Sparse Matrix Decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 8992-9010.
- [5] Wang, S., Selesnick, W.I., Cai, G., et al. (2019) Synthesis Versus Analysis Priors via Generalized Minimax-Concave Penalty for Sparsity-Assisted Machinery fault Diagnosis. *Mechanical Systems and Signal Processing*, **127**, 202-233. <https://doi.org/10.1016/j.ymsp.2019.02.053>
- [6] You, J., Jiao, Y., Lu, X., et al. (2019) A Nonconvex Model with Minimax Concave Penalty for Image Restoration. *Journal of Scientific Computing*, **78**, 1063-1086. <https://doi.org/10.1007/s10915-018-0801-z>
- [7] Cai, G., Selesnick, W.I., Wang, S., et al. (2018) Sparsity-Enhanced Signal Decomposition via Generalized Minimax-Concave Penalty for Gearbox Fault Diagnosis. *Journal of Sound and Vibration*, **432**, 213-234. <https://doi.org/10.1016/j.jsv.2018.06.037>
- [8] Shi, Y., Jiao, Y.L., et al. (2018) An Alternating Direction Method of Multipliers for MCP-penalized Regression with High-Dimensional Data. *Acta Mathematica Sinica, English Series*, **34**, 1892-1906. <https://doi.org/10.1007/s10114-018-7096-8>
- [9] Bello-Cruz, J.Y., Li, G.Y. and Nghia, T.T.A. (2021) On the Linear Convergence of Forward-Backward Splitting Method: Part I—Convergence Analysis. *Journal of Optimization Theory and Applications*, **188**, 378-401. <https://doi.org/10.1007/s10957-020-01787-7>
- [10] Drusvyatskiy, D. and Lewis, A.S. (2018) Error Bounds, Quadratic Growth, and Linear Convergence of Proximal Methods. *Mathematics of Operations Research*, **43**, 919-948. <https://doi.org/10.1287/moor.2017.0889>
- [11] Hu, Y.H., Li, C., Meng, K.W. and Yang, X.Q. (2021) Linear Convergence of Inexact Descent Method and Inexact Proximal Gradient Algorithms for Lower-Order Regularization Problems. *Journal of Global Optimization*, **79**, 853-883. <https://doi.org/10.1007/s10898-020-00955-3>
- [12] Dontchev, A.L. and Rockafellar, R.T. (2009) *Implicit Functions and Solution Mappings. A View from Variational Analysis*. Springer, New York. <https://doi.org/10.1007/978-0-387-87821-8>
- [13] Chieu, N.H., Hien, L.V., Nghia, T.T.A. and Tuan, H.A. (2021) Quadratic Growth and Strong Metric Subregularity of the Subdifferential via Subgradient Graphical Derivative. *SIAM Journal on Optimization*, **31**, 545-568. <https://doi.org/10.1137/19M1242732>
- [14] Li, M.H., Meng, K.W. and Yang, X.Q. (2023) Variational Analysis of Kurdyka-Łojasiewicz Property, Exponent and Modulus. arXiv: 2308.15760.
- [15] Bello-Cruz, Y., Li, G.Y. and Nghia, T.T.A. (2022) Quadratic Growth Conditions and Uniqueness of Optimal solution to Lasso. *Journal of Optimization Theory and Applications*, **194**, 167-190. <https://doi.org/10.1007/s10957-022-02013-2>
- [16] Clarke, F.H., Ledyaev, Y.S., Stern, R.J. and Wolenski, P.R. (1998) *Nonsmooth Analysis and Control Theory*. Springer, New York.