

稀疏加组稀疏优化问题的方向稳定点及其光滑化方法

苏妍妍, 彭定涛*

贵州大学, 数学与统计学院, 贵州 贵阳

收稿日期: 2022年9月26日; 录用日期: 2022年10月19日; 发布日期: 2022年10月28日

摘要

本文研究稀疏加组稀疏优化问题的非凸松弛模型, 其中惩罚项既含有稀疏惩罚, 又含有组稀疏惩罚, 对稀疏惩罚和组稀疏惩罚均采用折叠凹惩罚函数进行连续松弛, 得到复合非光滑非凸优化模型。为刻画此模型的最优性条件, 给出了其方向导数刻画和方向稳定点, 分析了方向稳定点的特征及其局部最优性质。为计算模型的方向稳定点, 构造了模型的光滑化逼近问题, 并证明了光滑化问题的一阶稳定点收敛于模型的方向稳定点, 为使用光滑方法计算模型的方向稳定点提供了理论保证。

关键词

稀疏加组稀疏优化问题, 非光滑非凸优化, 折叠凹惩罚函数, 光滑化方法, 方向稳定点

Directional Stationary Points of Sparse plus Group Sparse Optimization Problems with Its Smoothing Methods

Yanyan Su, Dingtao Peng*

School of Mathematics and Statistics, Guizhou University, Guizhou University, Guiyang Guizhou

Received: Sep. 26th, 2022; accepted: Oct. 19th, 2022; published: Oct. 28th, 2022

Abstract

In this paper, we study the nonconvex relaxation model for the sparse plus group sparse optimization problem, in which the penalty term contains both sparse penalty and group sparse penalty. As continuous relaxations, the folded concave penalty functions are used to relax both sparse penalty

*通讯作者。

and group sparse penalty, which results the compound nonsmooth and nonconvex optimization model. In order to characterize the optimality of the nonconvex relaxation problem, the directional derivative and the directional stationary point are introduced, and then the characteristics of the directional stationary points and its local optimality are analyzed. To calculate the directional stationary points of the relaxation model, the smoothing approximation problem is constructed, and it is proved that the stationary points of the smoothing problem converge to the directional stationary point of the relaxation problem, which provides a theoretical guarantee for the calculation of the directional stationary point of the relaxation problem by using the smooth methods.

Keywords

Sparse plus Group Sparse Optimization Problem, Nonconvex and Nonsmooth Optimization, Folded Concave Penalty Function, Smoothing Method, Stationary Points

Copyright © 2022 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

稀疏性是自然界中普遍存在的一种性质, 利用该性质能够实现对信息的充分压缩, 从而减少所需存储空间和信息的传输量, 或是从海量的信息中过滤噪声影响提炼有效信息, 随着大数据时代的到来, 稀疏优化在压缩感知、变量选择、图像恢复等领域引起了广泛的关注, 这些问题的主要目的是寻求欠定线性或非线性方程组的最稀疏解。

通常的稀疏性是对向量的每个分量而言的, 而组稀疏是把向量的分量进行分组再考查各组是否整体为零, 它体现为向量中非零分量或零分量会集中出现在某几个区域。组稀疏优化是一类具有分组结构的稀疏优化问题, 利用变量之间结构关系的先验信息对变量进行分组分块, 进而提高结果的可解释性和预测性能。对于一个 n 维变量 $x \in \mathbb{R}^n$, 我们可以利用先验信息将 x 分成 K 组: $x = (x_{(1)}^\top, \dots, x_{(K)}^\top)^\top \in \mathbb{R}^n$, 其中 $x_{(j)} = (x_{(j)1}, \dots, x_{(j)n_j})^\top \in \mathbb{R}^{n_j}$ 表示 x 的第 j 个分组, $x_{(j)i} \in \mathbb{R}$ 表示 $x_{(j)}$ 的第 i 个元素, 第 j 组中分量的个数为 n_j , $\sum_{j=1}^K n_j = n$ 。近年来, 组稀疏结构在机器学习、信号图像处理、数理统计、模式识别、生物信息学等领域引起了广泛的关注[1] [2] [3] [4] [5]。

在应用中, 有时真实信号/数据不仅具有稀疏结构, 同时也具有组稀疏结构, 这就需要考虑如下稀疏加组稀疏优化问题:

$$\min_{x \in \mathbb{R}^n} f(x) := \mathcal{L}(x) + \lambda_1 \|x\|_0 + \lambda_2 \|x\|_{2,0} \quad (1)$$

其中 $\lambda_1, \lambda_2 > 0$, $\mathcal{L}: \mathbb{R}^n \rightarrow \mathbb{R}$ 连续可微, $\|x\|_0 := \#\{x_{(j)i} : x_{(j)i} \neq 0, j = 1, \dots, K; i = 1, \dots, n_j\} = \{x \text{ 中非零分量的个数}\}$, 而 $\|x\|_{2,0} := \#\{x_{(j)} : \|x_{(j)}\| \neq 0, j = 1, \dots, K\} = \{x \text{ 中非零组的个数}\}$ 。

正则项中 ℓ_0 -范数与 $\ell_{2,0}$ -范数的存在使得上述稀疏加组稀疏优化问题本质上是一类非凸、非光滑、非 Lipschitz, 甚至是不连续的 NP 难问题[4] [6]。一种常见的处理方法是研究其松弛模型。 ℓ_1 -范数是 ℓ_0 -范数最常用的凸松弛, 受到很多研究, 人们提出了许多求解大规模 ℓ_1 正则问题的高效算法[7] [8] [9]。但 ℓ_1 -松弛容易导致过度松弛而产生有偏估计, 模型的解也不具备 oracle 性质[10] [11] [12]。Fan 和 Li [11] [12] 提出了一种折叠凹惩罚, 并证明了该非凸优化能够得到具有无偏性和 oracle 性质的解。随后研究者们给

出了一系列连续的非凸惩罚函数, 包括对数罚[11]、SCAD 罚(Smoothly Clipped Absolute Deviation Penalty) [11]、MCP (Minimax Concave Penalty) [13]、Capped- ℓ_1 罚[14] [15]、分数罚[16]、硬阈值罚(HTP) [17] [18] 和桥罚[19]等。

对稀疏加组稀疏优化问题(1), 本文考虑以下松弛模型:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) &:= \mathcal{L}(x) + \lambda_1 \sum_{i=1}^n \varphi_1(|x_i|) + \lambda_2 \sum_{j=1}^K \varphi_2(\|x_{(j)}\|) \\ &= \mathcal{L}(x) + \sum_{j=1}^K \left[\sum_{i=1}^{n_j} \lambda_1 \varphi_1(|x_{(j)i}|) + \lambda_2 \varphi_2(\|x_{(j)}\|) \right], \end{aligned} \tag{2}$$

其中 $\lambda_1, \lambda_2 > 0$, $\varphi_\ell: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ ($\ell = 1, 2$) 为满足以下条件的折叠凹罚函数:

φ_ℓ 是局部 Lipschitz 连续的, 在 $[0, \infty)$ 上单调不减, 且 $\varphi'_\ell(0) := \varphi'_\ell(0+) > 0$;

$\varphi_\ell(0) = 0$, 且当 $t > 0$ 时, $\varphi_\ell(t) > 0$ 。

本文结构如下, 在第二节中, 我们介绍方向稳定点和局部最优性质, 并给出上述松弛模型的方向导数的表达式和方向稳定点的特征; 第三节对松弛模型构造光滑化逼近函数, 并给出了光滑化问题的一阶稳定点的刻画及性质; 第四节研究光滑化问题一阶稳定点与松弛问题的方向稳定点的一致性, 为使用光滑化方法求解松弛模型提供理论保证。

符号标记: $\mathcal{J}(x) := \{j \in \{1, \dots, K\} : x_{(j)} \neq 0\}$ 表示 x 中非零分组的下标, $\mathcal{I}_j(x) := \{i \in \{1, \dots, n_j\} : x_{(j)i} \neq 0\}$ 表示 x 的第 j 组中非零元素的下标, $|\mathcal{I}_j(x)|$ 表示第 j 组中非零元素的个数, $x_{\mathcal{I}_j(x)}$ 表示第 j 组中所有非零元素按原来的顺序组成的向量, $\mathcal{D}(x) = \{d \in \mathbb{R}^n : d_{(j)i} = 0, \forall j \in \{1, \dots, K\}, i \notin \mathcal{I}_j(x)\}$, $\nabla_{(j)i} \mathcal{L}(x) = \frac{\partial \mathcal{L}(x)}{\partial x_{(j)i}}$,

$$\nabla_{(j)} \mathcal{L}(x) = \frac{\partial \mathcal{L}(x)}{\partial x_{(j)}} = \left(\frac{\partial \mathcal{L}(x)}{\partial x_{(j)1}}, \dots, \frac{\partial \mathcal{L}(x)}{\partial x_{(j)n_j}} \right)^\top.$$

2. 问题(2)的方向稳定点

2.1. 方向稳定点的局部最优性

首先, 我们介绍方向稳定点的定义[20] [21] [22] [23]。

定义 2.1 (方向稳定点) 设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 在给定点 $x^* \in \mathbb{R}^n$ 处方向可微, 若 $\forall d \in \mathbb{R}^n$, f 的方向导数

$$f'(x^*; d) = \lim_{t \downarrow 0} \frac{f(x^* + td) - f(x^*)}{t} \geq 0,$$

则称 x^* 为 f 的一个方向稳定点。

注意到, 若 f 在 x^* 处可微, 则 $\nabla f(x^*) = 0$, 进而 $f'(x^*; d) = \langle \nabla f(x^*), d \rangle = 0, \forall d \in \mathbb{R}^n$ 。

为了方便表示, 记 $m(x) = \|x\|$, 则任意 $x, d \in \mathbb{R}^n$, $m(x)$ 的方向导数为:

$$m'(x; d) = \begin{cases} \|d\|, & \|x\| = 0 \\ \frac{\langle x, d \rangle}{\|x\|}, & \|x\| \neq 0 \end{cases}. \tag{3}$$

特别的, 当 $n = 1$ 时, $m(x) = \|x\| = |x|$, 因此

$$m'(x; d) = \begin{cases} |d|, & x = 0 \\ \frac{xd}{|x|}, & x \neq 0 \end{cases}. \tag{4}$$

文献[24]中指出, 方向稳定点有以下局部最优性质。

引理 2.1 (方向稳定点的局部最优性) 设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 在 $x \in \mathbb{R}^n$ 上是局部 Lipschitz 连续且方向可微的, 则有下列结论:

- (i) 如果 \hat{x} 是 f 的局部极小点, 则 \hat{x} 也是 f 的方向稳定点。
- (ii) f 在 \hat{x} 满足一阶增长性条件, 即存在 \hat{x} 的一个邻域 \mathcal{V} 和 $\delta > 0$, 使得

$$f(x) \geq f(\hat{x}) + \delta \|x - \hat{x}\|, \forall x \in \mathcal{V},$$

等价于在 \hat{x} 满足

$$f'(\hat{x}; d) > 0, \forall d \in \mathbb{R}^n \setminus \{0\}.$$

2.2. 方向稳定点的特征

在本小节中, 我们给出了问题(2)方向导数的计算公式和方向稳定点的特征。

前述折叠凹惩罚函数大多可以写成凸差(Difference-of-Convex)形式[20], 即两个凸函数的差。我们的分析基于惩罚函数的凸差表示。因此, 引进以下假设。

假设 2.1 罚函数 $\varphi_\ell: \mathbb{R}_+ \rightarrow \mathbb{R}_+$, ($\ell=1, 2$) 具有凸差形式: $\varphi_\ell(t) \triangleq g_\ell(t) - h_\ell(t)$, 其中 $h_\ell(t) \triangleq \max_{1 \leq v \leq \bar{v}_\ell} \{h_{\ell v}(t)\}$, $g_\ell, h_{\ell v}$ ($1 \leq v \leq \bar{v}_\ell$) 在 $t \in (0, \infty)$ 上都是可微凸函数, 且满足:

$$g'_\ell(0) := g'_\ell(0+), \quad h'_{\ell v}(0) = h'_{\ell v}(0+) \quad (1 \leq v \leq \bar{v}_\ell).$$

在假设 2.1 下, 问题(2)可以改写成以下形式:

$$\min_{x \in \mathbb{R}^n} f(x) := \mathcal{L}(x) + \left\{ \lambda_1 \sum_{j=1}^K \sum_{i=1}^{n_j} \left[g_1(\|x_{(j)i}\|) - h_1(\|x_{(j)i}\|) \right] + \lambda_2 \sum_{j=1}^K \left[g_2(\|x_{(j)}\|) - h_2(\|x_{(j)}\|) \right] \right\}. \quad (5)$$

定理 2.1 若假设 2.1 成立, 问题(2)的方向导数有以下计算公式:

$$\begin{aligned} f'(x; d) := & \langle \nabla \mathcal{L}(x), d \rangle + \lambda_1 \sum_{j=1}^K \sum_{i=1}^{n_j} \left[g'_1(\|x_{(j)i}\|) - \max_{v_{(j)i} \in \mathcal{A}(x_{(j)i})} h'_{1v_{(j)i}}(\|x_{(j)i}\|) \right] m'(x_{(j)i}; d_{(j)i}) \\ & + \lambda_2 \sum_{j=1}^K \left[g'_2(\|x_{(j)}\|) - \max_{v_{(j)} \in \mathcal{A}(x_{(j)})} h'_{2v_{(j)}}(\|x_{(j)}\|) \right] m'(x_{(j)}; d_{(j)}), \end{aligned} \quad (6)$$

其中 $\mathcal{A}(x_{(j)i}) = \{v_\ell \in \{1, 2, \dots, \bar{v}_\ell\} : h_{v_\ell}(\|x_{(j)i}\|) = h(\|x_{(j)i}\|)\}$, $\mathcal{A}(x_{(j)}) = \{v_\ell \in \{1, 2, \dots, \bar{v}_\ell\} : h_{v_\ell}(\|x_{(j)}\|) = h(\|x_{(j)}\|)\}$ ($\ell=1, 2$)。

证明. [24] (定理 2.3) 中给出了组稀疏惩罚问题折叠凹松弛模型的方向导数的表达式, 在该表达式基础上, 加上稀疏惩罚项折叠凹松弛的方向导数表达式即可, 而后者可以通过方向导数的定义直接计算得到。□

下面引理说明, 在问题(2)的任意方向稳定点 \hat{x} 处, 对应于 \hat{x} 的非零元素的支撑集上损失函数 \mathcal{L} 的梯度值可以用 $g_\ell, h_{\ell v}$ ($\ell=1, 2$) 的导数刻画, 而对应于 \hat{x} 的非零组中元素的位置上, \mathcal{L} 的梯度值受到 $g_\ell(0), h_{\ell v}(0)$ 的控制。

引理 2.2 假设 2.1 成立, $\hat{x} \in \mathbb{R}^n$ 是问题(2)的一个方向稳定点, 则可以得到以下结论: $\forall v_{(j)} \in \mathcal{A}(\hat{x}_{(j)})$, $v_{(j)i} \in \mathcal{A}(\hat{x}_{(j)i})$,

- (i) $\forall j \in \mathcal{J}(\hat{x}), i \in \mathcal{I}_j(\hat{x})$

$$|\nabla_{(j)i} \mathcal{L}(\hat{x})| = |\hat{x}_{(j)i}| \left| \lambda_1 \frac{g'_1(\|\hat{x}_{(j)i}\|) - h'_{1v_{(j)i}}(\|\hat{x}_{(j)i}\|)}{|\hat{x}_{(j)i}|} + \lambda_2 \frac{g'_2(\|\hat{x}_{(j)}\|) - h'_{2v_{(j)}}(\|\hat{x}_{(j)}\|)}{\|\hat{x}_{(j)}\|} \right|.$$

特别的, 当 $\bar{v} = 1$ 时, $\forall j \in \mathcal{J}(\hat{x}), i \in \mathcal{I}_j(\hat{x})$

$$|\nabla_{(j)i} \mathcal{L}(\hat{x})| = |\hat{x}_{(j)i}| \left| \lambda_1 \frac{\varphi'_1(|\hat{x}_{(j)i}|)}{|\hat{x}_{(j)i}|} + \lambda_2 \frac{\varphi'_2(\|\hat{x}_{(j)}\|)}{\|\hat{x}_{(j)}\|} \right|.$$

(ii) 任取 $j \in \mathcal{J}(\hat{x})$, 满足 $\{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x}) \neq \emptyset$ (相当于 $|\mathcal{I}_j(\hat{x})| < n_j$ 或 $x_{\mathcal{I}_j(\hat{x})} \neq x_{(j)}$), 则

$$|\nabla_{(j)i} \mathcal{L}(\hat{x})| \leq \lambda_1 [g'_1(0) - h'_{1v_{(j)i}}(0)], \forall i \in \{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x}),$$

进而

$$\left| \sum_{i \in \{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x})} \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| \leq \lambda_1 \sum_{i \in \{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x})} [g'_1(0) - h'_{1v_{(j)i}}(0)].$$

特别的, 当 $\bar{v}_\ell = 1 (\ell = 1, 2)$ 时, $\forall j \in \mathcal{J}(\hat{x})$, 若 $\{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x}) \neq \emptyset$, 则

$$|\nabla_{(j)i} \mathcal{L}(\hat{x})| \leq \lambda_1 \varphi'_1(0), \forall i \in \{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x}), \tag{7}$$

进而

$$\left| \sum_{i \in \{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x})} \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| \leq \lambda_1 (n_j - |\mathcal{I}_j(\hat{x})|) \varphi'_1(0).$$

证明. (i) 根据定理 2.1, $\forall d \in \mathbb{R}^n$, $v_{(j)} \in \mathcal{A}(\hat{x}_{(j)})$, $v_{(j)i} \in \mathcal{A}(\hat{x}_{(j)i})$, 可以得到:

$$\begin{aligned} 0 &\leq f'(\hat{x}; d) \\ &= \langle \nabla \mathcal{L}(\hat{x}), d \rangle + \lambda_1 \sum_{j=1}^K \sum_{i=1}^{n_j} \left[g'_1(|\hat{x}_{(j)i}|) - \max_{v_{(j)i} \in \mathcal{A}(\hat{x}_{(j)i})} h'_{1v_{(j)i}}(|\hat{x}_{(j)i}|) \right] m'(\hat{x}_{(j)i}; d_{(j)i}) \\ &\quad + \lambda_2 \sum_{j=1}^K \left[g'_2(\|\hat{x}_{(j)}\|) - \max_{v_{(j)} \in \mathcal{A}(\hat{x}_{(j)})} h'_{2v_{(j)}}(\|\hat{x}_{(j)}\|) \right] m'(\hat{x}_{(j)}; d_{(j)}) \\ &\leq \langle \nabla \mathcal{L}(\hat{x}), d \rangle + \lambda_1 \sum_{j=1}^K \sum_{i=1}^{n_j} \left[g'_1(|\hat{x}_{(j)i}|) - h'_{1v_{(j)i}}(|\hat{x}_{(j)i}|) \right] m'(\hat{x}_{(j)i}; d_{(j)i}) \\ &\quad + \lambda_2 \sum_{j=1}^K \left[g'_2(\|\hat{x}_{(j)}\|) - h'_{2v_{(j)}}(\|\hat{x}_{(j)}\|) \right] m'(\hat{x}_{(j)}; d_{(j)}). \end{aligned} \tag{8}$$

特别地, 任取 $d \in \mathcal{D}(\hat{x})$, 则根据(3)和(4)知

$$m'(\hat{x}_{(j)i}; d_{(j)i}) = \begin{cases} |d_{(j)i}| = 0, & \hat{x}_{(j)i} = 0 \\ \frac{\hat{x}_{(j)i} d_{(j)i}}{|\hat{x}_{(j)i}|}, & \hat{x}_{(j)i} \neq 0 \end{cases}, \quad m'(\hat{x}_{(j)}; d_{(j)}) = \begin{cases} \|d_{(j)}\| = 0, & \|\hat{x}_{(j)}\| = 0 \\ \frac{\langle \hat{x}_{(j)}, d_{(j)} \rangle}{\|\hat{x}_{(j)}\|}, & \|\hat{x}_{(j)}\| \neq 0 \end{cases}. \tag{9}$$

因此, 得到

$$\begin{aligned} 0 &\leq \sum_{j=1}^K \sum_{i=1}^{n_j} \left\{ \nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 \left[g'_1(|\hat{x}_{(j)i}|) - h'_{1v_{(j)i}}(|\hat{x}_{(j)i}|) \right] m'(\hat{x}_{(j)i}; d_{(j)i}) \right\} \\ &\quad + \lambda_2 \sum_{j=1}^K \left[g'_2(\|\hat{x}_{(j)}\|) - h'_{2v_{(j)}}(\|\hat{x}_{(j)}\|) \right] m'(\hat{x}_{(j)}; d_{(j)}) \\ &= \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} \left\{ \nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 \left[g'_1(|\hat{x}_{(j)i}|) - h'_{1v_{(j)i}}(|\hat{x}_{(j)i}|) \right] \frac{\hat{x}_{(j)i} d_{(j)i}}{|\hat{x}_{(j)i}|} \right\} \\ &\quad + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \left[g'_2(\|\hat{x}_{(j)}\|) - h'_{2v_{(j)}}(\|\hat{x}_{(j)}\|) \right] \frac{\langle \hat{x}_{(j)}, d_{(j)} \rangle}{\|\hat{x}_{(j)}\|} \end{aligned}$$

$$= \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} d_{(j)i} \nabla_{(j)i} \mathcal{L}(\hat{x}) + \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} \hat{x}_{(j)i} \left\{ \frac{\lambda_1 \left[g'_1 \left(\left| \hat{x}_{(j)i} \right| \right) - h'_{1\nu(j)i} \left(\left| \hat{x}_{(j)i} \right| \right) \right]}{\left| \hat{x}_{(j)i} \right|} + \frac{\lambda_2 \left[g'_2 \left(\left\| \hat{x}_{(j)} \right\| \right) - h'_{2\nu(j)} \left(\left\| \hat{x}_{(j)} \right\| \right) \right]}{\left\| \hat{x}_{(j)} \right\|} \right\}$$

根据 $d \in \mathcal{D}(\hat{x})$ 的任意性, 可得 $\forall j \in \mathcal{J}(\hat{x}), i \in \mathcal{I}_j(\hat{x})$

$$0 = \nabla_{(j)i} \mathcal{L}(\hat{x}) + \hat{x}_{(j)i} \left\{ \frac{\lambda_1 \left[g'_1 \left(\left| \hat{x}_{(j)i} \right| \right) - h'_{1\nu(j)i} \left(\left| \hat{x}_{(j)i} \right| \right) \right]}{\left| \hat{x}_{(j)i} \right|} + \frac{\lambda_2 \left[g'_2 \left(\left\| \hat{x}_{(j)} \right\| \right) - h'_{2\nu(j)} \left(\left\| \hat{x}_{(j)} \right\| \right) \right]}{\left\| \hat{x}_{(j)} \right\|} \right\}.$$

因此 $\forall j \in \mathcal{J}(\hat{x}), i \in \mathcal{I}_j(\hat{x})$

$$\left| \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| = \left| \hat{x}_{(j)i} \right| \left| \lambda_1 \frac{\left[g'_1 \left(\left| \hat{x}_{(j)i} \right| \right) - h'_{1\nu(j)i} \left(\left| \hat{x}_{(j)i} \right| \right) \right]}{\left| \hat{x}_{(j)i} \right|} + \lambda_2 \frac{\left[g'_2 \left(\left\| \hat{x}_{(j)} \right\| \right) - h'_{2\nu(j)} \left(\left\| \hat{x}_{(j)} \right\| \right) \right]}{\left\| \hat{x}_{(j)} \right\|} \right|.$$

当 $\bar{\nu} = 1$ 时, 可得 $\forall j \in \mathcal{J}(\hat{x}), i \in \mathcal{I}_j(\hat{x})$

$$0 = \nabla_{(j)i} \mathcal{L}(\hat{x}) + \hat{x}_{(j)i} \left[\lambda_1 \frac{\varphi'_1 \left(\left| \hat{x}_{(j)i} \right| \right)}{\left| \hat{x}_{(j)i} \right|} + \lambda_2 \frac{\varphi'_2 \left(\left\| \hat{x}_{(j)} \right\| \right)}{\left\| \hat{x}_{(j)} \right\|} \right], \tag{10}$$

进而

$$\left| \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| = \left| \hat{x}_{(j)i} \right| \left| \lambda_1 \frac{\varphi'_1 \left(\left| \hat{x}_{(j)i} \right| \right)}{\left| \hat{x}_{(j)i} \right|} + \lambda_2 \frac{\varphi'_2 \left(\left\| \hat{x}_{(j)} \right\| \right)}{\left\| \hat{x}_{(j)} \right\|} \right|.$$

(ii) 任取满足(2)中条件的 j_0 , $\hat{x}_{(j_0)} = (\hat{x}_{(j_0)1}, \dots, \hat{x}_{(j_0)n_{j_0}})^\top$ 和任意取定 $i_0 \in \{1, \dots, n_{j_0}\} \setminus \mathcal{I}_{j_0}(\hat{x})$, 定义 d^+, d^- 如下:

$$d_{(j)i}^+ = \begin{cases} 1, & j = j_0, i = i_0 \\ 0, & \text{其它} \end{cases}, \quad d_{(j)i}^- = \begin{cases} -1, & j = j_0, i = i_0 \\ 0, & \text{其它} \end{cases},$$

则有 $\langle x_{(j_0)}, d_{(j_0)}^\pm \rangle = 0$ 且 $\|d_{(j)}^\pm\| = 0, \forall j \neq j_0$. 从而由(3)和(4)知

$$m'(\hat{x}_{(j_0)}; d_{(j_0)}^\pm) = \frac{\langle x_{(j_0)}, d_{(j_0)}^\pm \rangle}{\|x_{(j_0)}\|} = 0, \quad m'(\hat{x}_{(j_0)i_0}; d_{(j_0)i_0}^\pm) = |d_{(j_0)i_0}^\pm| = 1.$$

根据(8), 可得

$$\begin{aligned} 0 &\leq f'(\hat{x}; d^\pm) \\ &\leq \nabla_{(j_0)i_0} \mathcal{L}(\hat{x}) d_{(j_0)i_0}^\pm + \lambda_1 \left[g'_1 \left(\left| \hat{x}_{(j_0)i_0} \right| \right) - h'_{1\nu(j_0)i_0} \left(\left| \hat{x}_{(j_0)i_0} \right| \right) \right] m'(\hat{x}_{(j_0)i_0}; d_{(j_0)i_0}^\pm) \\ &\quad + \lambda_2 \left[g'_2 \left(\left\| \hat{x}_{(j_0)} \right\| \right) - h'_{2\nu(j_0)} \left(\left\| \hat{x}_{(j_0)} \right\| \right) \right] m'(\hat{x}_{(j_0)}; d_{(j_0)}^\pm) \\ &\leq \pm \nabla_{(j_0)i_0} \mathcal{L}(\hat{x}) + \lambda_1 \left[g'_1(0) - h'_{1\nu(j_0)i_0}(0) \right], \end{aligned}$$

因此

$$|\nabla_{(j_0)_{i_0}} \mathcal{L}(\hat{x})| \leq \lambda_1 [g'_1(0) - h'_{1\nu_{(j_0)_{i_0}}}(0)].$$

进而

$$\left| \sum_{i \in \{1, \dots, n_{j_0}\} \setminus \mathcal{I}_{j_0}(\hat{x})} \nabla_{(j_0)i} \mathcal{L}(\hat{x}) \right| \leq \sum_{i \in \{1, \dots, n_{j_0}\} \setminus \mathcal{I}_{j_0}(\hat{x})} |\nabla_{(j_0)i} \mathcal{L}(\hat{x})| \leq \lambda_1 \sum_{i \in \{1, \dots, n_{j_0}\} \setminus \mathcal{I}_{j_0}(\hat{x})} [g'_1(0) - h'_{1\nu_{(j_0)i}}(0)].$$

当 $\bar{\nu}_\ell = 1 (\ell = 1, 2)$ 时, $\forall j \in \mathcal{J}(\hat{x})$, 若 $\{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x}) \neq \emptyset$, 则

$$|\nabla_{(j)i} \mathcal{L}(\hat{x})| \leq \lambda_1 \varphi'_1(0),$$

进而

$$\left| \sum_{i \in \{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x})} \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| \leq \lambda_1 (n_j - |\mathcal{I}_j(\hat{x})|) [g'_1(0) - h'_1(0)] = \lambda_1 (n_j - |\mathcal{I}_j(\hat{x})|) \varphi'_1(0).$$

结论得证。□

当 $\varphi_\ell (\ell = 1, 2)$ 是 $(0, \infty)$ 上两个可微凸函数的差(例如 SCAD 罚、MCP、Capped- ℓ_1 罚)时, 下述推论给出问题(2)的方向导数在其方向稳定点 \hat{x} 处更加简洁的表达式, 以及当 $f'(\hat{x}; d) = 0$ 时, 方向 d 的特征。

推论 2.1 若假设 2.1 成立, 且 $\varphi_\ell(t) = g_\ell(t) - h_\ell(t)$, 其中 $g_\ell, h_\ell (\ell = 1, 2)$ 是 $(0, \infty)$ 上的可微凸函数。若 \hat{x} 是问题(2)的方向稳定点, 则下述结论成立:

(i) $\forall d \in \mathbb{R}^n$,

$$f'(\hat{x}; d) = \sum_{j=1}^K \sum_{i \in \mathcal{I}_j(\hat{x})} \left[\nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 \varphi'_1(0) |d_{(j)i}| \right] + \lambda_2 \sum_{j \notin \mathcal{J}(\hat{x})} \varphi'_2(0) \|d_{(j)}\|. \quad (11)$$

(ii) 若 $j \notin \mathcal{J}(\hat{x})$, 则下述不等式成立:

$$\begin{aligned} |\nabla_{(j)i} \mathcal{L}(\hat{x})| &\leq \lambda_2 \varphi'_2(0) + \lambda_1 \varphi'_1(0), \quad \forall i \in \{1, \dots, n_j\}, \\ 0 &\leq \lambda_2 \varphi'_2(0) \|d_{(j)}\| + \sum_{i=1}^{n_j} \left[\nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 \varphi'_1(0) |d_{(j)i}| \right], \quad \forall d \in \mathbb{R}^n. \end{aligned}$$

(iii) 若 $\lambda_1 \varphi'_1(0) > |\nabla_{(j)i} \mathcal{L}(\hat{x})|, \forall j \in \mathcal{J}(\hat{x}), \forall i \in \{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x})$, 则当方向 $d \in \mathbb{R}^n$ 满足 $f'(\hat{x}; d) = 0$ 时, 必有

$$d_{(j)i} = 0, \quad \forall j \in \mathcal{J}(\hat{x}), \forall i \in \{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x}).$$

(iv) 若 $\varphi'_2(0) > 0, \lambda_1 \varphi'_1(0) > |\nabla_{(j)i} \mathcal{L}(\hat{x})|, \forall j \in \mathcal{J}(\hat{x}), \forall i \in \{1, \dots, n_j\}$, 则当方向 $d \in \mathbb{R}^n$ 满足 $f'(\hat{x}; d) = 0$ 时, 必有

$$d_{(j)} = 0, \quad \forall j \notin \mathcal{J}(\hat{x}).$$

(v) 若 $\varphi'_2(0) > 0, \lambda_1 \varphi'_1(0) > |\nabla_{(j)i} \mathcal{L}(\hat{x})|, \forall j \in \{1, \dots, K\}, i \in \mathcal{I}_j(\hat{x})$, 则当方向 $d \in \mathbb{R}^n$ 满足 $f'(\hat{x}; d) = 0$ 时, 必有 $\hat{x}_{(j)i} = 0 \Rightarrow d_{(j)i} = 0$.

证明. (i) 因为 $\bar{\nu} = 1$, 根据(6)知: $\forall d \in \mathbb{R}^n$,

$$\begin{aligned} f'(\hat{x}; d) &= \langle \nabla \mathcal{L}(\hat{x}), d \rangle + \lambda_1 \sum_{j=1}^K \sum_{i=1}^{n_j} \left[g'_1(\|\hat{x}_{(j)i}\|) - h'_1(\|\hat{x}_{(j)i}\|) \right] m'(\hat{x}_{(j)i}; d_{(j)i}) \\ &\quad + \lambda_2 \sum_{j=1}^K \left[g'_2(\|\hat{x}_{(j)}\|) - h'_2(\|\hat{x}_{(j)}\|) \right] m'(\hat{x}_{(j)}; d_{(j)}). \end{aligned}$$

由(3)和(4)可得

$$\begin{aligned}
 f'(\hat{x}; d) &= \langle \nabla \mathcal{L}(\hat{x}), d \rangle + \lambda_1 \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} \left[g'_1(\|\hat{x}_{(j)i}\|) - h'_1(\|\hat{x}_{(j)i}\|) \right] \frac{\hat{x}_{(j)i} d_{(j)i}}{\|\hat{x}_{(j)i}\|} \\
 &\quad + \sum_{j=1}^K \sum_{i \notin \mathcal{I}_j(\hat{x})} \lambda_1 [g'_1(0) - h'_1(0)] |d_{(j)i}| + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} [g'_2(0) - h'_2(0)] \|d_{(j)}\| \\
 &\quad + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \left[g'_2(\|\hat{x}_{(j)}\|) - h'_2(\|\hat{x}_{(j)}\|) \right] \frac{\langle \hat{x}_{(j)}, d_{(j)} \rangle}{\|\hat{x}_{(j)}\|} \\
 &= \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} d_{(j)i} \left\{ \nabla_{(j)i} \mathcal{L}(\hat{x}) + \lambda_1 \hat{x}_{(j)i} \frac{g'_1(\|\hat{x}_{(j)i}\|) - h'_1(\|\hat{x}_{(j)i}\|)}{\|\hat{x}_{(j)i}\|} \right\} \\
 &\quad + \sum_{j=1}^K \sum_{i \notin \mathcal{I}_j(\hat{x})} \left\{ \nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 [g'_1(0) - h'_1(0)] |d_{(j)i}| \right\} + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} [g'_2(0) - h'_2(0)] \|d_{(j)}\| \\
 &\quad + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} \hat{x}_{(j)i} d_{(j)i} \frac{g'_2(\|\hat{x}_{(j)}\|) - h'_2(\|\hat{x}_{(j)}\|)}{\|\hat{x}_{(j)}\|} \\
 &= \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} d_{(j)i} \left\{ \nabla_{(j)i} \mathcal{L}(\hat{x}) + \hat{x}_{(j)i} \left[\lambda_1 \frac{\varphi'_1(\|\hat{x}_{(j)i}\|)}{\|\hat{x}_{(j)i}\|} + \lambda_2 \frac{\varphi'_2(\|\hat{x}_{(j)}\|)}{\|\hat{x}_{(j)}\|} \right] \right\} \\
 &\quad + \sum_{j=1}^K \sum_{i \notin \mathcal{I}_j(\hat{x})} \left[\nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 \varphi'_1(0) |d_{(j)i}| \right] + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \varphi'_2(0) \|d_{(j)}\|
 \end{aligned}$$

根据(10), 可得

$$f'(\hat{x}; d) = \sum_{j=1}^K \sum_{i \notin \mathcal{I}_j(\hat{x})} \left[\nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 \varphi'_1(0) |d_{(j)i}| \right] + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \varphi'_2(0) \|d_{(j)}\|.$$

(ii) 根据等式(11), $\forall d \in \mathbb{R}^n$,

$$\begin{aligned}
 f'(\hat{x}; d) &= \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} \left[\nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 \varphi'_1(0) |d_{(j)i}| \right] \\
 &\quad + \sum_{j \notin \mathcal{J}(\hat{x})} \left\{ \lambda_2 \varphi'_2(0) \|d_{(j)}\| + \sum_{i=1}^{n_j} \left[\nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 \varphi'_1(0) |d_{(j)i}| \right] \right\}.
 \end{aligned} \tag{12}$$

任取 $j_0 \notin \mathcal{J}(\hat{x})$, 任取 $i_0 \in \{1, \dots, n_{j_0}\}$, 定义 d^0 如下:

$$d^0_{(j)i} = \begin{cases} -\nabla_{(j_0)i_0} \mathcal{L}(\hat{x}), & j = j_0, i = i_0 \\ 0, & \text{其它} \end{cases}.$$

注意到 \hat{x} 是问题(2)的方向稳定点, 可得

$$\begin{aligned}
 0 \leq f'(\hat{x}; d^0) &= \lambda_2 \varphi'_2(0) \|d^0_{(j_0)}\| + \nabla_{(j_0)i_0} \mathcal{L}(\hat{x}) d^0_{(j_0)i_0} + \lambda_1 \varphi'_1(0) |d^0_{(j_0)i_0}| \\
 &= \left| \nabla_{(j_0)i_0} \mathcal{L}(\hat{x}) \right| \left[\lambda_2 \varphi'_2(0) - \left| \nabla_{(j_0)i_0} \mathcal{L}(\hat{x}) \right| + \lambda_1 \varphi'_1(0) \right].
 \end{aligned}$$

再根据 j_0, i_0 的任意性, 可得

$$\left| \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| \leq \lambda_2 \varphi'_2(0) + \lambda_1 \varphi'_1(0), \quad \forall j \notin \mathcal{J}(\hat{x}), \forall i \in \{1, \dots, n_j\}.$$

任取 $j_g \notin \mathcal{J}(\hat{x})$, 定义 d^s 如下:

$$d_{(j)}^g = \begin{cases} d_{j_g}, & j = j_g \\ 0, & \text{其它} \end{cases}.$$

根据(12), 可得

$$\begin{aligned} 0 \leq f'(\hat{x}; d^g) &= \lambda_2 \phi_2'(0) \left\| d_{(j_g)}^g \right\| + \sum_{i=1}^{n_{j_g}} \left[\nabla_{(j_g)i} \mathcal{L}(\hat{x}) d_{(j_g)i}^g + \lambda_1 \phi_1'(0) \left| d_{(j_g)i}^g \right| \right] \\ &= \lambda_2 \phi_2'(0) \left\| d_{(j_g)} \right\| + \sum_{i=1}^{n_{j_g}} \left[\nabla_{(j_g)i} \mathcal{L}(\hat{x}) d_{(j_g)i} + \lambda_1 \phi_1'(0) \left| d_{(j_g)i} \right| \right]. \end{aligned}$$

再由 j_g 的任意性, 知

$$0 \leq \lambda_2 \phi_2'(0) \left\| d_{(j)} \right\| + \sum_{i=1}^{n_j} \left[\nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 \phi_1'(0) \left| d_{(j)i} \right| \right], \quad \forall j \notin \mathcal{J}(\hat{x}), d \in \mathbb{R}^n.$$

(iii) 根据(12)知

$$\begin{aligned} f'(\hat{x}; d) &\geq \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \notin \mathcal{I}_j(\hat{x})} \left| d_{(j)i} \right| \left[\lambda_1 \phi_1'(0) - \left| \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| \right] \\ &\quad + \sum_{j \notin \mathcal{J}(\hat{x})} \left\{ \lambda_2 \phi_2'(0) \left\| d_{(j)} \right\| + \sum_{i=1}^{n_j} \left[\nabla_{(j)i} \mathcal{L}(\hat{x}) d_{(j)i} + \lambda_1 \phi_1'(0) \left| d_{(j)i} \right| \right] \right\}. \end{aligned} \tag{13}$$

根据不等式(7)和推论 3.1 (ii), 知

$$f'(\hat{x}; d) \geq \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \notin \mathcal{I}_j(\hat{x})} \left| d_{(j)i} \right| \left[\lambda_1 \phi_1'(0) - \left| \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| \right] \geq 0.$$

再由条件 $f'(\hat{x}; d) = 0$, 得

$$0 = \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \notin \mathcal{I}_j(\hat{x})} \left| d_{(j)i} \right| \left[\lambda_1 \phi_1'(0) - \left| \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| \right].$$

因 $\lambda_1 \phi_1'(0) > \left| \nabla_{(j)i} \mathcal{L}(\hat{x}) \right|, \forall j \in \mathcal{J}(\hat{x}), \forall i \in \{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x})$, 故

$$d_{(j)i} = 0, \quad \forall j \in \mathcal{J}(\hat{x}), \forall i \in \{1, \dots, n_j\} \setminus \mathcal{I}_j(\hat{x}).$$

(iv) 由(12)得

$$\begin{aligned} f'(\hat{x}; d) &\geq \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \notin \mathcal{I}_j(\hat{x})} \left| d_{(j)i} \right| \left[\lambda_1 \phi_1'(0) - \left| \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| \right] + \sum_{j \notin \mathcal{J}(\hat{x})} \sum_{i=1}^{n_j} \left| d_{(j)i} \right| \left[\lambda_1 \phi_1'(0) - \left| \nabla_{(j)i} \mathcal{L}(\hat{x}) \right| \right] \\ &\quad + \sum_{j \notin \mathcal{J}(\hat{x})} \lambda_2 \phi_2'(0) \left\| d_{(j)} \right\|. \end{aligned}$$

根据不等式(7)以及条件 $\lambda_1 \phi_1'(0) > \left| \nabla_{(j)i} \mathcal{L}(\hat{x}) \right|, \forall j \notin \mathcal{J}(\hat{x}), \forall i \in \{1, \dots, n_j\}$, 知

$$f'(\hat{x}; d) \geq \sum_{j \notin \mathcal{J}(\hat{x})} \lambda_2 \phi_2'(0) \left\| d_{(j)} \right\| \geq 0.$$

再由条件 $f'(\hat{x}; d) = 0$ 和 $\phi_2'(0) > 0$, 得 $d_{(j)} = 0, \forall j \notin \mathcal{J}(\hat{x})$ 。

(v)综合(iii)和(iv)即得。□

3. 问题(2)的光滑化模型

我们已知问题(2)的方向稳定点具有良好的局部最有性质, 但考虑到正则项的非凸非光滑性质, 直接计算问题(2)的方向稳定点是困难的。在实际应用中, 光滑逼近的方法广泛应用于科学计算和实验, 例如 [24] [25] [26] [27]。因此本文探讨使用光滑化方法来计算非光滑问题(2)的方向稳定点。

首先构造其光滑化模型。考虑到光滑函数的复合函数仍然是光滑函数, 可以通过分别将 $\varphi_\ell(t)$ ($\ell=1,2$) 和 $m(x)$ 进行光滑化, 得到连续可微的函数 $\tilde{\varphi}_{\ell\mu}(t)$ ($\ell=1,2$) 和 $\tilde{m}_\mu(x)$, 再将它们重新复合, 就得到了 $\varphi_\ell(\|x\|)$ 的光滑逼近函数 $\tilde{\varphi}_{\ell\mu} \circ \tilde{m}_\mu(x)$ 。

$m(x)$ 的光滑化: 取 $\tilde{m}_\mu(x) = \sqrt{\|x\|^2 + \mu} \geq 0$ 作为 $m(x) = \|x\|$ 的参数为 μ 的光滑化函数, 其中 $\mu \in (0, \infty)$ 。显然 $\tilde{m}_\mu(x)$ 关于 x 二阶连续可微, 关于 μ 单调递增。注意到, $\forall x, d \in \mathbb{R}^n$, $\tilde{m}'_\mu(x; d)$ 的方向导数和梯度为:

$$\tilde{m}'_\mu(x; d) = \langle \nabla \tilde{m}_\mu(x), d \rangle, \nabla \tilde{m}_\mu(x) = \frac{x}{\sqrt{\|x\|^2 + \mu}}. \tag{14}$$

特别地, 对 $x_{(j)i} \in \mathbb{R}$, 有

$$m(x_{(j)i}) = |x_{(j)i}|, \tilde{m}_\mu(x_{(j)i}) = \sqrt{(x_{(j)i})^2 + \mu}, \tag{15}$$

$$\tilde{m}'_\mu(x_{(j)i}) = \frac{x_{(j)i}}{\sqrt{(x_{(j)i})^2 + \mu}}, \tilde{m}'_\mu(x_{(j)i}; d_{(j)i}) = \frac{x_{(j)i}d_{(j)i}}{\sqrt{(x_{(j)i})^2 + \mu}}. \tag{16}$$

容易验证, $\tilde{m}_\mu(x)$ 满足如下收敛性质: $\forall x, d \in \mathbb{R}^n$,

(i) (光滑化函数的一致收敛性): $\lim_{x' \rightarrow x, \mu \downarrow 0} \tilde{m}_\mu(x') = m(x)$;

(ii) (方向导数的一致收敛与弱一致收敛性):

在 $x \neq 0$ 点, $\lim_{x' \rightarrow x, \mu \downarrow 0} \langle \nabla \tilde{m}_\mu(x'), d \rangle = \langle \nabla m(x), d \rangle = m'(x; d)$;

对 $x = 0$ 点, $\limsup_{x' \rightarrow 0, \mu \downarrow 0} \langle \nabla \tilde{m}_\mu(x'), d \rangle = \|d\| = m'(0; d)$ 。

受 Peng 和 Chen [24] 的启发, 对于 $\varphi_\ell(t)$ ($\ell=1,2$) 的光滑化函数, 本文作出下述假设。

假设 3.1 设 $\varphi_\ell(t)$ ($\ell=1,2$) 的光滑化函数 $\tilde{\varphi}_{\ell\mu}(t)$ ($\ell=1,2$) 满足以下条件:

(i) $\lim_{\mu \downarrow 0, t \rightarrow s} \tilde{\varphi}_{\ell\mu}(t) = \varphi_\ell(s)$ 。

(ii) 当 $s = 0$ 时, $\limsup_{\mu \downarrow 0, t \downarrow 0} \tilde{\varphi}'_{\ell\mu}(t) = \varphi'_\ell(0+) = \varphi'_\ell(0) \geq 0$; 当 $s \neq 0$ 时, $\lim_{\mu \downarrow 0, t \rightarrow s} \tilde{\varphi}'_{\ell\mu}(t) = \varphi'_\ell(s)$ 。

需要强调的是, 在许多情况下, 满足假设 3.1 的光滑化函数是存在的, 例如 Peng 和 Chen 在 [24] 中给出了 MCP、SCAD、Capped- ℓ_1 三类典型罚函数的光滑化函数, 并验证了它们满足假设 3.1。

若光滑化函数 $\varphi_\ell(t)$ ($\ell=1,2$) 满足假设 3.1, 则可得问题(2)的光滑化模型:

$$\min_{x \in \mathbb{R}^n} \tilde{f}_\mu(x) := \mathcal{L}(x) + \sum_{j=1}^K \left[\sum_{i=1}^{n_j} \lambda_1 \tilde{\varphi}_{1\mu} \circ \tilde{m}_\mu(x_{(j)i}) + \lambda_2 \tilde{\varphi}_{2\mu} \circ \tilde{m}_\mu(x_{(j)i}) \right]. \tag{17}$$

基于其连续可微性, 当 $\mu > 0$ 时, 对于任意方向 $d \in \mathbb{R}^n$, $\tilde{f}_\mu(x)$ 的方向导数为:

$$\tilde{f}'_\mu(x; d) = \langle \nabla \tilde{f}_\mu(x), d \rangle, \tag{18}$$

其中

$$\nabla \tilde{f}_\mu(x) = \nabla \mathcal{L}(x) + \lambda_2 \begin{pmatrix} \tilde{\varphi}'_{2\mu} \circ \tilde{m}_\mu(x_{(1)}) \nabla \tilde{m}_\mu(x_{(1)}) \\ \vdots \\ \tilde{\varphi}'_{2\mu} \circ \tilde{m}_\mu(x_{(k)}) \nabla \tilde{m}_\mu(x_{(k)}) \end{pmatrix} + \lambda_1 \begin{pmatrix} \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(x_{(1)1}) \tilde{m}'_\mu(x_{(1)1}) \\ \vdots \\ \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(x_{(1)n_1}) \tilde{m}'_\mu(x_{(1)n_1}) \\ \vdots \\ \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(x_{(K)1}) \tilde{m}'_\mu(x_{(K)1}) \\ \vdots \\ \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(x_{(K)n_K}) \tilde{m}'_\mu(x_{(K)n_K}) \end{pmatrix}. \tag{19}$$

问题(17)是光滑优化问题, 其稳定点 $\hat{x}^\mu \in \mathbf{R}^n$ 定义为使 $\nabla \tilde{f}_\mu(\hat{x}^\mu) = 0$ 的点。

以下定理给出了光滑化问题(17)稳定点的特征, 即在其稳定点处损失函数的梯度值 $\nabla \mathcal{L}$ 可以用对应的惩罚函数 $\tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu$ 的值来刻画。

定理 3.1 (光滑化问题稳定点的特征) 记 \hat{x}^μ 为光滑化参数为 $\mu > 0$ 时光滑化问题(17)的稳定点, 则

$$|\nabla_{(j)} \mathcal{L}(\hat{x}^\mu)| = |\hat{x}^\mu_{(j)}| \left| \lambda_1 \frac{\tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(j)})}{\sqrt{(\hat{x}^\mu_{(j)})^2 + \mu}} + \lambda_2 \frac{\tilde{\varphi}'_{2\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(j)})}{\sqrt{\|\hat{x}^\mu_{(j)}\|^2 + \mu}} \right|, \forall j \in \{1, \dots, K\}, \forall i \in \{1, \dots, n_j\}.$$

证明. 因为 \hat{x}_μ 满足 $0 = \nabla \tilde{f}_\mu(\hat{x}^\mu)$, 根据(14)和(19), 可得

$$\begin{aligned} 0 = \nabla \mathcal{L}(\hat{x}^\mu) + \lambda_2 & \begin{pmatrix} \tilde{\varphi}'_{2\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(1)}) \nabla \tilde{m}_\mu(\hat{x}^\mu_{(1)}) \\ \vdots \\ \tilde{\varphi}'_{2\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(K)}) \nabla \tilde{m}_\mu(\hat{x}^\mu_{(K)}) \end{pmatrix} + \lambda_1 \begin{pmatrix} \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(1)}) \tilde{m}'_\mu(\hat{x}^\mu_{(1)}) \\ \vdots \\ \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(1)n_1}) \tilde{m}'_\mu(\hat{x}^\mu_{(1)n_1}) \\ \vdots \\ \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(K)1}) \tilde{m}'_\mu(\hat{x}^\mu_{(K)1}) \\ \vdots \\ \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(K)n_K}) \tilde{m}'_\mu(\hat{x}^\mu_{(K)n_K}) \end{pmatrix} \\ = \nabla \mathcal{L}(\hat{x}^\mu) + & \begin{pmatrix} \lambda_2 \tilde{\varphi}'_{2\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(1)}) \frac{\hat{x}^\mu_{(1)1}}{\sqrt{\|\hat{x}^\mu_{(1)}\|^2 + \mu}} + \lambda_1 \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(1)1}) \frac{\hat{x}^\mu_{(1)1}}{\sqrt{(\hat{x}^\mu_{(1)1})^2 + \mu}} \\ \vdots \\ \lambda_2 \tilde{\varphi}'_{2\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(1)}) \frac{\hat{x}^\mu_{(1)n_1}}{\sqrt{\|\hat{x}^\mu_{(1)}\|^2 + \mu}} + \lambda_1 \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(1)n_1}) \frac{\hat{x}^\mu_{(1)n_1}}{\sqrt{(\hat{x}^\mu_{(1)n_1})^2 + \mu}} \\ \vdots \\ \lambda_2 \tilde{\varphi}'_{2\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(K)}) \frac{\hat{x}^\mu_{(K)1}}{\sqrt{\|\hat{x}^\mu_{(K)}\|^2 + \mu}} + \lambda_1 \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(K)1}) \frac{\hat{x}^\mu_{(K)1}}{\sqrt{(\hat{x}^\mu_{(K)1})^2 + \mu}} \\ \vdots \\ \lambda_2 \tilde{\varphi}'_{2\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(K)}) \frac{\hat{x}^\mu_{(K)n_K}}{\sqrt{\|\hat{x}^\mu_{(K)}\|^2 + \mu}} + \lambda_1 \tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(K)n_K}) \frac{\hat{x}^\mu_{(K)n_K}}{\sqrt{(\hat{x}^\mu_{(K)n_K})^2 + \mu}} \end{pmatrix}. \end{aligned}$$

因此,

$$|\nabla_{(j)} \mathcal{L}(\hat{x}^\mu)| = |\hat{x}^\mu_{(j)}| \left| \lambda_1 \frac{\tilde{\varphi}'_{1\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(j)})}{\sqrt{(\hat{x}^\mu_{(j)})^2 + \mu}} + \lambda_2 \frac{\tilde{\varphi}'_{2\mu} \circ \tilde{m}_\mu(\hat{x}^\mu_{(j)})}{\sqrt{\|\hat{x}^\mu_{(j)}\|^2 + \mu}} \right|, \forall j \in \{1, \dots, K\}, \forall i \in \{1, \dots, n_j\}.$$

□

4. 光滑化前后稳定点的一致性

使用光滑化方法求解非光滑优化问题的一个重要的问题是: 随着光滑参数趋于零, 光滑化问题的解

能否趋向于原问题的解, 即光滑化前后解的一致性问题。只有当解的一致性成立时, 才能保证光滑化问题的解趋向于原问题的解。

下面的定理说明本文的光滑化方法能够保证光滑化问题(17)的稳定点与问题(2)的方向稳定点具有一致性。

定理 4.1 (稳定点的一致性) 设光滑化函数 $\tilde{\varphi}_{1\mu}$ 和 $\tilde{\varphi}_{2\mu}$ 均满足假设 3.1, 并设 \hat{x}^{μ_k} 是光滑化问题(17)在光滑化参数 $\mu = \mu_k$ 时的稳定点, 则当 $\mu_k \rightarrow 0^+$ 时, $\{\hat{x}^{\mu_k}\}_{k=1}^{\infty}$ 的任意聚点都是问题(2)的方向稳定点。

证明. 设 \hat{x} 为 $\{\hat{x}^{\mu_k}\}_{k=1}^{\infty}$ 的任一聚点, 不妨设

$$\lim_{k \rightarrow \infty} \hat{x}^{\mu_k} = \hat{x}.$$

因 \hat{x}^{μ_k} 是 $\tilde{f}_{\mu_k}(x)$ 的稳定点, 根据 $\tilde{f}_{\mu_k}(x)$ 连续可微性和(18)和(19)可知, $\forall d \in \mathbb{R}^n$, 有

$$\begin{aligned} 0 &= \langle \nabla \tilde{f}_{\mu_k}(\hat{x}^{\mu_k}), d \rangle \\ &= \langle \nabla \mathcal{L}(\hat{x}^{\mu_k}), d \rangle + \lambda_1 \sum_{j=1}^K \sum_{i=1}^{n_j} \tilde{\varphi}'_{1\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) \tilde{m}'_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) d_{(j)i} \\ &\quad + \lambda_2 \sum_{j=1}^K \tilde{\varphi}'_{2\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}) \langle \nabla \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}), d_{(j)} \rangle \\ &= \langle \nabla \mathcal{L}(\hat{x}^{\mu_k}), d \rangle + \lambda_1 \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} \tilde{\varphi}'_{1\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) \tilde{m}'_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) d_{(j)i} \\ &\quad + \lambda_1 \sum_{j=1}^K \sum_{i \notin \mathcal{I}_j(\hat{x})} \tilde{\varphi}'_{1\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) \tilde{m}'_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) d_{(j)i} + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \tilde{\varphi}'_{2\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}) \langle \nabla \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}), d_{(j)} \rangle \\ &\quad + \lambda_2 \sum_{j \notin \mathcal{J}(\hat{x})} \tilde{\varphi}'_{2\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}) \langle \nabla \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}), d_{(j)} \rangle. \end{aligned} \tag{20}$$

令 $k \rightarrow \infty$, 则 $\mu_k \rightarrow 0$, $\hat{x}^{\mu_k} \rightarrow \hat{x}$ 。根据第三节中 $\tilde{m}_{\mu}(x)$ 的一致和弱一致收敛性质以及假设 3.1, 可得下述极限:

$$\begin{aligned} \forall j \in \mathcal{J}(\hat{x}), \forall i \in \mathcal{I}_j(\hat{x}): \lim_{k \rightarrow \infty} \tilde{m}'_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) d_{(j)i} &= m'(\hat{x}_{(j)i}; d_{(j)}), \lim_{k \rightarrow \infty} \tilde{\varphi}'_{1\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) = \varphi'_1(\|\hat{x}_{(j)i}\|); \\ \forall j \in 1, \dots, K, \forall i \notin \mathcal{I}_j(\hat{x}): \limsup_{k \rightarrow \infty} \tilde{m}'_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) d_{(j)i} &= m'(0; d_{(j)i}) = \|d_{(j)i}\| \geq 0, \limsup_{k \rightarrow \infty} \tilde{\varphi}'_{1\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) = \varphi'_1(0) > 0; \\ \forall j \in \mathcal{J}(\hat{x}): \lim_{k \rightarrow \infty} \langle \nabla \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}), d_{(j)} \rangle &= m'(\hat{x}_{(j)}; d_{(j)}), \lim_{k \rightarrow \infty} \tilde{\varphi}'_{2\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}) = \varphi'_2(\|\hat{x}_{(j)}\|); \\ \forall j \notin \mathcal{J}(\hat{x}): \limsup_{k \rightarrow \infty} \langle \nabla \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}), d_{(j)} \rangle &= m'(0; d_{(j)}) = \|d_{(j)}\| \geq 0, \limsup_{k \rightarrow \infty} \tilde{\varphi}'_{2\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}) = \varphi'_2(0) > 0. \end{aligned}$$

在(20)中利用上述极限, 根据定理 2.1 我们可以得到 $\forall d \in \mathbb{R}^n$

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty} \langle \nabla \tilde{f}_{\mu_k}(\hat{x}^{\mu_k}), d \rangle \\ &= \lim_{k \rightarrow \infty} \langle \nabla \mathcal{L}(\hat{x}^{\mu_k}), d \rangle + \lim_{k \rightarrow \infty} \lambda_1 \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} \tilde{\varphi}'_{1\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) \tilde{m}'_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) d_{(j)i} \\ &\quad + \lim_{k \rightarrow \infty} \lambda_1 \sum_{j=1}^K \sum_{i \notin \mathcal{I}_j(\hat{x})} \tilde{\varphi}'_{1\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) \tilde{m}'_{\mu_k}(\hat{x}^{\mu_k}_{(j)i}) d_{(j)i} \\ &\quad + \lim_{k \rightarrow \infty} \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \tilde{\varphi}'_{2\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}) \langle \nabla \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}), d_{(j)} \rangle \\ &\quad + \lim_{k \rightarrow \infty} \lambda_2 \sum_{j \notin \mathcal{J}(\hat{x})} \tilde{\varphi}'_{2\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}) \langle \nabla \tilde{m}_{\mu_k}(\hat{x}^{\mu_k}_{(j)}), d_{(j)} \rangle \end{aligned}$$

$$\begin{aligned}
 &\leq \lim_{k \rightarrow \infty} \langle \nabla \mathcal{L}(\hat{x}^{\mu_k}), d \rangle + \lim_{k \rightarrow \infty} \lambda_1 \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} \tilde{\varphi}'_{1\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}_{(j)i}^{\mu_k}) \tilde{m}'_{\mu_k}(\hat{x}_{(j)i}^{\mu_k}) d_{(j)i} \\
 &\quad + \lambda_1 \sum_{j=1}^K \sum_{i \in \mathcal{I}_j(\hat{x})} \lim_{k \rightarrow \infty} \tilde{\varphi}'_{1\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}_{(j)i}^{\mu_k}) \cdot \limsup_{k \rightarrow \infty} \tilde{m}'_{\mu_k}(\hat{x}_{(j)i}^{\mu_k}) d_{(j)i} \\
 &\quad + \lim_{k \rightarrow \infty} \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \tilde{\varphi}'_{2\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}_{(j)}^{\mu_k}) \langle \nabla \tilde{m}_{\mu_k}(\hat{x}_{(j)}^{\mu_k}), d_{(j)} \rangle \\
 &\quad + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \lim_{k \rightarrow \infty} \tilde{\varphi}'_{2\mu_k} \circ \tilde{m}_{\mu_k}(\hat{x}_{(j)}^{\mu_k}) \cdot \limsup_{k \rightarrow \infty} \langle \nabla \tilde{m}_{\mu_k}(\hat{x}_{(j)}^{\mu_k}), d_{(j)} \rangle \\
 &= \langle \nabla \mathcal{L}(\hat{x}), d \rangle + \lambda_1 \sum_{j \in \mathcal{J}(\hat{x})} \sum_{i \in \mathcal{I}_j(\hat{x})} \tilde{\varphi}'_1 \circ m(\hat{x}_{(j)i}) \cdot m'(\hat{x}_{(j)i}; d_{(j)i}) \\
 &\quad + \lambda_1 \sum_{j=1}^K \sum_{i \in \mathcal{I}_j(\hat{x})} \varphi'_1 \circ m(\hat{x}_{(j)i}) \cdot m'(\hat{x}_{(j)i}; d_{(j)i}) + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \varphi'_2 \circ m(\hat{x}_{(j)}) \cdot m'(\hat{x}_{(j)}; d_{(j)}) \\
 &\quad + \lambda_2 \sum_{j \in \mathcal{J}(\hat{x})} \varphi'_2 \circ m(\hat{x}_{(j)}) \cdot m'(\hat{x}_{(j)}; d_{(j)}) \\
 &= \langle \nabla \mathcal{L}(\hat{x}), d \rangle + \lambda_1 \sum_{j=1}^K \sum_{i=1}^{n_j} \varphi'_1(\|\hat{x}_{(j)i}\|) m'(\hat{x}_{(j)i}; d_{(j)i}) + \lambda_2 \sum_{j=1}^K \varphi'_2(\|\hat{x}_{(j)}\|) m'(\hat{x}_{(j)}; d_{(j)}) \\
 &= f'(\hat{x}; d)
 \end{aligned}$$

因此, \hat{x} 是问题(2)的方向稳定点。□

5. 总结

本文研究了稀疏加组稀疏优化问题的非凸松弛模型。给出了非凸松弛模型方向导数的刻画和方向稳定点, 分析了方向稳定点的特征及其局部最优性质。进一步构造了松弛模型的光滑化逼近问题, 并证明了光滑化问题的稳定点与松弛模型的方向稳定点具有一致性, 为后续使用光滑方法计算模型的方向稳定点提供了理论保障。

基金项目

国家自然科学基金项目(11861020, 12261020)、贵州省高层次留学人才创新创业择优资助重点项目([2018]03)、贵州省科技计划项目(ZK[2021]009, [2018]5781)、贵州省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., Meng, K., Qin, J. and Yang, X. (2017) Group Sparse Optimization via $l_{p,q}$ Regularization. *Journal of Machine Learning Research*, **18**, 960-1011.
- [3] Jiao, Y., Jin, B. and Lv, X. (2017) Group Sparse Recovery via the $l_0(l_2)$ Penalty: Theory and Algorithm. *IEEE Transactions on Signal Processing*, **65**, 998-1012. <https://doi.org/10.1109/TSP.2016.2630028>
- [4] 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>
- [5] 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>
- [6] Natarajan, B.K. (1995) Sparse Approximate Solutions to Linear Systems. *SIAM Journal on Computing*, **24**, 227-234. <https://doi.org/10.1137/S0097539792240406>
- [7] Li, X., Sun, D. and Toh, K.-C. (2018) A Highly Efficient Semismooth Newton Augmented Lagrangian Method for Solving Lasso Problems. *SIAM Journal on Optimization*, **28**, 433-458. <https://doi.org/10.1137/16M1097572>
- [8] Lin, M., Liu, Y. J., Sun, D. and Toh, K.C. (2019) Efficient Sparse Semismooth Newton Methods for the Clustered

- Lasso Problem. *SIAM Journal on Optimization*, **29**, 2026-2052. <https://doi.org/10.1137/18M1207752>
- [9] Zhang, Y., Zhang, N., Sun, D. and Toh, K.-C. (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>
- [10] Candès, E.J., Wakin, M.B. and Boyd, S.P. (2008) Enhancing Sparsity by Reweighted l_1 Minimization. *Journal of Fourier Analysis and Applications*, **14**, 877-905. <https://doi.org/10.1007/s00041-008-9045-x>
- [11] Fan, J. and Li, R. (2001) Variable Selection via Nonconvex Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- [12] 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>
- [13] Zhang, C.-H. (2010) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *Annals of Statistics*, **38**, 894-942. <https://doi.org/10.1214/09-AOS729>
- [14] Thi, L.E. and Cheng, S.O. (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>
- [15] Tong, Z. (2010) Analysis of Multi-Stage Convex Relaxation for Sparse Regularization. *Journal of Machine Learning Research*, **11**, 1081-1107.
- [16] Nikolova, M., Ng, M.-K., Zhang, S. and Ching, W.-K. (2008) Efficient Reconstruction of Piecewise Constant Images Using Nonsmooth Nonconvex Minimization. *SIAM Journal on Imaging Sciences*, **1**, 2-25. <https://doi.org/10.1137/070692285>
- [17] Bruckstein, A.M., Donoho, D.L. and Elad, M. (2009) From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Review*, **51**, 34-81. <https://doi.org/10.1137/060657704>
- [18] Knight, K. and Fu, W. (2000) Asymptotics for Lasso-Type Estimators. *Annals of Statistics*, **28**, 1356-1378. <https://doi.org/10.1214/aos/1015957397>
- [19] Huang, J., Ma, S., Xie, H. and Zhang, C.H. (2009) A Group Bridge Approach for Variable Selection. *Biometrika*, **96**, 339-355. <https://doi.org/10.1093/biomet/asp020>
- [20] 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>
- [21] Chang, T.H., Hong, M. and Pang, J.-S. (2017) Local Minimizers and Second-Order Conditions in Composite Piecewise Programming via Directional Derivatives.
- [22] 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>
- [23] Rochafellar, R.T. and Wets, R.J.-B. (2009) Variational Analysis. 3rd Edition, Springer-Verlag, Berlin.
- [24] 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>
- [25] Chen, X. (2012) Smoothing Methods for Nonsmooth, Nonconvex Minimization. *Mathematical Programming*, **134**, 71-99. <https://doi.org/10.1007/s10107-012-0569-0>
- [26] Chen, X., Niu, L. and Yuan, Y. (2013) Optimality Conditions and a Smoothing Trust Region Newton Method for Non-Lipschitz Optimization. *SIAM Journal on Optimization*, **23**, 1528-1552. <https://doi.org/10.1137/120871390>
- [27] Chen, X., Xu, F. and Ye, Y. (2010) Lower Bound Theory of Nonzero Entries in Solutions of l_2 - l_p Minimization. *SIAM Journal on Computing*, **32**, 2832-2852. <https://doi.org/10.1137/090761471>