Advances in Applied Mathematics 应用数学进展, **2022, 11(12), 9081-9095**
Published Online December 2022 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2022.1112958

基于非凸惩罚回归的参数估计和异常值检测

张尊皓，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：2022年11月26日；录用日期：2022年12月21日；发布日期：2022年12月30日

摘 要

本文基于最优化理论提出了一种实现多元线性回归模型的参数估计和异常值检测的方法。首先，建立了基于**Huber**损失函数和ℓ_0惩罚项的回归模型，为便于求解进一步将该模型中ℓ_0惩罚项松弛为**Capped-**ℓ_1惩罚；其次，刻画了松弛问题的方向稳定点，并建立了原问题和松弛问题的等价性。最后提出了松弛问题的光滑化模型，并证明了光滑模型与松弛问题稳定点的一致性。

关键词

Huber函数，稳健参数估计，异常值检测，非凸惩罚，光滑化方法

Parameter Estimation and Outliers Detection Based on Nonconvex Penalized Regression

Zunhao Zhang, Dingtao Peng*, Yanyan Su

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Nov. 26th, 2022; accepted: Dec. 21st, 2022; published: Dec. 30th, 2022

* 通讯作者。

文章引用: 张尊皓, 彭定涛, 苏妍妍. 基于非凸惩罚回归的参数估计和异常值检测[J]. 应用数学进展, 2022, 11(12): 9081-9095. DOI: 10.12677/aam.2022.1112958

Abstract

This paper presents a method to implement parameter estimation and outliers detection for multiple linear regression models based on optimization theory. First, a regression model based on the Huber loss function and the ℓ_0 penalty term is developed, and the ℓ_0 penalty term in this model is further relaxed to the Capped-ℓ_1 penalty to facilitate the solution; and second, the directional stability point of the relaxation problem is characterized, and the equivalence of the original and relaxation problems is established. Finally, we propose a smooth model for the relaxation problem and prove the consistency of the stable point of smooth model and the relaxation problem.

Keywords

Huber Function, Robust Parameter Estimation, Outliers Detection, Nonconvex Penalty, Smoothing Method

1. 前言

统计学的一个中心问题就是从有噪声的观察值中估计线性回归模型的回归参数. 考虑如下多元线性模型:

$$y = Ax + \omega, \tag{1}$$

其中$y = (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$是观测值, $A \in \mathbb{R}^{n\times p}$是设计矩阵, $\omega = (\omega_1, \ldots, \omega_n)^\top \in \mathbb{R}^n$是随机观测噪声. 根据已知的$A$和$y$, 为求回归参数$x$, 一种常用的方法是用最小二乘损失作为极小化目标函数:

$$\min_{x\in\mathbb{R}^p} ||y - Ax||^2. \tag{2}$$

然而, 在实际应用中, 观测数据y可能包含了较大的观测误差, 这种存在较大观测误差的数据通常称

为异常值 [1]. 最小二乘回归模型稳健性较差, 异常值会导致最小二乘估计结果与实际存在较大误差 [2], 因此需要更稳健的参数估计方以实现在数据包含异常值的情况下对参数进行鲁棒估计 [3–5]. 1982年, Cook提出均值漂移模型(MEAN-SHIFT MODEL [6]), 该模型把随机观测噪声项ω分成两个部分: $\omega = \mu + e$, 其中, $\mu = (\mu_1, \ldots, \mu_n)^\top \in \mathbb{R}^n$是较小数值尺度上的误差, $e = (e_1, \ldots, e_n)^\top \in \mathbb{R}^n$是较大数值尺度上的误差. 此时模型(1)就变为:

$$y = Ax + \mu + e, \tag{3}$$

其中x是待估的回归参数, 它反应y与x之间的多元线性回归关系; 向量e是异常值识别变量, 设计矩阵A中的每一组数据都对应e的一个分量e_i, e_i 的大小代表了异常值的大小. 对于模型(3), 为了在估计回归参数和进行异常值检测的同时进行变量选择, 许多研究者考虑对变量x和误差e施加ℓ_0 惩罚, 得到以下模型:

$$\min_{x,e} \frac{1}{2}||y - Ax - e||^2 + \lambda_1\|x\|_0 + \lambda_2\|e\|_0, \tag{4}$$

其中, $\|x\|_0$是指向量x的0范数, 即向量中非零元素的个数. 但问题(4)是一个非光滑、NP难的问题 [7], 不容易求解, 因此有学者考虑用其他惩罚函数松弛ℓ_0惩罚. Nguyen等人 [8]和Chen等人 [9]基于最小二乘损失函数和ℓ_1惩罚构造如下模型:

$$\min_{x,e} \frac{1}{2}||y - Ax - e||^2 + \lambda_1||x||_1 + \lambda_2||e||_1. \tag{5}$$

Katayama等人 [10]对x应用ℓ_1惩罚, 对e应用Adaptive-ℓ_1惩罚, 得到如下模型:

$$\min_{x,e} \frac{1}{2}||y - Ax - e||^2 + \lambda_1||x||_1 + \lambda_2 \sum_{i=1}^{n} w_i|e_i|. \tag{6}$$

问题(5)、(6)中采用的ℓ_1范数为ℓ_0范数的最紧凸松弛. 带有ℓ_1惩罚的回归模型又被称为Lasso回归, 同岭回归类似, Lasso回归是一种压缩估计, 它在回归拟合过程中不仅可以对回归参数进行压缩, 还可以把一些不重要的变量的回归系数压缩为零, 从而实现变量选择, 这是Lasso回归的优点, 但它也有一些缺点: 应用ℓ_1惩罚虽然可以使模型解得的估计量满足连续性, 但不满足无偏性 [11]. Fan和Li [11]提出: 好的正则函数应该使估计量满足无偏性、稀疏性、连续性、Oracle性质, 其中Oracle性质是指求解模型得到的估计量与Oracle解有相同渐进分布. 这些标准得到了学界的广泛认可. 研究表明, 几类折叠凹函数 [12, 13] 是满足上述良好性质的惩罚函数, 例如SCAD [11]、MCP [14]、Capped-ℓ_1 [15]. 此外, 上述模型中最小二乘损失函数会放大异常值对回归结果的影响, 无法在数据中存在异常值的时候得到精确的估计 [16]. Huber函数 [17]不仅保留了最小二乘损失函数所具有的光滑性质, 而且具有鲁棒性. 在包含重尾误差的高维数据处理中, Huber函数引起了广泛的关注 [18–20], 因此为了获取稳健的参数估计, 本文引入了Huber损失函数, 得到如下模型:

$$\min_{x,e} \frac{1}{n} \sum_{i=1}^{n} h_\alpha(a_i^T x + e_i - y_i) + \lambda_1||x||_0 + \lambda_2||e||_0, \tag{7}$$

其中, $\lambda_1>0,\lambda_2>0$是惩罚参数, $h_\alpha(t)$是Huber函数:

$$h_\alpha(t)=\begin{cases}\frac{t^2}{2}, & |t|\le\alpha\\ \alpha|t|-\frac{\alpha^2}{2}, & |t|>\alpha\end{cases},\quad \forall t\in\mathbb{R}.$$

基于折叠凹函数的优良性质, 我们将问题(7)中的ℓ_0惩罚松弛为Capped-ℓ_1惩罚, 考虑以下非凸松弛问题:

$$\min_{x,e}\frac{1}{n}\sum_{i=1}^{n}h_\alpha(a_i^Tx+e_i-y_i)+\lambda_1\sum_{j=1}^{p}\varphi(x_j)+\lambda_2\sum_{i=1}^{n}\varphi(e_i), \tag{8}$$

其中$\varphi(t)$是Capped-ℓ_1惩罚函数:

$$\varphi(t)=\min\{1,\frac{|t|}{v}\},\ \ v>0.$$

本文结构如下, 第二节分析松弛问题(8)的最优性条件, 给出其方向稳定点的刻画及下界性质; 第三节研究原问题(7)与松弛问题(8)解的等价性; 第四节, 提出松弛问题(8)的光滑化方法, 并探讨光滑问题与松弛问题(8)稳定点的一致性. 为通过使用光滑化方法求解松弛问题(8)提供理论保证.

为了下文描述的简便, 记

$$H(x,e)=\frac{1}{n}\sum_{i=1}^{n}h_\alpha(a_i^Tx+e_i-y_i),$$

$$\Phi_1(x)=\sum_{j=1}^{p}\varphi(x_j),\ \forall x\in\mathbb{R}^p,\quad \Phi_2(e)=\sum_{i=1}^{n}\varphi(e_i),\ \forall e\in\mathbb{R}^n.$$

2. 松弛问题(8)的方向稳定点

本节给出松弛问题(8)的方向稳定点的刻画.

2.1. 松弛问题(8)的方向稳定点

首先我们给出函数的方向导数和方向稳定点的定义.

定义 2.1 *设函数$f:\mathbb{R}^n\to\mathbb{R}$在$x$处方向可微, $x\in dom f$, $d\in\mathbb{R}^n$, 若极限*

$$f'(x;d)=\lim_{t\downarrow 0}\frac{f(x+td)-f(x)}{t} \tag{9}$$

存在, 则称该极限是函数f在点x处沿方向$d\in\mathbb{R}^n$的方向导数.

定义 2.2 称$(x^*, e^*) \in \mathbb{R}^{p+n}$是松弛问题(8)的方向稳定点, 若

$$H'(x^*, e^*; x - x^*, e - e^*) + \lambda_1 \Phi_1'(x^*; x - x^*) + \lambda_2 \Phi_2'(e^*; e - e^*) \geq 0, \quad \forall x \in \mathbb{R}^p,\ \forall e \in \mathbb{R}^n. \tag{10}$$

根据上述方向导数和方向稳定点的定义, 容易得到以下结论:

(1) 若f是可微的, 则$f'(x; d) = \langle \nabla f(x), d \rangle$.

(2) 设x^*是函数$f(x)$的方向稳定点, 若$f(x)$在x^*处可微, 则$\nabla f(x^*) = 0$, 进而$f'(x^*; d) = \langle \nabla f(x^*), d \rangle = 0,\ \forall d \in \mathbb{R}^n$.

下面给出问题(8)的方向稳定点的刻画.

定理 2.1 若$(x^*, e^*) \in \mathbb{R}^{p+n}$是松弛问题(8)的方向稳定点, 则

$$\frac{1}{n} \sum_{i=1}^{n} h_\alpha'(a_i^T x^* + e_i^* - y_i)[a_i^T (x - x^*) + (e_i - e_i^*)] + \lambda_1 \sum_{j=1}^{p} \varphi'(x_j^*; x_j - x_j^*) + \lambda_2 \sum_{i=1}^{n} \varphi'(e_i^*; e_i - e_i^*) \geq 0,$$
$$\forall x \in \mathbb{R}^p,\ \forall e \in \mathbb{R}^n,$$

其中, $h'(t)$是$Huber$函数的导数:

$$h_\alpha'(t) = \begin{cases} t, & |t| < \alpha, \\ sgn(t)\alpha, & |t| > \alpha, \end{cases} \quad \forall t \in \mathbb{R},$$

$sgn(t)$是符号函数:

$$sgn(t) = \begin{cases} -1, & t < 0, \\ 0, & t = 0, \\ 1, & t > 0, \end{cases} \quad \forall t \in \mathbb{R},$$

$\varphi'(t^*; t - t^*)$是$Capped$-ℓ_1函数在点t^*处关于方向$(t - t^*)$的方向导数:

$$\varphi'(t^*; t - t^*) = \begin{cases} \frac{|t|}{v}, & t^* = 0, \\ \frac{\mathrm{sgn}(t^*)(t - t^*)}{v}, & |t^*| \in (0, v), \\ \min\{0, \frac{\mathrm{sgn}(t^*)(t - t^*)}{v}\}, & |t^*| = v, \\ 0, & \text{otherwise}, \end{cases} \quad \forall t \in \mathbb{R}.$$

证明. 因为Huber函数是可微函数, 所以

$$\begin{aligned} H'(x^*, e^*; x - x^*, e - e^*) &= \langle \nabla_{x^*} H(x^*, e^*), x - x^* \rangle + \langle \nabla_{e^*} H(x^*, e^*), e - e^* \rangle \\ &= \frac{1}{n} \sum_{i=1}^{n} h_\alpha'(a_i^T x^* + e_i^* - y_i) a_i^T (x - x^*) + \frac{1}{n} \sum_{i=1}^{n} h_\alpha'(a_i^T x^* + e_i^* - y_i)(e_i - e_i^*) \\ &= \frac{1}{n} \sum_{i=1}^{n} h_\alpha'(a_i^T x^* + e_i^* - y_i)[a_i^T (x - x^*) + (e_i - e_i^*)]. \end{aligned} \tag{11}$$

根据方向导数的定义和函数$\Phi_1(x), \Phi_2(e)$的可分性, 得

$$\Phi_1'(x^*; x-x^*) = \sum_{j=1}^{p} \varphi'(x_j^*; x_j - x_j^*), \Phi_2'(e^*; e-e^*) = \sum_{i=1}^{n} \varphi'(e_i^*; e_i - e_i^*),$$

其中, 根据方向导数的定义和

$$\varphi(t) = \min\{1, \frac{|t|}{v}\} = \begin{cases} \frac{|t|}{v}, & |t| \le v, \\ 1, & |t| > v, \end{cases}$$

可以得到$\varphi'(t, t-t^*)$的表达式. 证毕. □

Peng等人 [21–23]证明了方向稳定点具有如下局部最优性质:

引理 2.1 *设函数$f: \mathbb{R}^n \to \mathbb{R}$在点$x^* \in \mathbb{R}^n$处是局部Lipschitz连续且方向可微的, 则*

(1) *若x^*是函数f的局部极小值点, 则x^*是函数f的方向稳定点;*

(2) *x^*是严格局部极小值点并满足一阶增长性条件, 即存在x^*的邻域$\mathcal{W}$和$\delta > 0$, 使得*

$$f(x) \ge f(x^*) + \delta \parallel x - x^* \parallel, \quad \forall x \in \mathcal{W},$$

当且仅当x^*满足

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

根据引理2.1, 因为松弛问题(8)的目标函数是局部Lipschitz连续且方向可微的, 因此x^*是问题(8)的局部极小值的必要条件是x^*是问题(8)的方向稳定点.

3. 下界性质及解的等价性

受文 [21–24]启发, 本节给出松弛问题(8)方向稳定点的下界性质, 并证明在一定条件下原问题(7)和松弛问题(8)的解等价.

3.1. 解的下界性质

首先定义如下指标集:

$P_1(x) = \{j : 0 < |x_j| < v\}$, $P_2(x) = \{j : |x_j| \ge v\}$, $P_1(x) \cup P_2(x) = \{j \in \{1, \ldots, p\} : x_j \ne 0\}$,
$P_3(e) = \{i : 0 < |e_i| < v\}$, $P_4(e) = \{i : |e_i| \ge v\}$, $P_3(e) \cup P_4(e) = \{i \in \{1, \ldots, n\} : e_i \ne 0\}$.

定义函数$M: \mathbb{R}^{p+n} \to \mathbb{R}$如下:

$$M(x, e) = \max\left\{ \frac{1}{\lambda_1 n} \left\| \left| \sum_{i=1}^{n} h_\alpha'(a_i^T x + e_i - y_i) a_i \right| \right\|_1, \ \frac{1}{\lambda_2 n} \sum_{i=1}^{n} \left| h_\alpha'(a_i^T x + e_i - y_i) \right| \right\}. \tag{12}$$

下面的定理表明: 在一定条件下, 松弛问题(8)的方向稳定点的非零分量具有一致的正下界.

定理 3.1 若$(x^*,e^*)\in\mathbb{R}^{p+n}$是松弛问题(8)的方向稳定点, 且满足$M(x^*,e^*)<\frac{1}{v}$, 则

(i) 要么$|x_j^*|\ge v$, 要么$|x_j^*|=0$, $j\in\{1,\dots,p\}$;

(ii) 要么$|e_i^*|\ge v$, 要么$|e_i^*|=0$, $i\in\{1,\dots,n\}$.

证明. 设$(x^*,e^*)\in\mathbb{R}^{p+n}(x_j^*\neq 0,\ \forall j;\ e_i^*\neq 0,\ \forall i)$是松弛问题(8)的方向稳定点. 记$x^*=(x_1^*,...,x_p^*)^\top$, $e^*=(e_1^*,...,e_n^*)^\top$. 只需证明$P_1(x^*)=\emptyset$ 和$P_3(e^*)=\emptyset$.

下面用反证法证明. 假设$P_1(x^*)\neq\emptyset$ 或$P_3(e^*)\neq\emptyset$.

(i)假设$P_1(x^*)\neq\emptyset$. 由定理2.1, 取$e=e^*$, 则

$$\frac{1}{n}\sum_{i=1}^{n}h_\alpha'(a_i^Tx^*+e_i^*-y_i)a_i^T(x-x^*)+\lambda_1\sum_{j=1}^{p}\varphi'(x_j^*;x_j-x_j^*)\ge 0,\quad \forall x\in\mathbb{R}^p. \tag{13}$$

$\forall i=1,...,p$, 记$a_i=(a_{i1}^*,...,a_{ip}^*)^\top$, 则

$$\frac{1}{n}\sum_{i=1}^{n}h_\alpha'(a_i^Tx^*+e_i^*-y_i)a_i^T(x-x^*)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}h_\alpha'(a_i^Tx^*+e_i^*-y_i)a_{ij}(x_j-x_j^*). \tag{14}$$

$\forall\varepsilon_1\in(0,1)$, 取$\widehat{x}=(\widehat{x}_1,\dots,\widehat{x}_p)^\top\in\mathbb{R}^p$满足

$$\widehat{x}_j=\begin{cases}(1-\varepsilon_1)x_j^*, & j\in P_1(x^*),\\ x_j^*, & j\in P_2(x^*).\end{cases}$$

由于$P_1(x^*)\neq\emptyset$, 故$(\widehat{x}-x^*)$不是零向量. 由于(13)对任意$x\in\mathbb{R}^p$均成立, 则

$$\begin{aligned}\sum_{i=1}^{n}h_\alpha'(a_i^Tx^*+e_i^*-y_i)a_i^T(\widehat{x}-x^*) &= \sum_{i=1}^{n}\sum_{j=1}^{p}h_\alpha'(a_i^Tx^*+e_i^*-y_i)a_{ij}(\widehat{x}_j-x_j^*)\\ &= \varepsilon_1\sum_{i=1}^{n}\sum_{j\in P_1(x^*)}h_\alpha'(a_i^Tx^*+e_i^*-y_i)a_{ij}(-x_j^*)\\ &\le \varepsilon_1\left\|\sum_{i=1}^{n}h_\alpha'(a_i^Tx^*+e_i^*-y_i)a_i\right\|_1\cdot\sum_{j\in P_1(\widehat{x})}|x_j^*|.\end{aligned} \tag{15}$$

另一方面,

$$\sum_{j=1}^{p}\varphi'(x_j^*;\widehat{x}_j-x_j^*)=\sum_{j\in P_1(\widehat{x})}\frac{sgn(x_j^*)(-\varepsilon_1x_j^*)}{v}=-\varepsilon_1\sum_{j\in P_1(\widehat{x})}\frac{sgn(x_j^*)(x_j^*)}{v}=-\frac{\varepsilon_1}{v}\sum_{j\in P_1(\widehat{x})}|x_j^*|. \tag{16}$$

由(13)-(16), 得

$$\lambda_1 \frac{\varepsilon_1}{v} \sum_{j\in P_1(\widehat{x})} |x_j^*| \le \frac{\varepsilon_1}{n} \left\| \sum_{i=1}^{n} h'_\alpha(a_i^T x^* + e_i^* - y_i) a_i \right\|_1 \sum_{j\in P_1(\widehat{x})} |x_j^*|,$$

由于 $\sum\limits_{j\in P_1(x^*)} |x_j^*| \neq 0$, 将其消去得

$$\frac{1}{v} \le \frac{1}{\lambda_1 n} \left\| \sum_{i=1}^{n} h'_\alpha(a_i^T x^* + e_i^* - y_i) a_i \right\|_1 \le M(x^*, e^*),$$

这与$M(x^*, e^*) < \frac{1}{v}$ 矛盾, 从而$P_1(x^*) = \emptyset$.

(ii)假设$P_3(e^*) \neq \emptyset$. 由定理2.1, 取$x = x^*$, 则

$$\frac{1}{n} \sum_{i=1}^{n} h'_\alpha(a_i^T x^* + e_i^* - y_i)(e_i - e_i^*) + \lambda_2 \sum_{i=1}^{n} \varphi'(e_i^*; e_i - e_i^*) \ge 0, \quad \forall e \in \mathbb{R}^n. \tag{17}$$

$\forall \varepsilon_2 \in (0,1)$, 取$\widehat{e} = (\widehat{e}_1, \ldots, \widehat{e}_n)^\top \in \mathbb{R}^n$满足

$$\widehat{e}_i = \begin{cases} (1-\varepsilon_2) e_i^*, & i \in P_3(e^*), \\ e_i^*, & i \in P_4(e^*). \end{cases}$$

由于$P_3(e^*)$非空, 故$(\widehat{e} - e^*)$不是零向量, 进而

$$\begin{aligned} \sum_{i=1}^{n} h'_\alpha(a_i^T x^* + e_i^* - y_i)(\widehat{e}_i - e_i^*) &= \varepsilon_2 \sum_{i\in P_3(e^*)} h'_\alpha(a_i^T x^* + e_i^* - y_i) e_i^* \\ &\le \varepsilon_2 \sum_{i\in P_3(e^*)} |h'_\alpha(a_i^T x^* + e_i^* - y_i)| \cdot |e_i^*|. \end{aligned} \tag{18}$$

另一方面,

$$\sum_{i=1}^{n} \varphi'(e_i^*; \widehat{e}_i - e_i^*) = \sum_{i\in P_3(e^*)} \varphi'(e_i^*; \widehat{e}_i - e_i^*) = -\frac{\varepsilon_2}{v} \sum_{i\in P_3(e^*)} sgn(e_i^*)(e_i^*) = -\frac{\varepsilon_2}{v} \sum_{i\in P_3(e^*)} |e_i^*|. \tag{19}$$

由(17)-(19), 得

$$\lambda_2 \frac{\varepsilon_2}{v} \sum_{i\in P_3(e^*)} |e_i^*| \le \frac{\varepsilon_2}{n} \sum_{i\in P_3(e^*)} |h'_\alpha(a_i^T x^* + e_i^* - y_i)| \cdot |e_i^*| \le \frac{\varepsilon_2}{n} \sum_{i=1}^{n} |h'_\alpha(a_i^T x^* + e_i^* - y_i)| \sum_{i\in P_3(e^*)} |e_i^*|,$$

$$\frac{1}{v} \le \frac{1}{\lambda_2 n} \sum_{i=1}^{n} |h'_\alpha(a_i^T x^* + e_i^* - y_i)|.$$

由于 $\sum\limits_{i\in P_3(e^*)} |e_i^*| \neq 0$, 将其消去得$\frac{1}{v} \le M(x^*, e^*)$, 与假设矛盾, 故$P_3(e^*) = \emptyset$. 证毕. □

3.2. 问题(7)与(8)解的等价性

下面借助定理3.1研究原问题(7)与松弛问题(8)的全局最优解之间的关系.

定理 3.2 设$(x^*, e^*) \in \mathbb{R}^{p+n}$ 满足$M(x^*, e^*) < \frac{1}{v}$, 则$(x^*, e^*)$ 是松弛问题(8)的全局最优解当且仅当(x^*, e^*)是原问题(7)的全局最优解.

证明. (1) 设$(x^*, e^*) \in \mathbb{R}^{p+n}$是松弛问题(8)的全局最优解, 由引理2.1, (x^*, e^*)是松弛问题(8)的方向稳定点. 由定理3.1, $\Phi_1(x^*) = \|x^*\|_0$, $\Phi_2(e^*) = \|e^*\|_0$. 注意到, $\forall (x, e) \in \mathbb{R}^{p+n}$, 总有$\Phi_1(x) \leq \|x\|_0, \Phi_2(e) \leq \|e\|_0$. 从而, $\forall (x, e) \in \mathbb{R}^{p+n}$, 有

$$\begin{aligned} H(x^*, e^*) + \lambda_1\|x^*\|_0 + \lambda_2\|e^*\|_0 &= H(x^*, e^*) + \lambda_1\Phi_1(x^*) + \lambda_2\Phi_2(e^*) \\ &\leq H(x, e) + \lambda_1\Phi_1(x) + \lambda_2\Phi_2(e) \\ &\leq H(x, e) + \lambda_1\|x\|_0 + \lambda_2\|e\|_0. \end{aligned}$$

因此, (x^*, e^*)是原问题(7)的全局最优解.

(2) 设$(x^*, e^*) \in \mathbb{R}^{p+n}$是原问题(7)的全局最优解, 但假设它不是松弛问题(8)的全局最优解. 设(x', e')是松弛问题(8)的全局最优解, 由定理3.1, 得$\Phi(x') = \|x'\|_0$, $\Phi(e') = \|e'\|_0$. 于是,

$$\begin{aligned} H(x', e') + \lambda_1\|x'\|_0 + \lambda_2\|e'\|_0 &= H(x', e') + \lambda_1\Phi_1(x') + \lambda_2\Phi_2(e') \\ &< H(x^*, e^*) + \lambda_1\Phi_1(x^*) + \lambda_2\Phi_2(e^*) \\ &\leq H(x^*, e^*) + \lambda_1\|x^*\|_0 + \lambda_2\|e^*\|_0. \end{aligned}$$

这与(x^*, e^*)是原问题(7)的全局最优解矛盾, 因此$(x^*, e^*) \in \mathbb{R}^{p+n}$是松弛问题(8)的全局最优解. □

4. 松弛问题(8)的光滑化方法

由于采用了非凸非光滑的正则, 松弛问题(8)是非光滑优化, 求解松弛问题(8)的非常典型的一类方法是光滑化方法. 光滑化方法一个核心问题是光滑化前后两个模型解的一致性.

本节先给出松弛问题(8)的光滑化模型, 然后证明光滑问题的稳定点与松弛问题(8)的方向稳定点具有一致性.

4.1. 问题(8)的光滑化模型

Capped-ℓ_1函数是一类非常典型的折叠凹函数, 它可以写为两个凸函数的差(DC) [21, 25], 其光滑化函数的构造基于其凸差表示:

$$\varphi(t) = \min\left\{1, \frac{|t|}{v}\right\} = 1 - \max\left\{0, 1 - \frac{|t|}{v}\right\} = g(t) - h(|t|), \tag{20}$$

其中

$$g(t)=1,\ h(|t|)=\max\left\{0,1-\frac{|t|}{v}\right\}.$$

记$m(t)=|t|$, 容易验证其方向导数为:

$$m'(t;d)=\begin{cases}|d|, & t=0,\\ \frac{t}{|t|}d, & t\neq 0.\end{cases}$$

由于$g(t)$是光滑的, 我们只需将$h(|t|)$光滑化. 已知光滑函数的复合函数仍是光滑的, 因此我们可以分别构造外函数$h(s)$与内函数$m(t)$的光滑函数$\widetilde{h}_\mu(s)$与$\widetilde{m}_\mu(t)$, 再将它们复合, 便可得到连续可微的光滑函数$\widetilde{h}_\mu(t)=\widetilde{h}_\mu\circ\widetilde{m}_\mu(t)$.

$m(t)$**的光滑化:** 取$m(t)$的光滑函数为:

$$\widetilde{m}_\mu(t)=\sqrt{t^2+\mu},\quad \mu>0, \tag{21}$$

其中μ是光滑化参数. $\widetilde{m}_\mu(t)$的导数和方向导数分别为:

$$\widetilde{m}'_\mu(t)=\frac{t}{\sqrt{t^2+\mu}},\quad \widetilde{m}'_\mu(t;d)=\widetilde{m}'_\mu(t)d=\frac{td}{\sqrt{t^2+\mu}}.$$

可以证明光滑函数$\widetilde{m}_\mu(t)$满足如下性质 [21]:

(1) (光滑函数的一致收敛性):

$$\lim_{w\to t,\mu\downarrow 0}\widetilde{m}_\mu(w)=m(t),\ \forall t\in\mathbb{R}.$$

(2) (方向导数的一致收敛性与弱一致收敛性):

$$\lim_{w\to t,\mu\downarrow 0}\widetilde{m}'_\mu(w;d)=\lim_{w\to t,\mu\downarrow 0}\widetilde{m}'_\mu(w)d=m'(t)d=m'(t,d),\ \forall t\in\mathbb{R}\setminus\{0\},\ \forall d\in\mathbb{R},$$

$$\limsup_{w\to 0,\mu\downarrow 0}\widetilde{m}'_\mu(w;d)=\limsup_{w\to 0,\mu\downarrow 0}\frac{td}{\sqrt{t^2+\mu}}=|d|=m'(0,d),\ \forall d\in\mathbb{R}.$$

$h(s)$**的光滑化:** 对$h(s)=\max\{0,1-\frac{s}{v}\}$, 可以采用$\widehat{s}(z,\mu)=\frac{1}{2}\left(z+\sqrt{z^2+\mu^2}\right)$来光滑化正部函数$\max\{0,z\}$ [21,26]. 再将$z=1-\frac{s}{v}$ 代入$\widehat{s}(z,\mu)$, 得到$h(s)$的光滑化函数为

$$\widetilde{h}_\mu(s)=\widehat{s}(1-\frac{s}{v},\mu)=\frac{1}{2}\left[1-\frac{s}{v}+\sqrt{(1-\frac{s}{v})^2+\mu^2}\right]. \tag{22}$$

容易证明光滑函数$\widetilde{h}_\mu(s)$具有如下性质:

(1) (光滑函数的一致收敛性):

$$\lim_{w\to s,\mu\downarrow 0}\widetilde{h}_\mu(w)=\frac{1}{2}\left(1-\frac{s}{v}+|1-\frac{s}{v}|\right)=h(s),\ \forall s\in\mathbb{R}.$$

(2) (方向导数的一致收敛性与弱一致收敛性):

$$\lim_{w\to s,\mu\downarrow 0} \widetilde{h}'_\mu(w;d) = h'(s,d),\ \forall s \in \mathbb{R}\setminus\{0\},\ \forall d \in \mathbb{R},$$

$$\limsup_{w\to 0,\mu\downarrow 0} \widetilde{h}'_\mu(w) = h'_+(0).$$

结合光滑函数$\widetilde{m}_\mu(x)$和光滑函数$\widetilde{h}_\mu(s)$, 我们得到松弛问题(8)的光滑化模型如下:

$$\min_{x,e}\ \widetilde{F}_\mu(x,e) := H(x,e) + \lambda_1 p + \lambda_2 n - \lambda_1 \sum_{j=1}^{p} \widetilde{h}_\mu \circ \widetilde{m}_\mu(x_j) - \lambda_2 \sum_{i=1}^{n} \widetilde{h}_\mu \circ \widetilde{m}_\mu(e_i). \tag{23}$$

4.2. 光滑问题(23)与松弛问题(8)稳定点的一致性

问题(23)是光滑的, 其稳定点是使$\nabla\widetilde{F}_\mu(x,e) = 0$的点.

本节证明光滑问题(23)的稳定点与松弛问题(8)的方向稳定点具有一致性.

定义指标集:

$$\mathcal{P}_1(x) = \{j \in \{1,\dots,p\} : x_j \neq 0\},\quad \mathcal{P}_2(x) = \{j \in \{1,\dots,p\} : x_j = 0\},$$

$$\mathcal{Q}_2(e) = \{i \in \{1,\dots,n\} : e_i \neq 0\},\quad \mathcal{Q}_2(e) = \{i \in \{1,\dots,n\} : e_i = 0\}.$$

定理 4.1 设在$\mu = \mu_k$时, $(x^*_{\mu_k}, e^*_{\mu_k})$是光滑问题(23)的稳定点, 则当光滑参数$\mu_k \downarrow 0$ $(k \to \infty)$时, $\{(x^*_{\mu_k}, e^*_{\mu_k})\}_{k=1}^\infty$的任意聚点都是松弛问题(8)的方向稳定点.

证明. 设(x^*, e^*)为点列$\{(x^*_{\mu_k}, e^*_{\mu_k})\}_{k=1}^\infty$的一个聚点, 不妨设

$$\lim_{k\to\infty} x^*_{\mu_k} = x^*,\ \lim_{k\to\infty} e^*_{\mu_k} = e^*.$$

因为$(x^*_{\mu_k}, e^*_{\mu_k})$是光滑问题(23)的稳定点, 故$\nabla\widetilde{F}_\mu(x^*_{\mu_k}, e^*_{\mu_k}) = 0$. 则 $\forall d^{(1)} \in \mathbb{R}^p$, $\forall d^{(2)} \in \mathbb{R}^n$, 有

$$\begin{aligned}
0 &= \widetilde{F}'_\mu(x^*_{\mu_k}, e^*_{\mu_k}; d^{(1)}, d^{(2)}) = \langle \nabla\widetilde{F}'_\mu(x^*_{\mu_k}, e^*_{\mu_k}), d^{(1)}, d^{(2)}\rangle \\
&= \langle \nabla H(x^*_{\mu_k}, e^*_{\mu_k}), d^{(1)}, d^{(2)}\rangle - \lambda_1 \sum_{j=1}^{p} \widetilde{h}'_\mu \circ \widetilde{m}_\mu(x^*_{\mu_k(j)})\widetilde{m}'_\mu(x^*_{\mu_k(j)})d^{(1)}_j \\
&\quad -\lambda_2 \sum_{i=1}^{n} \widetilde{h}'_\mu \circ \widetilde{m}_\mu(e^*_{\mu_k(i)})\widetilde{m}'_\mu(e^*_{\mu_k(i)})d^{(2)}_i \\
&= \langle \nabla H(x^*_{\mu_k}, e^*_{\mu_k}), d^{(1)}, d^{(2)}\rangle - \lambda_1 \left(\sum_{j\in\mathcal{P}_1(x^*)} + \sum_{j\in\mathcal{P}_2(x^*)}\right) \widetilde{h}'_\mu \circ \widetilde{m}_\mu(x^*_{\mu_k(j)})\widetilde{m}'_\mu(x^*_{\mu_k(j)})d^{(1)}_j \\
&\quad -\lambda_2 \left(\sum_{j\in\mathcal{Q}_1(e^*)} + \sum_{j\in\mathcal{Q}_2(e^*)}\right) \widetilde{h}'_\mu \circ \widetilde{m}_\mu(e^*_{\mu_k(i)})\widetilde{m}'_\mu(e^*_{\mu_k(i)})d^{(2)}_i.
\end{aligned} \tag{24}$$

根据光滑函数$\widetilde{m}_\mu(x)$和$\widetilde{h}_\mu(t)$的性质, 对等式(24)两边取极限$(k\to\infty)$, 则 $\forall d^{(1)}\in\mathbb{R}^p$, $\forall d^{(2)}\in\mathbb{R}^n$, 有

$$\begin{aligned}
0 &= \lim_{k\to\infty}\widetilde{F}'_\mu(x^*_{\mu_k},e^*_{\mu_k};d^{(1)},d^{(2)})\\
&= \lim_{k\to\infty}\langle\nabla H(x^*_{\mu_k},e^*_{\mu_k}),d^{(1)},d^{(2)}\rangle-\lambda_1\sum_{j\in\mathcal{P}_1(x^*)}\lim_{k\to\infty}\widetilde{h}'_\mu\circ\widetilde{m}_\mu(x^*_{\mu_k(j)})\widetilde{m}'_\mu(x^*_{\mu_k(j)})d^{(1)}_j\\
&\quad-\lambda_1\sum_{j\in\mathcal{P}_2(x^*)}\lim_{k\to\infty}\widetilde{h}'_\mu\circ\widetilde{m}_\mu(x^*_{\mu_k(j)})\widetilde{m}'_\mu(x^*_{\mu_k(j)})d^{(1)}_j-\lambda_2\sum_{i\in\mathcal{Q}_1(e^*)}\lim_{k\to\infty}\widetilde{h}'_\mu\circ\widetilde{m}_\mu(e^*_{\mu_k(i)})\widetilde{m}'_\mu(e^*_{\mu_k(i)})d^{(2)}_i\\
&\quad-\lambda_2\sum_{i\in\mathcal{Q}_2(e^*)}\lim_{k\to\infty}\widetilde{h}'_\mu\circ\widetilde{m}_\mu(e^*_{\mu_k(i)})\widetilde{m}'_\mu(e^*_{\mu_k(i)})d^{(2)}_i\\
&\le \lim_{k\to\infty}\langle\nabla H(x^*_{\mu_k},e^*_{\mu_k}),d^{(1)},d^{(2)}\rangle-\lambda_1\sum_{j\in\mathcal{P}_1(x^*)}\lim_{k\to\infty}\widetilde{h}'_\mu\circ\widetilde{m}_\mu(x^*_{\mu_k(j)})\widetilde{m}'_\mu(x^*_{\mu_k(j)})d^{(1)}_j\\
&\quad-\lambda_1\sum_{j\in\mathcal{P}_2(x^*)}\limsup_{k\to\infty}\widetilde{h}'_\mu\circ\widetilde{m}_\mu(x^*_{\mu_k(j)})\cdot\limsup_{k\to\infty}\widetilde{m}'_\mu(x^*_{\mu_k(j)})d^{(1)}_j\\
&\quad-\lambda_2\sum_{i\in\mathcal{Q}_1(e^*)}\lim_{k\to\infty}\widetilde{h}'_\mu\circ\widetilde{m}_\mu(e^*_{\mu_k(i)})\widetilde{m}'_\mu(e^*_{\mu_k(i)})d^{(2)}_i\\
&\quad-\lambda_2\sum_{i\in\mathcal{Q}_2(e^*)}\limsup_{k\to\infty}\widetilde{h}'_\mu\circ\widetilde{m}_\mu(e^*_{\mu_k(i)})\cdot\limsup_{k\to\infty}\widetilde{m}'_\mu(e^*_{\mu_k(i)})d^{(2)}_i\\
&= \langle\nabla H(x^*,e^*),d^{(1)},d^{(2)}\rangle-\lambda_1\sum_{j=1}^p h'\circ m(x^*_j)m'(x^*_j)d^{(1)}_j-\lambda_2\sum_{i=1}^n h'\circ m(e^*_i)m'(e^*_i)d^{(2)}_i\\
&= \langle\nabla H(x^*,e^*),d^{(1)},d^{(2)}\rangle-\lambda_1\sum_{j=1}^p h'(m(x^*_j))m'(x^*_j;d^{(1)}_j)-\lambda_2\sum_{i=1}^n h'(m(e^*_i))m'(e^*_i;d^{(2)}_i)\\
&= \langle\nabla H(x^*,e^*),d^{(1)},d^{(2)}\rangle+\lambda_1\sum_{j=1}^p[g'(m(x^*_j))-h'(m(x^*_j))]m'(x^*_j;d^{(1)}_j)\\
&\quad+\lambda_2\sum_{i=1}^n[g'(m(e^*_i))-h'(m(e^*_i))]m'(e^*_i;d^{(2)}_i)\\
&= H'(x^*,e^*;d^{(1)},d^{(2)})+\lambda_1\sum_{j=1}^p\varphi'(x^*_j;d^{(1)}_j)+\lambda_2\sum_{i=1}^n\varphi'(e^*_i;d^{(2)}_i)\\
&= H'(x^*,e^*;d^{(1)},d^{(2)})+\lambda_1\Phi'_1(x^*;d^{(1)})+\lambda_2\Phi'_2(e^*;d^{(2)}).
\end{aligned}$$

因此, (x^*,e^*)是松弛问题(8)的方向稳定点. □

5. 总结

基于多元线性回归模型、Huber损失函数和均值漂移模型, 本文对具有ℓ_0惩罚的原问题模型(7), 用Capped-ℓ_1惩罚进行松弛得到松弛问题(8), 定义了松弛问题(8)的方向稳定点, 分析了松弛问题(8)方向稳定点的下界性质, 并探讨了原问题(7)和松弛问题(8)解的等价性, 为求解松弛问题(8), 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. 下一步将通过数值实验和算例进

一步检验算法的实际效果.

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] Papageorgiou, G., Bouboulis, P. and Theodoridis, S. (2015) Robust Linear Regression Analysis—A Greedy Approach. *IEEE Transactions on Signal Processing*, **63**, 3872-3887. https://doi.org/10.1109/TSP.2015.2430840

[2] Huber, P. (1972) The 1972 Wald Lecture Robust Statistics: A Review. *Annals of Mathematical Statistics*, **43**, 1041-1067. https://doi.org/10.1214/aoms/1177692459

[3] Rousseeuw, P. and Leroy, A. (1987) Robust Regression and Outlier Detection. Wiley, New York, NY. https://doi.org/10.1002/0471725382

[4] Maronna, R., Martin, R. and Yohai, V. (2006) Robust Statistics: Theory and Methods. Wiley, New York, NY.

[5] Huber, P. (1981) Robust Statistics. Wiley, New York, NY.

[6] Cook, R. and Weisberg, S. (1982) Residuals and Influence in Regression. Chapman and Hall, New York, NY.

[7] Natarajan, B. (1995) Sparse Approximate Solutions to Linear Systems. *SIAM Journal on Computing*, **24**, 227-234. https://doi.org/10.1137/S0097539792240406

[8] Nguyen, N. and Tran, T. (2013) Robust Lasso with Missing and Grossly Corrupted Observations. *IEEE Transactions on Information Theory*, **59**, 2036-2058. https://doi.org/10.1109/TIT.2012.2232347

[9] Chen, J. and Liu, Y. (2019) Stable Recovery of Structured Signals from Corrupted Subgaussian Measurements. *IEEE Transactions on Information Theory*, **65**, 2976- 2994. https://doi.org/10.1109/TIT.2018.2890194

[10] Katayama, S. and Fujisawa, H. (2017) Sparse and Robust Linear Regression: An Optimization Algorithm and Its Statistical Properties. *Statistica Sinica*, **27**, 1243-1264. https://doi.org/10.5705/ss.202015.0179

[11] Fan, J. and Li, Y. (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

[12] Ong, C. and An, L. (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

[13] Peleg, D. and Meir, R. (2008) A Bilinear Formulation for Vector Sparsity Optimization. *Signal Processing*, **88**, 375-389. https://doi.org/10.1016/j.sigpro.2007.08.015

[14] Zhang, C. (2010) Nearly Unbiased Variable Selection under Minimax Concave Penalty. *Annals of Statistics*, **38**, 894-942. https://doi.org/10.1214/09-AOS729

[15] Zhang, T. (2010) Analysis of Multi-Stage Convex Relaxation for Sparse Regularization. *Journal of Machine Learning Research*, **11**, 1081-1107.

[16] Candés, E., Romberg, J. and Tao, T. (2006) Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. *IEEE Transactions on Information Theory*, **52**, 489-509. https://doi.org/10.1109/TIT.2005.862083

[17] Huber, P. (1964) Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, **35**, 73-101. https://doi.org/10.1214/aoms/1177703732

[18] Fan, J., Li, Q. and Wang, Y. (2017) Estimation of High Dimensional Mean Regression in the Absence of Symmetry and Light Tail Assumptions. *Journal of Royal Statistical Society, Series B*, **79**, 247-265. https://doi.org/10.1111/rssb.12166

[19] Yi, C. and Huang, J. (2017) Semismooth Newton Coordinate Descent Algorithm for Elastic-Net Penalized Huber Loss Regression and Quantile Regression. *Journal of Computational and Graphical Statistics*, **26**, 547-557. https://doi.org/10.1080/10618600.2016.1256816

[20] Sun, Q., Zhou, W. and Fan, J. (2020) Adaptive Huber Regression. *Journal of the American Statistical Association*, **115**, 254-265. https://doi.org/10.1080/01621459.2018.1543124

[21] 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

[22] Zhang, X. and Peng, D. (2022) Solving Constrained Nonsmooth Group Sparse Optimization via Group Capped-L1 Relaxation and Group Smoothing Proximal Gradient Algorithm. *Computational Optimization and Applications*, **83**, 801-844. https://doi.org/10.1007/s10589-022-00419-2

[23] 彭定涛, 唐琦, 张弦. 组稀疏优化问题精确连续Capped-L1松弛[J]. 数学学报, 2022, 65(2): 243-262.

[24] 罗孝敏, 彭定涛, 张弦. 基于MCP正则的最小一乘回归问题研究[J]. 系统科学与数学, 2021, 41(8): 2327-2337.

[25] Ahn, M., Pang, J. and Jack, X. (2017) Difference-of-Convex Learning: Directional Stationarity, Optimality, and Sparsity. *SIAM Journal on Optimization*, **27**, 1637-1665. https://doi.org/10.1137/16M1084754

[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