doi:10.16055/j.issn.1672-058X.2020.0005.006

两个绝对值优化问题解的等价性*

罗孝敏, 彭定涛**
(贵州大学 数学与统计学院,贵阳 550025)

摘 要:针对损失函数为最小一乘问题,惩罚项由基数函数定义的绝对值优化问题,提出用 MCP (Minimax Concave Penalty)非凸正则来连续逼近基数罚,得到一个精确连续的绝对值优化松弛问题。首先,证明了带基数罚的绝对值优化问题的全局最优解;其次,研究了带基数罚的绝对值优化问题与带 MCP 罚的绝对值优化松弛问题之间全局最优解的等价性;最后,证明了在一定的条件下这两个绝对值优化问题具有相同的全局最优解。

关键词:绝对值优化问题;MCP;最优解;等价性

中图分类号:O224 **文献标志码**:A **文章编号**:1672-058X(2020)05-0037-06

0 引 言

稀疏优化在金融、计量经济学、图像处理、模式识别、变量选择、变量空间降维、稀疏重构、投资组合等多个领域都具有重要的作用[1-4]。其目的是获得欠定线性非零分量个数尽可能少的解,即 L_0 罚问题:

$$\min_{x \in R^n} G(x):=f(x)+\lambda\|x\|_0$$

其中,$f(x)$ 表示损失函数,例如最小一乘、最小二乘、Huber 函数等,$\|x\|_0$ 表示向量 x 非零分量的个数。在文献[5]中已经证明了 L_0 罚是 NP 难的。Lasso 是常用来松弛 L_0 罚问题的一种方法。有研究证明当数据矩阵满足零空间性质或限制等距条件,损失函数为最小二乘时,Lasso 问题与 L_0 罚问题是等价的[6],但 Lasso 得到的结果常常是有偏估计量[7],并且 Lasso 模型解缺少 Oracle 性质[8-10]。而 MCP,SCAD,capped-L_1 等折叠凹罚具有很好的性质,如估计量的无偏性、连续性、稀疏性等[9]。注意到,MCP 罚产生的估计量不仅具有 Oracle 性质,而且在理论及数值上优于 capped-L_1,SCAD[1]。

考虑损失函数为最小一乘时,L_0 罚问题与松弛问题解的等价性是否成立。在统计意义下,最小一乘损失可以有效地处理离群值,这使得最小一乘损失具有鲁棒性,所以考虑用 MCP 罚来连续近似逼近 L_0 罚。另外,由于高维回归问题非凸正则化的复杂性,目前关于高维非凸正则的相关理论成果稀少。

收稿日期:2019-12-04;**修回日期**:2020-01-04.
* **基金项目**:国家自然科学基金项目资助(11861020);贵州省高层次留学人才创新创业择优资助重点项目资助([2018]03);贵州省科技计划项目资助([2018]5781)和贵州省青年科技人才成长项目资助([2018]121).
作者简介:罗孝敏(1993—),女,贵州遵义人,硕士研究生,从事稀疏优化研究.
** **通讯作者**:彭定涛(1979—),男,湖北十堰人,教授,博士生导师,从事最优化理论与方法研究. Email:dingtaopeng@126.com.

因为带基数罚的最小一乘问题往往归结为 n 个一维绝对值优化问题,因此对其对应的绝对值优化问题进行研究是研究带基数罚的最小一乘问题的基础。本文研究绝对值优化问题 MCP 松弛模型与 L_0 罚问题之间解的关系,讨论这两个问题是否具有相同的全局最优解,期望为研究高维回归问题非凸正则化提供一定的思路。

文章的结构如下,第一节主要介绍研究的两个优化模型,第二节证明了两个优化模型之间解的等价性。

1 优化模型

首先,考虑原始绝对值优化问题:

$$\min_{x\in\mathbf{R}} g(x):=|ax-b|+\lambda|x|_0 \tag{1}$$

这里 $a>0, b\in\mathbf{R}$, $|x|_0$ 表示一维情况下的 L_0 范数,即 $\forall x\in\mathbf{R}$,有

$$|x|_0=\begin{cases}0, x=0\\ 1, x\neq 0\end{cases}$$

式(1)的一维连续绝对值优化松弛问题为

$$\min_{x\in\mathbf{R}} \tilde{g}(x):=|ax-b|+\varphi(\theta,\lambda,x) \tag{2}$$

其中,罚函数

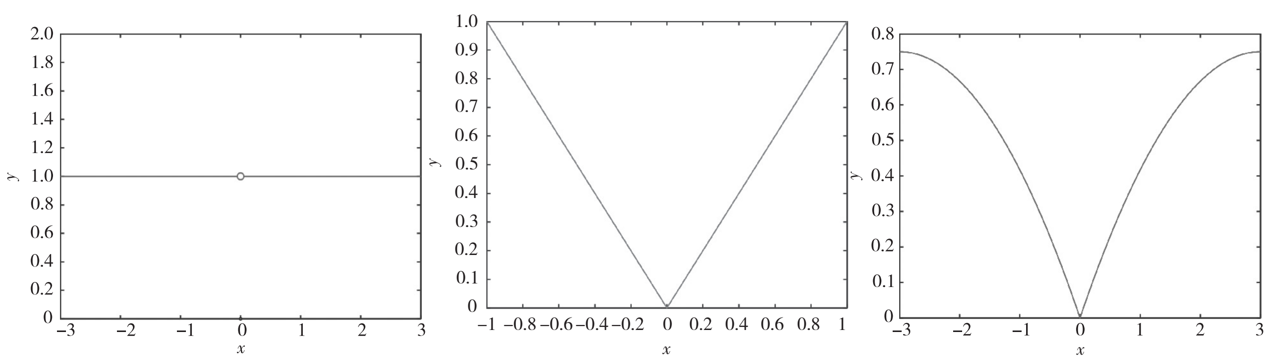
$$\varphi(\theta,\lambda,x)=\left(\lambda|x|-\frac{x^2}{2\theta}\right)I_{(|x|\leqslant\theta\lambda)}+\frac{1}{2}\theta\lambda^2 I_{(|x|>\theta\lambda)}$$

此处,$\theta>0$, $\lambda>0$ 是参数,$I_{(\cdot)}$ 为指示函数。注意到,MCP 罚函数满足下列的性质:

(i) 当 $x>0$ 时,$\varphi(\theta,\lambda,x)$ 是单调不减的光滑凹函数,并有 $0<\varphi'(\theta,\lambda,x)\leqslant\lambda$。

(ii) $x_i\in\mathbf{R}/\{0\}$ 有 $\varphi(\theta,\lambda,x_i)>0$,并且 $\varphi(\theta,\lambda,0)=0$。

图 1 是一维情形下 L_0, L_1, MCP 函数的图像,由 MCP 函数的定义及图 1 可知,当 $\theta\to+\infty$ 时,MCP 罚逐渐趋向于 L_1 罚;当 $\theta\to 1$ 时,MCP 罚逐渐逼近 L_0 罚;当 $|x_i|\leqslant\theta\lambda$, MCP 罚对特征系数进行压缩;当 $|x_i|>\theta\lambda$, MCP 罚对特征系数不再进行压缩。因此 MCP 罚是通过对于目标变量特征系数的压缩来实现近似无偏估计的。



(a) L_0　　(b) L_1　　(c) MCP

图 1 一维情形下 L_0, L_1, MCP 函数图像

Fig. 1 L_0, L_1, MCP functions image under some conditions

2 解的等价关系

主要分析式(1)与式(2)之间解(全局最优解)的等价性。

定理 1 当 $a>0, b\in\mathbf{R}, \lambda>0$ 时,则式(1)的全局最优解 x^* 满足:

$$x^*=\begin{cases}0, \lambda>|b|\\ 0 \text{ 或 } \dfrac{b}{a}, \lambda=|b|\\ \dfrac{b}{a}, \lambda<|b|\end{cases}$$

证明 当 $x=0$ 时,$g(x)=|ax-b|+\lambda|x|_0=|b|$;

当 $x\neq0$ 时，则有 $g\left(\frac{b}{a}\right)=\lambda\leqslant g(x)=|ax-b|+\lambda$，故当 $g(0)<g\left(\frac{b}{a}\right)$，即 $|b|<\lambda$，则 $x^*=0$；当 $g(0)>g\left(\frac{b}{a}\right)$，即 $|b|>\lambda$，则 $x^*=\frac{b}{a}$；当 $g(0)=g\left(\frac{b}{a}\right)$，即 $|b|=\lambda$ 时，同时达到最优。

定理 2　设 $a>0,\lambda\geqslant2a,\theta\lambda=2$，则对 $b\in\mathbf{R}$，g 的全局解是 $\tilde{g}$ 的全局最优解。

证明　由式(2)可知：

$$\tilde{g}(x)=\begin{cases}|ax-b|+\frac{1}{2}\theta\lambda^2, & |x|>\theta\lambda\\ |ax-b|+\lambda|x|-\frac{x^2}{2\theta}, & |x|\leqslant\theta\lambda\end{cases}$$

Ⅰ 对 $b\leqslant0$ 的情况讨论。

(i) $\frac{b}{a}\leqslant-\theta\lambda$。

① 当 $x\leqslant\frac{b}{a}$，即 $ax-b\leqslant0$，故

$$\tilde{g}(x):=b-ax+\frac{\theta\lambda^2}{2}$$

由于

$$\tilde{g}\left(\frac{b}{a}\right)=\frac{1}{2}\theta\lambda^2\leqslant\tilde{g}(x)$$

则 $\tilde{g}(x)$ 在此区间上的最优解 $x_1=\frac{b}{a}$，最优值

$$\tilde{g}\left(\frac{b}{a}\right)=\frac{1}{2}\theta\lambda^2$$

② 当 $\frac{b}{a}<x\leqslant-\theta\lambda$，即 $ax-b\geqslant0$，故

$$\tilde{g}(x):=ax-b+\frac{\theta\lambda^2}{2}$$

由于 $\tilde{g}\left(\frac{b}{a}\right)=\frac{1}{2}\theta\lambda^2<\tilde{g}(x)$，故最优解不在此区间上。

③ 当 $-\theta\lambda<x\leqslant0$，即 $ax-b\geqslant0$，故

$$\tilde{g}(x):=ax-b-\lambda x-\frac{x^2}{2\theta}=$$

$$-\frac{1}{2\theta}[x+\theta(\lambda-a)]^2+\frac{\theta(\lambda-a)^2}{2}-b$$

又因为

$$\tilde{g}(0)=-b<\tilde{g}(-\theta\lambda)=\frac{\theta\lambda^2}{2}-a\theta\lambda-b<\tilde{g}(x)$$

故在此区间上最优解 $x_3=0$，最优值为 $\tilde{g}(0)=-b$。

④ 当 $0<x\leqslant\theta\lambda$，即 $ax-b\geqslant0$，故

$$\tilde{g}(x):=ax-b+\lambda x-\frac{x^2}{2\theta}=$$

$$-\frac{1}{2\theta}[x-\theta(\lambda+a)]^2+\frac{\theta(\lambda+a)^2}{2}-b$$

由于

$$\tilde{g}(0)=-b<\tilde{g}(x)$$

故最优解不在此区间上。

⑤ 当 $x>\theta\lambda$，即 $ax-b\geqslant0$，故

$$\tilde{g}(x):=ax-b+\frac{\theta\lambda^2}{2}$$

由于

$$\tilde{g}(0)=-b<\tilde{g}(\theta\lambda)=a\theta\lambda-b+\frac{1}{2}\theta\lambda^2<\tilde{g}(x)$$

故最优解不在此区间上。

(ii) $-\theta\lambda<\frac{b}{a}\leqslant0$。

① 当 $x\leqslant-\theta\lambda$，即 $ax-b\leqslant0$，故

$$\tilde{g}(x):=b-ax+\frac{\theta\lambda^2}{2}$$

由于

$$\tilde{g}(0)=b<\tilde{g}(-\theta\lambda)=\frac{1}{2}\theta\lambda^2+a\theta\lambda+b\leqslant\tilde{g}(x)$$

故最优解不在此区间上。

② 当 $-\theta\lambda<x\leqslant\frac{b}{a}$ 时，则 $ax-b\leqslant0$，故

$$\tilde{g}(x):=b-ax-\lambda x-\frac{x^2}{2\theta}=$$

$$-\frac{1}{2\theta}[x+\theta(\lambda+a)]^2+\frac{\theta(\lambda+a)^2}{2}+b$$

由于

$$g(x)\geqslant\tilde{g}\left(\frac{b}{a}\right)=\frac{-\lambda b}{a}-\frac{b^2}{2a^2\theta}=$$

$$|b|\left(\frac{2a\theta\lambda-|b|}{2a^2\theta}\right)>|b|\left(\frac{2a\theta\lambda-a\theta\lambda}{2a^2\theta}\right)=$$

$$\frac{\lambda|b|}{2a}\geqslant|b|=\tilde{g}(0)$$

故最优解不在此区间上。

③ 当$\frac{b}{a}<x\leqslant 0$ 时,则 $ax-b\geqslant 0$,故

$$\tilde{g}(x):=ax-b-\lambda x-\frac{x^2}{2\theta}=$$

$$-\frac{1}{2\theta}[x-\theta(a-\lambda)]^2+\frac{\theta(a-\lambda)^2}{2}-b$$

又因为

$$g(x)>\tilde{g}\left(\frac{b}{a}\right)=\frac{-\lambda b}{a}-\frac{b^2}{2a^2\theta}=$$

$$|b|\left(\frac{2a\theta\lambda-|b|}{2a^2\theta}\right)>|b|\left(\frac{2a\theta\lambda-a\theta\lambda}{2a^2\theta}\right)=$$

$$\frac{\lambda|b|}{2a}\geqslant|b|=\tilde{g}(0)$$

故在此区间上最优解 $x_7=0$,最优值 $\tilde{g}(0)=-b$。

④ 当 $0<x\leqslant\theta\lambda$ 时,即 $ax-b\geqslant 0$,故

$$\tilde{g}(x):=ax-b+\lambda x-\frac{x^2}{2\theta}=$$

$$-\frac{1}{2\theta}[x-\theta(\lambda+a)]^2+\frac{\theta(\lambda+a)^2}{2}-b$$

又因为

$$\tilde{g}(0)=-b<\tilde{g}(x)$$

故 $\tilde{g}(x)$ 的最优解不在此区间上。

⑤ 当 $x>\theta\lambda$ 时,即 $ax-b\geqslant 0$,故

$$\tilde{g}(x):=ax-b+\frac{\theta\lambda^2}{2}$$

由于

$$\tilde{g}(0)=b<\tilde{g}(\theta\lambda)=a\theta\lambda-b+\frac{1}{2}\theta\lambda^2<\tilde{g}(x)$$

故 $\tilde{g}(x)$ 的最优解不在此区间上。

II 对 $b>0$ 的情况讨论。

(i) $0<\frac{b}{a}\leqslant\theta\lambda$。

① 当 $x<-\theta\lambda$,即 $ax-b\leqslant 0$,故

$$\tilde{g}(x):=b-ax+\frac{\theta\lambda^2}{2}$$

由于

$$\tilde{g}(0)=b<\tilde{g}(-\theta\lambda)=\frac{1}{2}\theta\lambda^2+a\theta\lambda+b<\tilde{g}(x)$$

故最优解不在此区间上。

② 当$-\theta\lambda\leqslant x\leqslant 0$,即 $ax-b\leqslant 0$,故

$$\tilde{g}(x):=b-ax-\lambda x-\frac{x^2}{2\theta}=$$

$$-\frac{1}{2\theta}[x+\theta(\lambda+a)]^2+\frac{\theta(\lambda+a)^2}{2}+b$$

由于

$$\tilde{g}(0)=b\leqslant\tilde{g}(x)$$

所以 $\tilde{g}(x)$ 在此区间上最优解 $x_{11}=0$,最优值 $\tilde{g}(0)=-b$。

③ $0<x\leqslant\frac{b}{a}$时,即 $ax-b\leqslant 0$,故

$$\tilde{g}(x):=b-ax+\lambda x-\frac{x^2}{2\theta}=$$

$$-\frac{1}{2\theta}[x-\theta(\lambda-a)]^2+\frac{\theta(\lambda-a)^2}{2}+b$$

由于

$$\tilde{g}(0)=b<\tilde{g}(x)$$

故 $\tilde{g}(x)$ 的最优解不在此区间上。

④ 当$\frac{b}{a}<x\leqslant\theta\lambda$ 时,即 $ax-b\geqslant 0$,故

$$\tilde{g}(x):=ax-b+\lambda x-\frac{x^2}{2\theta}=$$

$$-\frac{1}{2\theta}[x-\theta(a+\lambda)]^2+\frac{\theta(a+\lambda)^2}{2}+b$$

又因为

$$g(x)>\tilde{g}\left(\frac{b}{a}\right)=\frac{\lambda b}{a}-\frac{b^2}{2a^2\theta}=$$

$$b\left(\frac{2a\theta\lambda-b}{2a^2\theta}\right)>b\left(\frac{2a\theta\lambda-a\theta\lambda}{2a^2\theta}\right)=$$

$$\frac{\lambda b}{2a}\geqslant b=\tilde{g}(0)$$

故 $\tilde{g}(x)$ 的最优解不在此区间上。

⑤ 当 $x>\theta\lambda$ 时,即 $ax-b\geqslant 0$,故

$$\tilde{g}(x):=ax-b+\frac{\theta\lambda^2}{2}$$

由于

$$\tilde{g}(0)=b<\tilde{g}(\theta\lambda)=a\theta\lambda-b+\frac{1}{2}\theta\lambda^2<\tilde{g}(x)$$

故 $\tilde{g}(x)$ 的最优解不在此区间上。

(ii) $\frac{b}{a}>\theta\lambda$。

① 当 $x<-\theta\lambda$ 时,即 $ax-b\leqslant 0$,故

$$\tilde{g}(x):=b-ax+\frac{\theta\lambda^2}{2}$$

由于

$$\tilde{g}(0)=b<\tilde{g}(-\theta\lambda)=\frac{1}{2}\theta\lambda^2+a\theta\lambda+b<\tilde{g}(x)$$

故 $\tilde{g}(x)$ 的最优解不在此区间上。

② 当 $-\theta\lambda\leqslant x\leqslant 0$,即 $ax-b\leqslant 0$,故

$$\tilde{g}(x):=b-ax-\lambda x-\frac{x^2}{2\theta}=$$

$$-\frac{1}{2\theta}[x+\theta(\lambda+a)]^2+\frac{\theta(\lambda+a)^2}{2}+b$$

又因为

$$\tilde{g}(0)=b\leqslant\tilde{g}(x)$$

则 $\tilde{g}(x)$ 在此区间上的最优解 $x_{17}=0$,最优值 $\tilde{g}(0)=-b$。

③ 当 $0<x\leqslant\theta\lambda$ 时,即 $ax-b\leqslant 0$,故

$$\tilde{g}(x):=b-ax-\lambda x-\frac{x^2}{2\theta}=$$

$$-\frac{1}{2\theta}[x+\theta(\lambda+a)]^2+\frac{\theta(\lambda+a)^2}{2}+b$$

又因为

$$\tilde{g}(0)=b<\tilde{g}(x)$$

故 $\tilde{g}(x)$ 的最优解不在此区间上。

④ 当 $\theta\lambda<x<\frac{b}{a}$ 时,则 $ax-b<0$,故

$$\tilde{g}(x):=b-ax+\frac{\theta\lambda^2}{2}$$

由于

$$\tilde{g}(0)<\tilde{g}(-\theta\lambda)=\frac{\theta\lambda^2}{2}+a\theta\lambda+b<\tilde{g}(x)$$

故 $\tilde{g}(x)$ 的最优解不在此区间上。

⑤ 当 $x\geqslant\frac{b}{a}$ 时,则 $ax-b\geqslant 0$,故

$$\tilde{g}(x):=ax-b+\frac{\theta\lambda^2}{2}$$

由于

$$\tilde{g}\left(\frac{b}{a}\right)=\frac{\theta\lambda^2}{2}\leqslant\tilde{g}(x)$$

故 $\tilde{g}(x)$ 在此区间上最优解 $x_{18}=\frac{b}{a}$,最优值为 $\tilde{g}\left(\frac{b}{a}\right)=\frac{1}{2}\theta\lambda^2$。

综上所述,式(2)的全局最优解 $\hat{x}$ 如下:

$$\hat{x}=\begin{cases}0,\theta\lambda^2>2|b|\\ 0\text{ 或 }\frac{b}{a},\theta\lambda^2=2|b|\\ \frac{b}{a},\theta\lambda^2<2|b|\end{cases}$$

故当取 $\theta\lambda=2$ 时,有

$$\hat{x}=\begin{cases}0,\lambda>|b|\\ 0\text{ 或 }\frac{b}{a},\lambda=|b|\\ \frac{b}{a},\lambda<|b|\end{cases}$$

故由定理 1 可知 g 与 $\tilde{g}$ 的全局最优解相同。

3　结束语

针对损失函数为最小一乘问题,惩罚项由基数函数定义的绝对值优化问题,用 MCP 罚来连续逼近基数罚,得到一个精确连续的松弛问题,证明了在一定的条件下这两个优化问题具有相同的全局最优解。这为进一步研究高维最小一乘回归问题提供了启发。

参考文献(References):

[1] ZHANG C H. Nearly Unbiased Variable Selection under Minimax Concave Penalty[J]. The Annals of Statistics, 2010,38(2):894—942

[2] BRUCKSTEIN A M, DONOHO D L, ELAD M. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images[J]. Siam Review, 2009, 51(1):34—81

[3] NIKOLOVA M, Ng M K, TAM C P. Fast Nonconvex Nonsmooth Minimization Methods for Image Restoration and Reconstruction[J]. IEEE Transactions on Image Process, 2010,19(12):3073—3088

[4] WEN B,CHEN X J,PONG T L. Linear Convergence of

Proximal Gradient Algorithm With Extrapolation for a Class of Nonconvex Nonsmooth Minimization Problems[J]. SIAM Journal on Optimization,2017,27(1):124—125
[5] NATARAJAN B K. Sparse Approximate Solutions to Linear Systems[J]. SIAM Journal on Computing,1995,24(5): 227—234
[6] DONOHO D L. Compressed Sensing[J]. IEEE Transactions Information Theory,2006,52(4):1289—1306
[7] ZHANG C H, HUAN J. The Sparsity and Bias of the Lasso Selection in High-Dimensional Linear Regression[J]. The Annals of Statistics, 2008, 36(4):1567—1594
[8] FAN J Q, PENG H. Nonconcave Penalized Likelihood with a Diverging Number of Parameters[J]. The Annals of Statistics, 2004, 32(3):928—961
[9] LIU H C, YAO T, LI R Z,et al. Folded Concave Penalized Sparse Linear Regression: Sparsity, Statistical Performance and Algorithmic Theory for Local Solution[J]. Mathematical Programming, 2017,166(1-2):207—240
[10] FAN J Q, LI R Z. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties[J]. American Statistical Association, 2001,96(456): 1348—1360

The Equivalence of Solutions of Two Absolute Value Optimization Problems

LUO Xiao-min, PENG Ding-tao

(School of Mathametics and Statistics, Guizhou University, Guiyang 550025,China)

Abstract:This paper focuses on the absolute value optimization problem, where the loss function is least absolute deviation and the penalty term is defined by the cardinality function. In order to obtain an exact continuous relaxation problem, we use MCP (minimax concave penalty) to approximate the cardinality penalty. Firstly, we prove the global optimal solution of the absolute value optimization problem with cardinal penalty. Then we discuss the equivalence of the global optimal solution between the absolute value optimization problem with cardinal penalty and the absolute value optimization relaxation problem with MCP penalty. Finally, under some mild conditions, we proved that the two problems have the same optimal solutions.

Key words:absolute value optimization problem; MCP; optimal solution; equivalence

责任编辑:李翠薇

引用本文/Cite this paper:

罗孝敏,彭定涛.两个绝对值优化问题解的等价性[J].重庆工商大学学报(自然科学版),2020,37(5):37—42

LUO X M,PENG D T. The Equivalence of Solutions of Two Absolute Value Optimization Problems[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2020,37(5):37—42