Advances in Applied Mathematics 应用数学进展, 2021, 10(11), 3850-3864
Published Online November 2021 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2021.1011409

稀疏相位检索问题的光滑化方法研究

刘晓佳，彭定涛*

贵州大学，数学与统计学院，贵州 贵阳

收稿日期：2021年10月17日；录用日期：2021年11月7日；发布日期：2021年11月19日

摘 要

对稀疏相位检索问题，本文建立了稀疏最小一乘优化模型。首先，将不连续的稀疏正则函数松弛为连续的**Capped-L1**正则函数。利用方向导数和方向稳定点建立了该松弛问题的一阶最优性条件。进一步，分析了方向稳定点的下界性质，并建立了松弛问题与原始问题全局解的等价性。最后，本文建议采用光滑化方法求解该问题。构造了目标函数的光滑化函数，并建立了光滑问题与松弛问题解的一致性，为使用光滑化方法求解该问题提供了理论保证。

关键词

稀疏相位检索问题，最小一乘估计，方向稳定点，光滑化方法

Research on Smoothing Method for Sparse Phase Retrieval Problems

Xiaojia Liu, Dingtao Peng*

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Oct. 17^{th}, 2021; accepted: Nov. 7^{th}, 2021; published: Nov. 19^{th}, 2021

* 通讯作者。

文章引用: 刘晓佳, 彭定涛. 稀疏相位检索问题的光滑化方法研究[J]. 应用数学进展, 2021, 10(11): 3850-3864.
DOI: 10.12677/aam.2021.1011409

Abstract

For sparse phase retrieval problems, this paper constructed a class of sparse least absolute deviation optimization model. First, we used the continuous Capped-L1 regularization function to relax the discontinuous sparsity function. By virtue of directional derivatives and the directional stationary points, we provided the first-order optimality condition for the relaxed optimization problem. Furthermore, we derived the lower bound property of the directional stationary points, based on which we proved the equivalence between the original problem and the relaxed problem in the sense of global solutions. Finally, we advised using smoothing methods to solve the relaxed problem. We proposed a class of smooth function to approximate the objective function, and established the consistency of the solutions of the smoothing problem and the relaxed problem. Our result provides a theoretical basis for using smoothing methods to solve sparse phase retrieval problems.

Keywords

Sparse Phase Retrieval Problem, Least Absolute Estimate, Directional Stationary Point, Smoothing Method

1. 引言

1.1. 稀疏相位检索

相位检索在X-射线晶体学 [1]、电子显微镜 [2]、衍射成像 [3]、天文学 [4]等领域有着重要应用. 由于光波的振荡频率过高, 使得电子测量装置无法直接记录光波的相位信息, 通常只能观测到强度信息, 但相位信息通常携带了光波的绝大多数信息更具实用价值, 因此需要利用观测到的强度信息对相位进行恢复, 这就是相位检索. 具体来说, 对给定的观测数据$\{a_i, y_i\}_{i=1}^N$, 寻找满足$y_i = |\langle a_i, x\rangle|^2$, $i = 1, \cdots, N$ 的向量x. 由于噪声的存在, 通常考虑如下二次度量回归模型:

$$y_i = |\langle a_i, x\rangle|^2 + \varepsilon_i,\ i = 1, \cdots, N, \tag{1}$$

其中ε_i是随机误差项, 满足$E(\varepsilon_i)=0$, $E(\varepsilon_i^2)=\sigma^2$. 由于真实的信号$x_0$具有稀疏性（即$x_0$ 的大部分分量为零）, 根据这一先验信息, 人们有希望根据观测样本恢复出真实信号.

对于稀疏相位检索问题, 现存的文献(如 [5] [6] [7] [8] [9]) 大多是基于最小二乘估计方法进行求解, 即采用如下最小二乘损失函数:

$$\frac{1}{N}\sum_{i=1}^{N}((a_i^\top x)^2-y_i)^2. \tag{2}$$

然而, 当存在异常观测值或噪声趋向于重尾分布时, 最小二乘方法会放大异常数据在估计中的作用, 从而产生较大的偏差. 因此, 最小二乘估计的稳健性较弱. 相较而言, 最小一乘估计不会放大误差, 具有较好的稳健性和抗干扰能力. 于是文献 [10] [11] [12]等采用最小一乘估计方法处理相位检索问题, 即采用如下最小一乘损失函数:

$$\frac{1}{N}\sum_{i=1}^{N}|(a_i^\top x)^2-y_i|. \tag{3}$$

考虑到噪声重尾分布以及真实信号的稀疏性, 自然地是通过求解如下稀疏优化模型来处理稀疏相位检索问题:

$$\min_{x\in\mathbb{R}^n}\frac{1}{N}\sum_{i=1}^{N}|(a_i^\top x)^2-y_i|+\lambda\|x\|_0, \tag{4}$$

其中, $\|x\|_0$表示向量x非零分量的个数, $\lambda>0$是正则化参数. 然而, 该问题的损失函数和正则函数均是非凸、非光滑的, 正则函数甚至是非Lipscthiz、不连续的, 因此该问题的求解是NP难的.

基于$\|x\|_0$的松弛正则化技术, 我们得到问题(4)的松弛模型:

$$\min_{x\in\mathbb{R}^n}\frac{1}{N}\sum_{i=1}^{N}|(a_i^\top x)^2-y_i|+R(x), \tag{5}$$

其中, $R(x):\mathbb{R}^n\to\mathbb{R}$是$\|x\|_0$的某一松弛函数, 常见的松弛函数有$L_1$, SCAD, MCP, Capped-$L_1$等 [13]. 松弛问题是连续的, 但依然是非凸非光滑的, 对其求解仍十分困难.

对于形如问题(5)的非凸非光滑优化问题, 研究者们通常通过Clarke次微分定义critical点刻画其最优性条件, 如文献 [14] [15] [16] [17]. 利用函数的凸差表示, Pang等人 [15]提出了lifted稳定点的概念, 并证明了, 对于刻画解点的最优性, lifted稳定点优于critical点. Ahn等人 [18]通过一阶方向导数定义了凸差优化问题的方向稳定点, 建立了该问题的最优性条件. 对于凸差优化问题, Cui等人 [19]讨论了几类稳定点的关系, 结果表明, 通过方向导数定义的方向稳定点强于其它次微分定义的稳定点, 能更好地刻解的最优性. 因此, 本文尝试用方向导数来刻画松弛问题(5)的方向稳定点, 建立其一阶最优性条件.

另一个需要考虑的问题是: 原问题松弛问题是否等价? 对于基于最小二乘方法的线性回归优化模型, [14]对于L_0正则提出了一簇连续凸差松弛, 并分析了松弛问题与原始全局解的等价性. [20]分

析了MCP, SCAD, Capped-L_1 等类型的松弛问题与原L_0问题全局解的等价性以及局部解的包含关系. 对于L_0正则的全局Lipschitz 连续优化问题, [17]研究了Capped-L_1正则松弛问题与原L_0问题全局解、最优值的等价性和局部解的包含关系. [21]对组稀疏优化问题研究了组Capped-L_1正则松弛问题与原始组稀疏优化问题全局解、最优值的等价性和局部解的包含关系. 对于我们所考虑的基于最小一乘损失的二次度量回归优化问题, 它的松弛问题与原L_0问题的关系是怎样的, 需要进行深入的研究.

受 [17] [20] [21]等文的启发, 本文考虑如下Capped-L_1松弛问题:

$$\min_{x\in\mathbb{R}^n} \frac{1}{N}\sum_{i=1}^{N}|(a_i^\top x)^2 - y_i| + \lambda\sum_{i=1}^{n}\min\{1, |x_i|/v\}. \tag{6}$$

其中$\phi^{CapL1}(t) = \min\{1, |t|/v\}$是Capped-$L_1$松弛函数, $v > 0$是给定参数.

本文主要工作如下:

(1) 给出松弛问题目标函数的方向导数, 通过方向导数定义了松弛问题的方向稳定点, 刻画问题(6)的一阶最优性条件, 并且分析了该稳定点的具体形式. 此外, 还分析了方向稳定点的下界性质, 建立了问题(6) 与问题(4)的解的等价性理论.

(2) 由于问题(6) 中损失函数与正则函数都是非光滑的, 对于如何计算问题(6)的方向稳定点, 我们建议采用光滑化方法. 首先对问题的损失函数进行光滑逼近, 得到光滑的近似问题. 通过理论分析, 我们建立了光滑化近似问题与问题(6)解的一致性. 这为使用光滑化方法求解问题(6)提供了理论保证和可行方向.

1.2. 符号与结构

以下符号贯穿全文. 对任意$t \in \mathbb{R}$, $|t|$表示t的绝对值. 对于一个集合$\mathcal{C}$, $|\mathcal{C}|$ 表示集合中元素的个数. 符号函数sgn(t)定义为:

$$\mathrm{sgn}(t) := \begin{cases} 1, & t > 0, \\ 0, & t = 0, \\ -1, & t < 0. \end{cases}$$

令$x, y \in \mathbb{R}^n$, $x^T y = \sum_{i=1}^{n} x_i y_i$. $\|x\|_1$表示向量x的1-范数, $\|x\|$ 表示向量x的2-范数, $|x| = (|x_1|, \cdots, |x_n|)^T$. 此外, 记向量$x$的支撑集为$\Gamma(x) = \{i \in \{1, \cdots, n\} : x_i \neq 0\} = \Gamma_1(x) \cup \Gamma_2(x)$ 其中, $\Gamma_1 = \{i : |x_i| < v\}$, $\Gamma_2 = \{i : |x_i| \geq v\}$. 定义指标集$I_1(x) := \{i \in \{1, \cdots, N : (a_i^\top x)^2 - y_i \neq 0\}$, $I_2(x) := \{i \in \{1, \cdots, N\} : (a_i^\top x)^2 - y_i = 0\}$. 除此之外, 其它符号会在文中具体说明.

本文的结构如下: 第二部分, 利用方向导数定义了问题(6)的方向稳定点, 刻画了问题的一阶最优性条件. 我们给出了方向稳定点的具体表达以及下界性质. 此外, 基于这些理论, 证明了问题(4)与问题(6) 解的等价性. 第三部分, 构建问题(6)的一类光滑化函数, 建立了光滑逼近问题与问题(6)解的一致性, 为使用光滑化方法求解问题(6)提供了理论保证. 第四部分, 简单的总结.

2. 最优性条件

2.1. 模型(6)的方向稳定点

由 [13] [19]可知, 相比于改进的稳定点, C-稳定点, 准则点而言, 方向导数定义的方向稳定点能更好的刻画解的最优性. 本节, 我们定义问题的(6)方向稳定点, 分析其具体形式以及下界性质, 并且建立问题(4)与问题(6)的等价性关系. 首先, 我们给出方向导数与问题(6)方向稳定点的基本定义.

定义 2.1 设$f:\mathbb{R}^n\to\mathbb{R}$ 在点$x\in\mathbb{R}^n$ 处局部Lipschitz连续且方向可微. 则函数f在点x处沿方向$w\in\mathbb{R}^n$ 的方向导数定义为

$$f'(x;w):=\lim_{\tau\downarrow 0}\frac{f(x+\tau w)-f(x)}{\tau}.$$

定义 2.2 称$\widehat{x}\in\mathbb{R}^n$是模型(6)的方向稳定点, 若

$$F'(\widehat{x};x-\widehat{x})+\Phi'(\widehat{x};x-\widehat{x})\geq 0,\ \forall x\in\mathbb{R}^n.$$

Peng等人 [13]证明了方向稳定点具有如下局部最优性质.

定理 2.1 *[13] 令函数$f:\mathbb{R}^n\to\mathbb{R}$在点$\widehat{x}\in\mathbb{R}^n$处是局部Lipschitz连续且方向可微的, 有如下两条性质成立:*

(i)若$\widehat{x}$是函数f的局部最小值, 那么$\widehat{x}$是函数f的方向稳定点;

(ii)$\widehat{x}$是严格局部最小值并满足一阶增长性条件, 即存在$\widehat{x}$的领域$\mathcal{W}$和$\delta>0$, 使得

$$f(x)\geq f(\widehat{x})+\delta\parallel x-\widehat{x}\parallel,\forall x\in\mathcal{W}, \tag{7}$$

当且仅当$\widehat{x}$满足

$$f'(\widehat{x};x-\widehat{x})>0,\forall x\in\mathbb{R}^n\setminus\{\widehat{x}\}. \tag{8}$$

下面, 我们研究问题(6)的方向稳定点的具体表达.

定理 2.2 *若$\widehat{x}\in\mathbb{R}^n$是模型(6)的方向稳定点, 则*

$$\frac{2}{N}\left\{\sum_{i\in I_1}\mathrm{sgn}((a_i^\top\widehat{x})^2-y_i)\cdot a_i^\top\widehat{x}a_i^\top(x-\widehat{x})+\sum_{i\in I_2}|a_i^\top\widehat{x}a_i^\top(x-\widehat{x})|\right\}+\sum_{i=1}^n\phi'(\widehat{x}_i;x_i-\widehat{x}_i)\geq 0,$$

其中I_1, I_2定义见§*1.2*, 且

$$\phi'(\widehat{x}_i; x_i - \widehat{x}_i) = \begin{cases} \frac{\lambda |x_i|}{v} & \widehat{x}_i = 0, \\ \frac{\mathrm{sgn}(\widehat{x}_i)\lambda(x_i - \widehat{x}_i)}{v} & |\widehat{x}_i| \in (0, v), \\ \max\{0, \frac{\mathrm{sgn}(\widehat{x}_i)\lambda(x_i - \widehat{x}_i)}{v}\} & |\widehat{x}_i| = v, \\ 0 & \text{otherwise}. \end{cases} \tag{9}$$

证明. 由方向导数的定义, 函数Φ在点$\widehat{x}$处的方向导数为

$$\Phi'(\widehat{x}; x - \widehat{x}) = \sum_{i=1}^{n} \phi'(\widehat{x}_i; x_i - \widehat{x}_i), \tag{10}$$

其中$\phi'(\widehat{x}_i; x_i - \widehat{x}_i)$定义如(9)所示. 相比之下, 损失函数$F$在点$\widehat{x}$ 处的方向导数的计算较为复杂. 集合I_1, I_2的定义见符号说明. 由方向导数的定义,

$$\begin{aligned} F'(\widehat{x}; x - \widehat{x}) &= \frac{1}{N} \cdot \lim_{t \downarrow 0} \frac{\sum_{i=1}^{N} \{|(a_i^\top(\widehat{x} + t(x - \widehat{x})))^2 - y_i| - |(a_i^\top \widehat{x})^2 - y_i|\}}{t} \\ &= \frac{1}{N} \cdot \lim_{t \downarrow 0} \frac{\sum_{i=1}^{N} \{|(a_i^\top \widehat{x})^2 + t^2(a_i^\top(x - \widehat{x}))^2 + 2ta_i^\top \widehat{x} a_i^\top(x - \widehat{x}) - y_i| - |(a_i^\top \widehat{x})^2 - y_i|\}}{t} \end{aligned}$$

当$i \in I_1$时, $(a_i^\top \widehat{x})^2 \neq y_i$,

$$\begin{aligned} f_i'(\widehat{x}; x - \widehat{x}) &= \lim_{t \downarrow 0} \frac{\mathrm{sgn}((a_i^\top \widehat{x})^2 - y_i) \cdot \{t^2(a_i^\top(x - \widehat{x}))^2 + 2ta_i^\top \widehat{x} a_i^\top(x - \widehat{x})\}}{t} \\ &= 2\mathrm{sgn}((a_i^\top \widehat{x})^2 - y_i) \cdot a_i^\top \widehat{x} a_i^\top(x - \widehat{x}). \end{aligned} \tag{11}$$

当$i \in I_2$时, $(a_i^\top \widehat{x})^2 = y_i$, 于是

$$\begin{aligned} f_i'(\widehat{x}; x - \widehat{x}) &= \lim_{t \downarrow 0} \frac{|t^2(a_i^\top(x - \widehat{x}))^2 + 2ta_i^\top \widehat{x} a_i^\top(x - \widehat{x})|}{t} \\ &= 2|a_i^\top \widehat{x} a_i^\top(x - \widehat{x})|. \end{aligned} \tag{12}$$

于是, 由(11)(12)可得, 损失函数F 在点$\widehat{x}$处的方向导数

$$F'(\widehat{x}; x - \widehat{x}) = \frac{2}{N} \left\{ \sum_{i \in I_1} \mathrm{sgn}((a_i^\top \widehat{x})^2 - y_i) \cdot a_i^\top \widehat{x} a_i^\top(x - \widehat{x}) + \sum_{i \in I_2} |a_i^\top \widehat{x} a_i^\top(x - \widehat{x})| \right\}. \tag{13}$$

结合(10)与(13), 该结论成立. □

2.2. 方向稳定点的下界性质与解的等价性

下面分析模型(6)方向稳定点的下界性质. 令$L:\mathbb{R}^n\to\mathbb{R}$定义为

$$L(x):=\left\|\sum_{i=1}^{N}|a_i^\top x a_i|\right\|_1.$$

注意到, 当$v\to 0$时, $\phi'_-(v):=\lim_{t\uparrow v}\phi'(t)>0\to\infty$. 于是存在$x^*$ 以及充分小的v使得$F(x^*)>\Upsilon>c$, $\phi'_-(v)>(2/N)L(x^*)$成立, 其中$c>0$是某一常数. 选取参数Υ, v, λ以及x^*满足:

$$F(x^*)>\Upsilon,\ \phi'_-(v)>\frac{2}{N}L(x^*).$$

定理 2.3 设$\widehat{x}\in\mathbb{R}^n$是模型(6)的方向稳定点, 且

$$F(\widehat{x})\le\Upsilon,\ \phi'_-(v)>\frac{2}{N}L(x^*),$$

则对任意$i=1,\cdots,n$,

$$|\widehat{x}_i|\ge v \text{ 或 } |\widehat{x}_i|=0.$$

证明. 只需要证明集合$\Gamma_1=\emptyset$. 假设$\Gamma_1\neq\emptyset$. 因为$\widehat{x}\in\mathbb{R}^n$是模型(6)的方向稳定点,

$$F'(\widehat{x};x-\widehat{x})+\Phi'(\widehat{x};x-\widehat{x})\ge 0,\ \forall x\in\mathbb{R}^n.$$

由x的任意性, 可对x进行赋值, 令$x_j=\widehat{x}_j$, $j\in\Gamma_2$, 且$x_j\neq\widehat{x}_j$, $j\in\Gamma_1$. 于是,

$$F'(\widehat{x};x-\widehat{x})+\sum_{j\in\Gamma_1}\phi'(\widehat{x}_j;x_j-\widehat{x}_j)\ge 0. \tag{14}$$

另外, 对于所定义的x, 由(13)可得,

$$F'(\widehat{x};x-\widehat{x})=\frac{2}{N}\left\{\sum_{j\in\Gamma_1}\left[\sum_{i\in I_1}\operatorname{sgn}((a_i^\top\widehat{x})^2-y_i)\cdot a_i^\top\widehat{x}a_i\right]_j(x_j-\widehat{x}_j)+\sum_{j\in\Gamma_1}\left[\sum_{i\in I_2}|a_i^\top\widehat{x}a_i|\right]_j|x_j-\widehat{x}_j|\right\}.$$

注意到,

$$\sum_{j\in\Gamma_1}\left[\sum_{i\in I_1}\operatorname{sgn}((a_i^\top\widehat{x})^2-y_i)\cdot a_i^\top\widehat{x}a_i\right]_j(x_j-\widehat{x}_j)\le\sum_{j\in\Gamma_1}\left[\sum_{i\in I_1}|a_i^\top\widehat{x}a_i|\right]_j|x_j-\widehat{x}_j|.$$

于是,

$$\begin{aligned}F'(\widehat{x};x-\widehat{x}) &\le \frac{2}{N}\left\{\sum_{j\in\Gamma_1}\left[\sum_{i\in I_1}|a_i^\top\widehat{x}a_i|\right]_j|x_j-\widehat{x}_j|+\sum_{j\in\Gamma_1}\left[\sum_{i\in I_2}|a_i^\top\widehat{x}a_i|\right]_j|x_j-\widehat{x}_j|\right\}\\ &= \frac{2}{N}\left\{\sum_{j\in\Gamma_1}\left[\sum_{i=1}^{N}|a_i^\top\widehat{x}a_i|\right]_j|x_j-\widehat{x}_j|\right\}.\end{aligned} \tag{15}$$

令$x_j = \widehat{x}_j - \epsilon\widehat{x}_j$, $\epsilon > 0$, $j \in \Gamma_1$. 结合(14) 与(15), 得

$$\begin{aligned}\sum_{j\in\Gamma_1}\phi'(\widehat{x}_j)|\widehat{x}_j| &\leq \frac{2}{N}\left\{\sum_{j\in\Gamma_1}\left[\sum_{i=1}^{N}|a_i^\top\widehat{x}a_i|\right]_j|\widehat{x}_j|\right\}\\ &\leq \frac{2}{N}L(x^*)\sum_{j\in\Gamma_1}|\widehat{x}_j|.\end{aligned}$$

由于对任意$j \in \Gamma_1$, $\phi'(\widehat{x}_j) = \lambda/v$. 于是, $\lambda/v \leq (2/N)L(x^*)$, 与已知条件$\lambda/v > (2/N)L(x^*)$ 矛盾. 该结论成立. □

下面, 我们讨论问题(4)与问题(6)解的关系.

定理 2.4 设$\lambda/v \geq (2/N)L(x^*)$,

(i) *若$\widehat{x} \in \mathbb{R}^n$ 是问题(6)的全局最优解, 且满足$F(\widehat{x}) \leq \Upsilon$, 则$\widehat{x}$是问题(4) 的全局最优解.*

(ii) *若$\widehat{x} \in \mathbb{R}^n$ 是问题(4)的全局最优解, 且问题(6)的全局解x'存在并满足$F(x') \leq \Upsilon$, 则$\widehat{x}$是问题(6)的全局最优解.*

证明. (i) 设$\widehat{x} \in \mathbb{R}^n$是问题(6)的全局最优解, 则$\widehat{x} \in \mathbb{R}^n$是问题(6) 的方向稳定点. 由$F(\widehat{x}) \leq \Upsilon$及下界定理2.3, 得$\Phi(\widehat{x}) = \lambda\|\widehat{x}\|_0$. 于是

$$F(\widehat{x}) + \lambda\|\widehat{x}\|_0 = F(\widehat{x}) + \Phi(\widehat{x}) < F(x) + \Phi(x) \leq F(x) + \lambda\|x\|_0.$$

因此, $\widehat{x}$是问题(4)的全局最优解.

(ii) 设$\widehat{x} \in \mathbb{R}^n$是问题(4)的全局最优解, 但$\widehat{x} \in \mathbb{R}^n$不是问题(6)的全局最优解. 设$x'$ 是问题(6) 的全局最优解, 则$F(x') \leq \Upsilon$. 由定理已知条件及下界定理2.3, 得$\Phi(x') = \lambda\|x'\|_0$. 于是

$$F(x') + \lambda\|x'\|_0 = F(x') + \Phi(x') < F(\widehat{x}) + \Phi(\widehat{x}) \leq F(\widehat{x}) + \lambda\|\widehat{x}\|_0.$$

这与$\widehat{x} \in \mathbb{R}^n$是问题(4)的全局最优解矛盾. 因此$\widehat{x}$是问题(6)的全局最优解. □

注意到, 问题(6)的全局解x'是一定存在的, 这是因为改变全局解分量的符号, 不会改变问题(6)目标函数的取值, 这是由相位检索问题本身的性质决定的.

3. 光滑化方法

问题(6)是一非凸、非光滑优化问题, 直接求解是困难的. 本节我们构建问题(6)的一类光滑函数. 并证明了光滑问题的方向稳定点的任意聚点是原问题的方向稳定点. 这些理论为问题(6)的求解提供了理论保证.

回顾 [22], 绝对值函数$|t|$, $t \in \mathbb{R}$的一类光滑函数可以定义为

$$\widetilde{\theta}(t,\mu) := \begin{cases}|t| & \text{if } |t| \geq \mu,\\ \frac{t^2}{2\mu} + \frac{\mu}{2} & \text{if } |t| < \mu.\end{cases}$$

其中$\mu > 0$是光滑参数. 由此我们得到问题(6)中损失函数的光滑函数

$$\widetilde{F}(x,\mu) := \frac{1}{N}\sum_{i=1}^{N}\widetilde{f}_i(x,\mu), \tag{16}$$

其中, $\widetilde{f}_i(x,\mu) = \widetilde{\theta}((a_i^T x)^2 - y_i, \mu)$. 记$\{1,\cdots,N\} = \mathcal{C}_1^\mu(x) \cup \mathcal{C}_2^\mu(x) \cup \mathcal{C}_3^\mu(x)$, 其中$\mathcal{C}_1^\mu(x) = \{1,\cdots,N : |(a_i^T x)^2 - y_i| < \mu\}$, $\mathcal{C}_2^\mu(x) = \{1,\cdots,N : |(a_i^T x)^2 - y_i| > \mu\}$, $\mathcal{C}_3^\mu(x) = \{1,\cdots,N : |(a_i^T x)^2 - y_i| = \mu\}$.

下面, 我们分析光滑函数$\widetilde{F}(x,\mu)$的一些性质.

引理 3.1 *对于函数(16), 下述结论成立:*

(1) $\lim_{z\to x,\mu\downarrow 0}\widetilde{F}(z,\mu) = F(x)$.

(2) *(关于x的Lipschitz连续性) 存在一个常数$L > 0$ 使得对任意$\mu \in (0,\overline{\mu}]$, 使得$\nabla\widetilde{F}(x,\mu)$ 是Lipschitz 连续的, 且Lipschitz 常数是$L\mu^{-1}$.*

(3) *(关于μ的Lipschitz连续性)* $|\widetilde{F}(x,\mu_1) - \widetilde{F}(x,\mu_2)| \le |\mu_1 - \mu_2|$.

(4) *(一致性与弱一致性) 任意$i \in I_1(x)$, 有$\lim_{z\to x,\mu\downarrow 0}\langle\nabla\widetilde{f}_i(z,\mu), w\rangle = f_i'(x;w)$, $\forall\, w \in \mathbb{R}^n$成立; 任意$i \in I_2(x)$, 有$\limsup_{z\to 0,\mu\downarrow 0}\langle\nabla\widetilde{f}_i(z,\mu), w\rangle = f_i'(x;w)$, $\forall\, w \in \mathbb{R}^n$ 成立.*

证明. (1) 由$\mathcal{C}_1^\mu(x)$, $\mathcal{C}_2^\mu(x)$, $\mathcal{C}_3^\mu(x)$的定义得,

$$\begin{aligned}
|\widetilde{F}(z,\mu) - F(x)| &= \sum_{i=1}^{N}\left[\widetilde{f}_i(z,\mu) - f_i(x,\mu)\right] \\
&= \sum_{i=1}^{N}\left[\widetilde{f}_i(z,\mu) - \widetilde{f}_i(x,\mu) + \widetilde{f}_i(x,\mu) - f_i(x,\mu)\right] \\
&= \sum_{i=1}^{N}\left[\widetilde{f}_i(z,\mu) - \widetilde{f}_i(x,\mu))\right] + \sum_{i\in\mathcal{C}_1^\mu(x)}\left[\widetilde{f}_i(x,\mu) - f_i(x,\mu)\right] \\
&= \Pi_1 + \Pi_2.
\end{aligned}$$

由函数$\widetilde{f}$的连续性可知

$$\lim_{z\to x,\mu\downarrow 0}\Pi_1 = 0. \tag{17}$$

对于Π_2, 容易得

$$\sum_{i\in\mathcal{C}_1^\mu(x)}\left[\widetilde{f}_i(x,\mu) - f_i(x,\mu)\right] \le |\mathcal{C}_1^\mu(x)|\frac{\mu}{2} \to 0,\ \ (\mu \to 0). \tag{18}$$

结合(17)与(18), 于是结论(1)成立.

(2) 由光滑函数定义得, $\nabla\widetilde{F}(x,\mu) = \frac{1}{N}\sum_{i=1}^{N}\nabla\widetilde{f}_i(x,\mu)$.

当$i \in \mathcal{C}_1^{\mu}(x)$ 时,

$$\nabla \widetilde{f}_i(x,\mu) = 2\left(\frac{(a_i^T x)^2 - y_i}{\mu}\right) a_i a_i^T x, \tag{19}$$

当$i \in \mathcal{C}_2^{\mu}(x) \cup \mathcal{C}_3^{\mu}(x)$时,

$$\nabla \widetilde{f}_i(x,\mu) = 2\mathrm{sgn}((a_i^T x)^2 - y_i) a_i a_i^T x. \tag{20}$$

由文献 [23]的中值定理2.3.7得, 对任意$x, z \in \mathbb{R}^n$,

$$\nabla \widetilde{F}(x,\mu) - \nabla \widetilde{F}(z,\mu) \in (\partial(\nabla \widetilde{F}(u,\mu)))^T (x-z), \tag{21}$$

其中u是线段$[z,x]$上的某一点, $\partial(\nabla \widetilde{F}(u,\mu))$表示向量函数$\nabla \widetilde{F}(\cdot,\mu)$在点$u$处的Clarke次微分. 当$i \in \mathcal{C}_1^{\mu}(u)$ 时, $\partial(\nabla \widetilde{f}_i(u,\mu)) = \nabla^2 \widetilde{f}_i(u,\mu) = [\frac{2(a_i^T u)^2}{\mu} + 2] a_i a_i^T$. 当$i \in \mathcal{C}_2^{\mu}(u)$时, $\partial(\nabla \widetilde{f}_i(u,\mu)) = \nabla^2 \widetilde{f}_i(u,\mu) = 2\mathrm{sgn}((a_i^T x)^2 - y_i) a_i a_i^T$. 当$i \in \mathcal{C}_3^{\mu}(u)$时, 由Clarke次微分定义可知,

$$\partial(\nabla \widetilde{f}_i(u,\mu)) = \mathrm{con}\{2\mathrm{sgn}((a_i^T x)^2 - y_i), [\frac{2(a_i^T u)^2}{\mu} + 2]\} a_i a_i^T.$$

事实上, 对任意$a \in \mathrm{con}\{2\mathrm{sgn}((a_i^T x)^2 - y_i), [2(a_i^T u)^2 \mu^{-1} + 2]\}$, $a \le 2y_i \mu^{-1} + 4 \le (2y_i + \overline{\mu})\mu^{-1}$. 于是得

$$\partial(\nabla \widetilde{F}(u,\mu))) \subseteq \frac{1}{N}\left\{\sum_{i \in \mathcal{C}_1^{\mu}(u) \cup \mathcal{C}_1^{\mu}(u)} \nabla^2 \widetilde{f}_i(u,\mu) + \sum_{i \in \mathcal{C}_3^{\mu}(u)} \partial(\nabla \widetilde{f}_i(u,\mu))\right\}. \tag{22}$$

结合(21), (22) 可得, 存在$V \in \partial(\nabla \widetilde{F}(u,\mu))$使得下述关系成立:

$$\nabla \widetilde{F}(x,\mu) - \nabla \widetilde{F}(z,\mu) \quad = \quad V^T(x-z) \le (2\overline{y} + \overline{\mu})\mu^{-1} \lambda_{\max}(\sum_{i=1}^N a_i a_i^T)\|x-z\|,$$

其中$\overline{y} = \max\{y_i, i = 1, \cdots, N\}$. 注意到, $(2\overline{y} + \overline{\mu})\mu^{-1}\lambda_{\max}(\sum_{i=1}^N a_i a_i^T)$ 是一常数, 且与变量x无关, 于是结论(2)成立.

(3) 不失一般性, 假设$\mu_2 \ge \mu_1 > 0$. 对任意$i \in \mathcal{C}_2^{\mu_2}(x) \cup \mathcal{C}_3^{\mu_2}(x)$,

$$\widetilde{\theta}((a_i^T x)^2 - y_i, \mu_1) = \widetilde{\theta}((a_i^T x)^2 - y_i, \mu_2) = |(a_i^T x)^2 - y_i|.$$

因此,

$$\widetilde{f}_i(x,\mu_1) - \widetilde{f}_i(x,\mu_2) = 0,\ i \in \mathcal{C}_2^{\mu_2}(x) \cup \mathcal{C}_3^{\mu_2}(x). \tag{23}$$

对任意$i \in \mathcal{C}_2^{\mu_1}(x) \cup \mathcal{C}_3^{\mu_1}(x) \cap \mathcal{C}_1^{\mu_2}(x)$,

$$\widetilde{\theta}((a_i^T x)^2 - y_i, \mu_2) > \widetilde{\theta}((a_i^T x)^2 - y_i, \mu_1) = |(a_i^T x)^2 - y_i|. \tag{24}$$

此时, $((a_i^Tx)^2-y_i)^2/2\mu_2\le\mu_2^2/2\mu_2=\mu_2/2$, 且$|(a_i^Tx)^2-y_i|\ge\mu_1$, 于是由光滑函数定义及(24)得

$$\widetilde{\theta}((a_i^Tx)^2-y_i,\mu_2)-\widetilde{\theta}((a_i^Tx)^2-y_i,\mu_1)=\frac{((a_i^Tx)^2-y_i)^2}{2\mu_2}+\frac{\mu_2}{2}-|(a_i^Tx)^2-y_i|\le\mu_2-\mu_1.$$

因此,

$$\widetilde{f}_i(x,\mu_2)-\widetilde{f}_i(x,\mu_1)\le\mu_2-\mu_1,\ i\in\mathcal{C}_2^{\mu_1}(x)\cup\mathcal{C}_3^{\mu_1}(x)\cap\mathcal{C}_1^{\mu_2}(x). \tag{25}$$

对任意$i\in\mathcal{C}_1^{\mu_1}(x)$, $(a_i^Tx)^2-y_i<\mu_1\le\mu_2$,

$$\widetilde{\theta}((a_i^Tx)^2-y_i,\mu_1)=\frac{((a_i^Tx)^2-y_i)^2}{2\mu_1}+\frac{\mu_1}{2},\ \widetilde{\theta}((a_i^Tx)^2-y_i,\mu_2)=\frac{((a_i^Tx)^2-y_i)^2}{2\mu_2}+\frac{\mu_2}{2}.$$

于是,

$$\begin{aligned}\widetilde{\theta}((a_i^Tx)^2-y_i,\mu_2)-\widetilde{\theta}((a_i^Tx)^2-y_i,\mu_1)&=(\frac{1}{2\mu_2}-\frac{1}{2\mu_1})((a_i^Tx)^2-y_i)^2+\frac{\mu_2-\mu_1}{2}\\ \le\quad(\frac{1}{2\mu_2}-\frac{1}{2\mu_1})\mu_1^2+\frac{\mu_2-\mu_1}{2}&\le\frac{\mu_2^2}{2\mu_2}-\frac{\mu_1^2}{2\mu_1}+\frac{\mu_2-\mu_1}{2}=\mu_2-\mu_1.\end{aligned}$$

因此,

$$\widetilde{f}_i(x,\mu_2)-\widetilde{f}_i(x,\mu_1)\le\mu_2-\mu_1,\ i\in\mathcal{C}_1^{\mu_1}(x). \tag{26}$$

另一方面, $(\mathcal{C}_2^{\mu_2}(x)\cup\mathcal{C}_3^{\mu_2}(x))\cup(\mathcal{C}_2^{\mu_1}(x)\cup\mathcal{C}_3^{\mu_1}(x)\cap\mathcal{C}_1^{\mu_2}(x))\cup\mathcal{C}_1^{\mu_1}(x)=\{1,\cdots,N\}$, 结合(23), (25), (26)可得

$$\widetilde{F}(x,\mu_2)-\widetilde{F}(x,\mu_1)\le\frac{1}{N}N(\mu_2-\mu_1)=\mu_2-\mu_1.$$

结论(3)成立.

(4) 对任意$i\in I_1(x)$, 即$|(a_i^Tx)^2-y_i|\ne0$. 存在$\mu>0$使得$|(a_i^Tx)^2-y_i|>\mu$. 此时, $i\in\mathcal{C}_2^{\mu}(x)$. 有$\widetilde{f}_i=f_i$成立, 且它们在点$z_k$ 处可微. 于是得$\lim_{\mu\downarrow0}\langle\nabla\widetilde{f}_i(x,\mu),w\rangle=\langle\nabla f_i(x,\mu),w\rangle=f_i'(x;w)$. 此外, 由$\nabla\widetilde{f}_i$的连续性可知, 当$z\to x$时, $\nabla\widetilde{f}_i(z,\mu)\to\nabla\widetilde{f}_i(x,\mu)$. 结合上述关系得,

$$\begin{aligned}\lim_{z\to x,\mu\downarrow0}\langle\nabla\widetilde{f}(z,\mu),w\rangle&=\lim_{z\to x,\mu\downarrow0}\Big(\langle\nabla\widetilde{f}(z,\mu),w\rangle-\langle\nabla\widetilde{f}(x,\mu),w\rangle\Big)+\lim_{z\to x,\mu\downarrow0}\langle\nabla\widetilde{f}(x,\mu),w\rangle\\&=0+\lim_{\mu\downarrow0}\langle\nabla\widetilde{f}(x,\mu),w\rangle=f_i'(x;w).\end{aligned}$$

于是(4)中的第一条结论成立.

对任意$i\in I_2(x)$, 即$|(a_i^Tx)^2-y_i|=0$. 对任意$\mu>0$均有$|(a_i^Tx)^2-y_i|<\mu$. 令$\{z_k\}$是收敛到x的

任意点列. 若$|(a_i^T z_k)^2 - y_i| < \mu$, 即$i \in \mathcal{C}_1^\mu$, 结合式子(19), 则

$$\limsup_{\substack{z \to x, \mu \downarrow 0, \\ |(a_i^T z)^2 - y_i| < \mu}} \langle \nabla \widetilde{f}(z,\mu), w\rangle = \limsup_{\substack{z \to x, \mu \downarrow 0, \\ |(a_i^T z)^2 - y_i| < \mu}} 2(\frac{(a_i^T z)^2 - y_i}{\mu}) a_i^T z a_i^T w = 2|a_i^T z a_i^T w| = f_i'(x; w),$$

其中第二个等式成立是因为当k充分大时, 存在满足$|(a_i^T z_k)^2 - y_i| < \mu$, $|(a_i^T z_k)^2 - y_i| \to \mu$ 且$(a_i^T z)^2 - y_i$ 与$a_i^T z a_i^T w$同号的点列$\{z_k\}$. 若$|(a_i^T z_k)^2 - y_i| \geq \mu$, 则$i \in \mathcal{C}_2^\mu$. 此时$\widetilde{f}_i = f_i$, 且它们在点$z_k$处可微. 结合式子(20), 于是

$$\limsup_{\substack{z \to x, \mu \downarrow 0, \\ |(a_i^T z)^2 - y_i| \geq \mu}} \langle \nabla \widetilde{f}(z,\mu), w\rangle = \limsup_{\substack{z \to x, \mu \downarrow 0, \\ |(a_i^T z)^2 - y_i| \geq \mu}} 2\mathrm{sgn}((a_i^T z)^2 - y_i) a_i^T z a_i^T w = 2|a_i^T z a_i^T w| = f_i'(x; w),$$

其中第二个等式成立是因为当k充分大时, 存在满足$|(a_i^T z_k)^2 - y_i| \geq \mu$, 且$(a_i^T z)^2 - y_i$与$a_i^T z a_i^T w$ 同号的点列$\{z_k\}$. 综上所述(4)中的第二条结论成立. 结论得证. □

结合上述光滑化技术, 我们得到一个损失函数连续可微的优化问题:

$$\min_{x \in \mathbb{R}^n} \widetilde{F}(x,\mu) + \Phi(x). \tag{27}$$

若$\widehat{x}^\mu$是上述问题的方向稳定点, 则

$$\langle \nabla \widetilde{F}(\widehat{x}^\mu), x - \widehat{x}^\mu\rangle + \Phi'(\widehat{x}^\mu; x - \widehat{x}^\mu) \geq 0, \ \forall x \in \mathbb{R}^n,$$

即

$$\widetilde{F}'(\widehat{x}^\mu, x - \widehat{x}^\mu) + \Phi'(\widehat{x}^\mu; x - \widehat{x}^\mu) \geq 0, \ \forall x \in \mathbb{R}^n.$$

下面, 我们将研究点列$\{\widehat{x}^{\mu_k}\}$ 的聚点的性质, 其中$\{\widehat{x}^{\mu_k}\}$为方向稳定点列, $\mu_k > 0$, $k = 1, 2, \cdots$, 且当$k \to \infty$时, $\mu_k \to 0$.

定理 3.1 *($\{\widehat{x}^{\mu_k}\}$ 聚点的一致性) 令$\widehat{x}^{\mu_k}$是$\mu = \mu_k$时问题(27)的方向稳定点. 若$\mu_k > 0$, $k = 1, 2, \cdots$, 且当$k \to \infty$时, $\mu_k \to 0$. 则点列$\{\widehat{x}^{\mu_k}\}$的任意聚点为问题(6)的方向稳定点.*

证明. 令$\widehat{x}$为点列$\{\widehat{x}^{\mu_k}\}$ 的聚点. 不失一般性, 我们假设$\{\widehat{x}^{\mu_k}\}$收敛到$\widehat{x}$. 因为$\widehat{x}^{\mu_k}$是$\mu = \mu_k$ 时问题(27)的方向稳定点, 因此

$$\langle \nabla \widetilde{F}(\widehat{x}^{\mu_k}), x - \widehat{x}^{\mu_k}\rangle + \Phi'(\widehat{x}^{\mu_k}; x - \widehat{x}^{\mu_k}) \geq 0, \ \forall x \in \mathbb{R}^n.$$

由引理3.1中的(4)可知, 若$i \in I_1(\widehat{x})$, 则

$$\lim_{k \to \infty} \langle \nabla \widetilde{f}(\widehat{x}^{\mu_k}, \mu_k), w\rangle = f'(\widehat{x}; x - \widehat{x}).$$

若$i \in I_2(\widehat{x})$, 则

$$\limsup_{k \to \infty} \langle \nabla \widetilde{f}(\widehat{x}^{\mu_k}, \mu_k), w\rangle = f'(\widehat{x}; x - \widehat{x}).$$

因此, 对于任意$x \in \mathbb{R}^n$,

$$\begin{aligned}
0 &\leq \lim_{k\to\infty} \langle \nabla \widetilde{F}(\widehat{x}^{\mu_k}), x-\widehat{x}^{\mu}\rangle + \Phi'(\widehat{x}^{\mu_k}; x-\widehat{x}^{\mu_k}) \\
&= \lim_{k\to\infty} \sum_{i\in I_1(\widehat{x})} \left[\langle \nabla \widetilde{f}(\widehat{x}^{\mu_k}), x-\widehat{x}^{\mu}\rangle + \sum_{i\in I_2(\widehat{x})} \langle \nabla \widetilde{f}(\widehat{x}^{\mu_k}), x-\widehat{x}^{\mu}\rangle\right] + \Phi'(\widehat{x}; x-\widehat{x}) \\
&\leq \sum_{i\in I_1(\widehat{x})} \lim_{k\to\infty} \langle \nabla \widetilde{f}(\widehat{x}^{\mu_k}), x-\widehat{x}^{\mu}\rangle + \sum_{i\in I_2(\widehat{x})} \limsup_{k\to\infty} \langle \nabla \widetilde{f}(\widehat{x}^{\mu_k}), x-\widehat{x}^{\mu}\rangle + \Phi'(\widehat{x}; x-\widehat{x}) \\
&= \sum_{i=1}^{N} f_i'(\widehat{x}, x-\widehat{x}) + \Phi'(\widehat{x}; x-\widehat{x}).
\end{aligned}$$

于是$\widehat{x}$为问题(6)的方向稳定点, 结论得证. □

上述定理不仅建立了问题(6)与光滑问题(27)稳定点之间的联系, 同时也为使用光滑化方法求解该问题提供了理论保证.

4. 总结

本文对稀疏相位检索问题进行优化建模和分析. 基于模型稳健性考虑, 我们提出了稀疏最小一乘优化模型. 对于不连续的稀疏正则函数, 采用松弛方法, 得到了稀疏相位检索问题带Capped-L1正则的最小一乘优化模型. 利用方向稳定点, 刻画了非凸非光滑松弛问题的一阶最优性条件, 分析了方向稳定点的下界性质, 以此给出了松弛问题与原始问题全局解的等价性. 最后, 建议采用光滑化方法求解非光滑的松弛问题, 特别地, 我们建立了光滑问题与松弛正则问题解的一致性, 为使用光滑化方法求解该问题提供了理论保证.

基金项目

国家自然科学基金项目(11861020)、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)) 和贵州省青年科技人才成长项目([2018]121).

参考文献

[1] Harrison, R. (1993) Phase Problem in Crystallography. *Journal of the Optical Society of America A*, **10**, 1046-1055. https://doi.org/10.1364/JOSAA.10.001046

[2] Miao, J., Ishikawa, T. and Shen, Q. (2008) Extending X-Ray Crystallography to Allow the Imaging of Noncrystalline Materials, Cells, and Single Protein Complexes. *Annual Review of Physical Chemistry*, **59**, 387-410.
https://doi.org/10.1146/annurev.physchem.59.032607.093642

[3] Bunk, O., Diaz, A. and Pfeiffer, F. (2014) Diffractive Imaging for Periodic Samples: Retrieving One-Dimensional Concentration Profiles across Microfluidic Channels. *Acta Crystallographica*

Section A: Foundations of Crystallography, **63**, 306-314.
https://doi.org/10.1107/S0108767307021903

[4] Dainty, J. and Fienup, J. (1987) Phase Retrieval and Image Reconstruction for Astronomy. In: Stark, H., Ed., *Image Recovery: Theory and Application*, Academic Press, San Diego, 231-275.

[5] Cai, T., Li, X. and Ma, Z. (2016) Optimal Rates of Convergence for Noisy Sparse Phase Retrieval via Thresholded Wirtinger Flow. *The Annals of Statistics*, **44**, 2221-2251.
https://doi.org/10.1214/16-AOS1443

[6] Sun, J., Qu, Q. and Wright, J. (2018) A Geometric Analysis of Phase Retrieval. *Foundations of Computational Mathematics*, **18**, 1131-1198. https://doi.org/10.1007/s10208-017-9365-9

[7] Shechtman, Y., Beck, A. and Eldar, Y. (2014) GESPAR: Efficient Phase Retrieval of Sparse Signals. *IEEE Transactions on Signal Processing*, **62**, 928-938.
https://doi.org/10.1109/TSP.2013.2297687

[8] Candès, E., Li, X. and Soltanolkotabi, M. (2015) Phase Retrieval via Wirtinger Flow: Theory and Algorithms. *IEEE Transactions on Information Theory*, **61**, 1985-2007.
https://doi.org/10.1109/TIT.2015.2399924

[9] Yang, Z., Yang, L., Fang, E., *et al.* (2019) Misspecified Nonconvex Statistical Optimization for Sparse Phase Retrieval. *Mathematical Programming*, **176**, 545-571.
https://doi.org/10.1007/s10107-019-01364-5

[10] Davis, D., Drusvyatskiy, D. and Paquette, C. (2020) The Nonsmooth Landscape of Phase Retrieval. *IMA Journal of Numerical Analysis*, **40**, 2652-2695.
https://doi.org/10.1093/imanum/drz031

[11] Eldar, Y. and Mendelson, S. (2013) Phase Retrieval: Stability and Recovery Guarantees. *Applied and Computational Harmonic Analysis*, **36**, 473-494.
https://doi.org/10.1016/j.acha.2013.08.003

[12] Duchi, J. and Ruan, F. (2019) Solving (Most) of a Set of Quadratic Equalities: Composite Optimization for Robust Phase Retrieval. *Information and Inference: A Journal of the IMA*, **8**, 471-529. https://doi.org/10.1093/imaiai/iay015

[13] 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

[14] Le, T., Pham, D., Le, H., *et al.* (2015) Approximation Approaches for Sparse Optimization. *European Journal of Operational Research*, **244**, 26-46.
https://doi.org/10.1016/j.ejor.2014.11.031

[15] Pang, J., 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

[16] Gong, P., Zhang, C., Lu, Z., *et al.* (2013) A General Iterative Shrinkage and Thresholding Algorithm for Nonconvex Regularized Optimization Problems. *Proceedings of the 30th International Conference on Machine Learning*, **28**, 37-45.

[17] Bian, W. and Chen, X. (2020) A Smoothing Proximal Gradient Algorithm for Nonsmooth Convex Regression with Cardinality Penalty. *SIAM Journal on Numerical Analysis*, **58**, 858-883. https://doi.org/10.1137/18M1186009

[18] Ahn, M., Pang, J. 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

[19] Cui, Y., Pang, J. and Sen, B. (2018) Composite Difference-Max Programs for Modern Statistical Estimation Problems. *SIAM Journal on Optimization*, **28**, 3344-3374. https://doi.org/10.1137/18M117337X

[20] Soubies, E., Blanc-Féraud, L. and Aubert, G. (2017) A Unified View of Exact Continuous Penalties for $l_2 - l_0$ Minimization. *SIAM Journal on Optimization*, **27**, 2034-2060. https://doi.org/10.1137/16M1059333

[21] 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, 2021.

[22] 高岩. 非光滑优化[M]. 北京: 科学出版社, 2008.

[23] Clarke, F. (1990) Optimization and Nonsmooth Analysis. Wiley, New York. https://doi.org/10.1137/1.9781611971309