

# GEP-MS CRA for a Kind of Zero-Norm Regularized Composite Optimization Problem

Peiwen Lv

South China University of Technology, Guangzhou Guangdong  
Email: lvpw@foxmail.com

Received: Jan. 14<sup>th</sup>, 2019; accepted: Jan. 26<sup>th</sup>, 2019; published: Feb. 2<sup>nd</sup>, 2019

---

## Abstract

This article starts from the variational characterization of zero-norm, then changes such a combination optimization problem to an equivalent model which has bi-linear structure and global Lipschitz continuous. This article also designed multi-stage convex relaxation methods to solve the zero-norm regularized composite optimization problem, and analyzed the convergence for it.

## Keywords

Zero-Norm Regularization, MPEC Problem, Exact Penalty, Multi-Stage Convex Relaxation

---

# 一类零模正则复合优化问题的多阶段凸松弛法

吕佩雯

华南理工大学, 广东 广州  
Email: lvpw@foxmail.com

收稿日期: 2019年1月14日; 录用日期: 2019年1月26日; 发布日期: 2019年2月2日

---

## 摘要

本文从零模函数的变分刻画入手, 将这类带有组合性质的优化问题等价转化为具有拟双线性结构且全局 Lipschitz 连续的优化模型, 以此设计了求解零模正则化复合优化问题的多阶段凸松弛方法, 并对该方法

进行了收敛性分析。

**关键词**

零模正则化, MPEC问题, 精确罚, 多阶段凸松弛

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

**1. 引言**

在过去几十年, 高维回归在基因学、金融、图像处理、气候科学、传感器网络等领域有着广泛应用。经典的高维线性回归模型假设预测变量的维数  $n$  总是大于样本的数量  $m$ , 尽管有效的预测变量的维数  $s$  总是远小于  $n$  的。零模正则化在统计、机器学习、信号与图像处理、稀疏模型选择和误差修正、金融工程及量子计算等领域中有着广泛的应用。这种非凸正则则为稀疏恢复提供了一种有效的方法。然而, 在高维的问题中却很难分析。

**2. 零模正则复合优化问题**

**2.1. 零模正则问题**

设  $R^n$  是赋予了内积  $\langle \cdot, \cdot \rangle$  及诱导范数  $\|\cdot\|$  的有限维向量空间。考虑零模正则规划:

$$\hat{x} \in \arg \min_{x \in R^n} v \left[ \psi(Ax) + \langle c, x \rangle + \frac{\mu}{2} \|x\|^2 \right] + \|x\|_0 \tag{1}$$

其中  $v > 0$  是正则项系数,  $\psi: R^m \rightarrow (-\infty, +\infty]$  为连续可微强凸函数,  $A := [a_1^T, a_2^T, \dots, a_m^T]^T \in R^{m \times n}$  为设计矩阵,  $c$  为已知向量,  $\mu > 0$  为任意小的常数,  $x \in R^n$  为系数向量,  $\|x\|_0$  为向量  $x$  中非零元个数。该模型将为以后的逻辑回归研究作铺垫[1] [2]。记  $f(x) := \psi(Ax) + \langle c, x \rangle + \mu/2 \|x\|^2$ , 则函数  $f$  是全局 Lipschitz 连续的, 记其 Lipschitz 常数为  $L_f$  [3]。由于零模函数的不连续性和非凸性质, (1)这种问题通常是 NP 难的, 难以求得其全局最优解。一个常用的处理方法是使用凸松弛技术, 这种方法通过解一个或一系列在数值上易于处理的凸优化问题来产生一个理想的局部最优解或至少一个可行解。

**2.2. 零模问题的等价 Lipschitz 替代**

本文将从问题(1)的等价 MPEC 问题和这个 MPEC 问题的全局精确罚问题中得到解估计  $\hat{x}$ , 并证明  $\hat{x}$  可以通过求解一个精确罚问题得到。令  $\Phi$  代表闭适当凸函数  $\phi: R \rightarrow (-\infty, +\infty]$  且  $[0, 1] \subseteq \text{int}(\text{dom}\phi)$  的函数族, 其中  $\phi$  满足下面的条件:  $\phi(1) = 1$ ,  $1 > t^* := \arg \min_{0 \leq t \leq 1} \phi(t)$ ,  $\phi(t^*) = 0$  且这样的  $t^*$  是唯一的。

令  $\bar{t}$  为  $(\partial\phi)^{-1}\left(\frac{1}{1-t^*}\right) \cap [t^*, 1)$  中的最小元素, 由下面引理(参考文献[4]引理 5)可知这样的  $\bar{t}$  存在:

**引理 1:** 令  $\phi \in \Phi$ , 存在  $t_0 \in [0, 1)$  使得  $\frac{1}{1-t^*} \in \partial\phi(t_0)$ , 且对任意给定的  $\omega \geq 0$  和  $\rho > 0$ , 最优值

$$v^* := \min_{t \in [0, 1]} \{ \phi(t) + \rho\omega(1-t) \} \text{ 满足:}$$

$$\begin{cases} v^* = 1, & \text{若 } \rho\omega \in (\phi'_-(1), +\infty); \\ v^* \geq \frac{\rho\omega(1-t_0)}{\phi'_-(1)(1-t^*)}, & \text{若 } \rho\omega \in \left[\frac{1}{1-t^*}, \phi'_-(1)\right]; \\ v^* \geq \rho\omega(1-t_0), & \text{若 } \rho\omega \in \left[0, \frac{1}{1-t^*}\right). \end{cases}$$

下面将说明对任意  $\phi \in \Phi$ ，问题(1)可以重写为一个等价 MPEC 问题。不难验证：对任意给定的  $x \in R^n$  和  $\phi \in \Phi$ ，

$$\|x\|_0 = \min_{\theta \in R^n} \left\{ \sum_{i=1}^n \phi(\theta_i) : \|x\|_1 - \langle \theta, |x| \rangle = 0, 0 \leq \theta \leq e \right\} \quad (2)$$

因此问题(1)等价于：

$$\min_{x \in R^n, \theta \in R^n} \left\{ v f(x) + \sum_{i=1}^n \phi(\theta_i) : \langle e - \theta, |x| \rangle = 0, 0 \leq \theta \leq e \right\} \quad (3)$$

假设问题(1)有非空全局最优解集且最优值为  $\omega^*$ ，由文献[4]可得如下引理：

**引理 2:** 令  $\phi \in \Phi$ ，零模正则问题(1)等价于问题(3)。当  $x^*$  是问题(1)的全局最优解时， $(x^*, \max(\text{sign}(|x^*|), t^* e))$  是问题(3)的全局最优解且最优值等于  $\omega^*$ ；相反的，若  $(x^*, \theta^*)$  是问题(3)的全局最优解，那么  $x^*$  是问题(1)的全局最优解且  $\|x^*\|_0 = \sum_{i=1}^n \phi(\theta_i^*)$ 。

问题(3)的可行集中包含着如下互补约束： $e - \theta \geq 0$ ， $|x| \geq 0$  且  $\langle e - \theta, |x| \rangle = 0$ 。这说明零模正则化问题也是一个带有互补约束的数学规划问题(MPEC)，但是 MPEC 在优化中也是一类很难的问题。虽然(3)的目标函数比原问题(1)简单，但却含有非凸互补约束  $\langle e - \theta, |x| \rangle = 0$ ，这将比非凸目标函数更难处理。为解决这个非凸约束，考虑问题(3)的罚问题：

$$\min_{x \in R^n, \theta \in [0, e]} \left\{ v f(x) + \sum_{i=1}^n [\phi(\theta_i) + \rho(1-\theta_i)|x_i|] \right\} \quad (4)$$

其中  $\rho > 0$  是罚参数。下面的定理将说明问题(4)是问题(3)的全局精确罚，即它们有相同的全局最优解集[4]。

**定理 1:** 令  $\phi \in \Phi$ 。对所有  $\rho > \bar{\rho} = \nu L_f \frac{(1-t^*)\phi'_-(1)}{1-t^*}$ ，有  $S_\rho^* = S^*$ 。这里  $S_\rho^*$  是与  $\rho$  相关的问题(4)的全局最优解集， $S^*$  是问题(3)的全局最优解集。

由定理 1，零模正则问题(1)的解估计量  $\hat{x}$  可在  $\rho > \bar{\rho}$  时，由求解问题(4)得到，即：

$$(\hat{x}, \hat{\theta}) \in \min_{x \in R^n, \theta \in [0, e]} \left\{ v f(x) + \sum_{i=1}^n [\phi(\theta_i) + \rho(1-\theta_i)|x_i|] \right\}, \text{ 其中 } \rho > \bar{\rho}. \quad (5)$$

由于  $[0, 1] \subseteq \text{int}(\text{dom}\phi)$ ，函数  $\sum_{i=1}^n \phi(\theta_i)$  在  $[0, e]$  上是 Lipschitzian 连续的，所以问题(5)在其可行集  $R^n \times [0, e]$  上也是 Lipschitzian 连续的。对任意  $\phi \in \Phi$ ，定义相应的函数  $\phi: R \rightarrow (-\infty, +\infty]$ ：若  $t \in [0, 1]$ ， $\phi(t) = \phi(t)$ ；若  $t \notin [0, 1]$ ， $\phi(t) = +\infty$ 。问题(5)可以重新写为更简便的形式：

$$\min_{x \in R^n, \theta \in R^n} \left\{ v f(x) + \rho \|x\|_1 + \sum_{i=1}^n [\phi(\theta_i) - \rho \theta_i |x_i|] \right\}.$$

由共轭函数的定义和以上讨论，可以得到：

$$\hat{x} \in \arg \min_{x \in R^n} \left\{ v f(x) + \rho \|x\|_1 - \sum_{i=1}^n \phi^*(\rho |x_i|) \right\}, \text{ 其中 } \rho > \bar{\rho} \quad (6)$$

这意味着下面的函数是  $\|x\|_0$  的等价 DC 替代:  $\Theta(x) := \rho\|x\|_1 - \sum_{i=1}^n \phi^*(\rho|x_i|)$ 。

### 3. 多阶段凸松弛法

从上一节可知, 求解零模正则化问题的解估计  $\hat{x}$ , 对于  $\phi \in \Phi$  且  $\rho > \bar{\rho}$  时, 可由求解罚问题(5)得到。当给定  $v > 0$  时, 阈值  $\bar{\rho} > 0$  是容易计算得到的。对一个给定的  $\rho > \bar{\rho}$ , 虽然罚问题(5)是非凸的, 但是这种结构使得它比零模正则化问题更好解决。特别的, 当变量  $\theta$  选定时, 问题(5)退化为关于  $x$  的凸的极小化问题; 当变量  $x$  选定时, 问题(5)退化为关于  $\theta$  的凸的极小化问题, 后面将看到这样的问题是有闭式解的。基于此, 本文将通过使用多阶段凸松弛法交替求解问题(5)来得到解估计  $\hat{x}$  [5]。

#### 算法 1. (GEP-MSCRA 求解 $\hat{x}$ )

**Step 0:** 取  $\phi \in \Phi$ ,  $v > 0$  和初始点  $\theta^0 \in [0, \bar{t}e]$ 。置  $k := 1$ 。

**Step 1:** 令  $v^{k-1} = e - \theta^{k-1}$  并计算下面的极小化问题:  $x^k \in \arg \min_{x \in R^n} \left\{ f(x) + \lambda \sum_{i=1}^n v_i^{k-1} |x_i| \right\}$ 。 (7)

若  $k=1$ , 由  $x^1$  的取值选取参数  $\rho > \bar{\rho}$  并使  $\lambda = \rho v^{-1}$ 。

**Step 2:** 计算下面的极小化问题:  $\theta^k \in \arg \min_{\theta \in [0, e]} \left\{ \sum_{i=1}^n \left[ \phi(\theta_i) - \rho \theta_i |x_i^k| \right] \right\}$ 。 (8)

**Step 3:** 置  $k \leftarrow k + 1$ , 返回(S. 1)步。

**注 1:** 由函数  $\phi$  的定义,  $\theta^k$  是(8)的最优解当且仅当  $\theta_i^k \in \partial \phi^*(\rho|x_i^k|)$ ,  $i=1, \dots, n$ 。由于  $\phi^*$  是  $R$  上的凸函数, 次微分  $\partial \phi^*(\rho|x_i^k|)$  是容易求得的。因此, 多阶段凸松弛法每次迭代的计算难度都在于解决问题(7), 即一个加权  $L_1$ -范数正则化问题[6] [7] [8]。这是一类凸优化问题, 它可以通过标准的凸优化方法求解, 比如: 增广拉格朗日法, 内点法[9], IRLS-LARS [10], 路径跟踪法, 迭代加权最小二乘法等。对于多阶段凸松弛法, 下面的命题(参考文献[4]注释 3.1)说明一旦满足互补条件, 迭代点  $x^k$  就是原问题(1)的局部最优解。

**命题 1:** 对任意的  $k \geq 1$ ,  $(x^k, \theta^k)$  为算法 1 生成的迭代点。若  $\sum_{i=1}^n (1 - \theta_i^{k-1}) |x_i^k| = 0$ , 则  $x^k$  是原问题(1)的局部最优解; 且当  $\sum_{i=1}^n \phi(\theta_i^{k-1}) = \|x^k\|_0$  时, 对于  $\rho > \bar{\rho}$ ,  $(x^k, \max(\text{sign}(|x^k|), t^*e))$  是问题(5)的局部最优解。

### 4. 理论分析

本节内容主要关注在限制强凸条件下, 针对  $f(x) := \psi(Ax) + \langle c, x \rangle + \frac{\mu}{2} \|x\|^2$ , 多阶段凸松弛法的收敛性分析。假设  $\bar{x}$  是真实解, 本文将建立迭代点  $x^k$  到真实解  $\bar{x}$  的误差界, 并证明误差数列随着  $k$  的增长是严格下降的。记  $\bar{S} := \{i: \bar{x}_i \neq 0\}$ ,  $\bar{r} = |\bar{S}| = \|\bar{x}\|_0 < n$ , 则  $\bar{S}^c := \{i: \bar{x}_i = 0\}$ 。对  $\bar{S}$  和整数  $l > 0$ , 定义集合:

$$C(\bar{S}, l) := \left\{ x \in R^n : \exists S \supseteq \bar{S} \text{ 且 } |S| \leq l, \text{ 使得 } \sum_{i \in S^c} |x_i| \leq \frac{2}{1-l} \sum_{i \in S} |x_i| \right\}。$$

若矩阵  $A$  满足: 对任意  $x \in C$ ,  $\frac{\alpha}{2} \|Ax\|^2 \geq \kappa \|x\|^2$  成立, 则称矩阵  $A$  在集合  $C$  上满足  $\kappa$ -RSC, 其中  $\kappa > 0$ 。假设噪声向量  $\varepsilon$  中的元素是独立次高斯的, 则下面的假设成立:

**假设 1:** 假设  $\varepsilon_i (i=1, \dots, m)$  是独立次高斯分布的, 即: 存在  $\sigma > 0$ , 使得对所有  $i$  和  $t \in R$ ,

$$E[\exp(t\varepsilon_i)] \leq \exp\left(\frac{\sigma^2 t^2}{2}\right)。$$

为给出每一迭代步  $x^k$  到真实解  $\bar{x}$  的误差界, 需要将参考文献[4]引理 7.8 拓展为下面一些引理:

**引理 3:** 对于  $k \geq 1$ , 记  $\Delta^k = x^k - \bar{x}$ ,  $\hat{\varepsilon} = A^T \nabla \psi(A\bar{x}) + c + \mu\bar{x}$ 。若存在一个指标集  $S^{k-1} \supseteq \bar{S}$ , 使得

$\min_{i \in (S^{k-1})^c} \theta_i^{k-1} \leq \bar{t}$ , 则当  $\lambda \geq \frac{(3-\bar{t})\|\hat{\epsilon}\|_\infty}{1-\bar{t}}$  时, 下式成立:  $\sum_{i \in (S^{k-1})^c} |\Delta_i^k| \leq \frac{2}{1-\bar{t}} \sum_{i \in S^{k-1}} |\Delta_i^k|$ .

**证明:** 对于问题(7),  $x^k$  和  $\bar{x}$  分别代表最优解和可行解, 则有:

$$\psi(Ax^k) + \langle c, x^k \rangle + \frac{\mu}{2} \|x^k\|^2 + \lambda \sum_{i=1}^n v_i^{k-1} |x_i^k| \leq \psi(A\bar{x}) + \langle c, \bar{x} \rangle + \frac{\mu}{2} \|\bar{x}\|^2 + \lambda \sum_{i=1}^n v_i^{k-1} |\bar{x}_i|$$

由于  $\hat{\epsilon} = A^T \nabla \psi(A\bar{x}) + c + \mu\bar{x}$ , 所以:

$$\begin{aligned} & \psi(Ax^k) - \psi(A\bar{x}) + \langle c, \Delta^k \rangle + \frac{\mu}{2} (\|x^k\|^2 - \|\bar{x}\|^2) - \langle \hat{\epsilon}, \Delta^k \rangle \\ & \leq -\langle \hat{\epsilon}, \Delta^k \rangle + \lambda \sum_{i=1}^n v_i^{k-1} (|\bar{x}_i| - |x_i^k|) \\ & \leq \sum_{i=1}^n |\hat{\epsilon}_i| |\Delta_i^k| + \lambda \sum_{i \in \bar{S}} v_i^{k-1} (|\bar{x}_i| - |x_i^k|) - \lambda \sum_{i \in \bar{S}^c} v_i^{k-1} |x_i^k| \\ & \leq \sum_{i=1}^n |\hat{\epsilon}_i| |\Delta_i^k| + \lambda \sum_{i \in \bar{S}} v_i^{k-1} |\Delta_i^k| - \lambda \sum_{i \in (S^{k-1})^c} v_i^{k-1} |\Delta_i^k| \\ & \leq \sum_{i \in S^{k-1} \cup \bar{S}} |\hat{\epsilon}_i| |\Delta_i^k| + (\lambda + \|\hat{\epsilon}\|_\infty) \sum_{i \in \bar{S}} |\Delta_i^k| + (\|\hat{\epsilon}\|_\infty - \lambda(1-\bar{t})) \sum_{i \in (S^{k-1})^c} |\Delta_i^k| \end{aligned} \tag{9}$$

由于  $\psi$  在  $R^m$  上是强凸的, 即:  $\psi(Ax^k) - \psi(A\bar{x}) - \langle A^T \nabla \psi(A\bar{x}), \Delta^k \rangle \geq \alpha/2 \|\Delta^k\|^2 \geq 0$ , 且  $\mu/2 (\|x^k\|^2 - \|\bar{x}\|^2) - \langle \mu\bar{x}, \Delta^k \rangle = \mu/2 \|\Delta^k\|^2$ , 所以,

$$(\lambda(1-\bar{t}) - \|\hat{\epsilon}\|_\infty) \sum_{i \in (S^{k-1})^c} |\Delta_i^k| \leq \sum_{i \in S^{k-1} \cup \bar{S}} |\hat{\epsilon}_i| |\Delta_i^k| + (\lambda + \|\hat{\epsilon}\|_\infty) \sum_{i \in \bar{S}} |\Delta_i^k| \leq (\lambda + \|\hat{\epsilon}\|_\infty) \sum_{i \in S^{k-1}} |\Delta_i^k| \tag{10}$$

再由引理条件  $\lambda \geq \frac{(3-\bar{t})\|\hat{\epsilon}\|_\infty}{1-\bar{t}}$ , 即可得到不等式。证毕。

**引理 4:** 对于  $k \geq 1$ , 若存在一个指标集  $S^{k-1} \supseteq \bar{S}$ , 使得  $\min_{i \in (S^{k-1})^c} \theta_i^{k-1} \leq \bar{t}$ , 且矩阵  $A$  在集合  $C(\bar{S}, |S^{k-1}|)$  上

满足常数  $\gamma_k > 0$  的限制强凸条件(RSC), 则当  $\lambda \geq \frac{(3-\bar{t})\|\hat{\epsilon}\|_\infty}{1-\bar{t}}$  时, 下式成立:

$$\|\Delta^k\| \leq \frac{1}{\gamma_k} \left( \|\hat{\epsilon}_{S^{k-1}}\| + \lambda \sqrt{\sum_{i \in \bar{S}} (v_i^{k-1})^2} \right).$$

**证明:** 由引理 3 可知,  $\Delta^k \in C(\bar{S}, |S^{k-1}|)$ , 再由(9)式得到:

$$\begin{aligned} \gamma_k \|\Delta^k\|^2 & \leq (\gamma_k + \mu/2) \|\Delta^k\|^2 \leq \sum_{i=1}^n |\hat{\epsilon}_i| |\Delta_i^k| + \lambda \sum_{i \in \bar{S}} v_i^{k-1} |\Delta_i^k| - \lambda \sum_{i \in (S^{k-1})^c} v_i^{k-1} |\Delta_i^k| \\ & \leq \sum_{i=1}^n |\hat{\epsilon}_i| |\Delta_i^k| + \lambda \sum_{i \in \bar{S}} v_i^{k-1} |\Delta_i^k| - \lambda(1-\bar{t}) \sum_{i \in (S^{k-1})^c} |\Delta_i^k| \\ & \leq \sum_{i \in S^{k-1}} |\hat{\epsilon}_i| |\Delta_i^k| + \lambda \sum_{i \in \bar{S}} v_i^{k-1} |\Delta_i^k| \\ & \leq \sqrt{\sum_{i \in S^{k-1}} |\hat{\epsilon}_i|^2} \|\Delta^k\| + \lambda \sqrt{\sum_{i \in \bar{S}} (v_i^{k-1})^2} \|\Delta^k\| \end{aligned} \tag{11}$$

因此, 结论得证。证毕。

下面给出迭代点到真实解的误差界。由参考文献[4]定理 4.1 可得:

**定理 2:** 令  $\hat{\epsilon} = A^T \nabla \psi(A\bar{x}) + c + \mu\bar{x}$ , 假设矩阵  $A$  在集合  $C(\bar{S}, 1.5\bar{r})$  上满足常数  $\kappa > 0$  的限制强凸条件

(RSC), 且  $\frac{\sqrt{3}(4-2\bar{t})(1-t^*)}{(3-\bar{t})\kappa} \leq \nu \leq \frac{1-\bar{t}}{(3-\bar{t})\|\hat{\epsilon}\|_\infty}$ 。若  $\frac{\nu(3-\bar{t})\|\hat{\epsilon}\|_\infty}{1-\bar{t}} \leq \rho \leq \sqrt{\frac{(3-\bar{t})\kappa\nu}{\sqrt{3}(4-2\bar{t})(1-t^*)}}$ , 则

$$\|x^k - \bar{x}\| \leq \begin{cases} \frac{v^{-1}(4-2\bar{t})}{\kappa(3-\bar{t})} \sqrt{1.5\bar{r}}, & \text{若 } k=1; \\ \frac{\rho v^{-1}(4-2\bar{t})}{\kappa(3-\bar{t})} \sqrt{1.5\bar{r}}, & \text{若 } k=2,3,\dots \end{cases}$$

定理 2 虽然给出了多阶段凸松弛法产生的迭代点的误差界, 但是依然不清楚当前迭代点  $x^k$  的误差界是否优于先前迭代点  $x^{k-1}$ , 即误差界数列是否下降。下面的定理 3(参考文献[4]定理 4.2)对所有  $i \in \bar{S}$ , 通过  $I_\Lambda(i)$  和  $I_{F^k}(i)$  界定  $(v_i^{k-1})^2$  解决了这一问题, 其中  $\Lambda := \{i : |\bar{x}_i| \leq 2\phi'(1)/\rho\}$ ,

$F^k := \left\{i : \|x_i^k - |\bar{x}_i|\| \geq [(1-t^*)\rho]^{-1}\right\}$ ,  $k=1,2,\dots$  且  $F^0 = \bar{S}$ 。首先对  $i \in \bar{S}$ , 通过  $I_\Lambda(i)$  和  $I_{F^k}(i)$ , 给出  $(v_i^{k-1})^2$  的上界(参考文献[4]引理 9)。

**引理 5:** 对每一个  $k \geq 1$ , 令  $F^k$  为上面定义的指标集, 则有:  $\sqrt{\sum_{i \in \bar{S}} (v_i^{k-1})^2} \leq \sqrt{\sum_{i \in \bar{S}} I_\Lambda(i)} + \sqrt{\sum_{i \in \bar{S}} I_{F^{k-1}}(i)}$ 。

**定理 3:** 假设矩阵  $A$  在集合  $C(\bar{S}, 1.5\bar{r})$  上满足常数  $\kappa > 0$  的限制强凸条件(RSC), 若参数  $v$  和  $\rho$  的选取与定理 2 一样, 则对所有的  $k \geq 1$ :

$$\|x^k - \bar{x}\| \leq \frac{\sqrt{3}}{\kappa(\sqrt{3}-1)} \left[ \|\hat{\epsilon}_{\bar{S}}\| + \rho v^{-1} \sqrt{\sum_{i \in \bar{S}} I_\Lambda(i)} \right] + \left( \frac{1}{\sqrt{3}} \right)^{k-1} \|x^1 - \bar{x}\|。$$

由定理 3 可知:  $x^1$  的误差界与  $x^k$  ( $k \geq 2$ ) 的误差界紧密相关, 且若前者充分小, 则后者也将很小。然而, 若  $\rho$  选择恰当, 即使  $x^1$  的误差界很大,  $x^k$  ( $k \geq 2$ ) 的误差界也将很小。定理还说明误差数列随着  $k$  增大而递减, 递减到  $\frac{\sqrt{3}}{\kappa(\sqrt{3}-1)} \left[ \|\hat{\epsilon}_{\bar{S}}\| + \lambda \sqrt{\sum_{i \in \bar{S}} I_\Lambda(i)} \right]$ , 称之为统计误差。统计误差包含两部分: 第一部分是由噪声引起的, 第二部分是由  $\phi$ ,  $\rho$  和  $\lambda$  的选择产生的。

## 5. 结论

本文从零模函数的变分刻画入手, 将这类带有组合性质的优化问题等价转化为具有拟双线性结构且全局 Lipschitz 连续的优化模型, 以此设计了求解零模正则化复合优化问题的多阶段凸松弛方法, 并对该方法进行了收敛性分析。在理论方面, 本文证实了将文献[4]的理论结果即: 在比限制等距性质(RIP)更弱的限制强凸条件下, 定量刻画了多阶段凸松弛法每阶段最优解的误差界, 且每阶段最优解的误差界递减拓展到复合模型的有效性, 将为以后研究逻辑回归奠基理论基础。以后将证明当阶段数达到一定条件后, 误差界保持不变且精确识别真实解的支撑集。

## 致 谢

感谢贲树军师兄给予的帮助。

## 参考文献

- [1] Loh, P.L. and Wainwright, M.J. (2015) Regularized M-Estimators with Nonconvexity: Statistical and Algorithmic Theory for Local Optima. *Journal of Machine Learning Research*, **16**, 559-616.
- [2] Casasaya, J. and Llibre, J. (2009) Large-Scale Sparse Logistic Regression. ACM Press, New York, 547-556.
- [3] Rockafellar, R.T. (1970) Convex Analysis. Princeton University Press.
- [4] Bi, S.J. and Pan, S.H. (2017) GEP-MSRA for Computing the Group Zero-Norm Regularized Least Squares Estimator. *arXiv.org*.

- 
- [5] 贲树军. 低秩优化问题的多阶段凸松弛法研究[D]: [博士学位论文]. 广州: 华南理工大学, 2014.
- [6] Candes, E.J., Wakin, M.B. and Boyd, S.P. (2008) Enhancing Sparsity by Reweighted  $\ell_1$  Minimization. *Journal of Fourier Analysis and Applications*, **14**, 877-905. <https://doi.org/10.1007/s00041-008-9045-x>
- [7] Schmidt, M., Fung, G. and Rosales, R. (2007) Fast Optimization Methods for  $\ell_1$  Regularization: A Comparative Study and Two New Approaches. *Lnai*, **4701**, 286-297. [https://doi.org/10.1007/978-3-540-74958-5\\_28](https://doi.org/10.1007/978-3-540-74958-5_28)
- [8] Park, M.Y. and Hastie, T. (2008)  $\ell_1$  Regularized Path Algorithm for Generalized Linear Models. *Journal of the Royal Statistical Society: Series B*, **69**, 659-677. <https://doi.org/10.1111/j.1467-9868.2007.00607.x>
- [9] Koh, K., Kim, S. and Boyd, S. (2007) An Interior-Point Method for Large-Scale  $\ell_1$ -Regularized Logistic Regression. *Journal of Machine Learning Research*, **8**, 1519-1555.
- [10] Lee, S., Lee, H., Abbeel, P. and Ng, A.Y. (2006) Efficient  $\ell_1$  Regularized Logistic Regression. *National Conference on Artificial Intelligence*, **1**, 401-408.

#### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2163-1476, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [orf@hanspub.org](mailto:orf@hanspub.org)