Operations Research and Fuzziology 运筹与模糊学, 2023, 13(6), 7598-7611
Published Online December 2023 in Hans. https://www.hanspub.org/journal/orf
https://doi.org/10.12677/orf.2023.136745

一类混合稀疏组稀疏优化问题的邻近梯度算法

童兴华，彭定涛*，张 弦

贵州大学, 数学与统计学院, 贵州 贵阳

收稿日期：2023年10月20日；录用日期：2023年12月22日；发布日期：2023年12月29日

摘 要

本文研究了一类混合稀疏组稀疏优化问题，其中损失函数为光滑凸函数，正则项为稀疏ℓ_1范数与组稀疏$\ell_{\alpha,p}(\alpha \geq 1, p > 0)$范数的组合。首先，提出了邻近梯度算法求解此混合稀疏组稀疏优化问题。其次，分别讨论了凸$(p \geq 1)$和非凸$(0 < p < 1)$两种情况下算法的收敛性。最后，对于非凸问题，在$\alpha = 1, p = \frac{1}{2}, \frac{2}{3}$时给出组合惩罚项邻近算子的闭式解。本文结果为求解混合稀疏组稀疏优化问题提供了理论依据和可行途径。

关键词

混合稀疏组稀疏优化问题，邻近梯度算法，邻近算子，闭式解

Proximal Gradient Algorithm for a Class of Mixed Sparse and Group Sparse Optimization Problems

Xinghua Tong, Dingtao Peng*, Xian Zhang

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Oct. 20^{th}, 2023; accepted: Dec. 22^{nd}, 2023; published: Dec. 29^{th}, 2023

* 通讯作者。

文章引用: 童兴华, 彭定涛, 张弦. 文章标题[J]. 运筹与模糊学, 2023, 13(6): 7598-7611.
DOI: 10.12677/orf.2023.136745

Abstract

This paper investigates a class of mixed sparse and group sparse optimization problems in which the loss function is a smooth and convex function, and the regularization term is a combination of the sparse ℓ_1 norm and the group sparse $\ell_{\alpha,p}$ norm ($\alpha \geq 1, p > 0$). Firstly, we propose a proximal gradient algorithm to solve this class of mixed sparse optimization problems. Secondly, we discuss the convergence properties in both convex ($p \geq 1$) and non-convex ($0 < p < 1$) scenarios. For non-convex problems, closed-form solutions of the proximal operators to the mixed penalty can be obtained when $\alpha = 1$, $p = \frac{1}{2}, \frac{2}{3}$. The result provides a theoretical foundation and a feasible approach for solving mixed sparse and group sparse optimization problems.

Keywords

Mixed Sparse and Group Sparse Optimization Problems, Proximal Gradient Algorithm, Proximal Operator, Closed-Form Solution

1. 引言

本文考虑如下混合稀疏组稀疏优化问题:

$$\min_{x\in\mathbb{R}^n} F(x) = f(x) + \lambda_1||x||_1 + \lambda_2||x||^p_{\alpha,p}, \tag{1}$$

其中$\lambda_1, \lambda_2 > 0, p > 0, \alpha \geq 1$, f在$\mathbb{R}^n$上为梯度L-Lischiitz连续的凸函数. 将向量x 分为r 个不相交的分组, 即$x = \left(x^T_{G_1}, \dots, x^T_{G_r}\right)^T \in \mathbb{R}^n$, 令$n_i = \#\{j : j \in G_i\}$, $i = 1, 2, 3, \cdots r$, 对应的范数为:

$$||x||_1 = \sum_{i=1}^{n} |x_i|, \quad ||x||^p_{\alpha,p} = \sum_{i=1}^{r} ||x_{G_i}||^p_\alpha.$$

当$\lambda_2 = 0$时, 模型(1)退化为稀疏优化的Lasso模型; 当$\lambda_1 = 0, \alpha = 2, p = 1$时, 模型(1)退化为组稀疏

优化的组Lasso 模型.

稀疏优化是指决策变量具有稀疏性或者近似稀疏性的优化问题, 本质上是从一个欠定线性系统中寻找稀疏解的过程. 在信号重构、图像处理、模式识别、变量选择和生物信息学等领域具有广泛的应用 [1, 2].

对线性最小二乘组稀疏优化模型, Jain等 [3]提出了贪婪-IHT算法; Chen 等 [4] 考虑了三种带线性不等式约束的线性最小乘组稀疏优化模型并给出了三种模型的稳定点、局部最优解、全局最优解之间的关系; Pan和Chen [5]用截断折叠凹(Capped folded concave)类型的松弛函数来松弛组稀疏图像恢复问题,并给出了一个光滑惩罚算法. 对组Lasso问题, Zhang等 [6, 7]提了邻近对偶牛顿算法和增广拉格朗日方法. 对带组稀疏惩罚的无约束非凸优化, Peng和Chen [8]证明了在一定条件下,二阶方向稳定点一定是强局部最优解并设计了计算该稳定点的二阶光滑化算法. 对稀疏分组的SLOPE模型与OSCAR 模型, Luo等 [9]设计了高效稳定的半光滑牛顿增广格朗日法. 对带凸约束和组稀疏惩罚的优化问题, Chen和Toint [10]分析了它的高阶复杂度. Beck 和Hallak [11]对一种广义的组稀疏模型分析了最优性条件和组稀疏相关的邻近算子. Zhang 等 [12] 中通过引入组约束特征值条件的概念, 建立了基于组ℓ_p正则化的logistic 回归模型的全局恢复界, 并采用交替方向乘法器算法(ADMM) 求解模型. Cai T等 [13]研究了问题所需的样本量和误差估计的收敛速度, 为稀疏组Lasso问题的样本复杂度和误差估计提供了理论保证. 但上述研究均未考虑稀疏与组稀疏的混合优化.

混合稀疏组稀疏优化问题(1)是比稀疏或组稀疏更一般化的模型. 本文研究该模型的优化理论和算法.

2. 邻近梯度算法

对问题(1)中目标函数, 令$h(x) := \lambda_1||x||_1 + \lambda_2||x||^p_{\alpha,p}$, 使用邻近梯度方法, 其迭代格式为:

$$\begin{cases} z^k = x^k - t_k \nabla f(x^k), \\ x^{k+1} = \text{prop}_{t_k h}(z^k), \end{cases} \tag{2}$$

其中

$$\text{prop}_{t_k h}(z^k) = \arg\min_{x \in \text{dom} h} \left\{ h(x) + \frac{1}{2t_k} \left\| x - z^k \right\|^2 \right\}. \tag{3}$$

由于模型中p的取值为$p > 0$, 因此模型包含两种不同类型的优化, 即当$p \geq 1$ 时模型(1)是凸优化, 当$0 < p < 1$时模型(1)是非凸优化, 下面分别研究其性质.

2.1. 凸问题 ($p \geq 1$)

对于正则项$||x||^p_{\alpha,p}$, 由凸函数的保凸运算可知：当$\alpha \geq 1, p \geq 1$ 时$||x||^p_{\alpha,p}$为凸函数, 此时问题(1)是一个凸优化, 此时, 邻近梯度算法具有下述收敛性.

引理 2.1 [1] *设$\{x^k\}$是迭代格式(2)生成的序列, 取固定步长$t_k = t \in (0, \frac{1}{L}]$, 若$p \geq 1$, 则$\{x^k\}$收敛到*

问题(1)的全局极小点, 函数值的收敛速度为$\mathcal{O}(\frac{1}{k})$, 即:

$$F(x^k)-F(x^*)\le \frac{1}{2kt}\|x^0-x^*\|^2. \tag{4}$$

为了加快$\{x^k\}$的收敛速度, 在(2) 的基础上可引入Nesterov加速技术, 其迭代格式为:

$$\begin{cases} x^{k+1} &= \operatorname{prop}_{t_k h}(y^k-t_k\nabla f(y^k)),\\ t_{k+1} &= \frac{1+\sqrt{1+4t_k^2}}{2},\\ y^{k+1} &= x^k+(\frac{t_k-1}{t_{k+1}})(x^k-x^{k-1}). \end{cases} \tag{5}$$

引理 2.2 [14] 设$\{x^k\}$是迭代格式(5)生成的序列, 其中$y^1=x^0, t_1=1$, 若$p\ge 1$, 则$\{x^k\}$收敛到问题(1)的全局极小点, 函数值的收敛速度为$\mathcal{O}(\frac{1}{k^2})$, 即:

$$F(x^k)-F(x^*)\le \frac{2tL}{(k+1)^2}\|x^0-x^*\|^2. \tag{6}$$

2.2. 非凸问题 $(0<p<1)$

当$0<p<1$时, 正则项$\|x\|_{\alpha,p}^p$是非凸、非Lipschitz的, 此时, 迭代格式(2)生成的序列$\{x^k\}$ 可以收敛到问题(1)的critial 点, 为介绍critial点, 需要引进下半连续函数的极限次微分,下面介绍相关的概念.

下半连续函数的次微分定义如下.

定义 2.1 [15] 设$g:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ 是正常下半连续函数, 对任意$x\in\operatorname{dom} g$,

(1) g 在x 处的*Fréchet* 次微分记作$\widehat{\partial} g(x)$, 是由所有满足如下关系的向量$u\in\mathbb{R}^n$ 构成的集合

$$\liminf_{y\neq x, y\to x}\frac{g(y)-g(x)-\langle u, y-x\rangle}{\|y-x\|}\ge 0,$$

若$x\notin\operatorname{dom} g$, 则定义$\partial g(x)=\emptyset$.

(2) g 在x 处的极限次微分记作$\partial g(x)$, 如下定义

$$\partial g(x):=\left\{u\in\mathbb{R}^n:\exists x^k\to x, g\left(x^k\right)\to g(x), u^k\to u, u^k\in\widehat{\partial} g\left(x^k\right)\right\}.$$

定义 2.2 [15] 若$x^*\in\mathbb{R}^n$满足$0\in\partial g(x^*)$,则称x^* 为g 的*critical* 点.

引理 2.3 [15] 设$\{x^k\}$是迭代格式(2)生成的序列, 取固定步长$t=t_k$, 若$0<p<1$, 则$\{x^k\}$收敛到问题(1)的*critical*点.

3. 混合非凸惩罚项邻近算子的闭式解

使用邻近梯度算法求解问题(1)时, 关键的计算在于求解子问题(3), 我们希望得到子问题(3)的闭式解, 但是通常闭式解是无法计算的, 本节针对p的某些特殊取值($p=\frac{1}{2},\frac{1}{3}$), 求子问题(3)的闭式解.

注意到$h(x)=\lambda_1||x||_1+\lambda_2||x||_{\alpha,p}^p$与$\|x-z^k\|$都是按组可分的, 因此问题(3)是组可分的:

$$\operatorname{prop}_{t_k h}(z^k)=\operatorname{prop}_{t_k h_1}(z^k_{G_1})\times\cdots\times\operatorname{prop}_{t_k h_r}(z^k_{G_r}), \tag{7}$$

其中

$$\begin{aligned}\operatorname{prop}_{t_k h_i}(z^k_{G_i})&=\arg\min_{x\in\mathbb{R}^{n_i}}\left\{h_i(x)+\frac{1}{2t_k}\left\|x-z^k_{G_i}\right\|^2\right\},\qquad \forall i=1,2,...,r.\\&=\arg\min_{x\in\mathbb{R}^{n_i}}\left\{\lambda_1\|x\|_1+\lambda_2\|x\|_\alpha^p+\frac{1}{2t_k}\left\|x-z^k_{G_i}\right\|^2\right\}.\end{aligned}\tag{8}$$

结合问题(1)中目标函数的可分性, 以及对应的邻近算子的可分性, 我们给出梯度步$z^k=x^k-t_k\nabla f(x^k)$ 与x^{k+1}的二者支撑集的一致性, 首先给出如下命题:

命题 3.1 设向量$z,\hat{x}\in\mathbb{R}^{n_i}$, z 的支撑集为$S=\{i\mid z_i\neq 0\}$, 其补集为S^c, $\hat{x}$ 满足如下等式

$$\hat{x}\in\arg\min_{x\in\mathbb{R}^{n_i}}\left\{\lambda_1\|x\|_1+\lambda_2\|x\|_\alpha^p+\frac{1}{2t_k}\left\|x-z\right\|^2\right\}, \tag{9}$$

则$\hat{x}$ 的支撑集也为S.

证明: 设$\hat{x}$ 的支撑集为S'. 设$\mathbb{R}^{n_i}$中与指标集S对应的子空间为$\mathbb{R}^l$. 由题可知

$$\begin{aligned}\hat{x}&\in\arg\min_{x\in\mathbb{R}^{n_i}}\left\{\lambda_1\|x\|_1+\lambda_2\|x\|_\alpha^p+\frac{1}{2t_k}\left\|x-z\right\|^2\right\}\\&=\arg\min_{x\in\mathbb{R}^{n_i}}\left\{\lambda_1\|x\|_1+\lambda_2\|x\|_\alpha^p+\frac{1}{2t_k}\left\|x_S-z_S\right\|^2+\frac{1}{2t_k}\|x_{S^c}\|^2\right\}\\&=\arg\min_{x_S\in\mathbb{R}^l,x_{S^c}=0}\left\{\lambda_1\|x_S\|_1+\lambda_2\|x_S\|_\alpha^p+\frac{1}{2t_k}\left\|x_S-z_S\right\|^2\right\},\end{aligned}\tag{10}$$

由(10)的最优性条件可知, $\hat{x}_S$满足如下条件,

$$\lambda_1\operatorname{sign}(\hat{x}_S)+\lambda_2 p\|\hat{x}_S\|_\alpha^{p-\alpha}\begin{pmatrix}|\hat{x}_1|^{\alpha-1}\operatorname{sign}(\hat{x}_1)\\\vdots\\|\hat{x}_l|^{\alpha-1}\operatorname{sign}(\hat{x}_l)\end{pmatrix}+\frac{1}{t_k}(\hat{x}_S-z_S)=0. \tag{11}$$

由上式可知: $\operatorname{sign}(\hat{x}_S)=\operatorname{sign}(z_S)$, 因此, $S=S'$. □

结合邻近算子的可分性质与命题3.1, 有如下结论:

定理 3.1 对任意的$k>1$,$\{z^k\},\{x^{k+1}\}$是由迭代(2)所生成的序列, 则$\{z^k\},\{x^{k+1}\}$的支撑集一致, 即非零元素所在位置相同.

现定义如下函数

$$Q_{\alpha,p}(x)=\lambda_1\|x\|_1+\lambda_2\|x\|_\alpha^p+\frac{1}{2t_k}\|x-z\|^2. \tag{12}$$

现将问题放到支撑集S 对应的子空间上, 由定理3.1可得:

$$\min_{x\in\mathbb{R}^{n_i}}\{Q_{\alpha,p}(x)\}=\min_{x_S\in\mathbb{R}^l}\{Q_{\alpha,p}(x_S)\},\qquad l=|S|.$$

定义如下邻近算子

$$P_{\alpha,p}(z)\in\arg\min_{x\in\mathbb{R}^{n_i}}\{Q_{\alpha,p}(x)\}, \tag{13}$$

则$P_{\alpha,p}(z)$可分为两部分, 即

$$P_{\alpha,p}(z_S)=\arg\min_{x_S\in\mathbb{R}^l}\{Q_{\alpha,p}(x_S)\},\qquad P_{\alpha,p}(z_{S^c})=0.$$

由于函数$Q_{\alpha,p}(x_S)$仅在$x_S=0$处不可微, 故$P_{\alpha,p}(z_S)$ 取值为0 或者$\tilde{x}$, 其中$\tilde{x}$满足如下一阶最优性条件:

$$\lambda_1\mathrm{sign}(\tilde{x})+\lambda_2 p\|\tilde{x}\|_\alpha^{p-\alpha}\begin{pmatrix}|\tilde{x}_1|^{\alpha-1}\operatorname{sign}(\tilde{x}_1)\\ \vdots\\ |\tilde{x}_l|^{\alpha-1}\operatorname{sign}(\tilde{x}_l)\end{pmatrix}+\frac{1}{t}(\tilde{x}-z_S)=0. \tag{14}$$

为了求解方程(14)在$\alpha=1,p=1/2,2/3$时的解，我们接下来讨论三次方程根的分布，由于三次方程$x^3+bx^2+cx+d=0$ 可通过平移变换$y=x-\frac{b}{3}$ 去掉二次项，从而我们只需要考虑缺项的三次方程

$$x^3+ax+b=0,\qquad a,b\in\mathbb{R},ab\neq 0. \tag{15}$$

由卡丹公式可得方程的三个根分别为

$$\begin{aligned}
x_1&=\sqrt[3]{-\frac{b}{2}+\sqrt{\frac{b^2}{4}+\frac{a^3}{27}}}+\sqrt[3]{-\frac{b}{2}-\sqrt{\frac{b^2}{4}+\frac{a^3}{27}}},\\
x_2&=\sqrt[3]{-\frac{b}{2}+\sqrt{\frac{b^2}{4}+\frac{a^3}{27}}}\,w+\sqrt[3]{-\frac{b}{2}-\sqrt{\frac{b^2}{4}+\frac{a^3}{27}}}\,w^2,\\
x_3&=\sqrt[3]{-\frac{b}{2}+\sqrt{\frac{b^2}{4}+\frac{a^3}{27}}}\,w^2+\sqrt[3]{-\frac{b}{2}-\sqrt{\frac{b^2}{4}+\frac{a^3}{27}}}\,w.
\end{aligned} \tag{16}$$

其中

$$w=\frac{-1+\sqrt{3}i}{2},\qquad w^2=\frac{-1-\sqrt{3}i}{2}.$$

不妨设

$$k_1=\sqrt[3]{-\frac{b}{2}+\sqrt{\frac{b^2}{4}+\frac{a^3}{27}}},\qquad k_2=\sqrt[3]{-\frac{b}{2}-\sqrt{\frac{b^2}{4}+\frac{a^3}{27}}},\qquad \Delta=\frac{b^2}{4}+\frac{a^3}{27}.$$

则三个根可表示为

$$\begin{aligned} x_1&=k_1+k_2,\\ x_2&=-\frac{1}{2}(k_1+k_2)+\frac{\sqrt{3}i}{2}(k_1-k_2),\\ x_3&=-\frac{1}{2}(k_1+k_2)-\frac{\sqrt{3}i}{2}(k_1-k_2). \end{aligned}\tag{17}$$

引理 3.1 设方程*(15)*的判别式为$\Delta=\dfrac{b^2}{4}+\dfrac{a^3}{27}$，则方程*(15)*根的分布情况如下：

(1) 当$\Delta>0$ 时，方程*(15)*有一个实根和一对共轭虚根,

(2) 当$\Delta=0$ 时，方程*(15)*有一个单实根和一个二重实根,

(3) 当$\Delta<0$ 时，方程*(15)*有三个不相等的实根.

证明：

(1) 当$\Delta>0$时，可知$k_1\neq k_2$，从而方程*(15)* 只有一个实根x_1. 考虑函数$f(x)=x^3+ax+b$，易知当$x\to\pm\infty$ 时，有$f(x)\to\pm\infty$ 并且有$f(0)=b$，从而根据$f(x)$的函数图像可以推出如下结果:

  (i) 当$b>0$ 时，方程*(15)*有一个负实根,

  (ii) 当$b<0$ 时，方程*(15)*有一个正实根。

(2) 当$\Delta=0$时，可知$k_2-k_1=0$，从而方程有三个实根，即

$$x_1=2\sqrt[3]{-\frac{b}{2}},\qquad x_{2,3}=\sqrt[3]{\frac{b}{2}}.\tag{18}$$

显然可以得出

  (i) 当$b>0$ 时，方程*(15)*有一个负实根，两个相等的正实根,

  (ii) 当$b<0$ 时，方程*(15)*有两个相等的负实根，一个正实根。

(3) 当$\Delta<0$时，考虑函数$f(x)=x^3+ax+b$，有

$$f'(x)=3x^2+a=0\Longrightarrow x=\pm\sqrt{-\frac{a}{3}}.\tag{19}$$

由$\Delta<0$可得

$$\frac{b^2}{4}+\frac{a^3}{27}<0\Longrightarrow-\frac{a^3}{27}>\frac{b^2}{4}>0\Longrightarrow a<0.$$

因此(19)中的根在实数域中是有意义的，将两个根代入函数$f(x)$ 可得：

$$f(\sqrt{-\frac{a}{3}})=\frac{2a\sqrt{-a}}{3\sqrt{3}}+b,\qquad f(-\sqrt{-\frac{a}{3}})=\frac{-2a\sqrt{-a}}{3\sqrt{3}}+b.$$

又有

$$\begin{aligned}-\frac{a^3}{27}>\frac{b^2}{4}&\Longrightarrow\left(\frac{b}{2}\right)^2<\left(\frac{-a}{3}\right)^3\\&\Longrightarrow\frac{a\sqrt{-a}}{3\sqrt{3}}<\frac{b}{2}<-\frac{a\sqrt{-a}}{3\sqrt{3}}\\&\Longrightarrow\frac{2a\sqrt{-a}}{3\sqrt{3}}<b<-\frac{2a\sqrt{-a}}{3\sqrt{3}}.\end{aligned}$$

从而可得

$$f(\sqrt{\frac{-a}{3}})<0,\qquad f(-\sqrt{\frac{-a}{3}})>0.$$

由数轴穿根法结合函数$f(x)$的函数图像可得：

(i) 当$b>0$ 时，方程(15)有一个负实根, 两个正实根,

(ii) 当$b<0$ 时，方程(15)有一个正实根，两个负实根。

考虑方程(15)在三角函数形式下的根，对k_1 进行化简可得：

$$k_1^3=-\frac{b}{2}+i\sqrt{-\Delta},\quad k_2^3=-\frac{b}{2}-i\sqrt{-\Delta},\tag{20}$$

故虚数k_1^3 和k_2^3 的模均为

$$|k_1^3|=|k_2^3|=\sqrt{\left(-\frac{b}{2}\right)^2-\Delta}=\sqrt{-\left(\frac{a}{3}\right)^3}=r.\tag{21}$$

将虚数k_1^3 转化为极坐标形式

$$k_1^3=r\left(\frac{-\frac{b}{2}}{r}+i\frac{\sqrt{-\Delta}}{r}\right)=r\left(\cos\alpha+i\sin\alpha\right),\quad\alpha=\arccos\left(\frac{-b}{2r}\right).\tag{22}$$

同理，虚数k_2^3 的极坐标形式为

$$k_2^3=r\left(\cos\alpha-i\sin\alpha\right),\tag{23}$$

通过欧拉公式转换，可得

$$k_1=\sqrt[3]{r}\left(\cos\frac{\alpha}{3}+i\sin\frac{\alpha}{3}\right),\quad k_2=\sqrt[3]{r}\left(\cos\frac{\alpha}{3}-i\sin\frac{\alpha}{3}\right).\tag{24}$$

代入可得

$$k_1+k_2=2\sqrt[3]{r}\cos\frac{\alpha}{3},\quad k_1-k_2=\sqrt[3]{r}\sin\frac{\alpha}{3}i. \tag{25}$$

结合*(17)*, 有

$$x_1=2\sqrt[3]{r}\cos\frac{\alpha}{3}, \tag{26}$$

$$\begin{aligned}x_2&=-\frac{1}{2}(k_1+k_2)+\frac{\sqrt{3}i}{2}(k_1-k_2)\\&=-\sqrt[3]{r}\left(\frac{1}{2}\cos\frac{\alpha}{3}+\frac{\sqrt{3}}{2}\sin\frac{\alpha}{3}\right)\\&=-2\sqrt[3]{r}\cos\left(\frac{\pi}{3}-\frac{\alpha}{3}\right)=2\sqrt[3]{r}\cos\left(\frac{\alpha}{3}+\frac{2\pi}{3}\right),\end{aligned}$$

同理可得

$$\begin{aligned}x_3&=-\frac{1}{2}(k_1+k_2)-\frac{\sqrt{3}i}{2}(k_1-k_2)\\&=-2\sqrt[3]{r}\cos\left(\frac{\pi}{3}+\frac{\alpha}{3}\right)=2\sqrt[3]{r}\cos\left(\frac{\alpha}{3}+\frac{4\pi}{3}\right),\end{aligned}$$

由函数$arccos(t)$的性质可知, 当$0<t<1$时，$arccos(t)\in(0,\frac{\pi}{2})$, 当$-1<t<0$ 时，$arccos(t)\in(\frac{\pi}{2},\pi)$.

(*i*) 当$b>0$ 时，方程*(15)*有两个正实根x_1,x_3 , 有一个负实根x_2,

(*ii*) 当$b<0$ 时，方程*(15)*有一个正实根x_1 ，两个负实根x_2,x_3.

利用上述结果, 参考 [16]的思想, 下面得到$\alpha=1$, $p=\frac{1}{2},\frac{2}{3}$ 时子问题(3)的闭式解.

定理 3.2 设$a=\|z_S\|_1-\lambda_1 tl$, $b=\lambda_2 tl$, 邻近算子*(13)* 的解满足如下结论：

(*i*) 当$\alpha=1,p=\dfrac{1}{2}$时, 有

$$P_{1,1/2}(z_S)=\begin{cases}\tilde{x}, & Q_{1,1/2}(\tilde{x})<Q_{1,1/2}(0),\\\{0,\tilde{x}\}, & Q_{1,1/2}(\tilde{x})=Q_{1,1/2}(0),\\0, & Q_{1,1/2}(\tilde{x})>Q_{1,1/2}(0),\end{cases} \tag{27}$$

其中

$$\tilde{x}\ =\ \arg\min_{x\in\{\tilde{x}_1,\tilde{x}_2\}}\left\{Q_{1,1/2}(x)\right\}, \tag{28}$$

$$\tilde{x}_1=z_S-t(\lambda_1+\frac{\lambda_2}{2\eta_1})\mathrm{sign}(z_S),\qquad \tilde{x}_2=z_S-t(\lambda_1+\frac{\lambda_2}{2\eta_2})\mathrm{sign}(z_S), \tag{29}$$

$$\eta_1=\begin{cases}0, & \Delta_1>0\\ \sqrt[3]{\frac{b}{4}}, & \Delta_1=0\\ 2\sqrt[3]{r_1}\cos\left(\frac{\alpha_1}{3}\right), & \Delta_1<0\end{cases},\qquad \eta_2=\begin{cases}0, & \Delta_1>0\\ \sqrt[3]{\frac{b}{4}}, & \Delta_1=0\\ 2\sqrt[3]{r_1}\cos\left(\frac{\alpha_1}{3}+\frac{4\pi}{3}\right), & \Delta_1<0\end{cases}. \tag{30}$$

$$\Delta_1=\frac{b^2}{16}-\frac{a^3}{27},\ \alpha_1=\arccos\left(\frac{-b}{4r_1}\right),\ r_1=\sqrt{\left(\frac{a}{3}\right)^3}; \tag{31}$$

(ii) 当$\alpha=1,p=\dfrac{2}{3}$时，有

$$P_{1,2/3}(z_S)=\begin{cases}\tilde{\omega}, & Q_{1,2/3}(\tilde{\omega})<Q_{1,2/3}(0),\\ \{0,\tilde{\omega}\}, & Q_{1,2/3}(\tilde{\omega})=Q_{1,2/3}(0),\\ 0, & Q_{1,2/3}(\tilde{\omega})>Q_{1,2/3}(0),\end{cases} \tag{32}$$

其中

$$\tilde{\omega}=z_S-\frac{4\lambda_2 t\theta_1^{1/2}}{3\left(\theta_1^{3/2}+\sqrt{2a-\theta_1^3}\right)}\operatorname{sign}(z_S),$$

$$\theta_1^2=M=\begin{cases}\sqrt[3]{\frac{a^2}{2}+\sqrt{\frac{a^4}{4}-(\frac{8b}{9})^3}}+\sqrt[3]{\frac{a^2}{2}-\sqrt{\frac{a^4}{4}-(\frac{8b}{9})^3}}, & \Delta_2>0\\ 2\sqrt[3]{\frac{a^2}{2}}, & \Delta_2=0\\ 2\sqrt[3]{r_2}\cos\left(\frac{\alpha_2}{3}\right), & \Delta_2<0\end{cases}. \tag{33}$$

$$\Delta_2=\frac{a^4}{4}-(\frac{8b}{9})^3,\alpha_2=\arccos(\frac{a^2}{2r_2}),r_2=\sqrt{(\frac{8b}{9})^3}$$

证明：*(i)*当$\alpha=1,p=\dfrac{1}{2}$时，*(14)*可转化为

$$\lambda_1\operatorname{sign}(\tilde{x})+\frac{\lambda_2\operatorname{sign}(\tilde{x})}{2\|\tilde{x}\|_1^{1/2}}+\frac{1}{t}(\tilde{x}-z_S)=0, \tag{34}$$

化简为

$$z_S=|\tilde{x}|\operatorname{sign}(\tilde{x})+\frac{(2\lambda_1\|\tilde{x}\|_1^{\frac{1}{2}}+\lambda_2)t\operatorname{sign}(\tilde{x})}{2\|\tilde{x}\|_1^{\frac{1}{2}}},$$

两边同时取l_1 范数

$$\|z_S\|_1=\frac{\left\|2\|\tilde{x}\|_1^{\frac{1}{2}}|x|+(2\lambda_1\|\tilde{x}\|_1^{\frac{1}{2}}+\lambda_2)t\mathbf{1}\right\|_1}{2\|\tilde{x}\|_1^{\frac{1}{2}}}.$$

其中$\mathbf{1}$ 为l阶全*1*向量，由$a, b \in \mathbb{R}_+^l, \|a+b\|_1 = \|a\|_1 + \|b\|_1$，可得

$$2\|\tilde{x}\|_1^{\frac{3}{2}} - 2(\|z_S\|_1 - \lambda_1 tl)\|\tilde{x}\|_1^{\frac{1}{2}} + \lambda_2 tl = 0, \tag{35}$$

记$\eta = \|\tilde{x}\|_1^{\frac{1}{2}} > 0$，$a = \|z_S\|_1 - \lambda_1 tl$，$b = \lambda_2 tl$，方程*(35)*可转换成如下三次方程

$$\eta^3 - a\eta + \frac{1}{2}b = 0. \tag{36}$$

当$a \le 0$时，方程*(36)*显然没有正实根，可得$p_{1,\frac{1}{2}}(z_S) = 0$.

当$a > 0$，$b > 0$ 时,设$\Delta_1 = \frac{b^2}{16} - \frac{a^3}{27}$，$\alpha_1 = \arccos\left(\frac{-b}{4r}\right), r_1 = \sqrt{\left(\frac{a}{3}\right)^3}$，由引理*3.1* 可知，方程*(36)*可能有两个正实根

$$\eta_2 = \begin{cases} 0 & \Delta_1 > 0 \\ \sqrt[3]{\frac{b}{4}} & \Delta_1 = 0 \\ 2\sqrt[3]{r_1}\cos\left(\frac{\alpha_1}{3}\right) & \Delta_1 < 0 \end{cases}, \qquad \eta_2 = \begin{cases} 0 & \Delta_1 > 0 \\ \sqrt[3]{\frac{b}{4}} & \Delta_1 = 0 \\ 2\sqrt[3]{r_1}\cos\left(\frac{\alpha_1}{3} + \frac{4\pi}{3}\right) & \Delta_1 < 0 \end{cases}. \tag{37}$$

代入*(34)*可得：

$$\tilde{x}_1 = z_S - t(\lambda_1 + \frac{\lambda_2}{2\eta_1})\mathrm{sign}(z_S), \qquad \tilde{x}_2 = z_S - t(\lambda_1 + \frac{\lambda_2}{2\eta_2})\mathrm{sign}(z_S). \tag{38}$$

不妨设

$$\tilde{x} \quad = \quad \arg \min_{x \in \{\tilde{x}_1, \tilde{x}_2\}} \left\{Q_{1,1/2}(x_S)\right\}. \tag{39}$$

综上所述

$$P_{1,1/2}(z_S) = \begin{cases} \tilde{x}, & Q_{1,1/2}(\tilde{x}) < Q_{1,1/2}(0), \\ \{0, \tilde{x}\}, & Q_{1,1/2}(\tilde{x}) = Q_{1,1/2}(0), \\ 0, & Q_{1,1/2}(\tilde{x}) > Q_{1,1/2}(0). \end{cases} \tag{40}$$

*(ii)*当$\alpha = 1, p = \dfrac{2}{3}$时，*(14)*可转化为

$$\lambda_1 \mathrm{sign}(\tilde{x}) + \frac{2\lambda_2 \mathrm{sign}(\tilde{x})}{3\|\tilde{x}\|_1^{1/3}} + \frac{1}{t}(\tilde{x} - z_S) = 0, \tag{41}$$

与*(i)*中的处理相同，可得：

$$\|\tilde{x}\|_1^{\frac{4}{3}} - (\|z_S\|_1 - \lambda_1 tl)\|\tilde{x}\|_1^{\frac{1}{3}} + \frac{2}{3}\lambda_2 tl = 0.$$

记$\eta = \|\tilde{x}\|_1^{\frac{1}{3}} > 0, a = \|z_S\|_1 - \lambda_1 tl, b = \lambda_2 tl$ ，方程*(35)*可转换成如下四次方程

$$\eta^4 - a\eta + \frac{2}{3}b = 0, \tag{42}$$

因$b > 0$, 故当$a \le 0$时, 方程(42)显然没有实根, 可得$p_{1,\frac{2}{3}}(z_S) = 0$.

下面考虑$a > 0$的情形. 将(42)转化为两个二次方程的乘积形式:

$$\eta^4 - a\eta + \frac{2}{3}b = (\eta^2 + \theta_1\eta + \theta_2)(\eta^2 + \theta_3\eta + \theta_4). \tag{43}$$

使用待定系数法求解系数, 对比等号两边对应系数, 得

$$\theta_1 + \theta_3 = 0, \quad \theta_2 + \theta_4 + \theta_1\theta_3 = 0, \quad \theta_1\theta_4 + \theta_2\theta_3 = -a, \quad \theta_2\theta_4 = \frac{2}{3}b.$$

从而可得

$$\theta_3 = -\theta_1, \quad \theta_2 = \frac{1}{2}(\theta_1^2 + \frac{a}{\theta_1}), \quad \theta_4 = \frac{1}{2}(\theta_1^2 - \frac{a}{\theta_1}), \quad \theta_2\theta_4 = \frac{1}{4}(\theta_1^4 - \frac{a^2}{\theta_1^2}) = \frac{2}{3}b.$$

现在只需要求出θ_1^2 的值即可求出各个系数, 从而可求解方程(42). 令$M = \theta_1^2$, 将上面第四个等式化简之后可得:

$$M^3 - \frac{8}{3}bM - a^2 = 0. \tag{44}$$

因$a > 0, b > 0$, 设$\Delta_2 = \frac{a^4}{4} - (\frac{8b}{9})^3, \alpha_2 = \arccos(\frac{a^2}{2r_2}), r_2 = \sqrt{(\frac{8b}{9})^3}$, 通过引理3.1 可得, 方程(44)的解为

$$\theta_1^2 = M = \begin{cases} \sqrt[3]{\frac{a^2}{2} + \sqrt{\frac{a^4}{4} - (\frac{8b}{9})^3}} + \sqrt[3]{\frac{a^2}{2} - \sqrt{\frac{a^4}{4} - (\frac{8b}{9})^3}}, & \Delta_2 > 0 \\ 2\sqrt[3]{\frac{a^2}{2}}, & \Delta_2 = 0 \\ 2\sqrt[3]{r_2}\cos\left(\frac{\alpha_2}{3}\right), & \Delta_2 < 0 \end{cases}. \tag{45}$$

由系数之间的关系可得, 方程(42)等价于

$$\eta^2 + \theta_1\eta + \frac{1}{2}(\theta_1^2 + \frac{a}{\theta_1}) = 0 \quad \text{或} \quad \eta^2 + \theta_3\eta + \frac{1}{2}(\theta_1^2 - \frac{a}{\theta_1}) = 0.$$

因$\eta > 0$, 由求根公式可得:

$$\eta_1 = \frac{1}{2}\left(|\theta_1| + \sqrt{\frac{2a}{|\theta_1|} - \theta_1^2}\right), \quad \eta_2 = \frac{1}{2}\left(|\theta_1| - \sqrt{\frac{2a}{|\theta_1|} - \theta_1^2}\right), \tag{46}$$

容易看出$\eta_1 > \eta_2$,由数轴穿根法易知方程(42)在η_2 左边为正, 右边为负, 即η_2是函数的鞍点而不是最小值点, 故舍弃掉η_2, 联立公式(41) 可得:

$$\tilde{x} = z_S - \frac{4\lambda_2 t\theta_1^{1/2}}{3\left(\theta_1^{3/2} + \sqrt{2a - \theta_1^3}\right)}\operatorname{sign}(z_S). \tag{47}$$

4. 总结

本文研究了一类混合稀疏组稀疏优化问题，使用邻近梯度算法求解该模型，进一步分别讨论了山($p \geq 1$) 和非凸($0 < p < 1$) 两种情况的迭代点列的收敛性.因为该算法一个关键的步骤求解邻近子问题, 因此对于非凸问题,得到了$p = \frac{1}{2}, \frac{2}{3}$时邻近子问题的闭式解. 在求解邻近子问题的闭式解时, 参考了 [16] 中的做法, 其本质是将问题转化为求三次和四次方程的根, 然后对方程的根进行讨论从而得出闭式解, 由伽罗瓦理论可知五次以上方程没有通用根式解, 从而这种方法使用时对参数的选择具有一定的限制. 本文所研究的混合稀疏优化模型(1)相比 [16] 中的组稀疏模型多了一个ℓ_1范数正则项, 不仅要求分组之间稀疏, 组内也要做到稀疏, 从而导致求解邻近子问题的闭式解时不能简单将其分为零和非零两部分, 因此我们将子问题放到梯度步z^k 所对应的子空间上求解, 其理论依据是证明了梯度步z^k 和迭代值x^{k+1}二者支撑集的一致性. 本文结果为求解混合稀疏组稀疏优化问题提供了理论依据和可行方法.

基金项目

国家自然科学基金项目(12261020)、贵州省科技计划项目(ZK[2021]009, [2018]5781)、贵州省高层次留学人才创新创业择优资助重点项目([2018]03)、贵州省青年科技人才成长项目([2018]121).

参考文献

[1] 刘浩洋, 户将, 李勇锋, 文再文. 最优化: 建模, 算法与理论[M]. 北京: 高等教育出版社, 2020.

[2] 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

[3] Jain, P., Rao, N. and Dhillon, I.S. (2016) Structured Sparse Regression via Greedy Hard Thresholding. *Advances in Neural Information Processing Systems*, **29**, 1516-1524.

[4] Chen, X., Pan, L. and Xiu, N. (2023) Solution Sets of Three Sparse Optimization Problems for Multivariate Regression. *Journal of Global Optimization*, **87**, 347-371. https://doi.org/10.1007/s10898-021-01124-w

[5] Pan, L. and Chen, X. (2021) Group Sparse Optimization for Image Recovery Using Capped Folded Concave Functions. *SIAM Journal on Imaging Sciences*, **14**, 1-25. https://doi.org/10.1137/19M1304799

[6] 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

[7] Zhang, Y., Zhang, N., Sun, D. and Toh, K.C. (2020) A Proximal Point Dual Newton Algorithm for Solving Group Graphical Lasso Problems. *SIAM Journal on Optimization*, **30**, 2197-2220. https://doi.org/10.1137/19M1267830

[8] 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

[9] Luo, Z., Sun, D., Toh, K.C. and Xiu, N. (2019) Solving the OSCAR and SLOPE Models Using a Semismooth Newton-Based Augmented Lagrangian Method. *Journal of Machine Learning Research*, **20**, 1-25.

[10] Chen, X. and Toint, P.L. (2021) High-Order Evaluation Complexity for Convexly-Constrained Optimization with Non-Lipschitzian Group Sparsity Terms. *Mathematical Programming*, **187**, 47-78. https://doi.org/10.1007/s10107-020-01470-9

[11] Beck, 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

[12] Zhang, Y., Wei, C. and Liu, X. (2022) Group Logistic Regression Models with $l_{p,q}$ Regularization. *Mathematics*, **10**, Article 2227. https://doi.org/10.3390/math10132227

[13] Cai, T.T., Zhang, A.R. and Zhou, Y.C. (2022) Sparse Group Lasso: Optimal Sample Complexity, Convergence Rate, and Statistical Inference. *IEEE Transactions on Information Theory*, **68**, 5975-6002. https://doi.org/10.1109/TIT.2022.3175455

[14] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, **2**, 183-202. https://doi.org/10.1137/080716542

[15] Bolte, J., Sabach, S. and Teboulle, M. (2014) Proximal Alternating Linearized Minimization for Nonconvex and Nonsmooth Problems. *Mathematical Programming*, **146**, 459-494. https://doi.org/10.1007/s10107-013-0701-9

[16] Hu, Y., Li, C., Meng, K., Qin, J. and Yang, X. (2017) Group Sparse Optimization via lp, q Regularization. *The Journal of Machine Learning Research*, **70**, 960-1011.