**Operations Research and Fuzziology** 运筹与模糊学, 2024, 14(3), 694-706
Published Online June 2024 in Hans. https://www.hanspub.org/journal/orf
https://doi.org/10.12677/orf.2024.143306

# 线性等式约束广义Lasso问题的算法研究

**吴娟娟**

上海理工大学理学院，上海

收稿日期：2024年3月18日；录用日期：2024年6月19日；发布日期：2024年6月27日

## 摘 要

**随着大数据时代的到来，众多研究领域都涉及到优化问题的求解，其中Lasso问题的求解尤其受到学者们的广泛研究。针对Lasso问题的求解，学者们研发出众多算法。随着应用的场景不同以及对数据的要求不同，带有约束的广义Lasso问题逐渐受到人们关注。本文将已有的快速邻近点算法结合半光滑牛顿算法，应用到对一类含线性等式约束的广义Lasso问题进行求解，并在一定的假设条件下证明了该算法的收敛性。最后，通过数值实验证实了该算法的高效性。**

## 关键词

**Lasso问题，邻近点算法，半光滑牛顿算法**

# Algorithm Research on Generalized Lasso Problems with Linear Equality Constraints

**Juanjuan Wu**

College of Science, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 18th, 2024; accepted: Jun. 19th, 2024; published: Jun. 27th, 2024

## Abstract

**With the advent of the big data era, many research fields involve solving optimization problems, among which the solution of the Lasso problem has been widely studied by scholars. Scholars have developed numerous algorithms for solving the Lasso problem. As the application scenarios vary and the data requirements differ, the generalized Lasso problem with constraints has gradually attracted attention. This paper combines existing fast proximal point algorithms with a semi-smooth Newton algorithm to solve a class of generalized Lasso problems with linear equality constraints. The convergence of the algorithm is proved under certain assumptions. Finally, the efficiency of the algorithm is verified through numerical experiments.**

文章引用: 吴娟娟. 线性等式约束广义 Lasso 问题的算法研究[J]. 运筹与模糊学, 2024, 14(3): 694-706.
DOI: 10.12677/orf.2024.143306

**Keywords**

**Lasso Problem, Proximal Point Algorithm, Semi-Smooth Newton Algorithm**



## 1. 引言

零和约束 Lasso 问题在生物领域有着广泛的应用背景。这种约束模型适用于许多生物学问题，尤其是在基因组学和生物信息学领域。例如：文献[1]中考虑具有零和约束的 Lasso 问题，如下：

$$\begin{aligned}&\min_x \frac{1}{2}\|Ax-y\|^2+\lambda\|x\|_1\\&\text{s.t. } \sum_{i=1}^n x_i=0,\end{aligned} \tag{1}$$

其中$\|\cdot\|$表示欧式范数，$\|\cdot\|_1$表示$l_1$范数，$x\in R^n$为变量，$A\in R^{m\times n}$，$y\in R^m$，$\lambda\in R$均为给定量。在一些组成数据的回归模型中，即表示整体百分比或比例的数据中，需要这种特殊结构的约束。

由于约束的特殊性，该类优化问题通常出现在许多不同的领域，例如地质学、生物学、生态学和经济学。如：在微生物组分析中，数据集通常是标准化的，并产生成分数据[2] [3]。因此，设计求解该类问题的高效算法具有更加广泛的应用价值。为了使求解该类问题的算法具有广泛的应用性，现考虑求解如下更一般的等式约束广义 Lasso 问题：

$$\begin{aligned}&\min_x \frac{1}{2}\|Ax-y\|^2+h(x)\\&\text{s.t. } \mathcal{A}(x)=b,\end{aligned} \tag{2}$$

其中非光滑项函数$h(x):=\lambda\|x\|_1$，$\|\cdot\|$表示欧式范数，$\|\cdot\|_1$表示$l_1$范数，$x\in R^n$为变量$A\in R^{m\times n}$，$y\in R^m$，$\lambda\in R$均为给定量；约束中的$\mathcal{A}:R^n\to R^s$表示线性映射；向量$b\in R^s$。

近年来，许多学者利用该问题的特殊结构，建立了良好的算法用来求解模型(1)。例如，Lin 等人于 2014 年在[4]中提出了一种通过循环坐标下降的拉格朗日方法解决子问题；而 Altenbuchinge 等人于 2017 年在[5]中提出了基于变量随机选择的坐标下降策略。对于复合凸优化问题，Li [6]等人提出了一种快速邻近点算法求解复合凸优化问题模型，并基于对偶原理的半光滑牛顿算法高效稳定地求解了邻近点算法所涉及的重要子问题。

此外，考虑到约束套索的更一般形式，2016 年 Gaines 等人在[7]中分析了一种基于二次规划的方法和 ADMM 方法，2020 年 Deng 等人在[8]中提出了一种半光滑牛顿增广拉格朗日方法，文献[7] [9] [10]中设计了路径算法。

基于以上算法的启发，本文的目标是将邻近点半光滑牛顿算法推广到用于求解带线性等式约束 Lasso 问题(2)，并通过数值实验证实该算法求解该类带等式约束 Lasso 问题具有良好性能。

## 2. 理论知识

本节将讨论凸复合优化问题的一些稳定性性质。在之后的理论分析中可以很容易看出，这些稳定性性质是我们建立邻近点算法快速收敛的关键。

首先，给出邻近映射的定义，这将对分析邻近点算法起着至关重要的作用。

**定义 1** ([11] Definition 6.1) (proximal mapping)令 $\mathcal{H}$ 为有限维实值希尔伯特空间，$h: R \to R\cup\{+\infty\}$ 为凸且封闭的函数。给定参数 $t>0$，邻近映射定义如下：

$$\operatorname{prox}_{th}(x) := \arg\min_{u\in\mathcal{H}}\left\{h(u)+\frac{1}{2t}\|u-x\|^2\right\},\ \ \forall x\in\mathcal{H}.$$

接下来将给出 Moreau 包络的具体定义，如下：

**定义 2** ([11] Definition 6.52) (Moreau envelope)给定适当的闭凸函数 $h: R \to R\cup\{+\infty\}$ 且参数 $t>0$，则函数 $h$ 的 Moreau 包络为：

$$M_h^t(x) := \min_{u\in\mathcal{H}}\left\{h(u)+\frac{1}{2t}\|u-x\|^2\right\}.$$

其中参数 $t$ 称为光滑参数。

接下来，将给出有关非光滑性的相关分析的重要结论。

设 $\mathcal{X},\mathcal{Y}$ 为两个有限维实向量空间。设 $\mathcal{O}$ 为 $\mathcal{X}$ 的一个开集，且 $\Phi:\mathcal{O}\to\mathcal{Y}$ 为局部 Lipschitz 连续函数。由 Rademacher 定理([12] Theorem 9.60)可知，函数 $\Phi(\cdot)$ 在 $\mathcal{O}$ 中几乎处处 $F$ 可微。用 $\mathcal{D}_\Phi$ 表示集合 $\mathcal{O}$ 中函数 $\Phi$ 均 $F$ 可微的点集。用 $\partial_B\Phi(x)$ 表示函数 $\Phi$ 在点 $x\in\mathcal{O}$ 处的 $B$ 次微分，定义如下：

$$\partial_B\Phi(x) := \left\{\lim_{k\to\infty}\nabla\Phi\left(x^k\right): x^k\in\mathcal{D}_\Phi, x^k\to x\right\}.$$

进一步，函数 $\Phi$ 在点 $x$ 处的 Clarke 广义 Jacobian 矩阵是由 $\partial_B\Phi(x)$ 的凸包定义，即 $\partial\Phi(x) := co\left(\partial_B\Phi(x)\right)$，并且用 $\nabla\Phi\left(x^k;\Delta x\right)$ 表示函数 $\Phi$ 在点 $x$ 处沿着非零方向 $\Delta x\in\mathcal{X}$ 的方向导数。

下面回顾一些关于半光滑性的概念。这有助于建立算法局部收敛的超线性速度。类似于[13]和[14]，给出半光滑性的定义。

**定义 3 (半光滑性质)** [13] [14]令 $F: U\subseteq R^n\to R^m$ 是局部 Lipschitz 连续的。令 $\mathcal{M}(x)$ 为函数族，其中每个函数 $M(x;\cdot)$ 为 $R^n$ 到 $R^m$ 的映射，并且满足 $M(x;0)=0$。函数 $F$ 被称为在点 $\bar{x}\in U$ 处关于 $\mathcal{M}$ 是半光滑的，若 $F$ 在点 $\bar{x}\in U$ 处方向可微且对于任意 $M(\bar{x}+\Delta x;\cdot)\in\mathcal{M}(\bar{x}+\Delta x)$ 有

$$F(\bar{x}+\Delta x)-F(\bar{x})-M(\bar{x}+\Delta x;\Delta x)=o\left(\|\Delta x\|_2\right),$$

其中 $o\left(\|\Delta x\|_2\right)$ 表示当 $\Delta x\to 0$ 时 $\|\Delta x\|_2$ 的高阶无穷小量。

同样，函数 $F$ 被称为在点 $\bar{x}\in U$ 处关于 $\mathcal{M}$ 是强半光滑的，若 $F$ 在点 $\bar{x}\in U$ 处方向可微且对于任意 $M(\bar{x}+\Delta x;\cdot)\in\mathcal{M}(\bar{x}+\Delta x)$ 有

$$F(\bar{x}+\Delta x)-F(\bar{x})-M(\bar{x}+\Delta x;\Delta x)=O\left(\|\Delta x\|_2^2\right),$$

其中 $O\left(\|\Delta x\|_2^2\right)$ 表示当 $\Delta x\to 0$ 时 $\|\Delta x\|_2^2$ 的同阶无穷小量。

在上述理论知识的基础上，将进行优化问题的具体算法设计。

## 3. 求解等式约束广义 Lasso 问题(2)

由于问题(2)中 $l_1$ 范数的存在，使得问题不易求解，因此考虑采用邻近点算法(PPA)求解问题(2)，邻近点算法求解产生了问题(2)的 PPA 子问题，此时 PPA 子问题的目标函数中两个变量可以完全分离，又保证了其强凸性。

### 3.1. PPA 子问题

将问题(2)的目标函数 Moreau-Yosida 正则化后得 PPA 子问题：

$$\begin{aligned}&\min_{x}\frac{1}{2}\|Ax-y\|_2^2+h(x)+\frac{1}{2t}\left\|x-x^{k-1}\right\|_2^2\\&\text{s.t. } \mathcal{A}(x)=b,\end{aligned}\tag{3}$$

其中 $t>0$ 是正则化因子，$x^{k-1}\in R^n$。

由于问题(2)的目标函数是凸函数，所以加入正则项后可得目标函数 $\frac{1}{2}\|Ax-y\|_2^2+\lambda\|x\|_1+\frac{1}{2t}\left\|x-x^{k-1}\right\|_2^2$ 是强凸的，因此在约束准则 Slater 条件成立的情况下，强对偶性是成立的，即原始问题与对偶问题的对偶间隙为零，因此可以通过求解(3)的对偶问题的最优可行解，从而得到对应的原始问题(2)的最优可行解 $x^*$。

### 3.2. 对偶问题

首先给出 PPA 子问题(3)的对偶问题。通过引入松弛变量 $z$，问题(3)等价于

$$\begin{aligned}&\min_{x,z} f(x,z)=\frac{1}{2}\|z\|_2^2+h(x)+\frac{1}{2t}\left\|x-x^{k-1}\right\|_2^2\\&\text{s.t. } Ax-y-z=0,\\&\quad\ \mathcal{A}(x)=b.\end{aligned}\tag{4}$$

其拉格朗日函数为：

$$\mathcal{L}(x,z;\xi_1,\xi_2):=\frac{1}{2}\|z\|_2^2+h(x)+\frac{1}{2t}\left\|x-x^{k-1}\right\|_2^2+\langle Ax-y-z,\xi_1\rangle+\langle\mathcal{A}(x)-b,\xi_2\rangle,\tag{5}$$

其中 $\xi_1$，$\xi_2$ 分别是 $Ax-y-z=0$，$\mathcal{A}(x)=b$ 对应的拉格朗日乘子。关于 $x,z$ 极小化 $\mathcal{L}$ 得：

$$\begin{aligned}D_t(\xi_1,\xi_2)&:=\inf_{x,z}\mathcal{L}(x,z;\xi_1,\xi_2)\\&=\inf_{x,z}\frac{1}{2}\|z\|_2^2-\langle z,\xi_1\rangle+h(x)+\frac{1}{2t}\left\|x-x^{k-1}\right\|_2^2+\langle x,A^{\mathrm{T}}\xi_1\rangle-\langle y,\xi_1\rangle+\langle\mathcal{A}(x)-b,\xi_2\rangle\\&=\frac{1}{2}\|\bar z\|_2^2-\langle\bar z,\xi_1\rangle+h(\bar x)+\frac{1}{2t}\left\|\bar x-x^{k-1}\right\|_2^2+\langle\bar x,A^{\mathrm{T}}\xi_1\rangle-y,\xi_1+\langle\mathcal{A}(\bar x)-b,\xi_2\rangle\\&=\frac{1}{2}\|\bar z\|_2^2-\langle\bar z,\xi_1\rangle+h(\bar x)+\frac{1}{2t}\left\|\bar x-x^{k-1}\right\|_2^2+\langle\bar x,A^{\mathrm{T}}\xi_1\rangle-\langle y,\xi_1\rangle+\langle\mathcal{A}(\bar x),\xi_2\rangle-\langle b,\xi_2\rangle\\&=\frac{1}{2}\|\bar z\|_2^2-\langle\bar z,\xi_1\rangle+h(\bar x)+\frac{1}{2t}\left\|\bar x-x^{k-1}\right\|_2^2+\langle\bar x,A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\rangle-\langle y,\xi_1\rangle-\langle b,\xi_2\rangle,\end{aligned}\tag{6}$$

其中 $\bar x,\bar z$ 是极小化拉格朗日函数(5)中的变量 $x,z$ 得到的最优解，$\bar x,\bar z$ 的具体形式将通过如下的引理给出。

于是得到问题(4)的对偶问题：

$$\max_{\xi_1,\xi_2}D_t(\xi_1,\xi_2)\tag{7}$$

**引理 1** 设 $(\xi_1,\xi_2)$ 是问题(4)的对偶问题(7)的一组最优解，则存在 $(\bar x,\bar z)$ 是 PPA 子问题的等价问题(4)的一组最优解，且满足如下等式：

$$\begin{cases}\bar x=\operatorname{prox}_{th}\left(x^{k-1}-t\left(A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right)\right),\\ \bar z=\xi_1.\end{cases}\tag{8}$$

**证明：** 下面将给出具体的求解过程，分别极小化(5)中的广义拉格朗日函数中的变量 $x$ 和 $z$，可以得到

$$
\begin{aligned}
\bar{x} &= \arg\min_x \mathcal{L}(x,z;\xi_1,\xi_2) \\
&= \arg\min_x h(x)+\frac{1}{2t}\left\|x-x^{k-1}\right\|_2^2+\left\langle x,A^{\mathrm{T}}\xi_1\right\rangle+\left\langle x,\mathcal{A}^*(\xi_2)\right\rangle \\
&= \arg\min_x \left\{h(x)+\frac{1}{2t}\left\|x+t\left(A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right)-x^{k-1}\right\|_2^2\right\} \\
&= \mathrm{prox}_{th}\left(x^{k-1}-t\left(A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right)\right),
\end{aligned} \tag{9}
$$

$$
\begin{aligned}
\bar{z} &= \arg\min_z \mathcal{L}(x,z;\xi_1,\xi_2) \\
&= \arg\min_z \left\{\frac{1}{2}\|z\|_2^2-\langle z,\xi_1\rangle\right\} \\
&= \xi_1,
\end{aligned} \tag{10}
$$

其中(9)式中倒数第二个等式的解是 $h(x)$ 在 $x^{k-1}-t\left(A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right)$ 处的邻近算子。

□

由引理 1 可得原始变量 $x,z$ 都可以由最优对偶变量 $\xi_1,\xi_2$ 表示，所以问题(4)的对偶问题(7)的目标函数 $D_t(\xi_1,\xi_2)$ 中关于变量 $x$ 的部分

$$
\begin{aligned}
& h(\bar{x})+\frac{1}{2t}\left\|\bar{x}+t\left(A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right)-x^{k-1}\right\|_2^2 \\
&= h\left(\mathrm{prox}_{th}\left(x^{k-1}-t\left(A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right)\right)\right) \\
&\quad +\frac{1}{2t}\left\|\mathrm{prox}_{th}\left(x^{k-1}-t\left(A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right)\right)+t\left(A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right)-x^{k-1}\right\|_2^2 \\
&= M_h^t\left(x^{k-1}-t\left(A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right)\right)
\end{aligned} \tag{11}
$$

显然成立，其中(11)式中的第二个等式是根据定义 2 得到的。因此，函数(6)式可以简化为：

$$
\begin{aligned}
D_t(\xi_1,\xi_2) := & M_h^t\left(x^{k-1}-t\left(A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right)\right)+\left\langle A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2),x^{k-1}\right\rangle \\
& -\frac{1}{2}\|\xi_1\|_2^2-\frac{t}{2}\left\|A^{\mathrm{T}}\xi_1+\mathcal{A}^*(\xi_2)\right\|_2^2-\langle y,\xi_1\rangle-\langle b,\xi_2\rangle.
\end{aligned}
$$

由引理 1 可知，PPA 子问题的等价问题(4)的解 $x,z$ 可以由对偶变量 $\xi_1,\xi_2$ 表示，因此接下来将给出求解问题(4)的邻近点算法的框架以及算法的收敛性分析。

## 3.3. 邻近点算法设计

下面算法将给出求解问题(4)的邻近点算法。

**算法 1.** 邻近点算法(PPA)求解问题(4)

| | |
|---|---|
| 1 | 初始化：$(x^0,z^0)\in R^n\times R^n,\epsilon>0,t^0>0$； |
| 2 | for $j=0,1,2,\cdots$ do |
| 3 | 计算：$(\xi_1^{j+1},\xi_2^{j+1})\approx\arg\max D_{t_j}(\xi_1,\xi_2)$； |
| 4 | 计算：$(x^{j+1},z^{j+1})$ 的更新由(8)式得到； |
| 5 | if $\max\left\{\left\|x^{j+1}-x^j\right\|_2^2,\left\|z^{j+1}-z^j\right\|_2^2\right\}\le\epsilon$ then |
| 6 | 输出：$x^{j+1},z^{j+1}$； |

续表

| | |
|---|---|
| 7 | else |
| 8 | 更新：$t^{j+1}$； |
| 9 | end |
| 10 | end |

对于算法 1 中 $x^{j+1}, z^{j+1}$ 的更新的具体形式已在引理 1 中详细给出。

### 3.4. 邻近点算法收敛性分析

本小节将给出算法 1 的收敛结果。在此之前，提出一些假设和命题。将与问题(4)相关的 Moreau 包络定义为

$$f_t\left(x^{k-1}\right) := \min_{x,z}\left\{\frac{1}{2}\|z\|_2^2 + \lambda\|x\|_1 + \frac{1}{2t}\left\|x - x^{k-1}\right\|_2^2 : Ax - y - z = 0, \mathcal{A}(x) = b\right\}. \quad (12)$$

**命题 1** ([15] Proposition 2.2)当对偶问题(7)的极大值在点 $\left(\xi_1^*, \xi_2^*\right)$ 处取得，则以下条件等价：

1) $\left(x^*, z^*\right)$ 是优化问题(4)的极小值点；

2) $f\left(x^*, z^*\right) = f_t\left(x^{k-1^*}\right)$，其中 $f\left(x^*, z^*\right)$ 和 $f_t\left(x^{k-1^*}\right)$ 分别是问题(4)和(12)式的极小值。

命题 1 的成立，说明问题(4)取到最优点 $\left(x^*, z^*\right)$ 和(12)式取到极值点 $x^{k-1^*}$ 是同时成立的，且此时原问题和对偶问题的目标函数值也相等。

在分析算法 1 的收敛性和收敛速度之前，首先给出如下假设。

**假设 1** 问题(4)的最优解集 $V^*$ 非空，$\sum_{j=0}^{\infty} t^j = +\infty$ 且存在参数 $\eta > 0$，$\epsilon > 0$ 和 $\omega \ge 1$ 使得

$$f\left(x^*, z^*\right) + \eta\left(d(x,z)\right)^{\omega} \le f(x,z),\ \forall (x,z) \in R^n \times R^n,$$

且

$$d(x,z) \le \epsilon,$$

其中

$$d(x,z) = \min_{\left(x^*, z^*\right)\in V^*}\left(\left\|x - x^*\right\|_2^2 + \left\|z - z^*\right\|_2^2\right)^{\frac{1}{2}}.$$

**定理 1** [16]设 $\left\{\left(x^j, z^j\right)\right\}$ 是由邻近点算法 1 产生的序列。若 $t^j$ 满足 $\sum_{j=0}^{\infty} t^j = +\infty$，则有

$$f\left(x^j, z^j\right) \downarrow f\left(x^*, z^*\right),$$

且若问题(4)的最优解集 $V^*$ 非空，则序列 $\left\{\left(x^j, z^j\right)\right\}$ 收敛到问题(4)的最优解 $\left(x^*, z^*\right)$。

假设 1 成立时，由定理 1 可知算法 1 产生的序列 $\left\{\left(x^j, z^j\right)\right\}$ 收敛到 $\left(x^*, z^*\right)$，如下的定理可说明算法 1 是线性收敛的。

**定理 2** [16]令假设 1 中的 $\omega = 2$ 且对于任意的 $j$，都有 $\left(x^j, z^j\right) \ne \left(x^*, z^*\right)$ 成立，于是有：

1) 若 $\lim_{j\to\infty} t^j = c$，$c \in (0, +\infty)$，则

$$\limsup_{j\to\infty} \frac{d\left(x^{j+1}, z^{j+1}\right)}{d\left(x^j, z^j\right)} \le \frac{1}{1+\eta c};$$

2) 若 $\lim_{j\to\infty} t^j = +\infty$，则

$$\limsup_{j\to\infty} \frac{d\left(x^{j+1}, z^{j+1}\right)}{d\left(x^j, z^j\right)} = 0.$$

## 4. PPA 子问题(3)的求解

已知问题(2)的求解实际上等价于求解一系列 PPA 子问题(3)，以及 PPA 算法具有线性收敛性。由于 PPA 子问题(3)的等价问题(4)和其对偶问题(7)是零对偶间隙的，且等价问题(4)的解 $x, z$ 可以由对偶变量 $\xi_1, \xi_2$ 表示。因此，设计高效的算法求解对偶问题(7)非常关键，本节将进一步分析并设计求解对偶问题(7)的高效算法。

为便于分析，首先将对偶问题(7)写成如下等价形式：

$$\begin{aligned}\min_{\xi_1,\xi_2} \psi(\xi_1,\xi_2) &:= -D_t(\xi_1,\xi_2) \\ &= -M^t_{\lambda\|\cdot\|_1}\left(x^{k-1} - t\left(A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2)\right)\right) + \frac{1}{2}\|\xi_1\|_2^2 - \left\langle A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2), x^{k-1}\right\rangle \\ &\quad + \frac{t}{2}\left\|A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2)\right\|_2^2 + \langle y, \xi_1\rangle + \langle b, \xi_2\rangle.\end{aligned} \tag{13}$$

注意到，此时问题(13)的目标函数 $\psi(\xi_1,\xi_2)$ 是凸的且连续可微。

### 4.1. 梯度

为方便表示，记 $(\xi_1,\xi_2) = u$，$(\Delta\xi_1, \Delta\xi_2) = \Delta u$。

问题(13)的目标函数 $\psi(u)$ 关于变量 $\xi_1, \xi_2$ 的梯度 $\nabla_{\xi_1}\psi(u), \nabla_{\xi_2}\psi(u)$ 具体形式为：

$$\begin{aligned}\nabla_{\xi_1}\psi(u) &= A\left[x^{k-1} - t\left(A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2)\right) - \mathrm{prox}_{th}\left(x^{k-1} - t\left(A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2)\right)\right)\right] \\ &\quad + \xi_1 - Ax^{k-1} + tAA^{\mathrm{T}}\xi_1 + tA\left(\mathcal{A}^*(\xi_2)\right) + y \\ &= -A\mathrm{prox}_{th}\left(x^{k-1} - t\left(A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2)\right)\right) + \xi_1 + y,\end{aligned} \tag{14}$$

$$\begin{aligned}\nabla_{\xi_2}\psi(u) &= \mathcal{A}\left(x^{k-1} - t\left(A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2)\right) - \mathrm{prox}_{th}\left(x^{k-1} - t\left(A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2)\right)\right)\right) \\ &\quad - \mathcal{A}\left(x^{k-1}\right) + t\mathcal{A}\left(A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2)\right) + b \\ &= -\mathcal{A}\mathrm{prox}_{th}\left(x^{k-1} - t\left(A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2)\right)\right) + b,\end{aligned} \tag{15}$$

因此问题(13)的目标函数 $\psi(u)$ 的梯度 $\nabla\psi(u)$ 为：

$$\nabla\psi(u) := \left[\nabla_{\xi_1}\psi(u); \nabla_{\xi_2}\psi(u)\right]. \tag{16}$$

### 4.2. 广义 Hessian 矩阵

根据已知梯度，现在进一步分析问题(13)的目标函数 $\psi(u)$ 的广义 Hessian 矩阵。

考虑到 $\nabla\psi(u)$ 是 Lipschitz 连续的，则在 Clarke 意义下 $\nabla\psi$ 在点 $(\xi_1,\xi_2)$ 处的广义 Jacobian 矩阵是 $B$-次微分 $\partial_B(\nabla\psi(u))$ 的凸包，即

$$\partial(\nabla\psi(u)) = co\left(\partial_B(\nabla\psi(u))\right).$$

记

$$\partial(\nabla\psi(u))[(\Delta u)] := \left(\partial\left(\nabla_{\xi_1}\psi(u)\right)[(\Delta u)], \partial\left(\nabla_{\xi_2}\psi(u)\right)[(\Delta u)]\right). \tag{17}$$

对(17)式的分析如下：

为了方便表示，记

$$H_{\xi_1,\xi_2} := A^{\mathrm{T}}\Delta\xi_1 + \mathcal{A}^*(\Delta\xi_2), \tag{18}$$

$$\hat{x}(\xi_1,\xi_2) := x^{k-1} - t\left(A^{\mathrm{T}}\xi_1 + \mathcal{A}^*(\xi_2)\right), \tag{19}$$

由于 $\|x\|_1$ 在 $x=0$ 处不可导，因此邻近算子的导数需要使用次梯度来表示。当 $x$ 不等于 0 时，邻近算子的次梯度为其导数。当 $x=0$ 时，邻近算子的次梯度可以是任何介于 $[-1,1]$ 的值。因此，此处有

$$\left[\partial\mathrm{prox}_{t\lambda\|\cdot\|_1}\left(\hat{x}(\xi_1,\xi_2)\right)\left[H_{\xi_1,\xi_2}\right]\right]_i = \begin{cases} \left[H_{\xi_1,\xi_2}\right]_i, & \left[\hat{x}(\xi_1,\xi_2)\right]_i > t\lambda, \\ a\left[H_{\xi_1,\xi_2}\right]_i \text{ 且 } a\in[0,1], & \left[\hat{x}(\xi_1,\xi_2)\right]_i = t\lambda, \\ 0, & \left[\hat{x}(\xi_1,\xi_2)\right]_i < t\lambda. \end{cases}$$

其中 $\left[H_{\xi_1,\xi_2}\right]_i$ 和 $\left[\hat{x}(\xi_1,\xi_2)\right]_i$ 中的下标 $i=1,\cdots,n$ 表示对应向量的第 $i$ 个元素。于是有：

$$\partial\left(\nabla\psi(u)\right)\left[(\Delta u)\right] = \left[tA\partial\mathrm{prox}_{t\lambda\|\cdot\|_1}\left(\hat{x}(\xi_1,\xi_2)\right)\left[H_{\xi_1,\xi_2}\right] + \Delta\xi_1, t\mathcal{A}\partial\mathrm{prox}_{t\lambda\|\cdot\|_1}\left(\hat{x}(\xi_1,\xi_2)\right)\left[H_{\xi_1,\xi_2}\right]\right].$$

### 4.3. 最优性条件

已知问题(13)是一个凸的无约束优化问题，现通过如下的定理给出该问题的一阶最优性条件。

**定理 3** 假设最优点对 $(\xi_1^*,\xi_2^*) := u^*$ 是问题(13)的全局最优解，则满足如下条件：

$$\begin{cases} \nabla_{\xi_1}\psi(u^*) = 0, \\ \nabla_{\xi_2}\psi(u^*) = 0. \end{cases} \tag{20}$$

注意到(20)式可以写成

$$\phi(u^*) := \left(\nabla_{\xi_1}\psi(u^*), \nabla_{\xi_2}\psi(u^*)\right) = 0, \tag{21}$$

在点 $(\xi_1^k,\xi_2^k) := u^k$ 处，定义问题(13)的 KKT 误差为

$$\eta_{kkt}(u^k) := \left\|\phi(u^k)\right\|_\infty. \tag{22}$$

因此，问题(13)一阶最优性条件成立等价于

$$\eta_{kkt}(u^k) = 0,$$

所以可以用 KKT 误差 $\eta^k := \eta_{kkt}(u^k)$ 来衡量当前迭代点的最优性。

接下来给出问题(13)的二阶充分性条件。

**命题 2** [17]对于函数 $\nabla\psi$ 在点 $u$ 处的广义 Jacobian 矩阵 $\hat{H}\in\partial_B\left(\nabla\psi(u^*)\right)$。

1) 若 $u^*$ 为问题(13)的局部极小值点，则对任意非零 $s$ 有

$$s^{\mathrm{T}}\hat{H}s \ge 0;$$

2) 令 $u^*$ 满足 KKT 条件(20)。若对任意非零 $s$ 有二阶充分性条件

$$s^{\mathrm{T}}\hat{H}s \ge 0$$

成立，则 $u^*$ 为问题(13)的严格局部极小值点。

### 4.4. 半光滑牛顿算法

由于半光滑牛顿算法的局部快速收敛性，极大提高了子问题的求解效率，进一步可以提高算法 1 的整体效率。在本小节中，考虑采用半光滑牛顿算法求解子问题(13)。

由于 $l_1$ 范数的邻近算子是局部 Lipschitz 连续的，因此可以定义 $\mathrm{prox}_{t\lambda\|\cdot\|_1}$ 的 Clarke 广义 Jacobian，记为 $\partial\mathrm{prox}_{t\lambda\|\cdot\|_1}$。进一步定义目标函数 $\psi$ 的广义 Hessian，即 $\nabla\psi$ 的 Clarke 广义 Jacobian，记为 $\partial^2\psi$。由于 $\partial^2\psi$ 没有显式表达式，因此定义如下的替代映射：

$$\hat{\partial}^2\psi(\xi_1,\xi_2):=\left[tA\partial\mathrm{prox}_{t\lambda\|\cdot\|_1}\left(\hat{x}(\xi_1,\xi_2)\right)+\Delta\xi_1,t\mathcal{A}\partial\mathrm{prox}_{t\lambda\|\cdot\|_1}\left(\hat{x}(\xi_1,\xi_2)\right)\right].$$

对于给定的 $(\xi_1,\xi_2)$，有

$$\partial^2\psi(\xi_1,\xi_2)(d)=\hat{\partial}^2\psi(\xi_1,\xi_2)(d),$$

所以，$\hat{\partial}^2\psi$ 可以被认为是 $\partial^2\psi$ 的一个较好的替代映射。对于给定的 $(\xi_1,\xi_2)$，令

$$\hat{H}=\begin{pmatrix} tAUA^{\mathrm{T}}+\Delta\xi_1 \\ t\mathcal{A}U\mathcal{A}^* \end{pmatrix}$$

其中 $U\in\partial\mathrm{prox}_{t\lambda\|\cdot\|_1}\left(\hat{x}(\xi_1,\xi_2)\right)$，则 $\hat{H}\in\hat{\partial}^2\psi$。

具体来说，广义牛顿方向 $\hat{d}:=\left(\hat{d}_{\xi_1},\hat{d}_{\xi_2}\right)$ 可以通过求解无约束的二次规划

$$\min_{\hat{d}}\left\langle\nabla\psi\left(u^k\right),\hat{d}\right\rangle+\frac{1}{2}\left\langle d,\hat{H}^k\hat{d}\right\rangle \tag{23}$$

得到，其中 $u^k=\left(\xi_1^k,\xi_2^k\right)$，$\hat{H}^k$ 为问题(13)的目标函数 $\psi$ 在点 $u^k$ 处的广义 Hessian 矩阵。

根据目标函数的性质以及算法的理论，将问题转化为求解如下牛顿方程：

$$\hat{H}^k d=-\nabla\psi\left(u^k\right). \tag{24}$$

由于 $\hat{\partial}^2\psi\left(u^k\right)$ 中的映射是对称和半正定的，但它们仍然可能是奇异的。对此，设置

$$H^k=\hat{H}^k+\nu^k id, \tag{25}$$

其中 $\hat{H}^k\in\hat{\partial}^2\psi\left(u^k\right)$，$id:R^m\times R^s\to R^m\times R^s$ 为恒等映射，$\nu^k$ 定义为 $\nu^k:=\min\left(\bar{\delta},\delta\eta^k\right)$，$\bar{\delta}$ 和 $\delta$ 为两个正参数，并且 $\eta^k:=\eta_{kkt}\left(u^k\right)$ 为(21)式中定义的在点 $u^k$ 处的 KKT 误差。显然，当 $\eta^k\neq 0$ 时，$H^k$ 为正定的。

考虑使用半光滑牛顿算法求解子问题(13)，生成迭代为

$$u^{k+1}=u^k-\alpha^k\left(H^k\right)^{-1}\left[\left(\nabla\psi\left(u^k\right)\right)\right], \tag{26}$$

其中 $\alpha^k>0$ 为第 $k$ 步合适的步长，$\left(H^k\right)^{-1}$ 为广义 Hessian 矩阵 $H^k$ 的逆。

记

$$\left(\Delta\xi_{1N}^k,\Delta\xi_{2N}^k\right):=-\left(H^k\right)^{-1}\left[\left(\nabla_{\xi_1}\psi\left(u^k\right),\nabla_{\xi_2}\psi\left(u^k\right)\right)\right],$$

则 $\left(\Delta\xi_{1N}^k,\Delta\xi_{2N}^k\right)$ 为在 $u^k$ 处得到的正则化广义牛顿方向，它可以通过求解广义牛顿方程：

$$H^k\left[\left(\Delta\xi_{1N}^k,\Delta\xi_{2N}^k\right)\right]=-\left[\left(\nabla_{\xi_1}\psi\left(u^k\right),\nabla_{\xi_2}\psi\left(u^k\right)\right)\right] \tag{27}$$

得到。

对此，考虑到计算成本，通常采用共轭梯度(CG)算法来求解广义牛顿方程(27)的近似解。现在给出半光滑牛顿算法求解系统(24)的基本框架。

**算法 2.** 半光滑牛顿算法求解问题(24)

Input： $\varepsilon > 0, \sigma \in (0,1), \tau \in (0,1)$；

Given： $(\xi_1^0, \xi_2^0) \in R^m \times R^s$；

**Compute** $d_{\xi_1}^0$ **and** $d_{\xi_2}^0$ from (23)

**for** $k = 0,1,\cdots$ **do**

**if** $\max\{\eta^k, gap^k\} \le \varepsilon$ **then**

return $\xi_1^k$ and $\xi_2^k$；

**end**；

**Compute** $d_{\xi_1}^k$ and $d_{\xi_2}^k$ by solving (23)；

Set $\alpha = 1$；

**while** $\psi(\xi_1^k + \alpha^k d_{\xi_1}^k, \xi_2^k + \alpha^k d_{\xi_2}^k) - \psi(\xi_1^k, \xi_2^k) > \alpha\sigma\Delta^k$ **do**

Set $\alpha = \tau\alpha$；

**end**

Set $\alpha^k = \alpha$ and $\xi_1^{k+1} = \xi_1^k + \alpha^k d_{\xi_1}^k$ and $\xi_2^{k+1} = \xi_2^k + \alpha^k d_{\xi_2}^k$；

**end**

在下文中，交代了算法的终止条件。与[18]类似，考虑使用 *gap* 来表示原问题和对偶问题的间隙，以便于测量原目标与对偶目标之间的差距，即

$$gap = \frac{|probj - dobj|}{1 + |probj| + |dobj|},$$

其中 *probj* 和 *dobj* 分别表示原问题目标函数值和对偶问题目标函数值。

当 $\max\{\eta^k, gap^k\} \le \varepsilon$ 时，算法终止，其中 $\varepsilon > 0$ 为给定误差，$\eta^k$ 为(22)式中定义的 KKT 误差，*gap* 为点 $u^k$ 处的间隙。因此，在算法 2 中给出了求解问题(23)的算法。

### 4.5. 半光滑牛顿算法收敛性分

本小节将分析半光滑牛顿算法的收敛性，为方便表示，记 $u := (\xi_1, \xi_2)$。

**定义 4** 假设 $\nabla\psi$ 是局部 Lipschitz 连续的，则其广义 Jacobian 存在。取在 $u^k$ 点任意的广义 Jacobian 矩阵 $\hat{H}_k \in \partial(\nabla\psi(u^k))$，若 $\hat{H}_k$ 可逆，其基本的迭代格式为

$$u^{k+1} = u^k - \hat{H}_k^{-1}\nabla\psi(u^k). \tag{28}$$

**假设 2** 映射 $\nabla\psi$ 在最优点 $u^*$ 处是半光滑和所有 Jacobian 矩阵是非奇异的。

**引理 2** 如果假设 2 成立，则存在常数 $c > 0, \kappa > 0$ 和一个小邻域 $N(u^*, \epsilon_0)$ 使得对于任意的 $u \in N(u^*, \epsilon_0)$ 和 $\hat{H}_k \in \partial(\nabla\psi(u^k))$，下面结论成立：

1) $u^*$ 是一个孤立解；
2) $\hat{H}_k$ 是非奇异的并且 $\|\hat{H}_k^{-1}\| \le c$；
3) 局部误差届条件对于 $\nabla\psi(u)$ 在邻域 $N(u^*, \epsilon_0)$ 上成立，也就是说 $\|u - u^*\| \le \kappa\|\nabla\psi(u)\|$。

如下定理给出了半光滑牛顿算法的局部二次收敛性。

**定理 4** 设 $\nabla\psi(\xi_1, \xi_2)$ 具有半光滑性并且 $u^*$ 是优化问题(13)的最优解。那么迭代(28)是良定义的，且存

在一个小邻域 $N\left(u^*,\epsilon\right)$，使得对于任意的 $k$ 有 $u^k \in N\left(u^*,\epsilon\right)$，迭代(28)是超线性收敛的。如果 $\nabla\psi\left(u\right)$ 是强半光滑的，迭代(28)是二次收敛的。

**证明：**根据引理 2，迭代(28)式是良定义的。可以推出：

$$
\begin{aligned}
\left\|u^{k+1}-u^k\right\| &= \left\|u^k-\hat{H}_k^{-1}\nabla\psi\left(u^k\right)-u^*\right\| \\
&= \left\|\hat{H}_k^{-1}\right\|\left\|\nabla\psi\left(u^k\right)-\nabla\psi\left(u^*\right)-\hat{H}_k\left(u^k-u^*\right)\right\| \\
&= O\left(\left\|u^k-u^*\right\|^2\right),
\end{aligned}
$$

其中最后一个等式来源于强半光滑性。

□

## 5. 数值实验

本节将通过仿真数值实验来测试本文使用的邻近点对偶半光滑牛顿算法(记为 PPA_DSN)算法求解含等式约束广义 Lasso 问题的实际效果。算法采用 MatlabR2017a 编程实现，且所有实验均在 AMD Ryzen 7 6800HS Creator Edition 3.20 GHz，16.0 GB 内存的个人笔记本电脑上运行。同时，考虑选择邻近梯度算法(记为 PPA_DG)和交替方向乘子法算法(记为 ADMM_GLP)作为对比算法。算法 PPA_DSN 所需参数分别为：设计矩阵 $A$ 的规模：$m\times n$；观测向量 $y$ 的规模：$m$ 维；最大迭代时间 600 秒；最大误差 $\epsilon=1\times10^{-6}$。

对于所有算法，若下面任一条件满足，则终止程序：

1) 算法残差 $\eta_{kkt}\left(u^k\right)\le\epsilon$。

2) 达到最大迭代数 maxit = 200。

为进行实验观测算法 PPA_DSN 求解含等式约束广义 Lasso 问题的效率，首先设置含等式约束广义 Lasso 问题的样本数量为 $m$，特征数量为 $n$，设计矩阵 $A$ 的每个分量 $A_{ij}$ 均随机生成；设向量 $y=Ax^*+\varepsilon$ （$\varepsilon$ 表示随机扰动向量)，正则化参数 $\lambda=0.01$。

| 实验 | 实验参数 | | | | PPA_DSN | | | | | PPA_DG | | | | | ADMM_GLP | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 序号 | n | p | m | lamda | Res | iter | iterf | iterg | Time(s) | Res | iter | iterf | iterg | Time(s) | Res | iter | iterf | iterg | Time(s) |
| 1 | 1000 | 1 | 100 | 0.001 | 3.62E-07 | 5 | 38 | 24 | 0.015331433 | 2.82E-07 | 4 | 47 | 32 | 0.007541567 | 8.90E-07 | 42 | 42 | 42 | 1.6166703 |
| 2 | 1000 | 1 | 100 | 0.01 | 5.90E-07 | 5 | 48 | 34 | 0.023176033 | 7.89E-07 | 5 | 46 | 32 | 0.007654367 | 8.59E-07 | 66 | 66 | 66 | 2.587935033 |
| 3 | 1000 | 1 | 100 | 0.1 | 7.91E-07 | 8 | 108 | 56 | 0.037917567 | 6.22E-07 | 8 | 1190 | 142 | 0.095826967 | 9.69E-07 | 112 | 112 | 112 | 4.447577933 |
| 4 | 1000 | 1 | 200 | 0.001 | 1.56E-07 | 5 | 43 | 29 | 0.056898333 | 3.59E-07 | 5 | 56 | 41 | 0.031497733 | 8.90E-07 | 50 | 50 | 50 | 2.049566067 |
| 5 | 1000 | 1 | 200 | 0.01 | 6.59E-07 | 5 | 52 | 37 | 0.0844792 | 2.94E-07 | 5 | 62 | 47 | 0.036355433 | 9.07E-07 | 73 | 73 | 73 | 3.167149933 |
| 6 | 1000 | 1 | 200 | 0.1 | 5.50E-07 | 11 | 129 | 81 | 0.187601067 | 6.30E-07 | 13 | 5949 | 564 | 0.888628833 | 9.03E-07 | 124 | 124 | 124 | 5.680443167 |
| 7 | 1000 | 1 | 500 | 0.001 | 4.55E-07 | 5 | 56 | 35 | 0.233744933 | 2.54E-07 | 5 | 95 | 74 | 0.103985133 | 9.29E-07 | 42 | 42 | 42 | 2.027330067 |
| 8 | 1000 | 1 | 500 | 0.01 | 4.60E-07 | 6 | 69 | 48 | 0.388544667 | 3.05E-07 | 6 | 103 | 80 | 0.1103258 | 9.16E-07 | 75 | 75 | 75 | 3.425746133 |
| 9 | 1000 | 1 | 500 | 0.1 | 7.54E-07 | 12 | 95 | 65 | 0.558835267 | 7.80E-07 | 26 | 19721 | 1781 | 4.753055167 | 9.54E-07 | 112 | 112 | 112 | 5.125327733 |
| 10 | 1500 | 1 | 100 | 0.001 | 4.58E-07 | 4 | 37 | 21 | 0.039449267 | 4.27E-07 | 4 | 42 | 26 | 0.019645233 | 7.73E-07 | 45 | 45 | 45 | 5.767912867 |
| 11 | 1500 | 1 | 100 | 0.01 | 6.53E-07 | 5 | 45 | 29 | 0.0592851 | 8.17E-07 | 5 | 443 | 68 | 0.077204367 | 9.81E-07 | 58 | 58 | 58 | 7.220847667 |
| 12 | 1500 | 1 | 100 | 0.1 | 6.11E-07 | 8 | 127 | 58 | 0.1214738 | 3.91E-07 | 10 | 3613 | 353 | 0.5308892 | 9.53E-07 | 135 | 135 | 135 | 17.25143803 |
| 13 | 1500 | 1 | 200 | 0.001 | 2.92E-07 | 4 | 40 | 24 | 0.064613833 | 5.47E-07 | 4 | 418 | 66 | 0.1061745 | 9.33E-07 | 43 | 43 | 43 | 5.343830967 |
| 14 | 1500 | 1 | 200 | 0.01 | 3.14E-07 | 5 | 46 | 28 | 0.0749381 | 6.09E-07 | 5 | 58 | 40 | 0.042827233 | 8.37E-07 | 71 | 71 | 71 | 8.879709 |
| 15 | 1500 | 1 | 200 | 0.1 | 5.95E-07 | 11 | 113 | 60 | 0.1848666 | 5.14E-07 | 11 | 1613 | 191 | 0.3507282 | 8.99E-07 | 122 | 122 | 122 | 15.42100337 |
| 16 | 1500 | 1 | 500 | 0.001 | 6.72E-08 | 5 | 60 | 43 | 0.276458 | 5.27E-07 | 4 | 70 | 52 | 0.105014867 | 9.33E-07 | 45 | 45 | 45 | 5.9742503 |
| 17 | 1500 | 1 | 500 | 0.01 | 5.89E-07 | 6 | 68 | 50 | 0.4261725 | 6.32E-07 | 5 | 78 | 58 | 0.130361367 | 9.33E-07 | 73 | 73 | 73 | 10.4621667 |
| 18 | 1500 | 1 | 500 | 0.1 | 7.94E-07 | 16 | 189 | 116 | 0.9164076 | 8.19E-07 | 22 | 11628 | 1086 | 5.419929667 | 9.38E-07 | 135 | 135 | 135 | 21.16663143 |
| 19 | 2000 | 1 | 100 | 0.001 | 6.23E-08 | 4 | 34 | 19 | 0.047016567 | 2.36E-07 | 4 | 385 | 58 | 0.0865613 | 9.34E-07 | 35 | 35 | 35 | 10.13202087 |
| 20 | 2000 | 1 | 100 | 0.01 | 4.43E-07 | 5 | 46 | 30 | 0.071386 | 6.58E-07 | 5 | 44 | 28 | 0.027828467 | 9.00E-07 | 62 | 62 | 62 | 17.1633852 |
| 21 | 2000 | 1 | 100 | 0.1 | 5.90E-07 | 9 | 119 | 53 | 0.155295633 | 6.55E-07 | 9 | 2806 | 278 | 0.547644167 | 9.69E-07 | 185 | 185 | 185 | 56.01451583 |
| 22 | 2000 | 1 | 200 | 0.001 | 7.80E-07 | 4 | 51 | 35 | 0.137580967 | 3.93E-07 | 4 | 52 | 33 | 0.046277133 | 8.98E-07 | 50 | 50 | 50 | 13.5363313 |
| 23 | 2000 | 1 | 200 | 0.01 | 4.39E-07 | 5 | 44 | 27 | 0.097582833 | 3.31E-07 | 5 | 50 | 33 | 0.047583833 | 8.21E-07 | 59 | 59 | 59 | 16.1644428 |
| 24 | 2000 | 1 | 200 | 0.1 | 7.69E-07 | 11 | 126 | 61 | 0.238224633 | 6.67E-07 | 11 | 2009 | 223 | 0.531780767 | 9.57E-07 | 135 | 135 | 135 | 39.57247717 |
| 25 | 2000 | 1 | 500 | 0.001 | 4.59E-07 | 4 | 69 | 50 | 0.392750333 | 4.19E-07 | 4 | 68 | 49 | 0.125349367 | 9.00E-07 | 46 | 46 | 46 | 16.38394947 |
| 26 | 2000 | 1 | 500 | 0.01 | 4.84E-07 | 6 | 52 | 34 | 0.343057367 | 3.22E-07 | 6 | 71 | 52 | 0.1663328 | 9.46E-07 | 71 | 71 | 71 | 26.2426295 |
| 27 | 2000 | 1 | 500 | 0.1 | 8.18E-07 | 12 | 129 | 75 | 0.672192767 | 5.94E-07 | 13 | 3188 | 338 | 1.8347003 | 8.95E-07 | 138 | 138 | 138 | 39.91801887 |

**Figure 1.** In correspondence with the experimental data of three algorithms
**图 1.** 三种算法对比的实验数据

为了验证 PPA_DSN 算法的有效性和稳定性，我们进行了不同参数设置下的实验。下文提供 PPA_DSN 算法、邻近梯度算法(记为 PPA_DG)和交替方向乘子算法(记为 ADMM_GLP)的数值结果。我们针对零和约束 Lasso 问题(1)，取变量维数 $n=[1000,1500,2000]$，$m=[100,200,500]$，正则化参数 $\lambda=[0.001,0.01,0.1]$，最大迭代时间 600 秒，最大误差。在这些设置下，共进行了 27 组实验。考虑到实验数据均为随机生成的，每组实验重复了 3 次，并取数值结果的平均值。具体结果详见图 1，其中 Res、Iter、Iterf、Iterd 和 Time 分别表示残差 $\eta_{kkt}(u^k)$、迭代步数、目标函数计算次数、搜索方向计算次数和 CPU 运行时间(以秒为单位)。从图 1 可以看出，算法 PPA_DSN 成功求解了所有测试问题，并且在大多数情况下，CPU 耗时不超过 1 秒。

## 6. 总结

本文主要对一类含线性等式约束的广义 Lasso 问题进行算法研究。由于大多实际问题的求解过程中会涉及对样本数据的限制，如零和约束 Lasso 问题，因此考虑求解一类含线性等式约束的广义 Lasso 问题。受众多研究者先前对 Lasso 模型的算法研究的启发，本文考虑将邻近点半光滑牛顿算法推广到用于求解含线性等式约束 Lasso 问题，并通过数值实验与邻近梯度算法以及交替方向乘子法进行对比，证实了该算法求解该类含线性等式约束 Lasso 问题具有良好性能。

## 参考文献


[1] Cristofari, A. (2023) A Decomposition Method for Lasso Problems with Zero-Sum Constraint. *European Journal of Operational Research*, **306**, 358-369. https://doi.org/10.1016/j.ejor.2022.09.030

[2] Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V. and Egozcue, J.J. (2017) Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology*, **8**, Article 2224. https://doi.org/10.3389/fmicb.2017.02224

[3] Shi, P., Zhang, A. and Li, H. (2016) Regression Analysis for Microbiome Compositional Data. *The Annals of Applied Statistics*, **10**, 1019-1040. https://doi.org/10.1214/16-aoas928

[4] Lin, W., Shi, P., Feng, R. and Li, H. (2014) Variable Selection in Regression with Compositional Covariates. *Biometrika*, **101**, 785-797. https://doi.org/10.1093/biomet/asu031

[5] Altenbuchinger, M., Rehberg, T., Zacharias, H.U., Stämmler, F., Dettmer, K., Weber, D., *et al*. (2016) Reference Point Insensitive Molecular Data Analysis. *Bioinformatics*, **33**, 219-226. https://doi.org/10.1093/bioinformatics/btw598

[6] 郦旭东. 复合凸优化的快速邻近点算法[J]. 计算数学, 2020, 42(4): 385-404.

[7] Gaines, B.R., Kim, J. and Zhou, H. (2018) Algorithms for Fitting the Constrained Lasso. *Journal of Computational and Graphical Statistics*, **27**, 861-871. https://doi.org/10.1080/10618600.2018.1473777

[8] Deng, Z., Yue, M. and So, A.M. (2020). An Efficient Augmented Lagrangian-Based Method for Linear Equality-Constrained Lasso. *ICASSP* 2020—2020 *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), Barcelona, 4-8 May 2020, 5760-5764. https://doi.org/10.1109/icassp40776.2020.9053722

[9] Jeon, J., Kim, Y., Won, S. and Choi, H. (2020) Primal Path Algorithm for Compositional Data Analysis. *Computational Statistics & Data Analysis*, **148**, Article ID: 106958. https://doi.org/10.1016/j.csda.2020.106958

[10] Tibshirani, R.J. and Taylor, J. (2011) The Solution Path of the Generalized Lasso. *The Annals of Statistics*, **39**, 1335-1371. https://doi.org/10.1214/11-aos878

[11] Beck, A. (2017). First-Order Methods in Optimization. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611974997

[12] Rockafellar, R.T. and Wets, R.J.B. (2009) Variational Analysis. Springer Science & Business Media.

[13] Facchinei, F. and Pang, J.S. (2003) Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer. https://doi.org/10.1007/b97543

[14] Zhang, Y., Zhang, N., Sun, D. and Toh, K. (2018) 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

[15] Li, Q. (2012) Conjugate Gradient Type Methods for the Nondifferentiable Convex Minimization. *Optimization Letters*, **7**, 533-545. https://doi.org/10.1007/s11590-011-0437-5

[16] Bertsekas, D. (2015) Convex Optimization Algorithms. Athena Scientific.

[17] Shen, C., Xue, W., Zhang, L. and Wang, B. (2020) An Active-Set Proximal-Newton Algorithm for $\ell_1$ Regularized Op-

timization Problems with Box Constraints. *Journal of Scientific Computing*, **85**, Article No. 57. https://doi.org/10.1007/s10915-020-01364-0

[18] Cui, Y., Leng, C. and Sun, D. (2016) Sparse Estimation of High-Dimensional Correlation Matrices. *Computational Statistics & Data Analysis*, **93**, 390-403. https://doi.org/10.1016/j.csda.2014.10.001