

# A Projection Algorithm for Solving Optimization Problems with Sparsity Constraints and Closed Convex Set Constraints

Jun Sun, Biao Qu

School of Management, Qufu Normal University, Rizhao Shandong  
Email: sunjun901203@163.com

Received: Jul. 25<sup>th</sup>, 2017; accepted: Aug. 8<sup>th</sup>, 2017; published: Aug. 16<sup>th</sup>, 2017

---

## Abstract

In this paper, we mainly consider the optimization problem with sparsity constraints and closed convex set constraints. We design a gradient projection algorithm with Armijo step size rule, and prove that the sequence of the iteration generated by this algorithm can converge to an  $\alpha$ -stationary point of the problem. Finally, a numerical example is given to demonstrate the effectiveness of the algorithm.

## Keywords

Sparsity-Constrained, Closed Convex Set Constraints, Convergence,  $\alpha$ -Stationary Point

---

## 求解带有稀疏约束和闭凸集约束的优化问题的投影算法

孙 军, 屈 彪

曲阜师范大学管理学院, 山东 日照  
Email: sunjun901203@163.com

收稿日期: 2017年7月25日; 录用日期: 2017年8月8日; 发布日期: 2017年8月16日

---

## 摘 要

本文中, 我们考虑了带有稀疏约束和闭凸集约束的优化问题的求解。设计了一种带有Armijo步长规则的

梯度投影算法, 证明了此算法产生的迭代点列可以收敛到问题的一个 $\alpha$ -稳定点上。最后给出了数值例子验证了算法的有效性。

## 关键词

稀疏约束, 闭凸集约束, 收敛,  $\alpha$ -稳定点

Copyright © 2017 by authors 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

我们主要研究带有稀疏约束和闭凸集约束的优化问题。随着实际问题的发展需要, 稀疏优化问题的约束条件进一步加强, 除了对于自变量的稀疏约束外, 还有一些其他的约束条件, 如在像素强度, 频率计数等中要求 $x \geq 0$ 。因此, 对于稀疏约束和闭凸集约束最优化问题的研究还是有很大的理论意义和实际价值。

本文所研究的问题形式如下:

$$\min f(x) \quad \text{s.t. } x \in C_S \cap B, \quad (1)$$

这里 $B$ 是一个闭凸的对称集合。

用 $C_S$ 表示所有最多有 $s$ 个非零元素的向量全体:

$$C_S = \{x \in R^n : \|x\|_0 \leq s\},$$

这里,  $s \in \{1, 2, \dots, n\}$ ,  $\|\cdot\|_0$ 表示 $l_0$ 范数, 称集合 $C_S$ 为稀疏集合。

对于稀疏约束优化问题, 已经有了一些研究, Negahban [1]在凸优化问题中提出了一种严格强凸性(RSC)性质并且证明了这是问题唯一解存在的充分条件。Agarwal [2]改进了RSC并且定义了严格强光滑性(RSS)来保证一阶方法的线性收敛性。后来, 陆续有些学者[3] [4] [5] [6]对RSC和RSS性质进行了改进来保证唯一解的存在性。Bahmani [7]对于二阶连续可微的目标函数提出了严格稳定Hessian性(SRH)性质, 对于非光滑的目标函数提出了严格稳定线性(SRL)性质, 并且证明了这些是稀疏约束优化问题可以得到唯一解的充分条件。以上的所有这些性质都可以看作Candes和Tao [8] [9]在压缩传感中提出的严格等距性(RIP)的扩展或者松弛。这里RIP保证了压缩传感中带有线性目标函数的 $l_0$ 优化问题有唯一解。2008年, Bunea [10]中提出了凸松弛算法, 在一些适当的假设下, 能够通过凸松弛的方法来解决稀疏约束优化问题, 准确的来说, 就是用 $l_1$ 范数来近似 $l_0$ 范数。

Amir Beck和Nadav Hallak [11]在2014年提出了一些稀疏约束和闭凸集约束优化问题的一些最优性条件, 并给出了收敛到不同稳定点的一些算法。Pan和Xiu [12]中对于 $B = R_+^n$ 的情况下, 提出了支撑投影算法(GSP)。

本文具体组织如下, 在第2部分先回顾了一些预备知识, 给出了稀疏约束和闭凸集约束优化问题解存在的必要条件, 然后在第3部分设计了一种投影算法, 证明了算法可以收敛到问题(1)的一个 $\alpha$ -稳定点。最后, 第4部分给出了数值例子验证了我们算法的可行性和有效性。

## 2. 预备知识

**假设 1**  $f(x)$  是有下界的。即存在  $\gamma$ , 使得对于所有的  $x$ , 都有  $f(x) \geq \gamma$ 。

**假设 2**  $f(x)$  是二阶连续可微的且其梯度是 Lipschitz 连续的, 即:

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

对于一个稀疏向量  $x$ , 如果  $\|x\|_0 = s$ , 则称向量  $x$  有完全支撑集。对于一个指标集合  $T$ , 对于  $y \in C_S \cap B$ ,  $I_1(y) \subset T$  且  $|T| = s$ , 则称  $T$  是  $y$  的一个超级支撑集。

首先, 我们将  $\{1, 2, \dots, n\}$  所有的排列记为  $\Sigma_n$ 。对于一个给定的向量  $x \in \mathbb{R}^n$  和一个排列  $\sigma \in \Sigma_n$ , 向量  $x^\sigma$  定义为:

$$(x^\sigma)_i = x_{\sigma(i)}.$$

即向量  $x^\sigma$  是  $x$  根据  $\sigma$  的一个重新排列。如  $x = (4, 5, 6)^\top$ ,  $\sigma$  是一个排列:  $\sigma(1) = 3, \sigma(2) = 2, \sigma(3) = 1$ , 那么  $x^\sigma = (6, 5, 4)^\top$ 。

对于给定的指标集  $T \subset \{1, 2, \dots, n\}$ , 称

$$B_T = \{x \in \mathbb{R}^n : U_T x \in B\}.$$

为  $B$  在  $T$  下的限制。这里  $U_T$  表示单位矩阵  $I_n$  在指标集  $T$  下所对应的列所组成的子矩阵。定义  $x_R$  是向量  $x$  在标集  $R$  下的一个子向量。

举个例子:  $B = \{(x_1, x_2, x_3, x_4)^\top : x_1 + 2x_2 + 3x_3 + 4x_4 = 1\}$ , 则  $B_{\{1,2\}} = \{(x_1, x_2)^\top : x_1 + 2x_2 = 1\}$ 。

**定义 1 [11]** (排序排列) 对于向量  $x \in \mathbb{R}^n$ , 如果一个排列使得向量  $x$  的元素为一个非增序列, 那么称这个排列为排序排列。向量  $x$  的所有排序排列记为  $\bar{\Sigma}(x)$ 。

**定义 2 [11]** (类型 1 的对称集合) 设集合  $D \subseteq \mathbb{R}^n$ , 如果对于任意的向量  $x \in D$  和  $\sigma \in \Sigma_n$ , 我们有  $x^\sigma \in D$ , 则称  $D$  为类型 1 的对称集合。

**定义 3 [11]** (类型 2 的对称集合) 设  $D$  为类型 1 的对称集合, 如果对于  $x \in D$ ,  $\sigma \in \Sigma_n$  和  $y \in \{-1, 1\}^n$ , 向量  $x \circ y = (x_i y_i)^\top$  仍属于集合  $D$ , 那么称  $D$  为类型 2 对称的集合。

**定义 4 [11]** (非负集合) 对于  $D \subseteq \mathbb{R}^n$ , 如果对于任意的  $x \in D$ , 都有  $x \geq 0$ , 则称集合  $D$  为非负集合。

**定义 5 [11]** 定义下面的函数  $p: \mathbb{R}^n \rightarrow \mathbb{R}^n$ :

$$p(x) = \begin{cases} x & B \text{ 是非负类型 1 的对称集合} \\ |x| & B \text{ 是类型 2 的对称集合} \end{cases}$$

则称此函数为对称函数。

给定一个闭凸集  $B$  和一个向量  $x \in \mathbb{R}^n$ , 记  $x$  在  $C_S \cap B$  上的投影为:

$$P_{C_S \cap B}(x) = \arg \min \{\|z - x\|^2 : z \in C_S \cap B\}.$$

**引理 1 [11]** 设  $x \in \mathbb{R}^n$ , 假设  $y \in P_{C_S \cap B}(x)$ , 则对于  $y$  的任意超级支撑集  $T$ , 都有:

$$y_T \in P_{B_T}(x_T).$$

**定义 6 [11]** (基本可行点) 如果一个向量  $x \in C_S \cap B$ , 对于  $x$  的任意一个超级支撑集  $S$  和  $\alpha > 0$ , 都有:

$$x_S = P_{B_S}(x_S - \alpha \nabla_S f(x)),$$

则称向量  $x$  为问题(1)的一个基本可行点。

**定理 1 [11]** 如果  $x^*$  是问题(1)的一个最优解, 那么  $x^*$  是一个基本可行点。

**定义 7 [11]** ( $\alpha$ -稳定点) 取  $\alpha > 0$ , 对于  $x \in C_S \cap B$ , 如果有

$$x \in P_{C_S \cap B}(x - \alpha \nabla f(x)),$$

我们称  $x$  为问题(1)的一个  $\alpha$ -稳定点。

**引理 2 [11]** 若  $x^* \in C_S \cap B$  是问题(1)的一个  $\alpha$ -稳定点, 那么  $x^*$  是一个基本可行点。

**定理 2 [11]** 若  $x^*$  是问题(1)一个最优解, 假设 2.1 成立, 对于  $\frac{1}{\alpha} > L$ ,  $x^*$  是一个  $\alpha$ -稳定点。

**引理 3 [11]** (下降性引理) 假设 2 成立,  $L' > L$ , 有  $f(x) < h_{L'}(x, y)$ , 其中

$$h_{L'}(x, y) \equiv f(y) + \langle \nabla f(y), x - y \rangle + \frac{L'}{2} \|x - y\|^2. \tag{2}$$

### 3. 算法及收敛性分析

下面给出我们的算法:

算法 1:

步骤 0. 初始点  $x^0 \in R^n$ ,  $0 < \lambda < \frac{1}{4L}$ ,  $0 < \alpha_0 < \frac{1}{L}$ ,  $0 < \beta < 1$ ,  $\varepsilon > 0$ , 令  $k = 0$ 。

步骤 1. 计算  $x^{k+1} \in P_{C_S \cap B}(x^k - \alpha_k \nabla f(x^k))$ ,  $\alpha_k = \alpha_0 \beta^{m_k}$  并且  $m_k$  是使得

$$f(x^k(\alpha_0 \beta^m)) \leq f(x^k) - \frac{\lambda \|x^k(\alpha_0 \beta^m) - x^k\|^2}{(\alpha_0 \beta^m)^2} \tag{3}$$

成立的最小正整数  $m$ 。这里  $x^k(\alpha) = P_{C_S \cap B}(x^k - \alpha \nabla f(x^k))$ 。

步骤 2. 若  $\|x^{k+1} - x^k\| \leq \varepsilon$ , 则停止, 否则令  $k = k + 1$ , 转步骤 1。

**引理 4** 对于任意的  $L' > L$ , 则对任意满足下式的  $y \in C_S \cap B$ ,  $x \in R^n$ ,

$$y \in P_{C_S \cap B}\left(x - \frac{1}{L'} \nabla f(x)\right).$$

有:

$$f(x) - f(y) \geq \frac{L' - L}{2} \|x - y\|^2. \tag{4}$$

**证明:** 因为  $y \in P_{C_S \cap B}\left(x - \frac{1}{L'} \nabla f(x)\right)$ , 所以

$$y \in \arg \min_{z \in C_S \cap B} \left\| z - \left(x - \frac{1}{L'} \nabla f(x)\right) \right\| \tag{5}$$

又

$$\begin{aligned} h_{L'}(z, x) &= f(x) + \langle \nabla f(x), z - x \rangle + \frac{L'}{2} \|z - x\|^2 \\ &= \frac{L'}{2} \left\| z - \left(x - \frac{1}{L'} \nabla f(x)\right) \right\|^2 + f(x) - \frac{1}{2L'} \|\nabla f(x)\|^2 \end{aligned} \tag{6}$$

又  $f(x) - \frac{1}{2L'} \|\nabla f(x)\|^2$  关于  $z$  为常数, 故(6)式等价于  $y \in \arg \min_{z \in C_S \cap B} h_{L'}(z, x)$ , 即  $h_{L'}(y, x) < h_{L'}(x, x) = f(x)$ 。

通过引理 3, 我们有

$$f(x) - f(y) \geq f(x) - h_{L'}(y, x),$$

即

$$f(y) \leq h_{L'}(y, x)。$$

有

$$f(x) - f(y) \geq \frac{L-L'}{2} \|x-y\|^2。$$

**引理 5** 假设  $\{x^k\}_{k \geq 0}$  是由算法产生的迭代点列, 则  $\alpha_k \in \left[ \frac{1-\sqrt{1-4\lambda L}}{2L}, \frac{1+\sqrt{1-4\lambda L}}{2L} \right]$ , 且

$$f(x^{k+1}) \leq f(x^k) - \frac{\lambda \|x^{k+1} - x^k\|}{\alpha_k^2}。$$

**证明:** 根据步骤 1, 我们有  $x^{k+1} = P_{C_S \cap B}(x^k - \alpha_k \nabla f(x^k))$ , 根据引理 3 和引理 4, 由简单计算, 我们可以得出

$$\alpha_k \in \left[ \frac{1-\sqrt{1-4\lambda L}}{2L}, \frac{1+\sqrt{1-4\lambda L}}{2L} \right]。$$

并且根据(3), 我们有

$$f(x^{k+1}) \leq f(x^k) - \frac{\lambda \|x^{k+1} - x^k\|}{\alpha_k^2}。 \quad (7)$$

引理得证。

**定理 3** 设  $\{x^k\}_{k \geq 0}$  是由算法 1 产生的点列, 则

1.  $\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0$ ;
2.  $\{x^k\}_{k \geq 0}$  的任何一个聚点都是  $\alpha$ -稳定点。

**证明:** 1. 根据算法, 有  $f(x^k) - f(x^{k+1}) \geq \frac{\lambda \|x^{k+1} - x^k\|^2}{2\alpha_k^2}$ , 可得  $\{f(x^k)\}_{k \geq 0}$  是单调非增的, 故  $\{f(x^k)\}_{k \geq 0}$  是收敛的。所以有:

$$\sum_{k=0}^{\infty} \frac{\|x^{k+1} - x^k\|^2}{\alpha_k^2} \leq \frac{2}{\lambda} \sum_{k=0}^{\infty} (f(x^k) - f(x^{k+1})) < \infty。$$

所以  $\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - x^k\|}{\alpha_k} = 0$ 。根据引理 5, 知  $\{\alpha_k\}$  是有界的。所以  $\lim_{k \rightarrow \infty} \|x^{k+1} - x^k\| = 0$ 。

2. 假设  $x^*$  是点列  $\{x^k\}_{k \geq 0}$  的一个聚点。则存在一个子列  $\{x^{k_n}\}_{n \geq 0}$  收敛到  $x^*$ 。通过结论 1, 可知当  $n \rightarrow \infty$  时,  $x^{k_n+1} \rightarrow x^*$ 。又因为

$$x^{k_n+1} \in P_{C_S \cap B}(x^{k_n} - \alpha_{k_n} \nabla f(x^{k_n}))。$$

故对于任意的  $x \in C_S \cap B$ , 我们有

$$\|x - (x^{k_n} - \alpha_{k_n} \nabla f(x^{k_n}))\|^2 \geq \|x^{k_n+1} - (x^{k_n} - \alpha_{k_n} \nabla f(x^{k_n}))\|^2。 \tag{8}$$

根据引理 5, 知  $\{\alpha_k\}$  有界, 所以存在收敛子列, 不妨假设当  $n \rightarrow \infty$ , 有  $\alpha_{k_n} \rightarrow \alpha > 0$ 。

对不等式(8)两边同时取极限得

$$\|x - (x^* - \alpha \nabla f(x^*))\|^2 \geq \|x^* - (x^* - \alpha \nabla f(x^*))\|^2。$$

又因为  $C_S$  和  $B$  都是闭集,  $\{x^k\} \in C_S \cap B$ , 所以  $x^* \in C_S \cap B$ , 可以得出

$$x^* \in P_{C_S \cap B}(x^* - \alpha \nabla f(x^*))。$$

所以  $x^*$  是问题(1)的一个  $\alpha$ -稳定点。

注: 对于往一般稀疏集合和闭凸集的交上投影比较难以获得, 但是对于我们前面提到的类型 1 或者类型 2 的对称集合上的投影现在已经可以得到, 见文献[11]。

#### 4. 数值实验

在本节中, 我们给出数值例子验证了我们算法的有效性。所有的程序是在 **LENONVE ideapad**, **Windows 10 Inter(R) Core(TM)i5-6200U CPU @2.30 GHz 2.40 GHz and 4 GB** 内存的电脑上运行。

我们考虑下面这个例子:

$$f(x) = \frac{1}{2} \|Ax - b\|^2 \quad \text{s.t. } \|x\|_0 \leq s, \|x\|_2 \leq 0.25$$

在试验中, 取  $\varepsilon = 10^{-6}$ ,  $\lambda = \frac{1}{8L}$ ,  $\beta = 0.5$ 。

在表 1 中, 我们取  $A = \text{rand}(M \times N)$ ,  $b = \text{rand}(M \times 1)$ , 取  $s = 3$ 。在表中,  $k$  表示迭代步数, CPU time 表示迭代时间。

在表 2 中, 我们取  $A = \text{rand}(M \times N)$ ,  $b = \text{rand}(M \times 1)$ ,  $s = 10$ ,  $k$  表示迭代步数, CPU time 表示迭代时间。

我们从上面两个表格中可以看到, 对于矩阵  $A$  的不同维数和不同的稀疏度  $s$ , 我们的算法都有不错的效果。

在表 3 中, 我们取  $A = \text{rand}(M \times N)$ ,  $M = 1000$ ,  $N = 1500$ ,  $s = 10$ ,  $x^0$  表示初始点,  $k$  表示迭代步数, CPU time 表示迭代时间。

从表 3 中, 我们可以得出, 从不同的初始点开始, 都有较快的收敛速度。

**Table 1.** Numerical result  
**表 1.** 数值结果

$M$	$N$	$k$	CPU time
10	10	10	0.040297
100	100	10	0.044221
1000	1000	12	0.532451
1000	1500	14	1.048292

**Table 2.** Numerical result  
**表 2.** 数值结果

$M$	$N$	$k$	CPU time
100	100	10	0.041210
1000	1000	13	0.530339
1000	1500	14	1.006472

**Table 3.** Numerical result  
**表 3.** 数值结果

$x^0$	$k$	CPU time
$ones(N,1)$	13	0.986657
$10 * ones(N,1)$	15	0.969777
$50 * ones(N,1)$	14	0.971449
$rand(N \times 1)$	13	0.960487
$10 * rand(N \times 1)$	15	0.994397

## 5. 总结

本文主要研究了带有稀疏约束和闭凸集约束的优化问题。我们对文献[12]所提出的支撑投影算法(GSP)进行了推广,来求解本文所研究的优化问题。我们证明了由此算法产生的点列可以收敛到问题的一个 $\alpha$ -稳定点,并且用数值例子说明了算法的有效性。

## 基金项目

国家自然科学基金(11271226): 凸可行问题的松弛投影算法及其应用研究。

## 参考文献 (References)

- [1] Negahban, S., Ravikumar, P., Wainwright, M. and Yu, B. (2012) A Unified Framework for High-Dimensional Analysis of  $M$ -Estimators with Decomposable Regularizers. *Statistical Science*, **27**, 538-557. <https://doi.org/10.1214/12-STS400>
- [2] Agarwal, A., Negahban, S. and Wainwright, M. (2010) Fast Global Convergence Rates of Gradient Methods for High-Dimensional Statistical Recovery. *Advances in Neural Information Processing Systems*, **23**, 37-45.
- [3] Blumensath, T. (2013) Compressed Sensing with Nonlinear Observations and Related Nonlinear Optimization Problems. *IEEE Transactions on Information Theory*, **59**, 3466-3474. <https://doi.org/10.1109/TIT.2013.2245716>
- [4] Jalali, A., Johnson, C.C. and Ravikumar, P.K. (2011) On Learning Discrete Graphical Models Using Greedy Methods. *Advances in Neural Information Processing Systems*, **24**, 1935-1943.
- [5] Tewari, A., Ravikumar, P.K. and Dhillon, I.S. (2011) Greedy Algorithms for Structurally Constrained High Dimensional Problems. *Advances in Neural Information Processing Systems*, **24**, 882-890.
- [6] Yuan, X., Li, P. and Zhang, T. (2013) Gradient Hard Thresholding Pursuit for Sparsity-Constrained Optimization. *ICML*, 127-135.
- [7] Bahmani, S., Raj, B. and Boufounos, P. (2013) Greedy Sparsity-Constrained Optimization. *Journal of Machine Learning Research*, **14**, 807-841.
- [8] Candès, E.J. and Tao, T. (2005) Decoding by Linear Programming. *IEEE Transactions on Information Theory*, **51**, 4203-4215. <https://doi.org/10.1109/TIT.2005.858979>
- [9] Donoho, D.L. (2006) Compressed Sensing. *IEEE Transactions on Information Theory*, **52**, 1289-1306. <https://doi.org/10.1109/TIT.2006.871582>

- 
- [10] Bunea, F. (2008) Honest Variable Selection in Linear and Logistic Regression Models via  $l_1$  and  $l_1 + l_2$  Penalization. *Electronic Journal of Statistics*, **2**, 1153-1194. <https://doi.org/10.1214/08-EJS287>
- [11] Beck, A. and Hallak, N. (2015) On the Minimization over Sparse Symmetric Sets: Projections, Optimality Conditions, and Algorithms. *Mathematics of Operations Research*, **41**, 604-605.
- [12] Pan, L.L., Xiu, N.H. and Zhou, S.L. (2014) Gradient Support Projection Algorithm for Affine Feasibility Problem with Sparsity and Nonnegativity. *Mathematics*, **42**, 1439-1444.

**期刊投稿者将享受如下服务:**

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

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

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