Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 6489190, 7 pages
https://doi.org/10.1155/2020/6489190

Research Article

A Gradient Projection Algorithm with a New Stepsize for Nonnegative Sparsity-Constrained Optimization

Ye Li,[1,2] **Jun Sun**,[3] **and Biao Qu**[1]

[1]*Qufu Normal University, Rizhao 276826, Shandong, China*
[2]*Shandong Women's University, Jinan 250300, Shandong, China*
[3]*Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Ye Li; 35030@sdwu.edu.cn

Received 26 May 2020; Revised 30 July 2020; Accepted 8 August 2020; Published 26 August 2020

Guest Editor: Chuanjun Chen

Nonnegative sparsity-constrained optimization problem arises in many fields, such as the linear compressing sensing problem and the regularized logistic regression cost function. In this paper, we introduce a new stepsize rule and establish a gradient projection algorithm. We also obtain some convergence results under milder conditions.

1. Introduction

In this paper, we are mainly concerned with the nonnegative sparsity-constrained optimization problem (NN-SCO):

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in S \cap R_+^n, \end{aligned} \tag{1}$$

where $f: R^n \longrightarrow R$ is a continuously differential function with a lower bound. $S: = \{x \in R^n: \|x\|_0 \le s\}$ is a sparse set, where $s < n$ is a given integer regulating the sparsity level in x and R_+^n is the nonnegative orthant in R^n. $\|x\|_0$ is the l_0 norm of x, counting the number of nonzero elements in x. Many application problems can be translated into problem (1), such as the widely studied linear compressing sensing problem of $f(x) = (1/2)\|Ax - b\|^2$ with $A \in R^{m\times n}$ being a sensing matrix, $b \in R^m$ is the observation vector, and $\|\cdot\|$ is the Euclidean norm in R^n [1]. Problem (1) has also used to the regularized logistic regression cost function [2].

Recently, a great deal of work has been devoted to algorithms for sparsity-constrained optimization problem. Beck and Eldar [3] established the IHT algorithm which converges to L-stationary under the Lipchitz continuity of the gradient of objective function. Beck and Hallak [4] generalized these results to sparse symmetric sets. Lu [5] designed a nonmonotone algorithm for symmetric set constraint problems. Pan, Xiu, and Zhou [6, 7] established the B-stationary, C-stationary, and α-stationary based on the Bouligand tangent cone and Clarke tangent. Recently, Pan, Zhou, and Xiu [8] established the improved IHT algorithm (IIHT) for problem (1) by using Armijo line search. They proved that any accumulation point converged to an α-stationary point under the restricted strong smoothness of objective function which is weaker than the Lipchitz continuity of the gradient.

Inspired by the above literature studies, in this paper, we establish a gradient projection algorithm with a new stepsize. The new algorithm removes the condition of the restricted strong smoothness of objective function which makes it more applicable. Meanwhile, we prove the convergence of the algorithm.

The rest of this paper is organized as follows. In Section 2, we present some notations, definitions, and lemmas. In Section 3, we give the algorithm of (1) and prove the convergence properties.

2. Preliminaries

2.1. Notations. To make it easier to read, we give some used notations as follows:

$$\begin{aligned} &S = \{x \in R^n\colon \|x\|_0 \le s\},\\ &I_1(x) \equiv \{i \in \{1,\ldots,n\}\colon x_i \ne 0\},\\ &I_1(xy) = I_1(x) \cup I_1(y),\\ &P_{S\cap R_+^n}(x) := \operatorname{argmin}_{y\in S\cap R_+^n}\|x-y\|,\\ &I_0(x) \equiv \{i \in \{1,\ldots,n\}\colon x_i = 0\},\\ &|I_1(x)|\colon \text{the cardinality of } I_1(x). \end{aligned} \tag{2}$$

2.2. Definitions

Definition 1 (see [8]). Let $x^* \in S\cap R_+^n$ be a given feasible point of (1). We say that x^* is an α-stationary point, if there exists $\alpha > 0$ such that

$$x^* \in P_{S\cap R_+^n}(x^* - \alpha\nabla f(x^*)). \tag{3}$$

Definition 2 (see [9]). A function f is called $2s$-restricted strongly smooth ($2s$-RSS) with parameter $L_{2s} > 0$, and if for any $x, y \in R^n$ satisfying $|I_1(xy)| < 2s$, it holds that

$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle \le \frac{L_{2s}}{2}\|y-x\|^2. \tag{4}$$

Definition 3 (see [9]). A function f is called $2s$-restricted strongly convex ($2s$-RSC) with parameter $l_{2s} > 0$, and if for any $x, y \in R^n$ satisfying $|I_1(xy)| < 2s$, it holds that

$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle \ge \frac{l_{2s}}{2}\|y-x\|^2. \tag{5}$$

If and only if for any $x, y \in R^n$ and $|I_1(xy)| \le 2s$, we have

$$\left\|(\nabla f(x) - \nabla f(y))_{I_1(xy)}\right\| \ge l_{2s}\|x-y\|. \tag{6}$$

In particular, in (5), if $l_{2s} = 0$, the function f is called $2s$-restricted convex ($2s$-RC).

Definition 4 (see [10]). The projected gradient $\nabla_{S\cap R_+^n} f(x)$ of f is defined by

$$\begin{aligned} \nabla_{S\cap R_+^n} f(x) &= \left\|P_{T^C_{S\cap R_+^n}}(-\nabla f(x))\right\|,\\ &= \arg\min\left\{\|v + \nabla f(x)\| \,\middle|\, v \in T^C_{S\cap R_+^n}\right\},\\ &= \max\left\{\langle -\nabla f(x^k), v\rangle, \|v\| = 1\right\}. \end{aligned} \tag{7}$$

2.3. Lemmas

Lemma 1 (see [8]). *For $\alpha > 0$, vector $x^* \in S\cap R_+^n$ is an α-stationary point if and only if*

$$\nabla_i f(x^*)\begin{cases} = 0, & i \in I_1(x^*),\\ \ge \alpha M_s(x^*), & i \in I_0(x^*). \end{cases} \tag{8}$$

In particular, when $\|x^*\|_0 = s$, $x^* \ge 0$, $\nabla_i f(x^*) = \begin{cases} 0, & i \in I_1(x^*)\\ \ge -\alpha M_s(x^*), & i \in I_0(x^*) \end{cases}$.

When $\|x^*\|_0 < s$, $x^* \ge 0$, $\nabla_i f(x^*) = \begin{cases} 0, & i \in I_1(x^*)\\ \in R_+^n, & i \in I_0(x^*) \end{cases}$.

Lemma 2 (see [8]). $P_{S\cap R_+^n}(x) = P_S(P_{R_+^n}(x))$.

Lemma 3 (see [8]). *For any $x^* \in S\cap R_+^n$, we have*

$$T^C_{S\cap R_+^n}(x^k) = \operatorname{span}\{e_i, i \in I_1(x^k)\}, \tag{9}$$

where $e_i \in R^n$ is a vector whose i^{th} component is one and others are zeros.

3. Main Results

In this section, we establish a new algorithm which improves the IIHT algorithm for (1) and then we analyze its convergence properties. At first, let us develop the gradient projection algorithm with a new stepsize rule.

Algorithm 1

Step 1. Initialize $x^0 \in S\cap R_+^n$, $0 \le \theta \le 1$, and $\varepsilon > 0$, and set $k \Longleftarrow 0$.

Step 2. Compute $L_k = \sup_{\alpha>0}(\|\nabla f(x^k) - \nabla f(z^k(\alpha,\theta))\|/\|x^k - z^k(\alpha,\theta)\|)$, where

$$\begin{aligned} z^k(\alpha,\theta) &= x^k + \theta(x^k(\alpha) - x^k),\\ x^k(\alpha) &= P_{S\cap R_+^n}(x^k - \alpha\nabla f(x^k)). \end{aligned} \tag{10}$$

Step 3. Compute $x^{k+1} = P_{S\cap R_+^n}(x^k - \alpha_k\nabla f(x^k))$ where α_k satisfies $0 \le \alpha_k \le (1/3L_k)$.

Step 4. If $\|\nabla_{\Gamma_k} f(x^k)\| \le \varepsilon$, then stop; otherwise, set $k \Longleftarrow k+1$ and go to Step 2.

Next, let us list the following assumptions for convenience:

(1) For any $k > 0$, $L_k < +\infty$

(2) f is bounded below on $S\cap R_+^n$

Lemma 4. *Let the sequence $\{x^k\}$ be generated by Algorithm 1, and set $l_k = 3L_k$. Then, we have*

$$f(x^k) \le h_{l_k}(x^k, x^{k+1}), \tag{11}$$

where $h_{l_k}(x^k, x^{k+1}) = f(x^{k+1}) + \langle \nabla f(x^{k+1}), x^k - x^{k+1}\rangle + (l_k/2)\|x^k - x^{k+1}\|^2$.

Proof. Let

$$g(t) = f(x^{k+1} + t(x^k - x^{k+1})). \tag{12}$$

Then,

$$\begin{aligned} g(0) &= f\left(x^{k+1}\right), \\ g(1) &= f\left(x^{k}\right), \\ g'(t) &= \left(x^{k}-x^{k+1}\right)^{T}\nabla f\left(x^{k+1}+t\left(x^{k}-x^{k+1}\right)\right). \end{aligned} \tag{13}$$

Thus,

$$\begin{aligned} f\left(x^{k}\right)-f\left(x^{k+1}\right) &= g(1)-g(0), \\ &= \int_0^1 g'(t)\mathrm{d}t, \\ &= \int_0^1 \left(x^{k}-x^{k+1}\right)^{T}\nabla f\left(x^{k+1}+t\left(x^{k}-x^{k+1}\right)\right)\mathrm{d}t, \\ &= \int_0^1 \left(x^{k}-x^{k+1}\right)^{T}\nabla f\left(x^{k+1}\right)\mathrm{d}t \\ &\quad + \int_0^1 \left(x^{k}-x^{k+1}\right)^{T}\left(\nabla f\left(x^{k+1}+t\left(x^{k}-x^{k+1}\right)\right)-\nabla f\left(x^{k+1}\right)\right)\mathrm{d}t, \\ &= \int_0^1 \left(x^{k}-x^{k+1}\right)^{T}\nabla f\left(x^{k+1}\right)\mathrm{d}t \\ &\quad + \int_0^1 \left(x^{k}-x^{k+1}\right)^{T}\left(\nabla f\left(x^{k+1}+t\left(x^{k}-x^{k+1}\right)\right)-\nabla f\left(x^{k+1}\right)+\nabla f\left(x^{k}\right)-\nabla f\left(x^{k}\right)\right)\mathrm{d}t, \\ &= \int_0^1 \left(x^{k}-x^{k+1}\right)^{T}\nabla f\left(x^{k+1}\right)\mathrm{d}t + \int_0^1 \left(x^{k}-x^{k+1}\right)^{T}\left(\nabla f\left(x^{k}\right)-\nabla f\left(x^{k+1}\right)\right)\mathrm{d}t, \\ &\quad + \int_0^1 \left(x^{k}-x^{k+1}\right)^{T}\left(\nabla f\left(x^{k+1}+t\left(x^{k}-x^{k+1}\right)\right)-\nabla f\left(x^{k}\right)\right)\mathrm{d}t, \\ &\le \int_0^1 \left(x^{k}-x^{k+1}\right)^{T}\nabla f\left(x^{k+1}\right)\mathrm{d}t + \int_0^1 \left(x^{k}-x^{k+1}\right)^{T}\left(\nabla f\left(x^{k}\right)-\nabla f\left(x^{k+1}\right)\right)\mathrm{d}t, \\ &\quad + \int_0^1 \left\|x^{k}-x^{k+1}\right\|\left\|\left(\nabla f\left(x^{k+1}+t\left(x^{k}-x^{k+1}\right)\right)-\nabla f\left(x^{k}\right)\right)\right\|\mathrm{d}t, \\ &\le \left(x^{k}-x^{k+1}\right)^{T}\nabla f\left(x^{k+1}\right)+L_k\left\|x^{k}-x^{k+1}\right\|^2+\left\|x^{k}-x^{k+1}\right\|\int_0^1 L_k(1-t)\left\|x^{k}-x^{k+1}\right\|\mathrm{d}t, \\ &\le \left(x^{k}-x^{k+1}\right)^{T}\nabla f\left(x^{k+1}\right)+\frac{3L_k}{2}\left\|x^{k}-x^{k+1}\right\|^2, \\ &= \left(x^{k}-x^{k+1}\right)^{T}\nabla f\left(x^{k+1}\right)+\frac{l_k}{2}\left\|x^{k}-x^{k+1}\right\|^2. \end{aligned} \tag{14}$$

Then, (11) is tenable. □

Lemma 5. *We suppose $l \ge l_k$. For $x^k \in S \cap R_+^n$ and $x^{k+1} = P_{S\cap R_+^n}(x^k - (1/l)\nabla f(x^k))$, we have*

$$f\left(x^{k}\right)-f\left(x^{k+1}\right) \ge \sigma\left\|x^{k}-x^{k+1}\right\|^2, \tag{15}$$

where $\sigma = (l - l_k/2)$.

Proof. Since $x^{k+1} \in P_{S\cap R_+^n}(x^k - (1/l)\nabla f(x^k))$, by the definition of projection, we get

$$x^{k+1} \in \arg\min_{x\in S\cap R_+^n}\left\|x-\left(x^{k}-\frac{1}{l}\nabla f\left(x^{k}\right)\right)\right\|^2. \tag{16}$$

Moreover,

$$h_l(x, x^k) = f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{l}{2}\|x - x^k\|,$$
$$= \frac{l}{2}\left\|x - \left(x^k - \frac{l}{2}\nabla f(x^k)\right)\right\|^2 + f(x^k) - \frac{1}{2l}\|\nabla f(x^k)\|^2. \tag{17}$$

Because $f(x^k) - (1/2l)\|\nabla f(x^k)\|^2$ is a constant independent of x, we can get

$$x^{k+1} \in \arg\min_{x \in P_{S \cap R^n_+}} h_l(x, x^k). \tag{18}$$

Therefore,

$$h_l(x^{k+1}, x^k) \geq h_l(x^k, x^k) = f(x^k). \tag{19}$$

By Lemma 4, we get

$$f(x^{k+1}) \geq h_l(x^{k+1}, x^k). \tag{20}$$

Hence,

$$f(x^k) - f(x^{k+1}) \geq f(x^k) - h_{l_k}(x^{k+1}, x^k),$$
$$\geq h_l(x^{k+1}, x^k) - h_{l_k}(x^{k+1}, x^k), \tag{21}$$
$$= \frac{l - l_k}{2}\|x^{k+1} - x^k\|^2.$$

Let $\sigma = (l - l_k/2)$. We get

$$f(x^k) - f(x^{k+1}) \geq \sigma\|x^k - x^{k+1}\|^2. \tag{22}$$

□

Lemma 6. *Let the sequence $\{x^k\}$ be generated by Algorithm 1. Then,*

(1) $f(x^k) - f(x^{k+1}) \geq ((1/\alpha_k) - l_k)/2\|x^k - x^{k+1}\|^2$

(2) $\{f(x^k)\}$ is an increasing sequence, and when $k \longrightarrow \infty$, $\{f(x^k)\}$ converges

(3) $\|x^k - x^{k+1}\| \longrightarrow 0$

(4) for any $k = 0, 1, 2, \ldots$, if $x^k \neq x^{k+1}$, we have $f(x^{k+1}) < f(x^k)$

Proof

(1) Since $0 \leq \alpha_k \leq (1/3L_k)$, we get

$$\frac{1}{\alpha_k} \geq 3L_k = l_k. \tag{23}$$

Setting $l = (1/\alpha_k)$ in (15), formula (1) can be obtained.

(2) We can easily get that $\{f(x^k)\}$ is an increasing sequence by (15). Moreover, by the assumptions (H_2), we can get that $\{f(x^k)\}$ converges.

(3) Let $\mu = ((1/\alpha_k) - l_k/2)$ in (1). We can get

$$f(x^k) - f(x^{k+1}) \geq \mu\|x^k - x^{k+1}\|^2. \tag{24}$$

Summing over both sides of this inequality, we get

$$\sum_{k=1}^{\infty}\|x^k - x^{k+1}\|^2 \leq \sum_{k=1}^{\infty}\frac{2}{\mu}\left(f(x^k) - f(x^{k+1})\right),$$
$$= \frac{2}{\mu}\left(f(x^0) - \lim_{k \longrightarrow +\infty} f(x^k)\right). \tag{25}$$

Since f is bounded below, we get

$$\|x^k - x^{k+1}\| \longrightarrow 0. \tag{26}$$

(4) It easily can be got by (2). □

Lemma 7. *Let the sequence $\{x^k\}$ be generated by Algorithm 1. Suppose that the function f is 2s-RC. We have*

$$\|\nabla f(x^{k+1}) - \nabla f(x^k)\| \leq l_k\|x^{k+1} - x^k\|. \tag{27}$$

Proof. Because the sequence $\{x^k\}$ be generated by Algorithm 1, we get $|I_1(xy)| < 2s$. By Lemma 4 and Lemma 5 in reference [8], we can get

$$\|\nabla f(x^{k+1}) - \nabla f(x^k)\| \leq l_k\|x^{k+1} - x^k\|. \tag{28}$$

□

Theorem 1. *Let the sequence $\{x^k\}$ be generated by Algorithm 1. Then, the following results hold:*

(1) Any accumulation of sequence $\{x^k\}$ is an α-stationary point.

(2) If f is 2s-RC, the projected gradient sequence converges to zero, i.e.,

$$\lim_{k \longrightarrow +\infty}\|\nabla_{\Gamma_k} f(x^k)\| = 0. \tag{29}$$

Proof

(1) Suppose that x^* is an accumulation point of sequence $\{x^k\}$. Then, there exists a subsequence $\{x^{k_n}\}$ converges to x^*.

Because

$$\|x^{k_n}\| - \|x^{k_n+1} - x^{k_n}\| \leq \|x^{k_n+1}\| = \|x^{k_n+1} - x^{k_n} + x^{k_n}\|$$
$$\leq \|x^{k_n}\| + \|x^{k_n+1} - x^{k_n}\|, \tag{30}$$

we get

$$\lim_{n \longrightarrow +\infty} x^{k_n+1} = \lim_{n \longrightarrow +\infty} x^{k_n} = x^*. \tag{31}$$

Moreover,

$$x^{k_n+1} = P_{S \cap R^n_+ R^n}\left(x^k - \alpha_{k_n}\nabla f(x^{k_n})\right),$$
$$= P_S\left(P_{R^n_+}\left(x^k - \alpha_{k_n}\nabla f(x^{k_n})\right)\right). \tag{32}$$

We consider the next two cases: □

Case 1. For $i \in I_1(x^*)$, there must exist a sufficiently large index N and a constant $c_0 > 0$ such that

$$\min\left\{x_i^{k_n}, x_i^{k_n+1}\right\} \geq c_0 > 0. \tag{33}$$

By $P_{S\cap R_+^n} = P_S(P_{R_+^n})$ and (33), we can get

$$x_i^{k_n+1} = x_i^k - \alpha_{k_n}\nabla_i f\left(x^{k_n}\right). \tag{34}$$

Since

$$\lim_{n\longrightarrow+\infty}\inf \alpha_{k_n} > 0, \tag{35}$$

without loss of generality, we can suppose $\lim_{k\longrightarrow+\infty}\alpha_{k_n} > c$. Let $n \longrightarrow +\infty$. We get

$$x_i^* = x_i^* - c\nabla_i f(x^*), \tag{36}$$

i.e.,

$$\nabla_i f(x^*) = 0, \quad \forall i \in I_1(x^*). \tag{37}$$

Case 2. For $i \in I_0(x^*)$, we consider two subcases.

Subcase 1. When $\|x^*\|_0 = s$, we get

$$0 = x^* = \lim_{n\longrightarrow+\infty} x_i^{k_n+1} = P_S\left(P_{R_+^n}\left(x^{k_n} - \alpha_{k_n}\nabla f\left(x^{k_n}\right)\right)\right)_i. \tag{38}$$

Due to the property of the projections P_S and $P_{R_+^n}$, we have

$$\max\left\{x_i^{k_n} - \alpha_{k_n}\nabla_i f\left(x^{k_n}\right)\right), 0\right\} \leq M_s(x^*). \tag{39}$$

Thus,

$$x_i^{k_n} - \alpha_{k_n}\nabla_i f\left(x^{k_n}\right) \leq M_s(x^*). \tag{40}$$

Taking limits on both sides, we obtain

$$\nabla_i f(x^*) \geq -\frac{1}{c}M_s(x^*). \tag{41}$$

Subcase 2. When $\|x^*\|_0 < s$, suppose $\nabla_i f(x^*) < 0$, and we have

$$\lim_{n\longrightarrow+\infty}\left(x_i^{k_n} - \alpha_{k_n}\nabla_i f\left(x^{k_n}\right)\right) = -c\nabla_i f(x^*) > 0. \tag{42}$$

For all sufficiently large n, we have

$$\begin{aligned} P_{R_+^n}\left(x_i^{k_n} - \alpha_{k_n}\nabla_i f\left(x^{k_n}\right)\right) &= x_i^{k_n} - \alpha_{k_n}\nabla_i f\left(x^{k_n}\right), \\ &= -c\nabla_i f(x^*) > 0. \end{aligned} \tag{43}$$

Since $\|x^*\|_0 < s$, for all sufficiently large n, we have

$$\begin{aligned} x_i^{k_n+1} &= P_S\left(P_{R_+^n}\left(x^{k_n} - \alpha_{k_n}\nabla f\left(x^{k_n}\right)\right)\right)_i, \\ &= P_{R_+^n}\left(x^{k_n} - \alpha_{k_n}\nabla f\left(x^{k_n}\right)\right)_i > 0. \end{aligned} \tag{44}$$

which contradicts with $i \in I_0(x^*)$. Thus, $\nabla_i f(x^*) \geq 0$.

Summarizing the two cases, we obtain

$$\nabla_i f(x^*)\begin{cases} = 0, & i \in I_1(x^*), \\ \geq -\dfrac{1}{c}M_s(x^*), & i \in I_0(x^*). \end{cases} \tag{45}$$

Thus, x^* is an α-stationary point of (1).

(2) Set $\Gamma^k = I_1(x^k)$. By Lemma 3, we have

$$\begin{aligned} T_{S\cap R_+^n}^C\left(x^k\right) &= R_{\Gamma^k}^n, \\ &= \operatorname{span}\left\{e_i, i \in I_1\left(x^k\right)\right\}. \end{aligned} \tag{46}$$

By Definition 4, we have

$$\begin{aligned} \left\|P_{T_{S\cap R_+^n}^C}\left(-\nabla f\left(x^k\right)\right)\right\| &= \max\left\{\left\langle\left(-\nabla f\left(x^k\right)\right), v\right\rangle, \|v\| = 1\right\}, \\ &= \left\|\nabla_{\Gamma_k} f\left(x^k\right)\right\|. \end{aligned} \tag{47}$$

Moreover, the maximum value is taken at $\|v\| = 1$. For any $\varepsilon > 0$, there exists $v^k \in R_{\Gamma^k}^n$ and $\|v^k\| = 1$ satisfies

$$\left\|\nabla_{\Gamma_k} f\left(x^k\right)\right\| \leq \left\langle -\nabla f\left(x^k\right), v^k\right\rangle + \varepsilon. \tag{48}$$

Because $x^{k+1} = P_{S\cap R_+^n}(x^k - \alpha_k\nabla f(x^k))$ and $x_{\Gamma^{k+1}} = y_{\Gamma^{k+1}}$, $x \in P_{S\cap R_+^n}(y)$, we get

$$x_{\Gamma^{k+1}}^{k+1} = \left(x^k - \alpha_k\nabla f\left(x^k\right)\right)_{\Gamma^{k+1}}, \tag{49}$$

i.e.,

$$x_{\Gamma^{k+1}}^{k+1} - \left(x^k - \alpha_k\nabla f\left(x^k\right)\right)_{\Gamma^{k+1}} = 0. \tag{50}$$

Thus, for any $\varpi_{k+1} \in R_{\Gamma^{k+1}}^n$, we get

$$\left\langle\left(x^{k+1} - \left(x^k - \alpha_k\nabla f\left(x^k\right)\right), \varpi_{k+1} - x_{k+1}\right\rangle = 0. \tag{51}$$

Taking $\varpi_{k+1} = x_{k+1} + v_{k+1}$, we get

$$\left\langle\left(x^{k+1} - x^k + \alpha_k\nabla f\left(x^k\right), -v_{k+1}\right\rangle = 0. \tag{52}$$

By the Cauchy–Schwartz inequality, we get

$$\left\langle\alpha_k\nabla f\left(x^k\right), -v^{k+1}\right\rangle = \left\langle x^{k+1} - x^k, v^{k+1}\right\rangle \leq \left\|x^{k+1} - x^k\right\|, \tag{53}$$

i.e.,

$$-\left\langle\nabla f\left(x^k\right), v^{k+1}\right\rangle \leq \frac{\left\|x^{k+1} - x^k\right\|}{\alpha_k}. \tag{54}$$

By Lemma 7, we get

$$\begin{aligned}-\langle\nabla f\left(x^{k+1}\right), v^{k+1}\rangle &= -\langle\nabla f\left(x^{k+1}\right)-\nabla f\left(x^{k}\right), v^{k+1}\rangle \\ &\quad -\langle\nabla f\left(x^{k}\right), v^{k+1}\rangle, \\ &\le l_k\left\|x^{k+1}-x^{k}\right\|+\frac{\left\|x^{k+1}-x^{k}\right\|}{\alpha_k}.\end{aligned} \tag{55}$$

Taking limits on both sides and using Lemma 6, we have

$$\lim_{k\longrightarrow+\infty}\sup\langle-\nabla f\left(x^{k}\right), v^{k+1}\rangle\le 0. \tag{56}$$

By (32), we get

$$\lim_{k\longrightarrow+\infty}\left\|\nabla_{\Gamma_k} f\left(x^{k}\right)\right\|=0. \tag{57}$$

Theorem 2. *Let the sequence $\{x^k\}$ be generated by Algorithm 1. x^* is an accumulation point of the sequence $\{x^k\}$. Suppose $f(x)$ is $2s-$ RC, then the following results hold:*

(1) If $\|x^*\|_0<s$, then x^* is a global minimizer of (1)

(2) If $\|x^*\|_0=s$, then x^* is a local minimizer of (1)

Proof

(1) For $\forall x\in S\cap R^n_+$, we have $|I_1(xx^*)|=|I_1(x)\cup I_1(x^*)|\le 2s$. Since $f(x)$ is $2s-$ RC, by Definition 3, we have

$$\begin{aligned}f(x)&\ge f(x^*)+\langle\nabla f(x^*), x-x^*\rangle, \\ &= f(x^*)+\sum_{i\in I_1(x^*)}\nabla_i f(x^*)(x_i-x_i^*) \\ &\quad+\sum_{i\in I_0(x^*)}\nabla_i f(x^*)(x_i-x_i^*).\end{aligned} \tag{58}$$

Because x^* is an accumulation point of the sequence $\{x^k\}$. By Theorem 1, x^* is an α-stationary. By Lemma 1, we can get

$$f(x)\ge f(x^*). \tag{59}$$

Thus, x^* is a global minimizer of (1).

(2) If $\|x^*\|_0=s$, then $I_1(x^*)=I_1(x^k)$.

In fact, for all sufficiently large k, taking $0<\delta<\min\{x_i^*: i\in I_1(x^*)\}$, we get

$$\left\|x^k-x^*\right\|\le\delta. \tag{60}$$

For any $i\in I_1(x^*)$, we have

$$x_i^k=x_i^*-\left(x_i^*-x_i^k\right)\ge x_i^*-\left|x_i^*-x_i^k\right|>x_i^*-\delta>0. \tag{61}$$

Thus,

$$I_1(x^*)\subset I_1\left(x^k\right). \tag{62}$$

By $\|x^k\|_0=s$ and $|I_1(x^*)|=\|x^*\|_0=s$, we have

$$I_1(x^*)=I_1\left(x^k\right). \tag{63}$$

For any $x^k\in S\cap R^n_+$ satisfying $\|x^k-x^*\|\le\delta$, we have $|I_1(x^kx^*)|=|I_1(x^k)\cup I_1(x^*)|\le 2s$. Since $f(x)$ is $2s-$ RC, by Definition 3, Theorem 1, and Lemma 1, we have

$$\begin{aligned}f\left(x^k\right)&\ge f(x^*)+\langle\nabla f(x^*), x^k-x^*\rangle, \\ &= f(x^*)+\sum_{i\in I_1(x^*)}\nabla_i f(x^*)\left(x_i^k-x_i^*\right) \\ &\quad+\sum_{i\in I_0(x^*)}\nabla_i f(x^*)\left(x_i^k-x_i^*\right), \\ &\ge f(x^*).\end{aligned} \tag{64}$$

Thus, x^* is a local minimizer of (1). □

Theorem 3. *Let the sequence $\{x^k\}$ be generated by Algorithm 1. x^* is a limit of the sequence $\{x^k\}$. Suppose $f(x)$ is $2s-$ RSC with parameter l_{2s} and $\|x^*\|_0=s$, for all sufficiently large k, and we have*

$$\left\|x^{k+1}-x^*\right\|^2\le\left\|x^k-x^*\right\|^2,\quad 0<\rho<1, \tag{65}$$

where $\rho=1-(-2l_{2s}^2\alpha_k/L_k)+2l_{2s}^2\alpha_k^2$.

Proof. By Theorem 2, we get $x^k\longrightarrow x^*$. As $f(x)$ is $2s-$ RSC with parameter l_{2s}, for any $x, y\in R^n$ and $|I_1(xy)|\le 2s$, we have

$$\left\|(\nabla f(x)-\nabla f(y))_{I_1(xy)}\right\|\ge l_{2s}\|x-y\|. \tag{66}$$

Set $\Gamma^k=I_1(x^k)$ and $\Gamma^*=I_1(x^*)$. By Theorem 2, we get $\Gamma^k=\Gamma^*$. For all sufficiently large k, we have

$$\begin{aligned}\left\|\nabla_{\Gamma_*} f(x^*)\right\|&=\lim_{k\longrightarrow+\infty}\left\|\nabla_{\Gamma_k} f\left(x^k\right)\right\|=0, \\ x_{\Gamma^{k+1}}^{k+1}&=\left(x^k-\alpha_k\nabla f\left(x^k\right)\right)_{\Gamma^{k+1}}.\end{aligned} \tag{67}$$

For all sufficiently large k, we have

$$\begin{aligned}\left\|x^{k+1}-x^*\right\|^2&=\left\|\left(x_{\Gamma^*}^k-\alpha_k\nabla_{\Gamma^*} f\left(x^k\right)\right)-x_{\Gamma^*}^*+\alpha_k\nabla_{\Gamma^*} f(x^*)\right\|^2 \\ &=\left\|x^k-x^*\right\|^2-2\alpha_k\langle x^k-x^*, \nabla f\left(x^k\right)-\nabla f(x^*)\rangle \\ &\quad+\alpha_k^2\left\|\left(\nabla f\left(x^k\right)-\nabla f(x^*)\right)_{\Gamma^*}\right\|.\end{aligned} \tag{68}$$

Because $\|x^k-x^*\|\ge L_k\|\nabla f(x^k)-\nabla f(x^*)\|$, we get

$$\begin{aligned}\left\|x^{k+1}-x^*\right\|^2&\le\left\|x^k-x^*\right\|^2-\left(\frac{2\alpha_k}{L_k}-\alpha_k^2\right) \\ &\quad\cdot\left\|\left(\nabla f\left(x^k\right)-\nabla f(x^*)\right)_{\Gamma^*}\right\|, \\ &\le\left(1-\frac{2l_{2s}^2\alpha_k}{L_k}+l_{2s}^2\alpha_k^2\right)\left\|x^k-x^*\right\|^2.\end{aligned} \tag{69}$$

Since $0\le\alpha_k\le(1/3L_k)$ and $\sigma=(l-l_k/2)=(l-L_k/2)$, we have $0\le\alpha_k\le(1/2\sigma+3L_k)$. Thus,

$$\frac{\beta}{2\sigma+3L_k}\le\inf\alpha_k\le\frac{1}{2\sigma+3L_k}. \tag{70}$$

Setting $\alpha_* = (\beta/2\sigma + 3L_k)$, we get

$$\alpha_* \le \alpha_k \le \frac{1}{3L_k}. \tag{71}$$

Thus,

$$\begin{aligned} 1 - \frac{2l_{2s}^2\alpha_k}{L_k} + l_{2s}^2\alpha_k^2 &= 1 + l_{2s}^2\left(\alpha_k - \frac{1}{L_k}\right)^2 - \frac{l_{2s}^2}{L_k^2}, \\ &\le 1 + l_{2s}^2\left(\alpha_* - \frac{1}{L_k}\right)^2 - \frac{l_{2s}^2}{L_k^2}, \\ &= 1 - \frac{2l_{2s}^2\alpha_*}{L_k} + l_{2s}^2\alpha_*^2, \\ &= \rho^2. \end{aligned} \tag{72}$$

By $l_{2s} \le L_k$ and $\rho^2 = 1 + l_{2s}^2(\alpha_* - (1/L_k))^2 - (l_{2s}^2/L_k^2)$, we get $\rho > 0$.

From $0 < \beta < 1$ and $\rho^2 = 1 - (2l_{2s}^2\alpha_*/L_k) + l_{2s}^2\alpha_*^2$, we have $\rho < 1$.

Thus,

$$\lim_{k \longrightarrow +\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} \le \sqrt{\rho}, \tag{73}$$

where $0 < \sqrt{\rho} < 1$. Thus, the sequence $\{x^k\}$ is Q-linear convergence to x^*. □

4. Conclusions

In this paper, we are mainly concerned with the nonnegative sparsity-constrained optimization problem. We introduce a new stepsize rule and propose a new gradient projection algorithm to solve this problem. The new algorithm removes the condition of the restricted strong smoothness of objective function which makes the new algorithm more applicable. Meanwhile, we prove the convergence of the algorithm.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Acknowledgments

This project was supported by the National Science Foundation of Shandong Province (no. ZR2018MA019).

References

[1] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, Berlin, Germany, 2010.

[2] S. Bahmani, B. Raj, and P. Boufounos, "Greedy sparsity-constrained optimization," *The Journal of Machine Learning Research*, vol. 14, pp. 807–841, 2013.

[3] A. Beck and Y. C. Eldar, "Sparsity constrained nonlinear optimization: optimality conditions and algorithms," *SIAM Journal on Optimization*, vol. 23, no. 3, pp. 1480–150, 2013.

[4] A. Beck and N. Hallak, "On the minimization over sparse symmetric sets: projections, optimality conditions, and algorithms," *Mathematics of Operations Research*, vol. 41, no. 1, pp. 196–223, 2016.

[5] Z. Lu, "Optimization over sparse symmetric sets via a nonmonotone projected gradient method," 2015, https://arxiv.org/abs/1509.08581.

[6] L.-L. Pan, N.-H. Xiu, and S.-L. Zhou, "On solutions of sparsity constrained optimization," *Journal of the Operations Research Society of China*, vol. 3, no. 4, pp. 421–439, 2015.

[7] L. Pan, N. Xiu, and S. Zhou, "Gradient support projection algorithm for affine feasibility problem with sparsity and nonnegativity," 2014, https://arxiv.org/abs/1406.7178.

[8] L. Pan, S. Zhou, N. Xiu, and H.-D. Qi, "A convergent iterative hard thresholding for nonnegative sparsity optimization," *Pacific Journal of Optimization*, vol. 13, pp. 325–353, 2017.

[9] S. Negahban, P. Ravikumar, M. Wainwright, and B. Yu, "A united framework for high-dimensional analysis of M-estimators with decomposable regularizes," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, Vancouver, British Columbia, Canada, December 2009.

[10] C. Wang and B. Qu, "Convergence of the gradient projection method with a new stepsize rule," *OR Transactions*, vol. 6, pp. 36–44, 2002.