Research Article

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

Ye Li,1,2 Jun Sun,3 and Biao Qu1

1Qufu Normal University, Rizhao 276826, Shandong, China
2Shandong Women’s University, Jinan 250300, Shandong, China
3Beijing 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

Copyright © 2020 Ye Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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):

$$\min \ f(x)$$

s.t. \ $x \in S \cap R^n_+$

(1)

where $f: R^n \rightarrow R$ is a continuously differential function with a lower bound. $S = \{x \in R^n: \|x\|_0 \leq s\}$ 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 $\cdot \| \cdot$ 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 $\alpha$-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 $\alpha$-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:
Lemma 1

Let \( \alpha > 0 \), vector \( x^* \in S \cap R^p_+ \) be an \( \alpha \)-stationary point if and only if

\[
\nabla_i f(x^*) = 0, \quad i \in I_1(x^*), \\
\geq \alpha M_i(x^*), \quad i \in I_0(x^*).
\]

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

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

\[
T^C_{S \cap R^p_+}(x^*) = \text{span}\{e_i, i \in I_1(x^*)\},
\]

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

\textbf{Step 1.} Initialize \( x^0 \in S \cap R^p_+, 0 \leq \theta \leq 1, \) and \( e > 0 \), and set \( k \leftarrow 0 \).

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

\[
z^{k}(\alpha, \theta) = x^k + \theta(x^k(\alpha) - x^k), \\
x^k(\alpha) = P_{S \cap R^p_+}(x^k - \alpha \nabla f(x^k)).
\]

\textbf{Step 3.} Compute \( x^{k+1} = P_{S \cap R^p_+}(x^k - a_k \nabla f(x^k)) \) where \( a_k \) satisfies \( 0 \leq a_k \leq (1/3L_k) \).

\textbf{Step 4.} If \( \| \nabla f(x^k) \| \leq e \), then stop; otherwise, set \( k \leftarrow k + 1 \) 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 \( S \cap R^p_+ \)

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

\[
f(x^k) \leq h_k(x^k, x^{k+1}),
\]

where

\[
h_k(x^k, x^{k+1}) = f(x^{k+1}) + \langle \nabla f(x^{k+1}), x^k - x^{k+1} \rangle + \frac{(l_k/2)}{\| x^k - x^{k+1} \|^2}.
\]

Proof. Let

\[
g(t) = f(x^{k+1} + t(x^k - x^{k+1})).
\]

Then,
We suppose that (11) is tenable.

Thus,

\[
g(0) = f(x^{k+1}),
\]

\[
g(1) = f(x^k),
\]

\[
g'(t) = (x^k - x^{k+1})^T \nabla f(x^{k+1} + t(x^k - x^{k+1})).
\]

Then, (11) is tenable.

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

\[
x^{k+1} \in \arg\min_{x \in S \cap \mathbb{R}_+^n} \| x - (x^k - \frac{1}{l} \nabla f(x^k)) \|^2.
\]

Moreover,

\[
f(x^k) - f(x^{k+1}) \geq \sigma \| x^k - x^{k+1} \|^2,
\]

where \( \sigma = (1 - l_k/2) \).
Let the sequence \( x^{k+1} \) be generated by Algorithm 1. Then,

1. \( f(x^k) - f(x^{k+1}) \geq \frac{1}{2} \| x - x^k \|^2 - \frac{1}{2} \| x - x^k \|^2, \)

2. \( f(x^k) \) is an increasing sequence, and when \( k \to \infty \), \( f(x^k) \) converges.
3. \( \| x^k - x^{k+1} \| \to 0 \)
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

\[
\frac{1}{\alpha_k} \geq 3L_k = l.
\]

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 \( \{ H_k \} \), we can get that \( \{ f(x^k) \} \) converges.

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

\[
f(x^k) - f(x^{k+1}) \geq \frac{1}{2} \| x - x^k \|^2.
\]

Summing over both sides of this inequality, we get

\[
\sum_{k=1}^{\infty} \| x^k - x^{k+1} \|^2 \leq \sum_{k=1}^{\infty} \frac{2}{\mu} (f(x^k) - f(x^{k+1})).
\]

(25)

\[
= \frac{2}{\mu} \left(f(x^0) - \lim_{k \to \infty} f(x^k)\right).
\]

Since \( f \) is bounded below, we get

\[
\| x^k - x^{k+1} \| \to 0.
\]

(26)

(4) It easily can be got by (2).

\[\square\]

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

\[
\| \nabla f(x^{k+1}) - \nabla f(x^k) \| \leq l_k \| x^{k+1} - x^k \|.
\]

(27)

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

\[
\| \nabla f(x^{k+1}) - \nabla f(x^k) \| \leq l_k \| x^{k+1} - x^k \|.
\]

(28)

\[\square\]

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 \( \alpha \)-stationary point.

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

\[
\lim_{k \to \infty} \| \nabla f(x^k) \| = 0.
\]

(29)

Proof

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

Because

\[
\| x^k \| - \| x^{k^*} \| \leq \| x^{k^*} - x^k \| = \| x^{k^*} - x^k + x^k \| \leq \| x^k \| + \| x^{k^*} - x^k \|.
\]

(30)

we get

\[
\lim_{k \to \infty} x^{k^*} = \lim_{k \to \infty} x^k = x^*.
\]

(31)

Moreover,

\[
x^{k^*} = P_{\mathcal{P}(x)}(x^k - \alpha_k \nabla f(x^k)),
\]

(32)

We consider the next two cases:

\[\square\]
Case 1. For \( i \in I_1(x^*) \), there must exist a sufficiently large index \( N \) and a constant \( c_0 > 0 \) such that
\[
\min \{ x_{i,N}^k, x_{i,N}^{k+1} \} \geq c_0 > 0.
\] (33)

By \( \mathcal{P}_{S; K}^e = \mathcal{P}_S(\mathcal{P}_K) \) and (33), we get
\[
x_{i,N}^{k+1} = x_i^k - \alpha_{x^k} \nabla f(x^k).
\] (34)

Since
\[
\lim_{n \to \infty} \inf \alpha_{x_n} > 0,
\] (35)

without loss of generality, we can suppose \( \lim_{x \to \infty} \alpha_{x_n} = c. \) Let \( n \to + \infty \). We get
\[
x_i^* = x_i^k - c \nabla f(x^*),
\] (36)
i.e.,
\[
\nabla_i f(x^*) = 0, \quad \forall i \in I_1(x^*).
\] (37)

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

Subcase 1. When \( \|x^*\| = s \), we get
\[
0 = x^* = \lim_{n \to \infty} x_{i,N}^{k+1} = \mathcal{P}_S(\mathcal{P}_K(\mathcal{P}_e \mathcal{P}_K(\mathcal{P}_e(x^k - \alpha_{x^k} \nabla f(x^k))))).
\] (38)

Due to the property of the projections \( \mathcal{P}_S \) and \( \mathcal{P}_K \), we have
\[
\max \{ x_{i,N}^k - \alpha_{x^k} \nabla f(x^k), 0 \} \leq M_s(x^*).
\] (39)

Thus,
\[
x_i^* = x_i^k - \alpha_{x^k} \nabla f(x^*) \leq M_s(x^*).
\] (40)

Taking limits on both sides, we obtain
\[
\nabla_i f(x^*) \geq -\frac{1}{c} M_s(x^*).
\] (41)

Subcase 2. When \( \|x^*\| < s \), suppose \( \nabla_i f(x^*) < 0 \), and we have
\[
\lim_{n \to \infty} (x_{i,N}^k - \alpha_{x^k} \nabla f(x^k)) = -c \nabla_i f(x^*) > 0.
\] (42)

For all sufficiently large \( n \), we have
\[
\mathcal{P}_K(\mathcal{P}_e \mathcal{P}_K(\mathcal{P}_e(x^k - \alpha_{x^k} \nabla f(x^k)))) = x_i^k - \alpha_{x^k} \nabla f(x^k),
\] (43)

Since \( \|x^*\| < s \), for all sufficiently large \( n \), we have
\[
x_i^{k+1} = \mathcal{P}_S(\mathcal{P}_K(\mathcal{P}_e \mathcal{P}_K(\mathcal{P}_e(x^k - \alpha_{x^k} \nabla f(x^k))))),
\] (44)

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

Summarizing the two cases, we obtain
\[
\nabla_i f(x^*) = \begin{cases} 0, & i \in I_1(x^*), \\ \geq -\frac{1}{c} M_s(x^*), & i \in I_0(x^*). \end{cases}
\] (45)

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

(2) Set \( \Gamma^k = I_1(x^k) \). By Lemma 3, we have
\[
\mathcal{T}_{\mathcal{S} \cap \mathcal{R}^e}(x^k) = \mathcal{P}_K(\mathcal{P}_e(x^k), = \text{span}\{e_i, i \in I_1(x^k)\}
\] (46)

By Definition 4, we have
\[
\left\| \mathcal{P}_{\mathcal{S} \cap \mathcal{R}^e}(-\nabla f(x^k)) \right\| = \max \left\{ \langle -\nabla f(x^k), v \rangle, \|v\| = 1 \right\},
\] (47)

Moreover, the maximum value is taken at \( \|v\| = 1 \). For any \( \varepsilon > 0 \), there exists \( v^\varepsilon \in \mathcal{R}^e \) and \( \|v^\varepsilon\| = 1 \) satisfies
\[
\left\| \nabla_i f(x^*) \right\| \leq \langle -\nabla f(x^k), v^\varepsilon \rangle + \varepsilon.
\] (48)

Because \( x^{k+1} = \mathcal{P}_S(\mathcal{P}_K(\mathcal{P}_e(x^k - \alpha_{x^k} \nabla f(x^k)))) \) and \( x_{i,k+1} = y_{i,k+1}, \)
\( x \in \mathcal{P}_S \cap \mathcal{R}^e \) (y), we get
\[
x_i^{k+1} = (x^k - \alpha_{x^k} \nabla f(x^k))_{i,k+1},
\] (49)
i.e.,
\[
x_i^{k+1} - (x^k - \alpha_{x^k} \nabla f(x^k))_{i,k+1} = 0.
\] (50)

Thus, for any \( \omega_{k+1} \in \mathcal{R}^e \), we get
\[
\langle x^{k+1} - (x^k - \alpha_{x^k} \nabla f(x^k)), \omega_{k+1} - x_{k+1} \rangle = 0.
\] (51)

Taking \( \omega_{k+1} = x_{k+1} + v_{k+1} \), we get
\[
\langle x^{k+1} - (x^k - \alpha_{x^k} \nabla f(x^k)), -v_{k+1} \rangle = 0.
\] (52)

By the Cauchy–Schwarz inequality, we get
\[
\langle \alpha_{x^k} \nabla f(x^k), -v_{k+1} \rangle = \langle x^{k+1} - x^k, v_{k+1} \rangle \leq \left\| x^{k+1} - x^k \right\|.
\] (53)
i.e.,
\[
-\langle \nabla f(x^k), v_{k+1} \rangle \leq \frac{\|x^{k+1} - x^k\|}{\alpha_{x^k}}.
\] (54)

By Lemma 7, we get
\[-\langle \nabla f(x^{k+1}), x^{k+1} \rangle = -\langle \nabla f(x^k), x^k \rangle \]
\[\leq I_k \| x^{k+1} - x^k \| + \frac{\| x^{k+1} - x^k \|}{\alpha_k}. \]  
(55)

Taking limits on both sides and using Lemma 6, we have
\[\lim_{k \to +\infty} \sup (-\langle \nabla f(x^k), x^{k+1} \rangle) \leq 0. \]  
(56)

By (32), we get
\[\lim_{k \to +\infty} \| \nabla f(x^k) \| = 0. \]  
(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-RSC, 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 all \(x \in S \cap R^n_+\), we have \(|I_1(x(x^*))| = |I_1(x) \cup I_1(x^*)| \leq 2s\). Since \(f(x)\) is 2s-RSC, by Definition 3, we have
\[f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle, \]
\[= f(x^*) + \sum_{i \in I_1(x^*)} \nabla i f(x^*) (x_i - x_i^*), \]
\[+ \sum_{i \in I_1(x^*)} \nabla i f(x^*) (x_i - x_i^*). \]  
(58)

Because \(x^*\) is an accumulation point of the sequence \{x^k\}. By Theorem 1, \(x^*\) is an \(\alpha\)-stationary. By Lemma 1, we can get
\[f(x) \geq f(x^*). \]  
(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
\[\|x^k - x^*\| \leq \delta. \]  
(60)

For any \(i \in I_1(x^*)\), we have
\[x^k_i = x^*_i - (x^* - x^k) \geq x^*_i - (x^* - x^k) > x^*_i - \delta > 0. \]  
(61)

Thus,
\[I_1(x^*) \supset I_1(x^k). \]  
(62)

By \(\|x^k\|_0 = s\) and \(|I_1(x^*)| = \|x^*\|_0 = s\), we have
\[I_1(x^*) = I_1(x^k). \]  
(63)

For any \(x^k \in S \cap R^n_+\) satisfying \(\|x^k - x^*\| \leq \delta\), we have
\[|I_1(x^k(x^*))| = |I_1(x^k) \cup I_1(x^*)| \leq 2s. \]  
Since \(f(x)\) is 2s-RSC, by Definition 3, Theorem 1, and Lemma 1, we have
\[f(x^k) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle, \]
\[= f(x^*) + \sum_{i \in I_1(x^*)} \nabla f(x^*) (x_i^* - x_i^*), \]
\[\geq f(x^*). \]  
(64)

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

**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 \(I_{2s}\) and \(\|x^*\|_0 = s\), for all sufficiently large \(k\), and we have
\[\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2, \quad 0 < \rho < 1, \]  
(65)

where \(\rho = 1 - (-2L_k\alpha_k/L_k) + 2L_k^2\alpha_k^2\).

**Proof.** By Theorem 2, we get \(x^k \to x^*\). As \(f(x)\) is 2s-RSC with parameter \(I_{2s}\), for any \(x, y \in R^n\) and \(|I_1(xy)| \leq 2s\), we have
\[\| \nabla f(x) - \nabla f(y) \|_{I_1(xy)} \geq L_k \| x - y \|. \]  
(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
\[\| \nabla f(x^*) \| = \lim_{k \to +\infty} \| \nabla f(x^k) \| = 0, \]
(67)

For all sufficiently large \(k\), we have
\[\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \frac{2\rho \alpha_k}{L_k} \| \nabla f(x^*) \|_{\Gamma^*}^2, \]
(68)

Because \(\|x^k - x^*\| \geq L_k \| \nabla f(x^k) - \nabla f(x^*) \|\), we get
\[\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \left(\frac{2\rho \alpha_k}{L_k} \right)^2 \| \nabla f(x^*) \|_{\Gamma^*}^2, \]
(69)

Since \(0 \leq \alpha_k \leq (1/3L_k)\) and \(\sigma = (1 - L_k/2) = (1 - L_k/2), \) we have \(0 \leq \alpha_k \leq (1/2\sigma + 3L_k)\). Thus,
\[\beta \left(\frac{2\sigma}{2\sigma + 3L_k}\right) \| x^k - x^* \|^2 \leq \inf\alpha_k \leq \frac{1}{2\sigma + 3L_k}. \]  
(70)
Setting \( \alpha_* = (\beta/2\sigma + 3L_k) \), we get
\[
\alpha_* \leq \alpha_k \leq \frac{1}{3L_k}.
\] (71)

Thus,
\[
1 - \frac{2L_k^2}{L_k} + \frac{1}{2} \alpha_k^2 = 1 + \frac{1}{2} \left( \frac{1}{L_k} \right)^2 - \frac{1}{2L_k},
\]
\[
\leq 1 + \frac{1}{2} \left( \frac{1}{L_k} \right)^2 - \frac{1}{2L_k},
\]
\[
= 1 - \frac{2}{L_k} \frac{\alpha_*}{L_k} + \frac{1}{2} \alpha_*^2,
\]
\[
= \rho^2.
\] (72)

By \( L_k \leq L_k \) and \( \rho^2 = 1 + \frac{1}{2} \left( \alpha_* - \frac{1}{L_k} \right)^2 - \frac{1}{2L_k} \), we get \( \rho > 0 \).

From \( \beta < 1 \) and \( \rho^2 = 1 - \frac{2L_k^2}{L_k} \alpha_* + \frac{1}{2} \alpha_*^2 \), we have \( \rho < 1 \).

Thus,
\[
\lim_{k \to \infty} \left\| x_{k+1} - x_* \right\| \leq \sqrt{\rho},
\] (73)

where \( 0 < \sqrt{\rho} < 1 \). Thus, the sequence \( \{x_k\} \) is Q-linear convergence to \( x^* \).

\section*{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.

\section*{Data Availability}

No data were used to support this study.

\section*{Conflicts of Interest}

The authors declare that they have no conflicts of interest.

\section*{Authors’ Contributions}

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

\section*{Acknowledgments}

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

\begin{thebibliography}{10}
\end{thebibliography}