WHIOCE PUBLISHING PTE. LTD.
PROVIDING FIRST-CLASS SCIENTIFIC INFORMATION FOR TOP SCHOLARS

2025 Volume 3, Issue 8
Lecture Notes in Education, Arts, Management and Social Science
ISSN(Online): 2705-053X

A New Projected Gradient Method for Unconstrained Problems

Guoyan Liu
Jinjiang College, Jiangsu University, Zhenjiang 212100, Jiangsu, China

Abstract: In this paper, we propose a new gradient projection method for problems without constraints. Based on the steepest descent method, there is a projection matrix P constructed from the descending direction and the iterative direction [1–4]. We gave a general proof of the convergence for the new projected gradient method in this paper. In numerical experiments using the CUTEr test problem library, the new projected gradient method performs better than the classic fastest descent method.

Keywords: Projection matrix; Global convergence; Unconstrained optimization

Online publication: September 26, 2025

1. Introduction

Rosen's gradient projection method, published in 1960, is the first method for solving constrained nonlinear programming problems in the history of mathematical programming. The importance of this method in the literature of nonlinear programming stems from the fact that many more efficient algorithms developed later incorporated the basic ideas propounded by Rosen [5–8].

Fletcher and Reeves (1964) extended this method to solve nonlinear optimization problems [9].

$\min f(x), x \in R^n,$

Where f is continuously differentiable. Gradient methods start from an initial point x_0 and generate new iterates by the rule

$x_{k+1} = x_k - \alpha_k g_k,$

Where $g_k = \nabla f(x_k)^T$ is the gradient, and α_k is a stepsize computed by some line search algorithm. In this paper, we choose the α_k by Wolfe line search.

In the steepest descent (SD) method, which can be traced back to Cauchy (1847) [10,11], the 'exact step-length' is given by

$$\alpha_k \in \arg\min_{\alpha \in R} f(x_k - \alpha g_k),$$

It is well known that SD can be very slow when the Hessian of f is ill-conditioned at a local minimum [12]. In this case, the iterates slowly approach the minimum in a zigzag fashion.

Until a new idea is proposed by Barzilai and Borwein in 1988 [5],the basic idea is to regard the matrix $D(\alpha_k) = {}^{1}\!/_{\alpha_k} I$ as an approximation of the Hessian of f.

$$\alpha_k = \arg\min_{\alpha \in R} \left\| D(\alpha) s_{k-1} - g_{k-1} \right\|_2$$

Where $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = g_k - g_{k-1}$

Inspired by these algorithms, we propose a new algorithm—a new unconstrained projected gradient algorithm [13–15].

2. Unconstrained projection gradient method

Since the gradient projection method is based on the projection of the gradient on a subspace, some of its properties will be described here for convenience.

Definition 2.1: If a matrix $P_{n\times n}$ satisfies both equations of $P=P^T$ and $P^2=P$, it is defined as a projection matrix [16–18].

Properties of a projection matrix:

1. It is semi-definite;

2. Any matrix P is a projection matrix, and $I–P$ is also a projection matrix, which is the necessary and sufficient condition[19].

3. If matrix P is a projection matrix, let $Q = I–P$

$L(x) = \{Px|x \in R^n\}$, $L^{\perp}(x)=\{Qx|x \in R^n\}$

The linear subspaces $L(x)$ and $L^{\perp}(x)$ are orthogonal, so that the projection matrix P takes any vector $x \in R^n$ into orthogonal subspaces.

$x = p + q, p \in L, q \in L^{\perp}$.

In this paper, we constructed a projection matrix P as follows:

$P = (y^{k-1}, \delta^{k-1})[(y^{k-1}, \delta^{k-1})^T(y^{k-1}, \delta^{k-1})]^{-1}(y^{k-1}, \delta^{k-1})^T$

Where $\delta^{k-1} = x^k - x^{k-1}$, $y^{k-1} = g^k - g^{k-1}$, when they are non-collinear.

When δ^{k-1} and y^{k-1} are collinear, let $P=I$ that is a unit array [20,21].

2.1. Algorithm (the new projection gradient method)

Step1. Choose x_0, $\varepsilon > 0$, set $d^0 = -g^0$, $\alpha^0 = 1$, $x_1 = x^0 + d^0$, k:=0;

Step2. stop if x_k satisfies the stopping criterion of $\|g^k\| \le \varepsilon$; or turn step3.

Step3. Compute $\delta^{k-1} = x^k - x^{k-1}$, $y^{k-1} = g^k - g^{k-1}$ and $\cos < \delta^{k-1}, y^{k-1} >$;

Step4. If $\cos < \delta^{k-1}, y^{k-1} >= \pm 1$, set $P=I$;

or else $P = (y^{k-1}, \delta^{k-1})[(y^{k-1}, \delta^{k-1})^T(y^{k-1}, \delta^{k-1})]^{-1}(y^{k-1}, \delta^{k-1})^T$;

Step5. set $d^k = -Pg^k$, choose α^k to satisfy the Armijo-Goldstein-*Wolfe* conditions:

$f(x^k + \alpha^k d^k) \le f(x^k) + \sigma_1 \alpha^k g_k^T d^k$,

$g(x^k + \alpha^k d^k)^T d^k \ge \sigma_2 g_k^T d^k$;

Step6. $k:= k + 1$ reset $x^{k+1} = x^k + \alpha^k d^k$, continue with step1 [22,23].

2.2. Proof of the convergence

Before we prove the convergence for the new projection gradient method, we assumed as follows:

1. Choose the starting point x^0, the level set $L(x_0) = \{x \in R^n : f(x) \le f(x^0)\}$ is bounded.

2. Derivative of f is Lipschitz continuously differentiable in a neighborhood of the level set $L(x_0)$

3. let x^* denotes the unique minimum point in the level set $L(x^0)$.

4.$f(x)$ is uniformly convex and twice continuously differentiable, there exist positive constants m and M with the following property [24]:

$$m\|u\|^2 \le u^T G(x) u \le M\|u\|^2 \quad \forall x \in L(x^0), u \in R^n \tag{2.1}$$

$$\frac{m}{2}\|x-x^*\|^2 < f(x) - f(x^*) < \frac{M}{2}\|x-x^*\|^2 \tag{2.2}$$

$$m\|x-x^*\| < g(x) < M\|x-x^*\| \;\forall x \in R^n \tag{2.3}$$

Lemma 2.1: since the function $f(x)$ is satisfied with assumption 4, then

$$\frac{\|y^{k-1}\|}{\|\delta^{k-1}\|} \le M \tag{2.4}$$

$$\frac{\|\delta^{k-1}\|^2}{y^{k-1T}\delta^{k-1}} \le \frac{1}{m}$$

Proof: from assumption 4, we can know

$$\frac{d}{d\tau} g\left(x^{k-1} + \tau\delta^{k-1}\right) = G\left(x^{k-1} + \tau\delta^{k-1}\right)\delta^{k-1}$$

Use the integration by this equality,

$$y^{k-1} = g^k - g^{k-1} = \int_0^1 G(x^{k-1} + \tau\delta^{k-1})\delta^{k-1} d\tau$$

Then $\|y^{k-1}\| \le \|\delta^{k-1}\| \int_0^1 \left\|G\left(x^{k-1} + \tau\delta^{k-1}\right)\delta^{k-1}\right\| d\tau$

By $L(x^0)$ being a bounded closed convex set, $x^{k-1} + \tau\delta^{k-1} \in L\left(x^0\right)$, and $f(x)$ satisfies assumption 4, there must be $M > 0$, such that $\left\|G\left(x^{k-1} + \tau\delta^{k-1}\right)\right\| \le M$

Then we can have $\|y^{k-1}\| \le M\|\delta^{k-1}\|$ $\|y^{k-1}\| / \|\delta^{k-1}\| \le M$

Combine all of the above

$$\delta^{k-1T} y^{k-1} = \int_0^1 \delta^{k-1T} G\left(x^{k-1} + \tau\delta^{k-1}\right)\delta^{k-1} d\tau \ge m\|\delta^{k-1}\|^2 \tag{2.5}$$

namely $\dfrac{\|\delta^{k-1}\|^2}{y^{k-1T}\delta^{k-1}} \le \dfrac{1}{m}$

Lemma 2.2: If $f(x)$ is satisfied with assumption 4 [25–28], then

$$\|g(x)\|^2 \ge m\left[f(x) - f(x^*)\right], \forall x \in R^n \tag{2.6}$$

Proof: Since $f(x)$ is a convex function, then for any point $x \in R^n$

$$f(x) - f(x^*) \le g(x)^T\left(x - x^*\right) \le \|g(x)\|\|x - x^*\| \tag{2.7}$$

$$\frac{d}{d\tau} g\left(x^* + \tau\left(x - x^*\right)\right) = G(x^* + \tau\left(x - x^*\right))\left(x - x^*\right)$$

Use the integration by parts, then

$$g(x)-g(x^*)=\int_0^1 G\left(x^*+\tau\left(x-x^*\right)\right)\left(x-x^*\right)d\tau$$

we denote $\overline{G}=\int_0^1 G\left(x^*+\tau\left(x-x^*\right)\right)d\tau$

and point x^* is the minimum point of $f(x)$.
Since $\overline{G}$ is satisfied with formula (2.2), so

$$m\left\|x-x^*\right\|^2 \le \left(x-x^*\right)^T \overline{G}\left(x-x^*\right)=\left(x-x^*\right)^T g(x) \le \left\|x-x^*\right\|.\left\|g(x)\right\|$$

Thus, combine $\left\|x-x^*\right\| \le \left\|g(x)\right\|/m$ and formula (2.7), then formula (2.6) holds.

Lemma 2.3: Let $f(x)$ be satisfied by assumption 4, α^k is satisfied by *Wolfe* line search

$$f(x^k+\alpha^k d^k) \le f(x^k)+\sigma_1\alpha^k g(x^k)^T d^k \quad \sigma_1 \in \left(0,\frac{1}{2}\right) \tag{2.8}$$

$$g(x^k+\alpha^k d^k)^T d^k \ge \sigma_2 g(x^k)^T d^k, \sigma_2 \in \left(\sigma_1,1\right) \tag{2.9}$$

$$x^{k+1}=x^k+\alpha^k d^k \tag{2.10}$$

Then

$$f(x^k)-f(x^{k-1}) \le -\sigma_1 m c_1 \cos^2\theta_{k-1}\left(f(x^k)-f(x^*)\right) \tag{2.11}$$

$$f(x^k)-f(x^*) \le \left(1-\sigma_1 m c_1 \cos^2\theta_{k-1}\right)\left(f(x^k)-f(x^*)\right) \tag{2.12}$$

Where $c_1=\left(1-\sigma_2\right)/M \quad c_2=2\left(1-\sigma_1\right)/m$, θ_{k-1} is the Angle between d^k and $-g^k$.

Proof: divided into the following steps:
(1) Prove the inequality of formula 1.8 at the left side [29–31],
From Lemma 2.1, we know

$$\frac{y^{k-1T}\delta^{k-1}}{\left\|\delta^{k-1}\right\|^2} \le \frac{\left\|y^{k-1}\right\|.\left\|\delta^{k-1}\right\|}{\left\|\delta^{k-1}\right\|^2}=\frac{\left\|y^{k-1}\right\|}{\left\|\delta^{k-1}\right\|} \le M \tag{2.13}$$

By Wolfe Powell (2.9), then

$$y^{k-1T}\delta^{k-1}=g^{kT}\delta^{k-1}-g^{k-1T}\delta^{k-1} \ge -\left(1-\sigma_2\right)g^{k-1T}\delta^{k-1} \tag{2.14}$$

combine the formulas (2.13) and (2.14), which are hold below

$$\left\|\delta^{k-1}\right\|^2 \ge \frac{y^{k-1T}\delta^{k-1}}{M} \ge -\frac{\left(1-\sigma_2\right)}{M}g^{k-1T}\delta^{k-1}=\frac{1-\sigma_2}{M}\left\|g^{k-1}\right\|.\left\|\delta^{k-1}\right\|\cos\theta_{k-1}$$

So $\frac{1-\sigma_2}{M}\left\|g^{k-1}\right\|\cos\theta_{k-1} \le \left\|\delta^{k-1}\right\|$

(2) proof of the right-hand inequality of formula (2.2).
According to Taylor's expansion and the formula (2.11) of Wolff's Powell,

$$g^{k-1T}\delta^{k-1}+\frac{1}{2}\delta^{k-1T}G(\xi^{k-1})\delta^{k-1}=f(x^k)-f\left(x^{k=1}\right) \le \sigma_1 g^{k-1T}\delta^{k-1}$$

Where ξ^{k-1} is between x^k and x^{k-1}. so

$$\delta^{k-1T}G\left(\xi^{k-1}\right)\delta^{k-1} \le -2\left(1-\sigma_1\right)g^{k-1T}\delta^{k-1}$$

For the set $L(x^0)$ is bounded close convex sets, so$\xi^{k-1} \in L(x^0)$, from formula (2.1), it is clear

$$m\left\|\delta^{k-1}\right\|^2 \le \delta^{k-1T}G\left(\xi^{k-1}\right)\delta^{k-1}$$

Therefore, $m\left\|\delta^{k-1}\right\|^2 \le -2\left(1-\sigma_1\right)g^{k-1T}\delta^{k-1} = 2\left(1-\sigma_1\right)\left\|g^{k-1}\right\|.\left\|\delta^{k-1}\right\|\cos\theta_{k-1}$

So, $\left\|\delta^{k-1}\right\| \le \dfrac{2\left(1-\sigma_1\right)}{m}\left\|g^{k-1}\right\|\cos\theta_{k-1}$.

(3) proof for formula (2.11) and (2.12)

By the proof of Wolff's Powell and above all, it is clear that

$$f(x^k)-f(x^{k-1}) \le \sigma_1 g^{k-1T}\delta^{k-1} = -\sigma_1\left\|g^{k-1}\right\|.\left\|\delta^{k-1}\right\|\cos\theta_{k-1} \le -\sigma_1 c_1\left\|g^{k-1}\right\|^2\cos^2\theta_{k-1} \tag{2.15}$$

By Lemma 2.2

$$\left\|g^{k-1}\right\|^2 \ge m\left[f(x^{k-1})-f(x^*)\right]$$

Substituting into equation (2.15), it holds that

$$f(x^k)-f(x^{k-1}) \le -\sigma_1 m c_1\cos^2\theta_{k-1}\left[f(x^{k-1})-f(x^*)\right]$$

We have the required result [32–35].

The global convergence of the new projection gradient method.

Theorem 2.1: let x^0 is the initial point, and the $f(x)$ is satisfied by assumption condition 1 on the level set $L(x^0)$, then the iterated point series $\{x^k\}$ generated by the new algorithm converges to the optimal solution x^*.

Proof: Proof is divided into two steps

1. Assume for any arbitrary small $\varepsilon > 0$, there is an integer k_0 when $k\to\infty$ for $k>k_0$ such that $\cos\theta_k<\varepsilon$

Proof by contradiction:

We assume $\cos\theta_{k-1}\to 0$ when $\|g^k\|>\varepsilon$, $\|d^k\|<\varepsilon$

It holds that $g^{k^T}Pg^k\to 0$ when $\cos\theta_{k-1} = \dfrac{-g^{kT}d^k}{\left\|g^k\right\|.\left\|d^k\right\|} = \dfrac{g^{kT}Pg^k}{\left\|g^k\right\|\left\|d^k\right\|} \to 0$

thus the matrix P is singular$\|g^k\|>\varepsilon$ $k\to\infty$, and thus δ^{k-1} and y^{k-1} is linearly dependent on the iterative direction when $k>k_0$, so

$$\alpha^{k+1}d^{k+1} = \delta^{k+1} = l(g^{k+2}-g^{k+1}) = -\alpha^k g^{k+1}$$

then $g^{k+2}=0$

above all, we get $\cos\theta_{k-1}\to 0$ $g^k\to 0$, its local optimal solution is the global optimal solution of $f(x)$ when it is convex.

2. If $\cos\theta_{k-1}>\eta$ is true for any $\eta>0$ as , it is true that

$$f(x^k)-f(x^*) \le \left(1-\sigma_1 m c_1\cos^2\theta_{k-1}\right)\left(f(x^k)-f(x^*)\right)$$

for the sequence $\{f(x^k)\}$ is a decreasing monotone column.

Let $c_3 = 1-\sigma_1 c_1 m\eta^2$ then $0<c_3<1$

$$0 < f(x^k)-f\left(x^*\right) \le c_3\left(f(x^{k-1})-f(x^*)\right) \le c_3^2\left(f(x^{k=2})-f(x^*)\right)\cdots \le c_3^k\left(f(x^0)-f(x^*)\right)$$

as $k\to\infty$ $f(x^k)\to f(x^*)$ $x^k \to x^*$ the minimum point is unique.

2.3. Rate of convergence

Hypothesis 2.1: let $\lambda_{\min}$ is the minimal eigenvalue of the projection matrix P and $\lambda_{\min}<L$.

Lemma 2.4: for an arbitrary integer $k > 0, \left\|d^k\right\|^2 \leq \left\|g^k\right\|^2$， $g^{k^T} d^k \geq \lambda_{\min} \left\|g^k\right\|^2$.

Theorem 2.2: Under the assumption that conditions 1 and 2 are true, the iterated points series $\{x^k\}$ by algorithm 2.1 converges to x^* at least linearly [36].

Proof: we choose α^k to satisfy the below

$$f\left(x^k+\alpha^k d^k\right) \leq f(x^k)+\sigma_1 \alpha^k g^{kT} d^k \text{ , } g\left(x^k+\alpha^k d^k\right)^T d^k \geq \sigma_2 g^{kT} d^k$$

So,

$$f^k - f^{k+1} \geq -\sigma_1 \alpha^k g^{kT} d^k \geq \sigma_1 \alpha^k \lambda \min \left\|g^k\right\|^2 \qquad (2.19)$$

$$\left\|g^{k+1}-g^k\right\| = \frac{\left\|g^{k+1}-g^k\right\| \cdot \left\|d^k\right\|}{\left\|d^k\right\|} \geq \frac{\left(g^{k+1}-g^k\right)^T d^k}{\left\|d^k\right\|} \geq \frac{\left(\sigma_2-1\right) g_k^T d^k}{\left\|d^k\right\|} \geq \left(1-\sigma_2\right) \lambda_{\min} \left\|g^k\right\|$$

Obtained by formula (1.33)

$$L\alpha^k \left\|d^k\right\| \geq \left\|g^{k+1}-g^k\right\| \geq \left(1-\sigma_2\right) \lambda_{\min} \left\|g^k\right\|$$

$$\alpha^k \geq \frac{\left(1-\sigma_2\right) \lambda_{\min} \left\|g^k\right\|}{L\left\|d^k\right\|} \geq \frac{\left(1-\sigma_2\right) \lambda_{\min}}{L}$$

By formula (2.19) $f^k - f^{k+1} \geq -\sigma_1 \alpha^k g^{kT} d^k \geq \sigma_1 \alpha^k \lambda \min \left\|g^k\right\|^2 \geq \frac{\sigma_1 \left(1\text{-}\sigma_2\right) \lambda_{\min}^2}{L} \cdot \left\|g^k\right\|^2$

let $\frac{\sigma_1\left(1-\sigma_2\right) \lambda_{\min}^2}{L} = \zeta$ $(0<\zeta<1)$.

So the above formula can be resolved to:

$$f^k - f^{k+1} \geq \zeta \left\|g^k\right\|^2 \geq \zeta m^2 \left\|x^k - x^*\right\|^2 \geq \frac{2\zeta m^2}{M}\left(f^k - f^*\right) \geq \frac{\zeta m^2}{M+m^2}\left(f^k - f^*\right)$$

Let $\gamma = \frac{\zeta m^2}{M+m^2}$ $(0<\gamma<1)$

$$f^{k+1} - f^* \leq \left(1-\gamma\right)\left(f^k - f^*\right) \leq \left(1-\gamma\right)^2 \left(f^{k-1} - f^*\right) \leq \cdots \leq \left(1-\gamma\right)^k \left(f^1 - f^*\right)$$

$$\frac{m}{2}\left\|x^{k+1} - x^*\right\| \leq \left(1-\gamma\right)^k \left(f^1 - f^*\right)$$

The main content of this paper is shown above, we constructed a projection matrix P by the thought of the basic gradient projection method and the steepest descent method, it is an extension of the original projection gradient method. Under certain assumptions, we have proved the convergence and linear convergence rate of the new algorithm. the numerical results are verified in this paper. Through the obvious numerical values, the new algorithm is expected to be more convincing by comparison with the classic steepest descent method [37].

3. Numerical experiment

According to the characteristics of the algorithm proposed in this paper, we have done some numerical experiments and compared it with the classic fastest descent method. All the results are obtained in MATLAB 7.1.0, the selected test function is from the literature, and the running environment is an ASUS computer [20] N82J, 2.53vHz, 320GB HDD. The required gradient accuracy is 10e-006, which is taken when the step size is selected σ_1=0.1 σ_2=0.5.

Table 1. The number of iteration steps for two algorithms at different latitudes

Serial number	Dimension	New projection gradient method	Classic fastest descent
1	2	1	3
2	2	1	1
3	3	3	13
4	3	47	60
5	4	698	786
6	5	2609	2632
7	4	186	204
8	3	9	16
9	5	1565	1563
10	10	699	729
11	20	1202	1730
12	50	3002	5320
13	100	4538	10087
14	120	5385	10156

From the above numerical results, we can see that when the dimensions are 2, the speed of the new gradient projection method is approximately 3 times that of the classical fastest descent method. With the increase of the dimension of the test function, the number of iteration steps of the two algorithms is obviously larger, especially when the dimension reaches more than ten, and the iteration exceeds 1,000. However, when the dimension is more than 100, the new projection gradient method is twice as fast as the fastest descending method [38].

The above numerical results are obtained from the same initial point. In short, for low-dimensional test problems, the new projection gradient method proposed in this paper has significant advantages over the classical fastest descent method, and the iteration time is significantly less. As the dimension of the problem increases, the number of iterative steps of the two algorithms increases at the same time, but generally speaking, the new algorithm is better than the classical fastest descent method.

When the test function is selected, the classical method of fastest descent will show that there is no solution or the algorithm is not feasible under the same conditions, but the new projection gradient method rarely shows such a situation. It can be seen that the new projection gradient method can solve more problems than the classical fastest descent method, and the scope of use is also expanded [39,40].

Disclosure statement

The author declares no conflict of interest.

References

[1] Barzilai J, Borwein M, 1988, Two Point Step Gradient Methods. IMA,J.Numer.Anal, 8: 141–148.

[2] Fridlander A, Martinez MJ, Molina B, et al., 1999, Gradient Method with Retards and Generalizations. SIAM J.Numer. Anal., 36: 275–289.

[3] Raydan M, Svaiter BF, 2002, Relaxed Steepest Descent and Cauchy-Barzilai-Borwein Method. Compute Optimal Appl., 5: 167.

[4] Dai YH, Yuan Y, 2003, Alternate Minimization Gradient Method. IMA J.Anal. 23: 377–393.

[5] Dai YH, 2006, The cyclic Barzilai-Borwein Method for Unconstrained Optimization. IMA Journal of Numerical Analysis, 26: 504–627.

[6] Raydan M, 1993, On the Barzilai and Borwein Choice of Steplength for the Gradient Method. IMA J.Numer.Anal., 13: 321–326.

[7] Raydan M, 1997, The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem, SIAM J. Optim., 7(1): 26–33.

[8] Birgin EG, Chambouleyron I, Martinez JM, 1999, Estimation of the Optical Constants and the Thickness of Thin Films Using Unconstrained Optimization, J. Comput.Phys, 151: 862–880.

[9] Birgin EG, Evtushenko YG, 1998, Automatic Differentiation and Spectral Projected Gradient Methods for Optimal Control Problems. Optim Methods Shoftw, 10, 125–146.

[10] Birgin EG, Martinez JM, Mirada, 2000, Nonmonotone Spectral Projected Gradient Methods for Convex Sets. SIAM Journal on Optimization, 10(4): 1196–1211.

[11] Dai YH, Liao LZ, n.d., R-Linear Convergence of the Barzilai and Borwein Gradient Method. Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences Research.

[12] Fletcher R, 1999, Low Storage Methods for Unconstrained Optimization. Lectures in Applied Mathematics (AMS), 26: 165–179.

[13] Friedlander A, Martinez JM, Molina B, et al., 1999, Gradient Method with Retards and Generalization. J. Numer. Anal., 36, 275–289.

[14] Raydan M, n.d., On the Barzilai and Borwein Choice of Step Length for the Gradient Method. J Anal., 13199: 321–326.

[15] Yuan Y, 1991, A Modified BFGS Algorithm for Unconstrained Optimization. IMA J.Numer Anal, 11: 325–332.

[16] Yuan Y, 1993, Numerical Methods for Nonlinear Programming, Shanghai Scientific and Technical Publishers.

[17] Wang W, 2002, Conjugate Hierarchy Algorithm and Its Convergence for Linear Constrained Optimization Problems. Journal of Natural Science of Northeast Normal University, 34(2): 12–15.

[18] Ye L, 2005, A Memory Gradient Rosen Projection Algorithm Combining Conjugate Gradient Parameters for Linear or Nonlinear Constrained Optimization Problems. Journal of Sichuan University, 42(2): 652–660.

[19] Liang Y, Jian J, 2003, A Conjugate Gradient Method for Linear Constrained Optimization. Operations Research and Management, 12(2): 31–35.

[20] Calamai PH, More JJ, 1987, Projected Gradient Methods for Linearly Constrained Problems. Mathematical Programming, 39: 93–116.

[21] Zhang XS, 1985, On the Convergence of Rosen's Gradient Projection Method: Three-Dimensional, Case, Acta Mathematicae Applicate Sinica, 8: 125–128.

[22] Du DZ, Remarks on the Convergence of Rosen's Gradient Projection Method. MSRI Technique Report, 01718–86.

[23] He G, 1987, Proof of Convergence of the Rosen's Gradient Projection Method. Journal of Chengdu University of Science and Technology, 1(1987): 55–68.

[24] Polak E, 1969, On the Convergence of Optimization Algorithms. Revue Francaised matique Infor et de Recherche Operrationelle, 3(16): 17–34.

[25] Wang C, 1982, A Feasible Direction Method for Nonlinear Programming. Acta Mathematica Sinica, 25: 15–19.

[26] Bazaraa MS, Sherry CM, 1979, Nonlinear Programming: Theory and Algorithms, John Wiley & Sons, Inc.

[27] Grippo L, Ampariello FL, 1986, Lucidi A Nonmonotone Line Search Technique For Newton's Method. Journal on Numerical SIAM Analysis, 23: 707–716.

[28] Shi ZJ, 1996, A Class of Globally Convergent Conjugate Projection Gradient Method and Its Superlinear Convergence. Computational Mathematics, 11(4): 411–421.

[29] Zhu Z, 2004, A New Conjugate Projection Gradient Algorithm and Its Superlinear Convergence, Chinese Journal of Applied Mathematics, 27(1): 149–161.

[30] Liu WB, Dai YH, 1999, Minimization Algorithms Based on Supervisor and Searcher Co-Operation.I:--Faster and Robust Gradient Algorithms for Minimization Problems with Stronger Noises. Academy of Mathematic and Systems Sciences, Chinese Academy of Sciences.

[31] Hestenes MR, Stiefel EL, 1952, Methods of Conjugate Gradients for Solving Linear Systems. JRes Nat Bur Standards Sect. 5(49): 409–436.

[32] Fletcher R, Reeves C, 1964, Function Minimization by Conjugate Gradients. Comput, 7: 149–154.

[33] Toint PL, 1981, Towards an Efficient Sparsity Exploiting Newton Method for Minimization. Sparse Matrices and Their Uses, Academic Press London, England, 57–88.

[34] Steihaug T, 1983, The Conjugate Gradient Method and Trust Region in Large Scale Optimization. SIAM Journal on Numerical Analysis, 20(1983): 626–637.

[35] Dimitri P, Bertsekas DP, n.d., Projected Newton Methods For Optimization Problems With Simple Constraints SIAM J.Control and Optimization, 20(2): 2198.

[36] Hestenes MR, Stiefel EL, 1952, Methods of Conjugate Gradients for Solving Linear. JRes Nat Standards Sect, 49(5): 409–436.

[37] Fletcher R, Reeves C, 1964, Function Minimization by Conjugate Gradients. Compute, 7: 149–154.

[38] Bertsekas DP, 1976, On the Goldstein-Levitin-Polyak Gradient Projection Method. IEEE Transactions on Automatic Control, 21: 174–184.

[39] Dai YH, Zhang HC, 2001, Adaptive Two-Point Stepsize Gradient Algorithm. Numerical Algorithms, 27: 377–385.

[40] Zhang HC, Hager WW, 2004, A Nonmonotone Line Search Technique and Its Application to Unconstrained Optimization. SIAM Journal on Optimization, 14: 63–85.