



# Global and local R-linear convergence of a spectral projected gradient method for convex optimization with singular solution

Zhensheng Yu\*, Xinyue Gan

*College of Science, University of Shanghai for Science and Technology, Shanghai, 200093, P. R. China.*

Communicated by Y. J. Cho

---

## Abstract

In this paper, we propose a spectral projected gradient method for the convex optimization problem with singular solution. By solving the equivalent equation of the gradient function, this method combines the perturbed spectral gradient direction with the projection direction to generate the next iteration point. Under some mild conditions, we establish the global convergence and the local R-linear convergence rate under the local error bound condition. Preliminary numerical tests are given to show that the proposed method works well. ©2016 All rights reserved.

*Keywords:* Unconstrained optimization, spectral projected gradient, local error bound, R-linear convergence.

*2010 MSC:* 65K05, 90C30.

---

## 1. Introduction

In this paper, we consider the following unconstrained optimization problem:

$$\min f(x), \quad x \in \mathbb{R}^n, \quad (1.1)$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is continuously differentiable and its gradient denoted by  $\nabla f(x)$ .

At  $k$ th iteration, denote  $s_{k-1} = x_k - x_{k-1}$ ,  $y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1})$ . Quasi-Newton methods for unconstrained optimization [9, 10] obey the recursive iterative process

$$x_{k+1} = x_k + B_k^{-1} \nabla f(x_k).$$

---

\*Corresponding author

*Email address:* [zhsh-yu@163.com](mailto:zhsh-yu@163.com) (Zhensheng Yu)

The sequence of matrices  $\{B_k\}$  satisfy the secant equation

$$B_k s_{k-1} = y_{k-1}.$$

Let  $I$  denote the identical matrix in  $\mathbb{R}^{n \times n}$  and assume that we want a matrix  $B_k$  with a very simple structure that satisfies the secant equation. More precisely, we wish

$$B_k = \lambda_k I \text{ with } \lambda_k \in \mathbb{R}.$$

Then the secant equation becomes:

$$\lambda_k s_{k-1} = y_{k-1}.$$

In general, this equation may not be consistent. However, accepting the least-squares solution that minimizes

$$\|\lambda s_{k-1} - y_{k-1}\|^2,$$

we obtain

$$\lambda_k = \frac{\langle s_{k-1}, y_{k-1} \rangle}{\langle s_{k-1}, s_{k-1} \rangle},$$

here  $\langle \cdot, \cdot \rangle$  denotes the inner product for given two vectors. This formula defines the most popular spectral gradient method for unconstrained optimization with the search direction:

$$d_k = -\frac{1}{\lambda_k} \nabla f(x_k).$$

The method was first introduced by Barzilai and Borwein [2], and the convergence for quadratic functions was established by Raydan [20], and a global scheme was discussed for nonquadratic functions [21], which used a variant of the nonmonotone line search of Grippo et al. [12]. Due to its simplicity and numerical efficiency, the spectral gradient as well as the spectral projected gradient method has been applied successfully to finding local minimizers of large scale problems [3, 4, 7, 11, 14, 15, 19, 27], for more details, see the recent paper [5] and references therein.

Recently, the local convergence analysis of the spectral gradient methods has also been concerned. However, the analysis results are often provided for convex quadratics. For two dimension convex quadratics, Barzilai and Borwein [2] established the R-superlinear convergence of the method. Under a restrictive condition, Molina and Raydan [19] established the Q-linear rate of the (preconditioned) spectral gradient method. The R-linear rate for any dimension strictly convex quadratics was established by Dai and Liao [8]. Under the condition that  $f(x)$  is three times continuously differentiable and the generated sequence  $\{x_k\}$  converges to  $x^*$  with the positive definite Hessian matrix assumption, that is,  $\nabla^2 f(x^*)$  is positive definite, Liu and Dai [18] proved that the R-linear rate of spectral gradient method for general function.

Besides the restricted quadratic or strictly convex condition, to obtain the local convergence rate, one often assumes that the iteration sequence convergence to some  $x^*$  of the solution. Moreover, the solution  $x^*$  is required to be an isolated solution.

To remove the above mentioned restricted conditions and motivated by the projection method for nonlinear equations [23, 24, 27, 28], especially the technique in [27], in this paper, we design a perturbed spectral projected gradient method for problem (1.1). Here the perturbation means that the gradient of the objective function is computed or supplied with some error. This can happen when computing the gradient involves solving a complex subproblem or the problem data is corrupted. One example is the dual ascent method arisen in the Lagrangian dual function of the constrained optimization problem, see Solodov ([22], page 266).

Compared with the existing spectral gradient methods for unconstrained optimization, our method enjoys the following properties:

- The method generates a bounded sequence automatically.

- The strong global convergence is guaranteed.
- The method is R-linearly convergence without the nonsingular assumption.

Note that in our local convergence analysis (see Section 3), we use a local error bound condition. It is worth pointing out that the local error bound condition is a weaker condition than the isolated solution. Using this condition, Yamashita and Fukushima [25], Kanzow, Yamashita and Fukushima [13], Zhou and Toh [28], Wang and Wang [24] considered the local convergence rate of the Levenberg Marquardt method and Newton method for nonlinear equations problems. Zhang, Wu and Zhang [26] considered the trust region method for unconstrained optimization. Li, Fukushima, Qi and Yamashita [16], Li and Li [17] considered the regularization method for convex optimization.

The organization of this paper is as follows. In Section 2, we propose our algorithm and analyze the global convergence. In Section 3, we establish the R-linear convergence of the algorithm under a local error-bound condition. The numerical tests and comparison are given in Section 4. In Section 5, we give the conclusion of this paper.

## 2. Algorithm and global convergence

Combining the perturbed spectral gradient method and the projection technique, we describe the algorithm as follows:

### Algorithm 2.1.

Step 0: Choose an initial point  $x_1 \in \mathbb{R}^n$ , parameters  $\beta, \sigma, \eta \in (0, 1)$ ,  $r > 0$  and set  $\theta_1 = 1$ ,  $k = 1$ .

Step 1: If  $\|\nabla f(x_k)\| = 0$ , then stop.

Step 2: Compute the search direction  $d_k$  by

$$d_k = -\theta_k \nabla f(x_k) + e_k, \quad (2.1)$$

where  $\|e_k\| \leq \eta \theta_k \|\nabla f(x_k)\|$ .

Step 3: Find the trial point  $z_k = x_k + \alpha_k d_k$ , where  $\alpha_k = \beta^{m_k}$  with  $m_k$  being the smallest nonnegative integer  $m$  such that

$$-\langle \nabla f(x_k + \beta^m d_k), d_k \rangle \geq \sigma \beta^m \|d_k\|^2. \quad (2.2)$$

If  $\nabla f(z_k) = 0$ , stop.

Step 4: Compute

$$x_{k+1} = x_k - \zeta_k \nabla f(z_k), \quad (2.3)$$

where

$$\zeta_k = \frac{\langle \nabla f(z_k), x_k - z_k \rangle}{\|\nabla f(z_k)\|^2}. \quad (2.4)$$

Step 5: Compute  $s_k = x_{k+1} - x_k$ ,  $y_k = \nabla f(x_{k+1}) - \nabla f(x_k) + r s_k$  and

$$\theta_{k+1} = \frac{\langle s_k, s_k \rangle}{\langle s_k, y_k \rangle}. \quad (2.5)$$

Set  $k := k + 1$ , and go to Step 1.

*Remark 2.2.* In Step 5, the vector  $y_k$  is different from the standard definition in the spectral gradient method, which is motivated by the idea of [15]. We will show that it plays an important role in the proof of the local R-linear convergence rate in Section 3.

To establish the global convergence of Algorithm 2.1, we make the following assumptions.

**Assumptions**

(A1) The solution set of (1.1), denoted by  $S$ , is nonempty; the function  $f(x)$  is assumed a convex function on  $\mathbb{R}^n$ , which means that  $\nabla f(x)$  is monotone on  $\mathbb{R}^n$ :

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$$

for all  $x, y \in \mathbb{R}^n$ .

(A2) The gradient  $\nabla f(x)$  is Lipschitz continuous on  $\mathbb{R}^n$ , that is, there exists a constant  $L > 0$  such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

To analyze the global convergence of Algorithm 2.1, we assume that it generates an infinite sequence  $\{x_k\}$ .

The following lemma shows that Algorithm 2.1 is well defined.

**Lemma 2.3.** *Suppose that Assumption (A2) holds. Then the Algorithm 2.1 is well-defined.*

*Proof.* Suppose that there exists  $k_0 \geq 1$  such that (2.2) is not satisfied for any nonnegative integer  $m$ , that is,

$$-\langle \nabla f(x_{k_0} + \beta^m d_{k_0}), d_{k_0} \rangle < \sigma \beta^m \|d_{k_0}\|^2, \quad \forall m \geq 1.$$

Let  $m \rightarrow \infty$ , then by the continuity of  $\nabla f(x)$ , we have

$$-\langle \nabla f(x_{k_0}), d_{k_0} \rangle \leq 0. \tag{2.6}$$

From Step 1, Step 2 and Step 5, we know that

$$\nabla f(x_k) \neq 0, \quad d_k \neq 0, \quad \forall k \geq 1.$$

By the remark of [27], we know that

$$\frac{1}{L+r} \leq \theta_k \leq \frac{1}{r}. \tag{2.7}$$

Hence by (2.1), we have

$$\begin{aligned} -\langle \nabla f(x_k), d_k \rangle &= \theta_k \langle \nabla f(x_k), \nabla f(x_k) \rangle - \langle \nabla f(x_k), e_k \rangle \\ &\geq (1 - \eta)\theta_k \|\nabla f(x_k)\|^2 > 0, \end{aligned} \tag{2.8}$$

which contradicts (2.6), the contradiction deduces our desired result. □

The following result shows that the sequences  $\{x_k\}$  and  $\{z_k\}$  are bounded, the proof is very similar to that of [23], we describe here especially some inequalities since we will use some them in the local convergence analysis.

**Lemma 2.4.** *Suppose Assumptions (A1) and (A2) hold, sequences  $\{x_k\}$  and  $\{z_k\}$  are generated by Algorithm 2.1. Then we have*

- (i)  $\{x_k\}$  and  $\{z_k\}$  are both bounded;
- (ii)  $\lim_{k \rightarrow \infty} \|x_k - z_k\| = 0$  and  $\lim_{k \rightarrow \infty} \|x_{k+1} - x_k\| = 0$ .

*Proof.* From (2.2), we have

$$\langle \nabla f(z_k), x_k - z_k \rangle = -\alpha_k \langle \nabla f(z_k), d_k \rangle \geq \sigma \alpha_k^2 \|d_k\|^2 = \sigma \|x_k - z_k\|^2 > 0. \tag{2.9}$$

Let  $x^* \in S$ , then we have

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &= \|x_k - \zeta_k \nabla f(z_k) - x^*\|^2 \\ &= \|x_k - x^*\|^2 - 2\zeta_k \langle \nabla f(z_k), x_k - x^* \rangle + \zeta_k^2 \|\nabla f(z_k)\|^2. \end{aligned} \tag{2.10}$$

By the monotonicity of gradient  $\nabla f(x)$  and  $x^* \in S$ , it holds that

$$\begin{aligned} \langle \nabla f(z_k), x_k - x^* \rangle &= \langle \nabla f(z_k), x_k - z_k \rangle + \langle \nabla f(z_k), z_k - x^* \rangle \\ &\geq \langle \nabla f(z_k), x_k - z_k \rangle + \langle \nabla f(x^*), z_k - x^* \rangle \\ &= \langle \nabla f(z_k), x_k - z_k \rangle. \end{aligned} \tag{2.11}$$

From (2.4),(2.9), (2.10) and (2.11), we have

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 2\zeta_k \langle \nabla f(z_k), x_k - z_k \rangle + \zeta_k^2 \|\nabla f(z_k)\|^2 \\ &= \|x_k - x^*\|^2 - \frac{\langle \nabla f(z_k), x_k - z_k \rangle^2}{\|\nabla f(z_k)\|^2} \\ &\leq \|x_k - x^*\|^2 - \frac{\sigma^2 \|x_k - z_k\|^4}{\|\nabla f(z_k)\|^2}. \end{aligned} \tag{2.12}$$

Hence the sequence  $\{\|x_k - x^*\|\}$  is decreasing and convergent, moreover, the sequence  $\{x_k\}$  is bounded. Using the continuity of  $\nabla f(x)$ , we know that there exists a constant  $M > 0$ , such that

$$\|\nabla f(x_k)\| \leq M, \quad \forall k \geq 1. \tag{2.13}$$

By the Cauchy-Schwartz inequality, the monotonicity of  $\nabla f(x)$  and (2.9), we have

$$\|\nabla f(x_k)\| \geq \frac{\langle \nabla f(x_k), x_k - z_k \rangle}{\|x_k - z_k\|} \geq \frac{\langle \nabla f(z_k), x_k - z_k \rangle}{\|x_k - z_k\|} \geq \sigma \|x_k - z_k\|. \tag{2.14}$$

From (2.13) and (2.14), we obtain that the sequence  $\{z_k\}$  is also bounded, without loss of generality, we assume that  $\|\nabla f(z_k)\| \leq M$ . It follows from (2.12) that

$$\frac{\sigma^2}{M^2} \sum_{k=1}^{\infty} \|x_k - z_k\|^4 \leq \sum_{k=1}^{\infty} (\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2) < \infty,$$

which implies that

$$\lim_{k \rightarrow \infty} \|x_k - z_k\| = 0.$$

Using (2.3) and Cauchy-Schwarz inequality, we obtain that

$$\begin{aligned} \|x_{k+1} - x_k\| &= \|(x_k - \zeta_k \nabla f(z_k)) - x_k\| \\ &= \|\zeta_k \nabla f(z_k)\| = \frac{\langle \nabla f(z_k), x_k - z_k \rangle}{\|\nabla f(z_k)\|} \\ &\leq \|x_k - z_k\|. \end{aligned} \tag{2.15}$$

Thus,

$$\lim_{k \rightarrow \infty} \|x_{k+1} - x_k\| = 0.$$

□

**Theorem 2.5.** Under Assumptions (A1) and (A2), the sequence  $\{x_k\}$  generated by Algorithm 2.1 converges to a solution of problem (1.1).

*Proof.* Since  $z_k = x_k + \alpha_k d_k$ , it holds from Lemma 2.4

$$\lim_{k \rightarrow \infty} \alpha_k \|d_k\| = \lim_{k \rightarrow \infty} \|x_k - z_k\| = 0. \tag{2.16}$$

From (2.1) and (2.7), we have

$$\|d_k\| \geq \|-\theta_k \nabla f(x_k)\| - \|e_k\| \geq \frac{(1-\eta)\|\nabla f(x_k)\|}{(L+r)}. \tag{2.17}$$

So if  $\liminf_{k \rightarrow \infty} \|d_k\| = 0$ , we obtain by (2.17) that

$$\liminf_{k \rightarrow \infty} \|\nabla f(x_k)\| = 0.$$

The continuity of  $\nabla f(x)$  implies that the sequence  $\{x_k\}$  has some accumulation point  $x^*$  such that  $\nabla f(x^*) = 0$ , that is,  $x^* \in S$ . From (2.12), it holds that  $\{\|x_k - x^*\|\}$  converges, and since  $x^*$  is an accumulation point of  $\{x_k\}$ , it must hold that  $\{x_k\}$  converges to  $x^*$ . Now assume

$$\liminf_{k \rightarrow \infty} \|d_k\| > 0,$$

then by (2.16), it holds that

$$\lim_{k \rightarrow \infty} \alpha_k = 0. \tag{2.18}$$

This implies that

$$-\langle \nabla f(x_k + \frac{\alpha_k}{\beta} d_k), d_k \rangle < \sigma \frac{\alpha_k}{\beta} \|d_k\|^2. \tag{2.19}$$

Since  $\{x_k\}$  and  $\{d_k\}$  are bounded, let  $k \rightarrow \infty$  in the above inequality and assume  $\{x_k\} \rightarrow \bar{x}$  and  $\{d_k\} \rightarrow \bar{d}$ , we obtain

$$-\langle \nabla f(\bar{x}), \bar{d} \rangle \leq 0,$$

this contradicts to (2.8), the contradiction deduces our desired result. This completes the proof.  $\square$

### 3. Convergence rate

In this section, we analyze the local convergence rate of our method. By Theorem 2.5, we assume  $\{x_k\} \rightarrow x^* \in S$ . Let  $dist(x, S)$  denote the distance from  $x$  to  $S$ , to this end, we make the following error bound assumption:

#### Assumption

(A3) For  $x^* \in S$ , there exist positive constants  $\delta, c_1$  such that

$$c_1 dist(x, S) \leq \|\nabla f(x)\|, \forall x \in N(x^*, \delta) \cap S. \tag{3.1}$$

Here  $N(x^*, \delta) = \{\zeta \in \mathbb{R}^n \mid \|\zeta - x^*\| \leq \delta\}$  be the closed ball centered at  $x^*$  with radius  $\delta > 0$ . Assumption (A3) is a local error bound condition and known to be much weaker than the more standard nonsingularity assumption. By (2.7), we know that  $\frac{1}{L+r} \leq \theta_k \leq \frac{1}{r}$ , hence from (2.1) we have

$$\|d_k\| \leq \frac{1+\eta}{r} \|\nabla f(x_k)\|. \tag{3.2}$$

Based on this inequality, we can easily obtain the following conclusion.

**Lemma 3.1.** *Under Assumptions (A1)-(A3), suppose that  $x_k \in N(x^*, \frac{1}{2}\delta)$ . Then we have*

$$\|d_k\| \leq \frac{L(1+\eta)}{r} dist(x_k, S). \tag{3.3}$$

*Proof.* Let  $\mu_k \in S$  be a closest solution to  $x_k$ . That is,  $\|x_k - \mu_k\| = dist(x_k, S)$ . Since  $x_k \in N(x^*, \frac{1}{2}\delta)$ , we have

$$\|\mu_k - x^*\| \leq \|\mu_k - x_k\| + \|x_k - x^*\| \leq \|x_k - x^*\| + \|x_k - x^*\| \leq \delta,$$

hence  $\mu_k \in N(x^*, \delta)$ . By (3.2),

$$\|d_k\| \leq \frac{1+\eta}{r} \|\nabla f(x_k) - \nabla f(\mu_k)\| = \frac{L(1+\eta)}{r} dist(x_k, S).$$

$\square$

**Lemma 3.2.** *Under Assumptions (A1)-(A3), if  $r \geq (L + \sigma)(1 + \eta)$ , then for all  $k$  large enough, we have  $z_k = x_k + d_k$ .*

*Proof.* By Theorem 2.5, the sequence  $\{\nabla f(x_k)\}$  converges to 0. By Lemma 3.1, the sequence  $\{d_k\}$  converges also to 0. Hence by the Lipschitz continuity of  $\nabla f(x)$ , we have

$$\nabla f(x_k + d_k) = \nabla f(x_k) + R_k,$$

with  $\|R_k\| \leq L\|d_k\|$ . By the definition of  $d_k$  and inequality (2.7), we have

$$-\langle \nabla f(x_k), d_k \rangle \geq \frac{1}{(1 + \eta)\theta_k} \|d_k\|^2 \geq \frac{r}{(1 + \eta)} \|d_k\|^2,$$

and therefore, we get

$$-\langle \nabla f(x_k + d_k), d_k \rangle \geq \left(\frac{r}{1 + \eta} - L\right) \|d_k\|^2.$$

since  $r \geq (L + \sigma)(1 + \eta)$ , we have  $\frac{r}{1 + \eta} - L \geq \sigma$  and  $\left(\frac{r}{1 + \eta} - L\right) \|d_k\|^2 \geq \sigma \|d_k\|^2$ , then by Step 3 in Algorithm 2.1, we know that  $x_k + d_k$  satisfies the inequality (2.2) and therefore  $z_k = x_k + d_k$ .  $\square$

Now, we introduce the main result in this section.

**Theorem 3.3.** *Suppose that  $\{x_k\}$  converges to  $x^* \in S$  and Assumptions (A1)-(A3) hold. Then the whole sequence  $\{x_k\}$  converges to  $x^*$  R-linearly.*

*Proof.* Let  $\omega_k := \operatorname{argmin}\{\|x_k - \omega\| \mid \omega \in S\}$ . From (2.12), we obtain

$$\|x_{k+1} - \omega_k\|^2 \leq \|x_k - \omega_k\|^2 - \frac{\langle \nabla f(z_k), x_k - z_k \rangle^2}{\|\nabla f(z_k)\|^2}. \quad (3.4)$$

Hence for sufficiently large  $k$ , we have from Assumption (A2) that

$$\begin{aligned} \|\nabla f(z_k)\| &= \|\nabla f(z_k) - \nabla f(\omega_k)\| \leq L\|z_k - \omega_k\| \\ &\leq L(\|x_k - z_k\| + \|x_k - \omega_k\|) \\ &\leq L(\|d_k\| + \|x_k - \omega_k\|) \\ &\leq L\left(1 + \frac{L(1 + \eta)}{r}\right) \operatorname{dist}(x_k, S) \\ &\doteq c_2 \operatorname{dist}(x_k, S). \end{aligned}$$

By the definition of  $d_k$ ,

$$\|\nabla f(x_k)\| \leq \left(\frac{L + r}{1 - \eta}\right) \|d_k\| \doteq c_3 \|d_k\|.$$

On the other hand, from Lemma 3.2, (3.4) and Assumption (A3), we have

$$\langle \nabla f(z_k), x_k - z_k \rangle \geq \sigma \|d_k\|^2 \geq \frac{\sigma}{c_3^2} (\|\nabla f(x_k)\|^2) \geq \frac{c_1^2 \sigma}{c_3^2} \operatorname{dist}^2(x_k, S). \quad (3.5)$$

It follows from (3.4) and (3.5) that

$$\operatorname{dist}^2(x_{k+1}, S) \leq \|x_{k+1} - \omega_k\|^2 \leq \left(1 - \frac{\sigma^2 c_1^4}{c_3^4 c_2^2}\right) \operatorname{dist}^2(x_k, S),$$

which implies that the sequence  $\{\operatorname{dist}(x_k, S)\}$  converges to 0 Q-linearly. Therefore, the sequence  $\{x_k\}$  converges to  $x^*$  R-linearly. This completes the proof.  $\square$

*Remark 3.4.* Note that to obtain the linear convergence, it is necessary to let the parameter  $\frac{\sigma^2 c_1^4}{c_3^4 c_2^2} < 1$ .

#### 4. Numerical results

In this section, we test the efficiency of our method on some test problems. The algorithms were coded in Matlab 2013a and run on a personal computer with 2.93GHZ CPU processor. The parameters are set as follows:  $\beta = 0.5$ ,  $\sigma = 10^{-2}$ ,  $\eta = 10^{-2}$  and  $e_k$  is generated randomly and satisfies the condition in Step 2. We use  $\|\nabla f(x_k)\| \leq 10^{-6}$  as the stopping criterion.

In Sect. 3, the R-linear convergence of the proposed algorithm has been proved theoretically under the local error bound condition  $\|\nabla f(x)\| \geq c_1 \text{dist}(x, S)$ . In what follows, we first examined the local convergence of the method on the following test problem [26]:

$$f(x_1, x_2) = (x_1 - 4x_2)^2.$$

The solution set of this problem is  $\{(x_1, x_2) | x_1 - 4x_2 = 0\}$ . We set the initial point to  $(-5000, 5000)$  and try to seek the minimum and the parameter  $r = 0.1$ . It is easy to verify that the Hessian is singular at any solution, and the local error bound condition is satisfied when  $c_1 \in [0, \sqrt{34}]$ . Numerical results indicate that the sequence generated by the algorithm converges to  $x = (3.52941176e + 03, -0.88235294e + 03)$ , which is an optimal solution. The iterative step  $x_k$  and the norm of gradient at every iteration are recorded in Table 1. These results indicate that our spectral gradient projection method converges quickly when  $x_k$  approaches the optimal solution. Figure 1 gives the behavior of the iteration for this problem.

Table 1: Test results for local convergence

<i>Iteration k</i>	$x_k$	$\ \nabla f(x_k)\ $
1	(-5000, 5000)	2.06155281e+05
2	(-4.21875000e+03, 1.87500000e+03)	9.66352881e+04
3	(-3.53143312e+03, -0.87426749e+03)	2.83365636e+02
4	(-3.52942320e+03, -0.88230717e+03)	1.60402414
5	(-3.52941181e+03, -0.88235272e+03)	0.00761685
6	(-3.52941176e+03, -0.88235294e+03)	6.59860170e-06

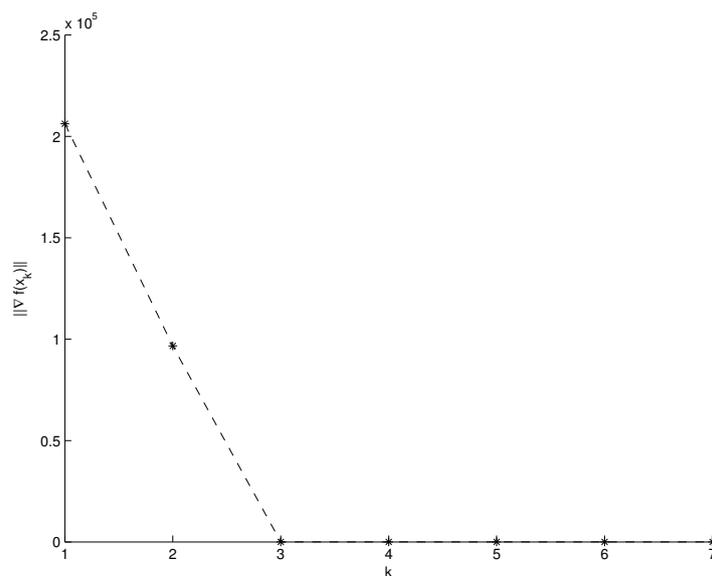


Figure 1: Behavior of the iteration.

We then test the global convergence of the method for some large scale problems. The problems are taken from [1] or the CUTE collection established by Bongartz, Conn, Gould and Toint[6]. Here the function name is the same as that of [1], for example, we write Dig7 to denote the Diagonal 7 function in [1]. Since the parameter  $r$  plays an important role in the convergence analysis, in Table 2, we give the test results for different choice of the parameter  $r$ . with perturbed parameter  $\eta = 0.01$ . Here, we give the comparison results for  $r = 0.1$  and  $r = 0.01$  and we use  $n$  to denote the dimension of the problem,  $IT$  denote the number of iteration and Time denote the CPU time used (in second). From Table 2, we can see that for our test problems, the behaviors of the algorithm for  $r = 0.1$  are somewhat better than these of  $r = 0.01$ , which is different from the behaviors form the method to monotone equation in [27], which show that the method works well for smaller parameter  $r$ .

Table 2: Test results for different  $r$ 

<i>Problem</i>	$r=0.1$			$r=0.01$	
	$n$	$IT$	Time	$IT$	Time
Raydan2	1000	11	0.0637	21	0.0818
	10000	11	0.1189	22	0.2050
	100000	12	0.5762	24	1.2157
	1000000	12	5.6008	26	13.6712
Dig5	1000	10	0.0358	25	0.0464
	10000	11	0.1057	27	0.2657
	100000	12	0.8347	29	2.4134
	1000000	12	8.5868	30	24.0679
Quadratic QF1	1000	28	0.0263	27	0.0132
	10000	26	0.1034	29	0.1542
	100000	26	1.0666	27	1.1297
	1000000	30	14.2717	28	13.3337
Generalized Quadratic	1000	25	0.0391	24	0.0368
	10000	26	0.1710	24	0.1267
	100000	28	1.1538	26	1.1262
	1000000	29	13.2829	27	13.2831
Dig7	1000	27	0.0519	26	0.0483
	10000	28	0.2240	27	0.2300
	100000	28	1.8972	27	1.8361
	1000000	30	22.1964	29	22.5995
Dig8	1000	15	0.0331	22	0.0465
	10000	16	0.1435	23	0.2265
	100000	17	1.2033	25	1.9743
	1000000	18	13.0988	27	22.7253
Extended Penalty	1000	68	0.0633	68	0.4261
	10000	85	0.7879	87	0.9253
	100000	97	4.5619	97	4.5481
Full Hessian FH1	1000	6	0.0150	6	0.0499
	10000	7	0.0940	7	0.0971
	100000	7	0.7974	7	0.7814
	1000000	8	9.1848	8	9.1205

## 5. Conclusion

In this paper, we establish the R-linear convergence of a spectral projected gradient method for unconstrained optimization with singular solution under a local error bound condition. We obtain the R-linearly convergence rate of the proposed method and give some numerical tests to show the efficiency of the proposed method. It is worth pointing out the convergence rate analysis is based on the assumption for some parameters  $r \geq (L + \sigma)(1 + \eta)$ , so how to weaken those conditions to obtain the rate of convergence worth further discussing.

## Acknowledgment

This work is supported by Innovation Program of Shanghai Municipal Education Commission (No.14YZ094).

## References

- [1] N. Andrei, *An Unconstrained Optimization Test Functions Collection*, Adv Model Optim., **10** (2008), 147–161. 4
- [2] J. Barzilai, J. M. Borwein, *Two Point Step size Gradient Methods*, IMA J. Numer. Anal., **8** (1988), 141–148. 1
- [3] E. G. Birgin, J. M. Martínez, M. Raydan, *Nonmonotone Spectral Projected Gradient Methods for Convex Sets*, SIAM J. Optim., **10** (2000), 1196–1211. 1
- [4] E. G. Birgin, J. M. Martínez, M. Raydan, *Encyclopedia of Optimization: Spectral Projected Gradient Methods*, Springer, **2008** (2008), 3652–3659. 1
- [5] E. G. Birgin, J. M. Martnez, M. Raydan, *Spectral Projected Gradient Methods: Review and Perspectives*, J. Stati. Software, **6** (2014), 1–21. 1
- [6] I. Bongartz, A. R. Conn, N. I. M. Gould, P. L. Toint, *CUTE: Constrained and Unconstrained Testing Environments*, ACM TOMS, **21** (1995), 123–160. 4
- [7] Y.-H. Dai, R. Fletcher, *Projected Barzilai-Borwein Methods for Large-Scale Box-Constrained Quadratic Programming*, Numer. Math., **100** (2005), 21–47. 1
- [8] Y.-H. Dai, L.-Z. Liao, *R-linear Convergence of the Barzilai and Borwein Gradient Method*, IMA J. Numer. Anal., **22** (2002), 1–10. 1
- [9] J. E. Dennis, J. J. More, *Quasi-Newton Methods, Motivation and Theory*, SIAM Rev., **19** (1977), 46–89. 1
- [10] J. E. Dennis, R. B. Schnabel, *Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, NJ, (1983). 1
- [11] M. A. Diniz-Ehrhardt, M. A. Gomes-Ruggiero, J. M. Martínez, S. A. Santos, *Augmented Lagrangian Algorithm Based On the Spectral Projected Gradient Method for Solving Nonlinear Programming Problems*, J. Optim. Theory Appl., **123** (2004), 497–517. 1
- [12] L. Grippo, F. Lampariello, S. Lucidi, *A Nonmonotone Line Search Technique for Newton's Methods*, SIAM J. Numer. Anal., **23** (1986), 707–716. 1
- [13] C. Kanzow, N. Yamashita, M. Fukushima, *Levenberg-Marquardt Methods for Constrained Nonlinear Equations with Strong Local Convergence Properties*, J. Comput. Appl. Math., **172** (2004), 375–397. 1
- [14] W. La Cruz, J. M. Martínez, M. Raydan, *Spectral Residual Method Without Gradient Information for Solving Large-scale Nonlinear Systems of Equations*, Math Comput., **75** (2006), 1449–1466. 1
- [15] W. La Cruz, M. Raydan, *Nonmonotone Spectral Methods for Large-scale Nonlinear Systems*, Optim Method Soft., **18** (2003), 583–599. 1, 2.2
- [16] D.-H. Li, M. Fukushima, L. Qi, N. Yamashita, *Regularized Newton Method for Convex Minimization Problems with Singular Solutions*, Comput. Optim. Appl., **26** (2004), 131–147. 1
- [17] Y.-J. Li, D.-H. Li, *Truncated Regularized Newton Method for Convex Minimizations*, Comput. Optim. Appl., **43** (2009), 119–131. 1
- [18] W. B. Liu, Y. H. Dai, *Minimization Algorithms Based on Supervisor and Searcher Cooperation*, J. Optim. Theory. Appl., **111** (2001), 359–379. 1
- [19] B. Molina, M. Raydan, *Preconditioned Barzilai-Borwein Method for the Numerical Solution of A Partial Differential Equation*, Numer. Algorithm, **13** (1996), 45–60. 1
- [20] M. Raydan, *On the Barzilai and Borwein Choice of Step Length for the Gradient Method*, IMA J. Numer. Anal., **13** (1993), 321–326. 1
- [21] M. Raydan, *The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem*, SIAM J. Optim., **7** (1997), 26–33. 1
- [22] M. V. Solodov, B. F. Svaiter, *Descent Methods with Linesearch in the Presence of Perturbations*, J. Comput. Appl. Math., **80** (1997), 265–275. 1

- [23] M. V. Solodov, B. F. Svaiter, *A Globally Convergent Inexact Newton Method for Systems of Monotone Equations: in Reformulation: Piecewise Smooth, Semi-smooth and Smoothing Methods*, Kluwer Acad. Publ., (1999), 355–369. 1, 2
- [24] C. Wang, Y. Wang, C. Xu, *A Projection Method for A System of Nonlinear Monotone Equations with Convex Constraints*, Math. Meth. Oper. Res., **66** (2007), 33–46. 1
- [25] N. Yamashita, M. Fukushima, *On the Rate of Convergence of the Levenberg-Marquardt Method: in Topics in numerical analysis*, Comput. Suppl., 15, Springer, Vienna, (2001), 237–249. 1
- [26] J. L. Zhang, L. Y. Wu, X. S. Zhang, *A Trust Region Method for Optimization Problem with Singular Solutions*, Appl. Math. Optim., **56** (2007), 379–394. 1, 4
- [27] L. Zhang, W. J. Zhou, *Spectral Gradient Projection Method for Solving Nonlinear Monotone Equations*, J. Comput. Appl. Math., **196** (2006), 478–484. 1, 2, 4
- [28] G. Zhou, K. C. Toh, *Superlinear Convergence of a Newton-type Algorithm for Monotone Equations*, J. Optim. Theory Appl., **125** (2005), 205–221. 1