Global Convergence of a Modified Spectral FR Conjugate Gradient Method Without Line Search

Shou-qiang Du\textsuperscript{1,2} *, Yuan-yuan Chen\textsuperscript{2}

\textsuperscript{1} School of Management, University of Shanghai for Science and Technology
Shanghai 200093, China
\textsuperscript{2} College of Mathematics, Qingdao University
Qingdao Shandong, 266071, China.

(Received 20 April 2007, accepted 9 May 2007)

Abstract: The nonlinear conjugate gradient method is widely used in unconstrained optimization. However, the line search is very difficult or expensive sometimes. In this paper, we proper a spectral FR method without line search, the global convergence of this method is also given.

Key words: Unconstrained optimization; line search; global convergence

1 Introduction

Nonlinear conjugate gradient methods are well suited for large-scale optimization problems due to the simplicity of their iteration and low memory requirements. The nonlinear conjugate gradient method is designed to solve the following unconstrained optimization problem

\[
\min_{x \in \mathbb{R}^n} f(x),
\]  

(1)

where \( f : \mathbb{R}^n \to \mathbb{R} \) is continuously differentiable. The iterative formula of the conjugate gradient method is given by

\[
x_{k+1} = x_k + \alpha_k d_k,
\]

(2)

where the step-length \( \alpha_k \) is obtained by carrying out some line search, and the direction \( d_k \) is defined by

\[
d_k = \begin{cases} 
eg g_k, & k = 1; \\
- g_k + \beta_k d_{k-1}, & k \geq 2,
\end{cases}
\]

where \( \beta_k \) is a scalar and \( g_k = \nabla f(x_k) \). The original nonlinear conjugate gradient method proposed by Fletcher – Reeves (FR conjugate gradient method)\cite{1}, in which \( \beta_k \) is defined by

\[
\beta_k^{FR} = \frac{\| g_k \|^2}{\| g_{k-1} \|^2},
\]

(3)

when applied to strictly quadratic objective functions, this method reduces to the linear conjugate gradient method provided that \( \alpha_k \) is the exact minimizer. Zoutendijk\cite{2} proved that the FR method with exact line search is globally convergent. Al – Baali\cite{3} proved the global convergence of FR method under the strong Wolfe line search

\[
f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k,
\]

\[
|d_k^T g(x_k + \alpha_k d_k)| \leq |g_k^T d_k|,
\]

where \( \delta \) is a scalar and \( g_k = \nabla f(x_k) \).
where \( 0 < \delta < \sigma < \frac{1}{2} \). Dai and Yuan\[4\], showed that the FR conjugate gradient method are globally convergent if the following line search conditions are satisfied
\[
\begin{align*}
  f(x_k) - f(x_k + \alpha_k d_k) &\geq -\delta \alpha_k g_k^T d_k, \\
  \sigma_1 g_k^T d_k &\leq g(x_k + \alpha_k d_k)^T d_k \leq -\sigma_2 g_k^T d_k,
\end{align*}
\]
where \( \sigma \geq 0, 0 < \delta < \sigma < 1, 0 < \sigma_2 < 1, \sigma_1 + \sigma_2 \leq 1 \).

Recently, Sun and Zhang\[5\] introduced a different method, it takes fixed step length by formula
\[
\alpha_k = -\frac{\delta g_k^T d_k}{\|d_k\|_{Q_k}},
\]
where \( \|d_k\|_{Q_k} = \sqrt{d_k^T Q_k d_k}, \delta \in (0, \frac{\nu_{\min}}{\mu}) \) and \( \frac{\delta \mu}{\nu_{\min}} < 1 \), \( \mu \) is a Lipstchtz constant, which will be given by the following Assumption. \( \{Q_k\} \) is a sequence of positive definite matrices satisfying
\[
\nu_{\min} d^T d \leq d^T Q_k d \leq \nu_{\max} d^T d,
\]
where all \( d \in \mathbb{R}^n \), \( \nu_{\min} \) and \( \nu_{\max} \) are positive constants.

On the other hand, Birgin and Martinez\[6\] proposed a spectral conjugate gradient method by combining conjugate gradient method and spectral gradient method in \[7\]. The direction \( d_k \) is given by the following way
\[
d_k = -\theta_k g_k + \beta_k d_{k-1},
\]
where \( \theta_k \) is a parameter and
\[
\beta_k = \frac{(\theta_k y_{k-1} - s_{k-1})^T g_k}{d_{k-1}^T y_{k-1}},
\]
where \( y_{k-1} = g_k - g_{k-1}, s_{k-1} = x_k - x_{k-1} \) and \( \theta_k \) is taken to be the spectral gradient
\[
\theta_k = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}.
\]

Unfortunately, the spectral conjugate gradient method can not guarantee to generate descent directions\[6\]. So in paper\[8\], based on the FR formula, they take modification to the FR method such that the direction generated is always a descent direction. The \( d_k \) be defined by
\[
d_k = \begin{cases} 
  -g_k, & k = 0; \\
  -\theta_k g_k + \beta_k^{FR} d_{k-1}, & k > 0,
\end{cases}
\]
where \( \beta_k^{FR} \) is specified by (3) and
\[
\theta_k = \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2},
\]
where \( y_k = g_{k+1} - g_k \).

Based on the idea of them, we propose a spectral FR conjugate gradient method without line search.

In the next section, we present the spectral FR conjugate gradient Algorithm. Some mild conditions are also given. In section3, the global convergence will be given. Finally, we will give some Remarks about this method.

### 2 Algorithm and lemmas

First, we will give the following assumptions on objective function \( f(x) \), which have been used often in the literature to analyze the global convergence of conjugate gradient methods with inexact line searches.

Assumption

(i) The level set \( L = \{ x \in \mathbb{R}^n | f(x) \leq f(x_0) \} \) is bounded.
(ii) In some neighborhood $U$ of $L$, $f(x)$ is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $\mu > 0$ such that

$$
\|g(x_{k+1}) - g(x_k)\| \leq \mu \|x_{k+1} - x_k\|, \forall x_k, x_{k+1} \in U.
$$

The algorithm has the following steps:

**Step0.** Given $x_0 \in R^n$, set $d_0 = -g_0$, $k := 0$. If $g_0 = 0$, then stop.

**Step1.** Find $\alpha_k > 0$ satisfying (4). By (2), $x_{k+1}$ is given. If $g_{k+1} = 0$, then stop.

**Step2.** Compute $d_k$ by (5).

**Step3.** Set $k := k + 1$, go to Step1.

**Lemma 1** Suppose direction $d_k$ is given by (5), then we have

$$d_k^T g_k = -\|g_k\|^2,$$

(7) holds for any $k \geq 0$.

**Proof.** We show by induction that (7) holds for all $k$. For $k = 0$, (7) is clearly true as $d_0 = -g_0$.

Now, assume for some $k \geq 1$, $d_{k-1}^T g_{k-1} = -\|g_{k-1}\|^2$, then by (3), (5) and (6), we have

$$
g_k^T d_k = -\theta_k \|g_k\|^2 + \frac{d_{k-1}^T g_{k-1}}{\|g_{k-1}\|^2} d_{k-1}^T g_k
$$

$$= -\frac{d_{k-1}^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \|g_k\|^2 + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} d_{k-1}^T g_k
$$

$$= \frac{\|g_k\|^2}{\|g_{k-1}\|^2} (-d_{k-1}^T g_k + d_{k-1}^T g_{k-1} + d_{k-1}^T g_k)
$$

$$= \frac{\|g_k\|^2}{\|g_{k-1}\|^2} d_{k-1}^T g_{k-1}
$$

$$= \frac{\|g_k\|^2}{\|g_{k-1}\|^2} (-\|g_{k-1}\|^2)
$$

$$= -\|g_k\|^2.
$$

Therefore, (7) is still true with $k - 1$ replaced $k$, so by induction Lemma is proved.

**Lemma 2** Suppose that Assumption holds. And $x_k$ is given by the Algorithm, then

$$\sum_{k \geq 1} \|g_k\|^4 \|d_k\|^2 < +\infty.
$$

**Proof** Similar to Lemma 4 in paper[5]. By the mean-value theorem we have

$$f(x_{k+1}) - f(x_k) = g^T (x_{k+1} - x_k),$$

where $\bar{g} = \nabla f(\bar{x})$ for some $\bar{x} \in [x_k, x_{k+1}]$.

By the Cauchy–Schwartz inequality, (2), (4) and Assumption, we know

$$f(x_{k+1}) - f(x_k) = \bar{g}^T (x_{k+1} - x_k)
$$

$$\leq \|\bar{g} - g_k\| \|x_{k+1} - x_k\| + \|g_k\| \|x_{k+1} - x_k\|
$$

$$\leq \alpha_k g_k^T d_k + \mu \|x_{k+1} - x_k\|^2
$$

$$\leq \alpha_k g_k^T d_k + \mu \|d_k\|^2
$$

$$= \alpha_k g_k^T d_k (1 - \frac{\mu \delta}{\nu_{\min} \|d_k\|^2})
$$

$$\leq -\delta (1 - \frac{\mu \delta}{\nu_{\min} \|d_k\|^2}).
$$

IINS homepage: http://www.nonlinearscience.org.uk/
By Assumption, the above formula and Lemma 1, we have
\[
\|g_k\|^4 \|d_k\|^2 = \frac{(g_k^T d_k)^2}{\|d_k\|^2} \leq \nu_{\text{max}} \frac{(g_k^T d_k)^2}{\|Q_k\|^2}
\]
\[
\leq \frac{\nu_{\text{max}}}{\delta(1 - \frac{\mu}{\nu_{\text{min}}})} [f(x_k) - f(x_{k+1})].
\]
So we have \(\sum_{k=1}^{\infty} \|g_k\|^4 < +\infty\). This completes the proof.

3 Convergence analysis for the algorithm

From the above Lemmas and Assumption, we give the following Theorem of global convergence for the spectral FR conjugate gradient method.

**Theorem 3** Consider the spectral FR conjugate gradient Algorithm, suppose that Assumptions hold. Then we have
\[
\liminf_{k \to \infty} \|g_k\| = 0.
\]

**Proof.** Suppose by contradiction that there exists a positive constant \(\epsilon > 0\) such that
\[
\|g_k\| \geq \epsilon, \quad (8)
\]
holds for all \(k \geq 0\).

From (5), we have
\[
\|d_k\|^2 = (\beta_{FR}^k)^2 \|d_{k-1}\|^2 - 2\theta_k g_k^T d_k - \theta_k^2 \|g_k\|^2. \quad (9)
\]
Dividing both sides of (9) by \((g_k^T d_k)^2\), by (3), (7) and (8), we have
\[
\frac{\|d_k\|^2}{\|g_k\|^4} = \frac{\|d_k\|^2}{(g_k^T d_k)^2} = \frac{\|g_k\|^2}{\|g_k\|^2} \frac{\|d_{k-1}\|^2}{\|g_k\|^2} + \frac{2\theta_k}{\|g_k\|^2} - \frac{\theta_k^2}{\|g_k\|^2}
\]
\[
= \frac{\|d_{k-1}\|^2}{\|g_k\|^4} - \frac{(\theta_k-1)^2}{\|g_k\|^2} + \frac{1}{\|g_k\|^2}
\]
\[
\leq \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + \frac{1}{\|g_k\|^2}. \quad (10)
\]
Then we have \(\frac{\|d_k\|^2}{\|g_k\|^4} \leq \sum_{i=0}^{k-1} \|g_i\|^2 \leq \frac{k}{\epsilon^2}. \quad \)So we obtain
\[
\sum_{k=1}^{\infty} \|g_k\|^4 \geq \epsilon^2 \sum_{k=1}^{\infty} \frac{1}{k} = +\infty.
\]
Which contradicts Lemma 2. Therefor, we have \(\liminf_{k \to \infty} \|g_k\| = 0\).

4 Discussion

From the above global convergence of the method, we disclosed an property of the FR method, i.e. its global convergence can be guaranteed by taking a fixed step length rather than following a set of exact or inexact line search. This step length might practical in cases that the exact or inexact line search is sometimes expansive to use. However, this method select sequence \(\{Q_k\}\) in certain flexibility, it would not guarantee the descent property. So we can use the method of paper [9].
(1) Choose \( \{Q_k\} \) by
\[
Q_k = \delta \mu_k I (\mu_k > 0),
\] and \( \mu_k \) is satisfied \( 1 - \frac{\mu_k}{\mu} > 0 \), \( I \) is identity matrices.
(2) Choose \( \{Q_k\} \) by Limited memory BFGS method.

By the above method, we also can give the corresponding spectral FR method. The global convergence analysis is similar to the above Lemma and Theorem, so we omit them. And we also can consider Quasi-Newton methods with super linear convergence for unconstrained optimization[10].

Optimization is a very active branch which is always to apply in engineering and economics[11-13]. So the above algorithm can also be used for them.

Acknowledgements

This work was supported by National Science fundation of China (under grant:10671126) and Shanghai leading academic discipline project(under grant:T0502).

References


IINS homepage:http://www.nonlinearscience.org.uk/