AMO - Advanced Modeling and Optimization, Volume 16, Number 1, 2014

On the global complexity bounds of two nonlinear conjugate gradient methods for nonconvex optimization

Chaohui Liang[1], Fei Wang[2] and Li Zhang [1]

Abstract

In this paper, we study global complexity bounds of the MHS method in [8] and the MFR method in [9] for nonconvex optimization. The global complexity bound for an iterative method solving unconstrained optimization of f is an upper bound to the number of iterations required to get an approximate solution such that $\|\nabla f(x)\| \leq \epsilon$. We show that the global complexity bounds of the MHS method and the MFR methods are $O(\epsilon^{-(2+2r)})$ and $O(a^{\epsilon^{-2}})$, respectively, where $r \geq 0$ and $a > 1$ are two constants.

Keywords. The MHS method; the MFR method; the global complexity bound.

AMS subject classification 2010. 90C30, 65K05.

1 Introduction

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

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

where $f : R^n \to R$ is a smooth function and its gradient $g(x) \triangleq \nabla f(x)$ is available. Throughout the paper, we denote $g_k = \nabla f(x_k)$, $s_k = x_{k+1} - x_k$ and $y_{k-1} = g_k - g_{k-1}$.

Recently, global complexity bounds for some iterative methods have been studied for unconstrained optimization problems by several authors in [1, 3, 4, 5, 6, 7]. The global complexity bound of an iterative method for unconstrained optimization of an objective function f is an upper bound to the number of iterations required to get an approximate solution such that $\|\nabla f(x)\| \leq \epsilon$, where ϵ is a given small positive constant. So far, some authors have given some complexity bounds of the steepest method and Newton type methods for unconstrained optimization [1, 3, 4, 5, 6, 7]. However, to the best of our knowledge, no complexity bounds for nonlinear conjugate gradient methods have been investigated. In this paper, we are going to discuss the global complexity bounds for two existing nonlinear conjugate gradient methods, one is the MHS method in [8] and the other is the MFR method in [9].

Let us first recall these two methods. In [8], Zhang et al. presented a modified Hestenes-Stiefel (MHS) nonlinear conjugate gradient method for solving (1.1). The steps of the MHS method are described as follows.

The MHS method:

[1]Department of Mathematics, Changsha University of Science and Technology, Changsha 410004, China.

[2]Department of Mathematics, Hunan City University, Yiyang 413000, China.

AMO - Advanced Modeling and Optimization. ISSN: 1841-4311

Step 0. Choose $x_0 \in R^n$, $r \geq 0$, $t > 0$, $\delta \in (0,1)$, $\rho \in (0,1)$ and $\epsilon > 0$. Let $k := 0$.

Step 1. If $\|g_k\| \leq \epsilon$, then stop. Otherwise go to Step 1.

Step 2. Compute d_k by (1.2)-(1.3) below, that is,

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k^{MHS} d_{k-1} - \theta_k^{MHS} z_{k-1}, & \text{if} \quad k \geq 1, \end{cases} \tag{1.2}$$

where

$$\beta_k^{MHS} = \frac{g_k^T z_{k-1}}{d_{k-1}^T z_{k-1}}, \quad \theta_k^{MHS} = \frac{g_k^T d_{k-1}}{d_{k-1}^T z_{k-1}}, \tag{1.3}$$

$$z_{k-1} = y_{k-1} + \Big(\max\{0, -\frac{d_{k-1}^T y_{k-1}}{d_{k-1}^T s_{k-1}}\} + t\|g_{k-1}\|^r \Big) s_{k-1}. \tag{1.4}$$

Step 3. Compute α_k by the following Armijo line search (1.5), that is, compute $\alpha_k = \max\{1, \rho^1, \rho^2, \ldots\}$ such that

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k. \tag{1.5}$$

Step 4. Set $x_{k+1} = x_k + \alpha_k d_k$.

Step 5. Let $k := k + 1$ and go to Step 1.

Remark 1.1 By the definition of z_k, it is clear that

$$d_{k-1}^T z_{k-1} \geq t\|g_{k-1}\|^r d_{k-1}^T s_{k-1} > 0, \tag{1.6}$$

which ensures that the MHS method is well defined.

In [9], Zhang et al. proposed a modified Fletcher-Reeves (MFR) nonlinear conjugate gradient method for solving (1.1), which is given below.

The MFR method:

Step 0. Choose $x_0 \in R^n$, $\delta \in (0,1)$, $\rho \in (0,1)$ and $\epsilon > 0$. Let $k := 0$.

Step 1. If $\|g_k\| \leq \epsilon$, then stop. Otherwise go to Step 1.

Step 2. Compute d_k by (1.7)-(1.8) below, that is,

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -\theta_k^{MFR} g_k + \beta_k^{FR} d_{k-1}, & \text{if} \quad k \geq 1, \end{cases} \tag{1.7}$$

where

$$\theta_k^{MFR} = \frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2}, \quad \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}. \tag{1.8}$$

Step 3. Compute α_k by the Armijo line search (1.5).

Step 4. Set $x_{k+1} = x_k + \alpha_k d_k$.

Step 5. Let $k := k + 1$ and go to Step 1.

Remark 1.2 It is easy to see that the MHS method and the MFR method satisfy the following sufficient descent condition

$$g_k^T d_k = -\|g_k\|^2. \tag{1.9}$$

This shows that they are descent methods.

The MHS method and the MFR method were proven to be globally convergent for nonconvex optimization [8, 9] when the line search (1.5) is used. In the next section and Section 3, we shall investigate the global complexity bounds for the MHS method and the MFR method, respectively.

2 The global complexity bound for the MHS method

In this section, we investigate the global complexity bound for the MHS method. To this end, we first give the following standard assumptions.

Assumption 1.

(i) The level set $\Omega = \{x|\ f(x) \leq f(x_0)\}$ is bounded.

(ii) In some neighborhood N of Ω, the gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in N. \tag{2.1}$$

It is clear that $x_k \in \Omega$ for all $k \geq 0$ and the sequence $\{f(x_k)\}$ is decreasing and therefore converges. Moreover, Assumption 1 implies that there exists a positive constant M such that

$$\|g(x)\| \leq M, \quad \forall x \in N. \tag{2.2}$$

From (1.4), (2.1) and (2.2), we have

$$\begin{aligned} \|z_k\| &\leq \|y_k\| + \Big(\frac{|d_k^T y_k|}{d_k^T s_k} + t\|g_k\|^r\Big)\|s_k\| \\ &\leq L\|s_k\| + \Big(\frac{L\|d_k\|\|s_k\|}{\|d_k\|\|s_k\|} + tM^r\Big)\|s_k\| \\ &= M_1\|s_k\|, \end{aligned} \tag{2.3}$$

where $M_1 = 2L + tM^r$.

The following result gives an estimation to the stepsize α_k from below, whose proof is standard.

Lemma 2.1. [8, 9] *Let Assumption 1 hold and the sequence $\{x_k\}$ be generated by the MHS method or the MFR method. Then there exists a constant $C_1 > 0$ such that*

$$\alpha_k \geq C_1 \frac{\|g_k\|^2}{\|d_k\|^2}. \tag{2.4}$$

The following theorem gives an estimation of the global complexity bound for the MHS method.

Theorem 2.1. *Suppose that Assumption 1 hold. Let the sequence $\{x_k\}$ be generated by the MHS method. Let J be the first iterative such that $\|g_J\| \leq \epsilon$. Then,*

$$J \leq C_2 \epsilon^{-(2+2r)}, \tag{2.5}$$

where C_2 is a positive constant.

Proof. By (1.2)-(1.3), (2.1), (1.6) and (2.3), we get

$$\begin{aligned}\|d_k\| &\leq \|g_k\| + |\beta_k^{MHS}|\|d_{k-1}\| + |\theta_k^{MHS}|\|z_{k-1}\| \\ &\leq \|g_k\| + \frac{2\|g_k\|\|z_{k-1}\|\|d_{k-1}\|}{d_{k-1}^T z_{k-1}} \\ &\leq \|g_k\| + \frac{2M_1\|g_k\|}{t\|g_{k-1}\|^r} \\ &= \Big(1 + \frac{2M_1}{t\|g_{k-1}\|^r}\Big)\|g_k\|,\end{aligned}$$

which implies that

$$\frac{\|g_k\|}{\|d_k\|} \geq \frac{t\|g_{k-1}\|^r}{t\|g_{k-1}\|^r + 2M_1}.$$

This together with (2.2) yields that

$$\frac{\|g_k\|}{\|d_k\|} \geq \frac{t\|g_{k-1}\|^r}{tM^r + 2M_1}. \tag{2.6}$$

Since the sequence $\{f(x_k)\}$ decreases and converges, we suppose that $\lim_{k\to\infty} f(x_k) = f_{min}$. Then from the line search (1.5), (1.9), Lemma 2.1 and (2.6), we obtain

$$\begin{aligned}f(x_0) - f_{min} &\geq f(x_0) - f(x_k) = \sum_{j=0}^{k-1}\big(f(x_j) - f(x_{j+1})\big) \\ &\geq \sum_{j=0}^{k-1}(-\delta\alpha_j g_j^T d_j) \geq \sum_{j=0}^{k-1}\delta C_1 \frac{\|g_j\|^2}{\|d_j\|^2}\|g_j\|^2 \\ &\geq \sum_{j=1}^{k-1} C_1\delta\Big(\frac{t\|g_{j-1}\|^r}{tM^r + 2M_1}\Big)^2\|g_j\|^2 \\ &\geq \sum_{j=1}^{k-1}\frac{C_1\delta t^2}{(tM^r + 2M_1)^2}\Big(\min_{0\leq j\leq k-1}\|g_j\|\Big)^{(2+2r)} \\ &= C_3(k-1)\Big(\min_{0\leq j\leq k-1}\|g_j\|\Big)^{(2+2r)},\end{aligned}$$

where $C_3 = \frac{C_1\delta t^2}{(tM^r+2M_1)^2}$. Hence, we have

$$\min_{0\leq j\leq k-1}\|g_j\| \leq \Big(\frac{f(x_0) - f_{min}}{C_3(k-1)}\Big)^{\frac{1}{2+2r}}.$$

Therefore,

$$k \geq \frac{f(x_0) - f_{min}}{C_3}\epsilon^{-(2+2r)} + 1 \Rightarrow \min_{0\leq j\leq k-1}\|g_j\| \leq \epsilon.$$

This finishes the proof. □

Remark 2.1 This theorem shows that the global complexity bound of the MHS method is $O(\epsilon^{-(2+2r)})$. We also can see that the smaller the parameter r is, the better this bound becomes.

3 The global complexity bound for the MFR method

In this section, we study the global complexity bound for the MFR method.

From (1.7) and (1.9), we have

$$\|d_k\|^2 = (\beta_k^{FR})^2\|d_{k-1}\|^2 + 2\theta_k^{MFR}\|g_k\|^2 - (\theta_k^{MFR})^2\|g_k\|^2.$$

This and (1.8) yield that

$$\begin{aligned} \frac{\|d_k\|^2}{\|g_k\|^4} &= \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} - \frac{(\theta_k^{MFR}-1)^2}{\|g_k\|^2} + \frac{1}{\|g_k\|^2} \\ &\le \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + \frac{1}{\|g_k\|^2} \\ &\le \sum_{j=0}^{k} \frac{1}{\|g_j\|^2}, \end{aligned}$$

which shows that

$$\frac{\|g_k\|^4}{\|d_k\|^2} \ge \frac{1}{\sum_{j=0}^{k} \frac{1}{\|g_j\|^2}}. \tag{3.1}$$

Now we present the theorem on the global complexity bound for the MFR method.

Theorem 3.1. *Suppose that Assumption 1 hold. Let the sequence $\{x_k\}$ be generated by the MFR method. Let J be the first iterative such that $\|g_J\| \le \epsilon$. Then,*

$$J \le C_4 a^{\epsilon^{-2}}, \tag{3.2}$$

where C_4 and $a > 1$ are two positive constants.

Proof. Similar to Theorem 2.1, from (3.1), we have

$$\begin{aligned} f(x_0) - f_{min} &\ge \sum_{j=0}^{k-1} \big(f(x_j) - f(x_{j+1})\big) \\ &\ge \delta C_1 \sum_{j=0}^{k-1} \frac{\|g_j\|^4}{\|d_j\|^2} \ge \delta C_1 \sum_{j=0}^{k-1} \frac{1}{\sum_{l=0}^{j} \frac{1}{\|g_l\|^2}} \\ &\ge \delta C_1 \Big(\sum_{j=0}^{k-1} \frac{1}{j+1}\Big)\big(\min_{0\le j\le k-1} \|g_j\|\big)^2 \\ &= \delta C_1 H_k \big(\min_{0\le j\le k-1} \|g_j\|\big)^2, \end{aligned}$$

where $H_k = \sum_{j=1}^{k} \frac{1}{j}$ is the sum of the first k-terms of the Harmonic series. It is well-know [2] that

$$H_k = \ln k + \gamma + \eta_k, \tag{3.3}$$

where γ is the Euler constant and $\eta_k \approx \frac{1}{2k}$ converges to 0. Therefore, we have

$$\min_{0\le j\le k-1} \|g_j\| \le \Big(\frac{f(x_0) - f_{min}}{\delta C_1 H_k}\Big)^{1/2},$$

$$H_k \geq \frac{f(x_0) - f_{min}}{C_1\delta}\epsilon^{-2} \Rightarrow \min_{0\leq j\leq k-1} \|g_j\| \leq \epsilon.$$

Hence, by (3.3), we have

$$\ln k \geq \frac{f(x_0) - f_{min}}{C_1\delta}\epsilon^{-2} - \gamma - \eta_k \Rightarrow \min_{0\leq j\leq k-1} \|g_j\| \leq \epsilon,$$

which implies that there exist two constants $C_4 > 0$ and $a > 1$ such that

$$k \geq C_4 a^{\epsilon^{-2}}.$$

The proof is then complete.

□

4 Conclusions

We have studied the global complexity bound of two existing nonlinear conjugate gradient methods for the nonconvex unconstrained optimization problem (1.1). The bound is useful when we want to estimate in advance the worst computational cost for a given accuracy of a solution. How to estimate the global complexity bounds for other conjugate gradient methods solving nonconvex optimization is our future work.

Acknowledgements This work was supported by the Project of the Scientific Research Fund (11C0251 and 13B137) of the Hunan Provincial Education Department and the NSF (14JJ3084) of Hunan Province.

References

[1] C. Cartis, N. I. M. Gould and Ph. L. Toint, On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization. SIAM J. Optim., 20 (2010) 2833-2852.

[2] http : //en.wikipedia.org/wiki/Harmonic_series_(mathematics).

[3] Y. Nesterov, Introductory Lectures on Convex Optimization. Kluwer Academic, Dordrecht, 2004.

[4] Y. Nesterov and B. T. Polyak, Cubic regularization of Newton method and its global performance. Math. Program., Ser. A, 108 (2006) 177-205.

[5] R. A. Polyak, Regularized Newton method for unconstrained convex optimization. Math. Program., Ser. B, 120 (2009) 125-145.

[6] K. Ueda and N. Yamashita, Convergence properties of the regularized Newton method for the unconstrained nonconvex optimization. Appl. Math. Optim., 62 (2010) 27-46.

[7] K. Ueda and N. Yamashita, A regularized Newton method without line search for unconstrained optimization. Technical Report 2009-007, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, 2009.

[8] L. Zhang, W. Zhou and D. Li, Some descent three-term conjugate gradient methods and their global convergence, Optim. Meth. Softw., 22 (2007) 697-711.

[9] L. Zhang, W. Zhou and D. Li, Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search, Numer. Math., 104 (2006) 561-572.