Journal of the Operations Research
Society of Japan
2004, Vol. 47, No. 2, 63-72

GLOBAL CONVERGENCE RESULTS OF A NEW THREE-TERM MEMORY GRADIENT METHOD

Sun Qingying Liu Xinhai
University of Petroleum

(Received March 5, 2002; Received January 7, 2004)

Abstract In this paper, a new class of three-term memory gradient methods with Armijo-like step size rule for unconstrained optimization is presented. Global convergence properties of the new methods are discussed without assuming that the sequence $\{x_k\}$ of iterates is bounded. Moreover, it is shown that, when $f(x)$ is pseudo-convex (quasi-convex) function, this new method has strong convergence results. Combining FR, PR, HS methods with our new method, FR, PR, HS methods are modified to have global convergence property. Numerical results show that the new algorithms are efficient.

Keywords: Nonlinear programming, three-terms memory gradient method, Armijo-like step size rule, convergence, numerical experiment

1. Introduction

Consider the following unconstrained problem

$$\min\{f(x) : x \in R^n\}, \tag{1}$$

where $f : R^n \to R$ is a continuously differentiable function.

In [2], the memory gradient algorithm for problem (1) was first presented. Compared with the ordinary gradient method, this algorithm has the advantage of high speed. Cragg and Levy [1] made a generalization of the memory gradient algorithm and presented a method called the super-memory gradient algorithm which from numerical experience has been shown to be much more rapidly convergent, in general, than the memory gradient algorithm.

In this paper, we consider a new three-terms memory gradient method for problem (1) whose search directions are defined by

$$d_k = -\nabla f(x_k) + \beta_k d_{k-1} + \alpha_k d_{k-2}, \tag{2}$$

and

$$x_{k+1} = x_k + \lambda_k d_k, \tag{3}$$

where β_k and α_k are parameters and λ_k is a step-size obtained by means of a one-dimensional search. Conditions are given on β_k and α_k to ensure that d_k is a sufficient descent direction at the point x_k of iterate. Global convergence properties of the new class of three terms memory gradient methods with Armijo-like step size rule are discussed without assuming that the sequence $\{x_k\}$ of iterates is bounded. Moreover, it is shown that, when $f(x)$ is pseudo-convex (quasi-convex) function, this new method has strong convergence results.

Combining FR, PR, HS methods with our new method, FR, PR, HS methods are modified to have global convergence property. Numerical results show that the new algorithms are efficient.

In Section 2, we present a new method. We start the convergence analysis of the new method in Section 3. The convergence properties for generalized convex functions are discussed in Section 4. Finally, a detailed list of the test problems that we have used is given in Section 5.

2. The New Three-term Memory Gradient Algorithm

Consider the three-term memory gradient method (2) and (3). Let $S_k = -\nabla f(x_k) + \beta_k d_{k-1}$. In order to ensure that d_k is a sufficient descent direction, we assume that

$$\begin{cases} \nabla f(x_k)^T \nabla f(x_k) > |\beta_k \nabla f(x_k)^T d_{k-1}|, \\ \|\nabla f(x_k)^T S_k\| \geq (1 + \triangle_1^k)|\beta_k| \cdot \|\nabla f(x_k)\| \cdot \|d_{k-1}\| \end{cases} \tag{4}$$

and

$$\begin{cases} |\nabla f(x_k)^T S_k| > |\alpha_k \nabla f(x_k)^T d_{k-2}|, \\ |\nabla f(x_k)^T d_k| \geq (1 + \triangle_2^k)|\alpha_k| \cdot \|\nabla f(x_k)\| \cdot \|d_{k-2}\| \end{cases} \tag{5}$$

where $\triangle_1^k > 0$, $\triangle_2^k > 0$ are parameters.

Condition (4) plays a vital role in choosing β_k, and a new choice for β_k is given by

$$\beta_k \in [-\underline{\beta}_k(\triangle_1^k), \overline{\beta}_k(\triangle_1^k)], \tag{6}$$

$$\overline{\beta}_k(\triangle_1^k) = \frac{1}{(1 + \triangle_1^k) + \cos\theta_k} \cdot \frac{\|\nabla f(x_k)\|}{\|d_{k-1}\|}, \tag{7}$$

$$\underline{\beta}_k(\triangle_1^k) = \frac{1}{(1 + \triangle_1^k) - \cos\theta_k} \cdot \frac{\|\nabla f(x_k)\|}{\|d_{k-1}\|}, \tag{8}$$

where θ_k is the angle between $\nabla f(x_k)$ and d_{k-1}.

Condition (5) plays a vital role in choosing α_k, and a new choice for α_k is given by

$$\alpha_k \in [-\underline{\alpha}_k(\triangle_1^k, \triangle_2^k), \overline{\alpha}_k(\triangle_1^k, \triangle_2^k)], \tag{9}$$

$$\overline{\alpha}_k(\triangle_1^k, \triangle_2^k) = \frac{1 + \triangle_1^k}{2 + \triangle_1^k} \cdot \frac{1}{(1 + \triangle_2^k) + \cos\overline{\theta}_k} \cdot \frac{\|\nabla f(x_k)\|}{\|d_{k-2}\|}, \tag{10}$$

$$\underline{\alpha}_k(\triangle_1^k, \triangle_2^k) = \frac{1 + \triangle_1^k}{2 + \triangle_1^k} \cdot \frac{1}{(1 + \triangle_2^k) - \cos\overline{\theta}_k} \cdot \frac{\|\nabla f(x_k)\|}{\|d_{k-2}\|}, \tag{11}$$

where $\overline{\theta}_k$ is the angle between $\nabla f(x_k)$ and d_{k-2}.

The new three-terms memory gradient algorithm (NTMG):

Data: $\forall\ x_1 \in R^n$, $d_0 = 0$, $\triangle_1^0 > 0$, $\triangle_2^0 > 0$, $\mu_1, \mu_2 \in (0, 1)$ and $\mu_1 \leq \mu_2$, γ_1, $\gamma_2 > 0$, $\gamma_2 < 1$.

Step1: Compute$\nabla f(x_1)$, if $\nabla f(x_1) = 0$, and x_1 is a stationary point of (1), stop; else set $d_1 = -\nabla f(x_1), k := 1$, and go to step2.

Step2: $x_{k+1} = x_k + \lambda_k d_k$, the step size λ_k is chosen so that

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \mu_1 \lambda_k \nabla f(x_k)^T d_k, \tag{12}$$

and

$$\lambda_k \geq \gamma_1 \text{ or } \lambda_k \geq \gamma_2 \lambda_k^* > 0, \tag{13}$$

where λ_k^* satisfies

$$f(x_k + \lambda_k^* d_k) > f(x_k) + \mu_2 \lambda_k^* \nabla f(x_k)^T d_k, \tag{14}$$

Step3: Compute $\nabla f(x_{k+1})$. if $\|\nabla f(x_{k+1})\| = 0$, and x_{k+1} is a stationary point of (1), stop; else let $k := k+1$, $\triangle_1^k \geq \triangle_1^0$, $\triangle_2^k \geq \triangle_2^0$, and go to step4.

Step4: Let $d_k = -\nabla f(x_k) + \beta_k d_{k-1} + \alpha_k d_{k-2}$,
where $\beta_k \in [-\underline{\beta}_k(\triangle_1^k), \overline{\beta}_k(\triangle_1^k)]$, $\alpha_k \in [-\underline{\alpha}_k(\triangle_1^k, \triangle_2^k), \overline{\alpha}_k(\triangle_1^k, \triangle_2^k)]$, go to step 2.

Remark We can give the new choice of the parameter β_k:

$$\beta_k = \operatorname{argmin}\{|\beta - \beta_k^{FR}||\beta \in [-\underline{\beta}_k(\triangle_1^k), \overline{\beta}_k(\triangle_1^k)]\};$$
$$\beta_k = \operatorname{argmin}\{|\beta - \beta_k^{PR}||\beta \in [-\underline{\beta}_k(\triangle_1^k), \overline{\beta}_k(\triangle_1^k)]\};$$
$$\beta_k = \operatorname{argmin}\{|\beta - \beta_k^{HS}||\beta \in [-\underline{\beta}_k(\triangle_1^k), \overline{\beta}_k(\triangle_1^k)]\};$$

where $\beta_k^{FR} = \| g_k \|^2 / \| g_{k-1} \|^2$ (Fletcher-Reeves), $\beta_k^{PR} = g_k^T(g_k - g_{k-1})/\| g_{k-1} \|^2$ (Polak-Ribiere), $\beta_k^{HS} = (g_k^T(g_k - g_{k-1}))/d_{k-1}^T(g_k - g_{k-1})$ (Hestenes-Stiefel), and three classes of new methods are established, denoted by NTFR, NTPR, NTHS, respectively. In particular, we can take $\alpha_k = 0$ in NTMG, NTFR, NTPR, NTHS methods, and four classes of new methods are established, denoted by NCG, NFR, NPR, NHS, respectively.

Lemma 1 If x_k is not a stationary point for problem (1), then $\| d_k \| \leq c_1 \|\nabla f(x_k)\|$, where $c_1 = 1 + \frac{1}{\triangle_1^0} + \frac{1}{\triangle_2^0}$.

PROOF. It follows from the definition of d_k. □

Lemma 2 If x_k is not a stationary point for problem (1), then d_k is a descent direction, i.e. $\nabla f(x_k)^T d_k \leq -c_2 \cdot \|\nabla f(x_k)\|^2$, where $c_2 = \frac{1+\triangle_1^0}{2+\triangle_1^0} \cdot \frac{1+\triangle_2^0}{2+\triangle_2^0}$.

PROOF. For $k = 1$, it is clear that $d_1 = -\nabla f(x_1)$ is a descent direction. For $k \geq 2$, by using assumption (4) and the definition of S_k, we have

$$\begin{aligned} \nabla f(x_k)^T S_k &= -\|\nabla f(x_k)\|^2 + \beta_k \cdot \nabla f(x_k)^T d_{k-1} \\ &\leq -\|\nabla f(x_k)\|^2 + |\beta_k \cdot \nabla f(x_k)^T d_{k-1}| \\ &\leq -\|\nabla f(x_k)\|^2 + \|\nabla f(x_k)\|^2 \\ &= 0. \end{aligned}$$

It follows from (4) that

$$\begin{aligned} \nabla f(x_k)^T S_k &\leq -\|\nabla f(x_k)\|^2 + |\beta_k \cdot \nabla f(x_k)^T d_{k-1}| \\ &\leq -\|\nabla f(x_k)\|^2 + \frac{1}{1+\triangle_1^k} |\nabla f(x_k)^T S_k|. \end{aligned}$$

The above inequality and $|\nabla f(x_k)^T S_k| = -\nabla f(x_k)^T S_k$ imply that

$$\nabla f(x_k)^T S_k \leq -\frac{1+\triangle_1^k}{2+\triangle_1^k} \cdot \|\nabla f(x_k)\|^2. \tag{15}$$

Since for $k = 2$, d_2 is identical with s_2, the result follows from equation (15). For $k \geq 3$, it follows from (5) and the definition of d_k and (15) that

$$\nabla f(x_k)^T d_k \leq -\frac{1+\triangle_2^k}{2+\triangle_2^k} \cdot |\nabla f(x_k)^T S_k| \leq -\frac{1+\triangle_1^k}{2+\triangle_1^k} \cdot \frac{1+\triangle_2^k}{2+\triangle_2^k} \cdot \|\nabla f(x_k)\|^2.$$

By using $\frac{1+\triangle_1^k}{2+\triangle_1^k} \geq \frac{1+\triangle_1^0}{2+\triangle_1^0}$, for $\triangle_1^k \geq \triangle_1^0$ and $\frac{1+\triangle_2^k}{2+\triangle_2^k} \geq \frac{1+\triangle_2^0}{2+\triangle_2^0}$, for $\triangle_2^k \geq \triangle_2^0$, we obtain that $\nabla f(x_k)^T d_k \leq -\frac{1+\triangle_1^k}{2+\triangle_1^k} \cdot \frac{1+\triangle_2^k}{2+\triangle_2^k} \cdot \|\nabla f(x_k)\|^2$. □

3. Convergence Analysis

Throughout this paper, let $\{x_k\}$ denote the sequence generated by (NTMG). If $\nabla f(x_k) = 0$ for a finite integer k, x_k is a stationary point of (1). In what follows, we assume that (NTMG) generates an infinite sequence. We now present our global convergence results.

Theorem 1 Suppose that $f(x) \in C^1$. Then:

(i) either $f(x_k) \to -\infty$ or $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$;

(ii) either $f(x_k) \to -\infty$ or $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$, if ∇f is uniformly continuous on R^n .

Proof. Since for all k, $\nabla f(x_k)^T d_k < 0$, we have$f(x_{k+1}) < f(x_k)$, which implies that $\{f(x_k)\}$ is a monotonically decreasing sequence. If $f(x_k) \to -\infty$, then we complete the proof. Therefore, in the following discussion, we assume that $\{f(x_k)\}$ is a bounded set.

Suppose (i) is not true. Then, there exists $\varepsilon > 0$ such that, for all k,

$$\nabla f(x_k) \geq \varepsilon. \tag{16}$$

It follows from Lemma 2, (12) and (16) that

$$f(x_k + 1) - f(x_k) \leq \mu_1 \lambda_k \nabla f(x_k)^T d_k \leq -c_2 \lambda_k \mu_1 \|\nabla f(x_k)\|^2 \leq -c_2 \lambda_k \mu_1 \varepsilon \|\nabla f(x_k)\|. \tag{17}$$

The above inequality and the boundedness of $\{f(x_k)\}$ imply that

$$\sum_{k=1}^{\infty} \lambda_k \|\nabla f(x_k)\| < +\infty. \tag{18}$$

It follows from Lemma 1 and (2) that, for all k,

$$\|x_{k+1} - x_k\| = \lambda_k \|d_k\| \leq c_1 \lambda_k \|\nabla f(x_k)\|.$$

The above inequalities and (18) yield $\sum_{k=1}^{\infty} \|x_{k+1} - x_k\| < +\infty$, which yields that$\{x_k\}$ is convergent, say to a point x_*. From (16), (18), we have

$$\lim_{k\to\infty} \lambda_k = 0. \tag{19}$$

It follows from Lemma 1, the convergence of $\{x_k\}$ and $f(x) \in C^1$ that $\{d_k\}$ is bounded. Without loss of generality, we may assume that there exists an index set $K \subset \{1, 2, ...\}$ such that $\lim_{k\to\infty, k\in K} d_k = d_*$. It follows from (13) and (19) that, when $k (k \in K)$ is large enough, we have $\lambda_k < \gamma_1$, and hence it follows from (13) that, $\lambda_k \geq \gamma_2 \lambda_k^*$, where λ_k^* satisfies (14), i.e. $f(x_k + \lambda_k^* d_k) - f(x_k)/\lambda_k^* \geq \mu_2 \lambda_k^* \nabla f(x_k)^T d_k$. Taking the limit for $k \in K$, we have

$$\nabla f(x_*)^T d_* \geq \mu_2 \nabla f(x_*)^T d_*. \tag{20}$$

By using (20) and $\mu_2 \in (0,1)$, we obtain that

$$\nabla f(x_*)^T d_* = 0. \tag{21}$$

It follows from Lemma 2 and (21) that $\|\nabla f(x_*)\| = 0$, which contradicts (16). This completes the proof of (i).

Suppose that there exist an infinite index set $K_1 \subset \{1,2,...\}$ and a positive scalar $\varepsilon > 0$ such that, for all $k \in K_1$,

$$\nabla f(x_k) > \varepsilon. \tag{22}$$

It follows from Lemma 2 and (12)that

$$f(x_k) - f(x_{k+1}) \geq -\mu_1 \lambda_k \nabla f(x_k)^T d_k \geq c_2 \lambda_k \mu_1 \|\nabla f(x_k)\|^2. \tag{23}$$

By using (22) and (23), we obtain that $\lambda_k \leq \mu_1^{-1}\epsilon^{-2}c_2^{-1}(f(x_k) - f(x_{k+1}), \forall k \in K_1$.

The boundedness of $\{f(x_k)\}$ and the monotonically decreasing property imply that $\{f(x_k)\}$ is convergent. Thus,

$$\limsup_{k\to\infty, k\in K_1} \lambda_k \leq \limsup_{k\to\infty, k\in K_1} \mu_1^{-1}\epsilon^{-2}c_2^{-1}(f(x_k) - f(x_{k+1}),$$

which yields that

$$\limsup_{k\to\infty, k\in K_1} \lambda_k = 0. \tag{24}$$

It follows from (22) and (23) that $\lambda_k \nabla f(x_k) \leq \mu_1^{-1}\epsilon^{-1}c_2^{-1}(f(x_k) - f(x_{k+1}))$, and $\limsup_{k\to\infty, k\in K_1} \lambda_k \nabla f(x_k) \leq \limsup_{k\to\infty, k\in K_1} \mu_1^{-1}\epsilon^{-1}c_2^{-1}(f(x_k) - f(x_{k+1}))$. Hence,

$$\limsup_{k\to\infty, k\in K_1} \lambda_k \nabla f(x_k) = 0. \tag{25}$$

It follows from Lemma 1 and (25) that

$$\limsup_{k\to\infty, k\in K_1} \lambda_k \|d_k\| \leq \limsup_{k\to\infty, k\in K_1} c_1 \lambda_k \|\nabla f(x_k)\|.$$

i.e.

$$\limsup_{k\to\infty, k\in K_1} \lambda_k \|d_k\| = 0. \tag{26}$$

It follows from (24) that, when $k (k \in K_1)$ is large enough, we have $\lambda_k < \gamma_1$, and hence it follows from (13) that, $\lambda_k \geq \gamma_2 \lambda_k^*$, where λ_k^* satisfies (14). Now set $x_{k+1}^* = x_k + \lambda_k^* d_k$. It follows from (24), (26) and $\lambda_k \geq \gamma_2 \lambda_k^*$, ($k \in K_1$ is large enough) that $\lim_{k\to\infty, k\in K_1} \lambda_k^* = 0$ and $\lim_{k\to\infty, k\in K_1} \lambda_k^* \|d_k\| = 0$. Hence, $\lim_{k\to\infty, k\in K_1} \|x_{k+1}^* - x_k\| = 0$.

Let $\rho_k^* = \frac{f(x_{k+1}^*) - f(x_k)}{\lambda_k^* \nabla f(x_k)^T d_k}$, $k \in K_1$, it follows from (14) that

$$\rho_k^* < \mu_2 < 1, k \in K_1. \tag{27}$$

It follows from Lemmas 1, 2 and (22) that

$$\begin{aligned}
\limsup_{k\to\infty, k\in K_1} |\rho_k^* - 1| &= \limsup_{k\to\infty, k\in K_1} |\frac{\nabla f(\xi_k^*)^T (\lambda_k^* d_k)}{\lambda_k^* \nabla f(x_k)^T d_k} - 1| \\
&= \limsup_{k\to\infty, k\in K_1} |\frac{(\nabla f(\xi_k^*) - \nabla f(x_k))^T d_k}{\nabla f(x_k)^T d_k}| \leq \limsup_{k\to\infty, k\in K_1} \frac{\|\nabla f(\xi_k^*) - \nabla f(x_k)\| \cdot \|d_k\|}{|\nabla f(x_k)^T d_k|} \\
&\leq \limsup_{k\to\infty, k\in K_1} \frac{\|\nabla f(\xi_k^*) - \nabla f(x_k)\| \cdot c_1 \cdot \|\nabla f(x_k)\|}{c_2 \cdot \|\nabla f(x_k)\|^2} \\
&\leq \limsup_{k\to\infty, k\in K_1} \frac{\|\nabla f(\xi_k^*) - \nabla f(x_k)\| \cdot c_1}{c_2 \cdot \varepsilon} = 0,
\end{aligned} \tag{28}$$

where $\xi_k^* = x_k + \vartheta_k(x_{k+1}^* - x_k), 0 < \vartheta_k < 1, k \in K_1$.
Hence (28) establishes that $\rho_k^* \geq \mu_2$ for all $k \in K_1$ sufficiently large. This is the desired contradiction because (27) guarantees that $\rho_k^* < \mu_2$. This yields (ii). □

4. Convergence Properties for Generalized Convex Functions

In this section, we discuss the convergence properties of (NTMG) for generalized convex functions. As shown in the following, parameters $\triangle_1^k, \triangle_2^k$ play an important role in our analysis. We make the following assumption:

(Q) For any integer k,

$$\triangle_1^k \geq \max\{\triangle_1^0, \frac{1+\|x_k\|}{f(x_{k-1})-f(x_k)}\|\nabla f(x_k)\|\},$$
$$\triangle_2^k \geq \max\{\triangle_2^0, \frac{1+\|x_k\|}{f(x_{k-1})-f(x_k)}\|\nabla f(x_k)\|\}.$$

Thus we have the following results.

Lemma 3 Suppose that (Q) holds and $f(x) \in C^1$. Let $\lambda_0 = \sup\{\lambda_k, k = 1, 2, ...\}$ and suppose that $\lambda_0 < +\infty$. If $f(x)$ is a quasi-convex function and the solution set of problem (1) is nonempty, then $\{x_k\}$ is a bounded sequence, each accumulation point x_* of which is a stationary point of problem (1) and $\lim_{k\longrightarrow\infty} x_k = x_*$.

PROOF. Note that for all $x \in R^n$ and all k,

$$\begin{aligned}
\|x_{k+1}-x\|^2 &= \|x_k - x\|^2 + 2(x_{k+1}-x, x_k - x) + \|x_{k+1}-x_k\|^2 \\
&= \|x_k - x\|^2 + 2\lambda_k(d_k, x_k - x) + \lambda_k^2\|d_k\|^2 \\
&= \|x_k - x\|^2 + 2\lambda_k(-\nabla f(x_k) + \beta_k d_{k-1} + \alpha_k d_{k-2}, x_k - x) + \lambda_k^2\|d_k\|^2 \\
&\leq \|x_k - x\|^2 + 2\lambda_k(\nabla f(x_k), x - x_k) \\
&\quad +2\lambda_k|\beta_k|\|d_{k-1}\|\|x_k - x\| + 2\lambda_k|\alpha_k|\|d_{k-2}\|\|x_k - x\| + \lambda_k^2\|d_k\|^2 \\
&\leq \|x_k - x\|^2 + 2\lambda_k(\nabla f(x_k), x - x_k) \\
&\quad +4\lambda_k\frac{f(x_{k-1})-f(x_k)}{1+\|x_k\|}(\|x_k\| + \|x\|) + \lambda_k^2\|d_k\|^2 \\
&\leq \|x_k - x\|^2 + 2\lambda_k(\nabla f(x_k), x - x_k) \\
&\quad +4\lambda_0(1+\|x\|)(f(x_{k-1})-f(x_k)) + \lambda_k^2\|d_k\|^2.
\end{aligned} \tag{29}$$

It follows from Lemma 1, Lemma 2 and (12) that

$$\|d_k\|^2 \leq c_1^2\|\nabla f(x_k)\|^2; \tag{30}$$
$$\|\nabla f(x_k)\|^2 \leq c_2^{-1}(-\nabla f(x_k))^T d_k); \tag{31}$$
$$-\lambda_k\nabla f(x_k))^T d_k \leq \mu_1^{-1}((f_k) - f(x_{k+1})). \tag{32}$$

By using (29), (30), (31), (32) and the above inequality, we obtain that

$$\begin{aligned}
\|x_{k+1}-x\|^2 &\leq \|x_k - x\|^2 + 2\lambda_k(\nabla f(x_k), x - x_k) \\
&\quad +4\lambda_0(1+\|x\|)(f(x_{k-1})-f(x_k)) + \lambda_0\lambda_k c_1^2 c_2^{-1}(-\nabla f(x_k)^T d_k) \\
&\leq \|x_k - x\|^2 + 2\lambda_k(\nabla f(x_k), x - x_k) \\
&\quad +4\lambda_0(1+\|x\|)(f(x_{k-1})-f(x_k)) + \lambda_0 c_1^2 c_2^{-1}\mu_1^{-1}(f(x_k) - f(x_{k-1}))). \\
&= \|x_k - x\|^2 + 2\lambda_k(\nabla f(x_k), x - x_k) \\
&\quad +m_1(x)(f(x_{k-1})-f(x_k)) + m_2(f(x_k) - f(x_{k-1}))),
\end{aligned} \tag{33}$$

where $m_1(x) = 4\lambda_0(1 + \|x\|), m_2 = \lambda_0\mu_1^{-1}c_1^2c_2^{-1}$.
Because the solution set of problem (1) is nonempty, we can choose $y \in R^n$ satisfying $f(y) \le f(x_k)$. Since $f(x)$ is a quasi-convex function, we have

$$(\nabla f(x_k), y - x_k) \le 0. \tag{34}$$

It follows from (33), (34) that

$$\|x_{k+1} - y\|^2 + m_1(y)f(x_k) + m_2 f(x_{k+1}) \le \|x_k - y\|^2 + m_1(y)f(x_{k-1}) + m_2 f(x_k)$$

which implies the sequence $\{\|x_k - y\|^2 + m_1(y)f(x_{k-1}) + m_2 f(x_k)\}$ is descent. Since we have assumed that the solution set of problem (1) is nonempty, and so $\inf\{f(x_k) : k = 1, 2, ...\} > -\infty$ both sequence $\{f(x_k)\}$ and $\{\|x_k - y\|^2 + m_1(y)f(x_{k-1}) + m_2 f(x_k)\}$ are bounded from below and converge. Therefore, the sequence $\{\|x_k - y\|^2\}$ converges and $\{x_k\}$ is bounded. This implies that $\{x_k\}$ has an accumulation point x_* and that there exists an index set $K_1 \subset \{1, 2, ...\}$ such that $\lim_{k\to\infty, k\in K_1} x_k = x_*$, and $\lim_{k\to\infty, k\in K_1} f(x_k) = f(x_*)$. It follows from the above equation and the fact $\{f(x_k)\}$ is a monotonically decreasing sequence implies $\lim_{k\to\infty, k\in K_1} f(x_{k-1}) = f(x_*)$. Therefore, we have

$$\lim_{k\to\infty} \{\|x_k - x_*\|^2 + m_1(x_*)f(x_{k-1}) + m_2 f(x_k)\}$$

$$= \lim_{k\to\infty, k\in K_1} \{\|x_k - x_*\|^2 + m_1(x_*)f(x_{k-1}) + m_2 f(x_k)\}$$

$$= [m_1(x_*) + m_2]f(x_*),$$

which implies $\lim_{k\to\infty} x_k = x_*$. From Theorem 1 the limit point x_* is a stationary point of problem (1). □

Theorem 2 Suppose that (Q) holds and $f(x) \in C^1$. Let $\lambda_0 = \sup\{\lambda_k, k = 1, 2, ...\}$ and suppose that $\lambda_0 < +\infty$. If $f(x)$ is a pseudo-convex function, then:

(i) $\{x_k\}$ is a bounded sequence if and only if the solution set of problem (1) is nonempty;

(ii) $\lim_{k\to\infty} f(x_k) = \inf\{f(x) : x \in R^n\}$;

(iii) If the solution set of problem (1) is nonempty, then any accumulation point x_* of $\{x_k\}$ is an optimal solution of problem (1) and $\lim_{k\to\infty} x_k = x_*$.

PROOF. Since $f(x)$ is pseudo-convex, it is quasi-convex and a stationary point of problem (1) is also an optimal solution of problem (1).
First, we will show part (i). If $\{x_k\}$ is a bounded sequence, then it follows from Theorem 1 that there exists an index set $K_2 \subset \{1, 2, ...\}$ and a point $x_* \in R^n$ such that $\lim_{k\to\infty, k\in K_1} x_k = x_*$, and x_* is a stationary point of problem (1), and is also an optimal solution of problem (1). Conversely, if the solution set of problem (1) is nonempty, then it follows from Lemma 3 that $\{x_k\}$ is a bounded sequence.

Next, we will prove (ii). We prove this conclusion by the following three cases (a), (b), (c).

(a) $\lim_{k\to\infty} f(x_k) = \inf\{f(x) : k = 1, 2, ...\} = -\infty$; It follows from $\{f(x_k)\}$ is a descent sequence, and $\lim_{k\to\infty} f(x_k) = \inf\{f(x) : k = 1, 2, ...\} \ge \inf\{f(x) : x \in R^n\}$.

(b) $\{x_k\}$ is bounded: It follows from (i) of this theorem that the solution of problem (1) is nonempty, and there exists an index set $K_3 \subset \{1, 2, ...\}$ and a point $x_* \in R^n$ such that $\lim_{k\to\infty, k\in K_1} x_k = x_*$, it follows from Theorem 1 that x_* is a stationary point of problem (1), and is also an optimal solution of problem (1).

(c) $\inf\{f(x) : k = 1, 2, ...\} > -\infty$; and $\{x_k\}$ is unbounded: Suppose that there exists $\bar{x} \in R^n$, $\varepsilon > 0$, and k_1 such that for all $k \geq k_1$, $f(x_k) > f(\bar{x}) + \varepsilon$. Since $f(x)$ is a pseudo-convex function, we have $(\nabla f(x_k), \bar{x} - x_k) \leq 0$, for all $k \geq k_1$. Setting $x = \bar{x}$ in (33) that

$$\|x_{k+1} - \bar{x}\|^2 + m_1(\bar{x})f(x_k) + m_2 f(x_{k-1}) \leq \|x_k - \bar{x}\|^2 + m_1(\bar{x})f(x_{k-1}) + m_2 f(x_k),$$

which implies the sequence $\{\|x_k - \bar{x}\|^2 + m_1(\bar{x})f(x_{k-1}) + m_2 f(x_k)\}$ is descent. Since we have assumed that $\inf\{f(x) : k = 1, 2, ...\} > -\infty$; both sequence $\{f(x_k)\}$ and $\{\|x_k - \bar{x}\|^2 + m_1(\bar{x})f(x_{k-1}) + m_2 f(x_k)\}$ are bounded from below and converge. Therefore, the sequence $\{\|x_k - \bar{x}\|^2\}$ converges and $\{x_k\}$ is bounded, which contradicts our assumption.

(iii) immediately follows from Lemma 3. □

Corollary 1 Suppose that (Q) holds and $f(x) \in C^1$. Let $\lambda_0 = \{\sup\{\lambda_k, k = 1, 2, ...\}$ and suppose that $\lambda_0 < +\infty$. If $f(x)$ is a convex function, then:

(i) $\{x_k\}$ is a bounded sequence if and only if the solution set of problem (1) is nonempty;

(ii) $\lim_{k\to\infty} f(x_k) = \inf\{f(x) : x \in R^n\}$.

(iii) If the solution set of problem (1) is nonempty , then any accumulation point x_* of $\{x_k\}$ is an optimal solution of problem (1) and $\lim_{k\to\infty} x_k = x_*$.

Proof. Since $f(x)$ is convex, it is pseudo-convex. It immediately follows from Theorem 2. □

Corollary 2 Suppose that (Q) holds and $f(x) \in C^1$. Let $\lambda_0 = \sup\{\lambda_k, k = 1, 2, ...\}$ and suppose that $\lambda_0 < +\infty$. If $f(x_)$ is a quasi-convex function, then either the solution set of problem (1) is empty or any accumulation point x_* of $\{x_k\}$ is a stationary point of problem (1) and $\lim_{k\to\infty} x_k = x_*$.

Proof. It immediately follows from Lemma 3. □

Note that Wei and Jiang [4] has obtained a similar result to Corollary 1 for gradient descent method with convex function.

5. Numerical Experiments

We choose three numerical examples from [3], and report some numerical results by using the new methods in this paper. We take $\triangle_1^0 = 0.067$, $\triangle_2^0 = 3$, $\alpha_k = \overline{\alpha}_k(\triangle_1^k, \triangle_2^k)$, $\mu_1 = \mu_2 = \mu = 0.25$, $\beta = 1/2.9$, $\gamma = 1$, (NTMG $\beta_k = \overline{\beta}_k(\triangle_1^k)$.) We denote by "IT" the number of iterations, by "f_{opt}" the objective function value at the solution, by "T" computational time, by "3.6461(-3)" "3.6461 " etc. The following is the numerical results.

Example 1

$$\begin{aligned} f(x) &= 10(x_1^2 - x_2)^2 + (1 - x_1)^2 + 9(x_4 - x_3^2)^2 + (1 - x_3)^2 \\ &\quad +10.1((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1) \end{aligned}$$

$$x_1 = (-3, -1, -3, -1)^T; \; x_{\text{opt}} = (1, 1, 1, 1); f(x_{\text{opt}}) = 0.$$

$$\|\nabla f(x_k)\| \leq 10^{-1}, \; 10^{-2}$$

Example 2 $f(x) = \sum_{i=1}^{N/2}[(x_{2i} - x_{2i-1}^2)^2 + (1 - x_{2i-1})^2]$;

$$x_1 = (-1.2, 1, -1.2, 1, ..., -1.2, 1)^T; -x_{\text{opt}} = (1, 1, ..., 1); f(x_{\text{opt}}) = 0.$$

$$\|\nabla f(x_k)\| \leq 10^{-1}, \; 10^{-2}, \text{N=120}$$

Table 1: Numerical results of example 1

Method(M=1)	IT	T	f_{opt}
NTMG	13, 37	0.0600s, 0.1099s	7.4247(-4), 9.3087(-6)
NTFR	17, 35	0.5900s, 0.5999s	4.6057(-3), 3.6098(-5)
NTPR	12, 119	0.0499s, 0.2200s	7.6747(-4), 4.2176(-5)
NTHS	13, 21	0.0000s, 0.0400s	7.6751(-4), 7.6750(-5)
FR	51, 73	0.2800s, 0.4400s	2.0677(-4), 1.5005(-6)
PR	15, 22	0.0500s, 0.0600s	1.7343(-4), 5.1071(-6)
HS	18, 26	0.0500s, 0.0600s	3.2442(-3), 2.0893(-6)
NCG	20, 50	0.0499s, 0.0500s	4.5151(-3), 2.0094(-5)
NFR	23, 59	0.0590s, 0.1100s	4.5809(-3), 3.4529(-5)
NPR	49, 81	0.3300s, 0.3800s	6.9121(-3), 6.3727(-6)
NHS	26, 52	0.0590s, 0.1100s	3.0037(-3), 5.2729(-5)

Table 2: Numerical results of example 2

Method(M=1)	IT	T	f_{opt}
NTMG	8, 11	14.6599s, 19.5000s	5.8984(-3), 7.9117(-6)
NTFR	8, 11	14.6700s, 19.3800s	6.1615(-4), 1.3811(-5)
NTPR	9, 14	16.2600s, 23.5600s	1.6195(-3), 6.9608(-5)
NTHS	9, 25	16.1000s, 39.6499s	4.7874(-3), 1.4845(-5)
FR	13, 19	39.2699s, 56.3499s	2.0765(-3), 1.4603(-5)
PR	9, 11	26.4200s, 32.1299s	8.0624(-3), 4.1389(-4)
HS	9, 11	26.4800s, 32.3933s	1.1136(-3), 8.9999(-5)
NCG	12, 16	19.0600s, 24.9900s	1.7409(-2), 4.2189(-4)
NFR	12, 19	35.3099s, 55.8100s	3.2288(-2), 1.7576(-4)
NPR	14, 15	41.0899s, 43.8800	1.3835(-3), 1.2374(-5)
NHS	17, 23	49.8799s, 67.3900s	2.3040(-2), 1.3199(-5)

Example 3

$f(x) = \sum_{i=1}^{N/4}[(x_{4i-1} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^2 + 10(x_{4i-3} - x_{4i})^2]$;

$$x_1 = (3, -1, 0, -3, -3, -1, 0, -3, ..., -3, -1, 0, 3)^T; -x_{\text{opt}} = (0, 0, ..., 0);\ f(x_{\text{opt}}) = 0.$$

$$\|\nabla f(x_k)\| \leq 10^{-1},\ 10^{-2}, \text{N=60}$$

Table 3: Numerical results of example 3

Method(M=1)	IT	T	f_{opt}
NTMG	54, 82	24.5000s, 36.8000s	4.4338(-3), 1.2339(-4)
NTFR	57, 231	42.0699s, 102.0500s	7.8181(-3), 3.5408(-4)
NTPR	40, 124	18.6700s, 55.7500s	6.0185(-3), 2.4653(-4)
NTHS	37, 81	17.2541s, 36.7450s	2.2546(-3), 1.2546(-4)
FR	44, 74	39.4400s, 66.2400s	8.2052(-4), 3.2891(-5)
PR	30, 70	26.0299s, 60.7500s	7.2680(-3), 6.3319(-5)
HS	33, 41	29.6200s, 35.5900s	3.3625(-3), 3.3124(-6)
NCG	55, 131	24.6099s, 57.9500s	5.1785(-3), 2.7273(-4)
NFR	64, 129	55.4799s, 111.930s	6.4145(-3), 2.7230(-4)
NPR	40, 144	34.7099s, 125.000s	2.3033(-3), 3.1234(-4)
NHS	33, 94	41.5199s, 82.0099s	5.2568(-3), 2.9378(-4)

The numerical results indicate the proposed new methods have performance superior to the classical FR, PR, HS algorithms with Armijo-like step size rule, especially in the total amount of computational time. Moreover, the new methods are stable, and attractive for large-scale optimization problems.

Acknowledgment

The author wishes to express his thanks to referees for their very helpful comments.

References

[1] E. E. Cragg and A. V. Levy: Study on supermemory gradient method for the minimization of functions. *Journal of Optimization Theory and Applications*, **4** (1969) 191-205.

[2] A. Miele and J. W. Cantrell: Study on a memory gradient method for the minimization of functions. *Journal of Optimization Theory and Applications*, **3** (1969) 457-470.

[3] D. Touati-Ahmed and C. Storey: Efficient hybrid conjugate gradient techniques. *Journal of Optimization Theory and Applications*, **64** (1990) 379-397.

[4] Z. Wei, L. Qi, and H. Jiang: Some convergence properties of descent methods. *Journal of Optimization Theory and Applications*, **95** (1997) 177-188.

Sun Qingying
Depart. of Applied Maths
University of Petroleum
Dongying, 257061, P.R.CHINA
E-mail: sunqingying01@163.com