J. Korean Math. Soc. **47** (2010), No. 1, pp. 63–70
DOI 10.4134/JKMS.2010.47.1.063

A CLASS OF NONMONOTONE SPECTRAL MEMORY GRADIENT METHOD

Zhensheng Yu, Jinsong Zang, and Jingzhao Liu

Abstract. In this paper, we develop a nonmonotone spectral memory gradient method for unconstrained optimization, where the spectral stepsize and a class of memory gradient direction are combined efficiently. The global convergence is obtained by using a nonmonotone line search strategy and the numerical tests are also given to show the efficiency of the proposed algorithm.

1. Introduction

In this paper, we consider the unconstrained optimization problem

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

where $f(x) : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable with gradient function $\nabla f(x) = g(x)$.

The spectral gradient method was originally proposed by Barzilai and Borwein [1] for quadratic function. The main feature of this method is that only gradient directions are used at each line search whereas a non-monotone strategy [6] guarantees global convergence. Compared with the classical steepest descent method, the spectral gradient method requires less computational work and speeds up the convergence greatly, hence it is suitable for large scale problems. Due to its simplicity and numerical efficiency, the spectral gradient method has now received a good deal of attention in optimization community. Recently, the method has been extended successfully to unconstrained and constrained optimization, see for examples [2, 3, 4, 10].

In [2], Birgin and Martínez combined the spectral gradient and conjugate gradient(CG) idea and proposed a spectral conjugate gradient method for unconstrained optimization. The numerical performance showed that the method is more efficient than the CG method. Memory gradient methods have the same

Received February 19, 2008; Revised November 24, 2008.
2000 *Mathematics Subject Classification.* 90C30, 65K05.
Key words and phrases. unconstrained optimization, spectral memory gradient method, nonmonotone technique, global convergence.
This work is supported by National Natural Science Foundation of China (No.10671126) and Shanghai Leading Academic Discipline Project (No. S30501).

property as that of CG methods, but compared with CG methods, the main difference is that they can use the information of the previous iterations more sufficiently and hence they are helpful to design algorithms with quick convergence rate. The research for memory gradient method can be found in [5, 8, 9].

Motivated by the idea of [2], in this paper, we consider combining the spectra gradient method with a class of memory gradient method [9] and propose a spectral memory gradient method for problem (1). A nonmonotone Arimijo-type line search (which is motivated by the nonmonotone trust region technique [12]) is employed to obtain the iterate sequence. Under certain conditions, the strong global convergence of the proposed method is obtained, the numerical tests are also given to show the efficiency of the proposed algorithm.

The paper is organized as follows: In Section 2, we propose our algorithm and show its global convergence; In Section 3, we report the numerical tests.

2. Algorithm model and global convergence

In this section, we describe the nonmonotone spectral memory gradient algorithm, we first give our memory gradient direction, where the parameter β_k comes from the definition in Y. Narushima and H. Yabe [9]:

Let d_k^1 be defined by

$$d_k^1 = -g_k + \beta_k d_{k-1},$$

where

$$\beta_k^{NY} = \begin{cases} 0, \text{ if } g_k^T d_{k-1} \leq 0, \\ \frac{\|g_k\|^2}{(g_k^T d_{k-1} + \|g_k\| \|d_{k-1}\|)}, \text{ otherwise.} \end{cases}$$

Our spectral memory gradient direction is defined as

$$d_k = \theta_k d_k^1. \tag{2}$$

The definition of θ_k is defined in the following algorithm.

Algorithm 2.1.

Step 0. Given positive integer M and constant $\varepsilon > 0,\ 1 > \sigma_2 > \sigma_1 > 0$, $\gamma \in (0,1), \delta > 0,\ \theta_{\max} > \theta_{\min} > 0$, choose $\theta_0 \in [\theta_{\min}, \theta_{\max}],\ k = 0$.

Step 1: Compute g_k, if $\|g_k\| \leq \varepsilon$, stop.

Step 2: Compute d_k by (2).

Step 3.1: Set $\lambda_k = \frac{-\delta g_k^T d_k}{\|d_k\|^2}$.

Step 3.2: Set $x_+ = x_k + \lambda_k d_k$.

Step 3.3: Compute the largest index $l(k)$ such that

$$f(x_{l(k)}) = \max_{\max\{k-M+1,\ 0\} \leq j \leq k} f(x_j).$$

Compute

$$\rho_{1,k} = \begin{cases} \dfrac{f(x_{l(k)}) - f(x_+)}{\displaystyle\sum_{j=l(k)}^{k} -\lambda_j g_j^T d_j}, & \text{if } k > 0, \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

and

$$\rho_{2,k} = \frac{f(x_k) - f(x_+)}{-\lambda_k g_k^T d_k}, \tag{4}$$

set

$$\rho_k = \max\{\rho_{1,k}, \rho_{2,k}\}. \tag{5}$$

If $\rho_k \geq \gamma$, then set $x_{k+1} = x_+,\ s_k = x_{k+1} - x_k,\ y_k = g(x_{k+1}) - g(x_k)$, and go to Step 4.

If $\rho_k \geq \gamma$ does not hold, define $\lambda_{\text{new}} \in [\sigma_1 \lambda_k, \sigma_2 \lambda_k]$, set $\lambda_k = \lambda_{\text{new}}$, and go to Step 3.2.

Step 4. Compute $b_k = \langle s_k, y_k \rangle$, if $b_k \leq 0$, set $\theta_{k+1} = \theta_{\max}$, else, compute

$$a_k = \langle s_k, s_k \rangle \quad \text{and} \quad \theta_{k+1} = \min\{\theta_{\max}, \max\{\theta_{\min}, a_k/b_k\}\}.$$

Step 5: Set $k := k + 1$, go to Step 1.

To establish the global convergence, we make the following assumptions:

Assumption 2.1.

A1: $f(x)$ is bounded below on the level set $\mathcal{L} = \{x \mid f(x) \leq f(x_0)\}$.

A2: The gradient function is Lipschitz continuous on an open set Ω that contains $\mathcal{L}$, i.e., there exists a constant L such that for all $x, y \in \Omega$

$$\|g(x) - g(y)\| \leq L\|x - y\|.$$

The following result gives the descent property of the direction d_k.

Lemma 1. *Let d_k be generated by Algorithm 2.1. Then we have*

$$g_k^T d_k \leq -\frac{\theta_{\min}}{2}\|g_k\|^2. \tag{6}$$

Proof. If $\beta_k = 0$, then $d_k = -\theta_k g_k$, since $\theta_k \geq \theta_{\min}$, we have

$$g_k^T d_k = -\theta_k \|g_k\|^2 \leq -\frac{\theta_{\min}}{2}\|g_k\|^2.$$

If $\beta_k > 0$, then it follows from the definition of d_k that

$$\begin{aligned} g_k^T d_k &= \theta_k(-\|g_k\|^2 + \beta_k g_k^T d_{k-1}) \\ &\leq \theta_k \left(-\|g_k\|^2 + \frac{\|g_k\|^2}{2g_k^T d_{k-1}} g_k^T d_{k-1} \right) \\ &\leq -\frac{\theta_k}{2}\|g_k\|^2 \\ &\leq -\frac{\theta_{\min}}{2}\|g_k\|^2. \end{aligned}$$

□

Lemma 2. *Let $\{x_k\}$ be generated by Algorithm 2.1. If Assumption 2.1 holds, then there exists a positive constant μ such that*

$$\lambda_k \geq \mu \frac{-g_k^T d_k}{\|d_k\|^2}. \tag{7}$$

Proof. If $\lambda_k = \frac{-\delta g_k^T d_k}{\|d_k\|^2}$, then conclusion is obvious.

If $\lambda_k \leq \frac{-\delta g_k^T d_k}{\|d_k\|^2}$, then there exists a $\pi \in [\frac{1}{\sigma_2}, \frac{1}{\sigma_1}]$ such that for $x_+ = x_k + \pi\lambda_k$, $\rho_{1,k} \leq \rho_k \leq \gamma$, i.e.,

$$f(x_k + \pi\lambda_k d_k) - f(x_k) \geq \gamma\pi\lambda_k g_k^T d_k. \tag{8}$$

By the mean-value theorem, there exists $t_k \in [0, 1]$ such that

$$\begin{aligned} f(x_k + \pi\lambda_k d_k) - f(x_k) &= \int_0^{\pi\lambda_k} [g(x_k + t_k\pi\lambda_k d_k) - g(x_k)]^T d_k dt + \pi\lambda_k g_k^T d_k \\ &\leq \pi^2\lambda_k^2 L|d_k\|^2 + \pi\lambda_k g_k^T d_k \end{aligned}$$

which together with yields

$$\gamma\pi\lambda_k g_k^T d_k \leq \pi^2\lambda_k^2 L|d_k\|^2 + \pi\lambda_k g_k^T d_k,$$

i.e.,

$$\pi^2\lambda_k^2 L|d_k\|^2 \geq (\gamma - 1)\pi\lambda_k g_k^T d_k.$$

Hence

$$\lambda_k \geq \frac{(1-\gamma)}{2L\pi} \frac{-g_k^T d_k}{\|d_k\|^2}.$$

If $\mu = \min\{\delta, \frac{(1-\gamma)}{2L\pi}\}$, we get (7). □

Define

$$p(k) = \begin{cases} l(k), & \text{if } \rho_k = \rho_{1,k} \\ k, & \text{if } \rho_k = \rho_{2,k}. \end{cases}$$

We will call iteration $p(k)$ the reference iteration associated with iteration k, then it is easy to obtain the following result.

Lemma 3. *For each k, we have*

$$f(x_{p(k)}) - f(x_{k+1}) \geq -\gamma \sum_{j=p(k)}^{k} \lambda_j g_j^T d_j. \tag{9}$$

Theorem 1. *Let $\{x_k\}$ be an infinite sequence generated by Algorithm 2.1, and Assumption 2.1 holds, then we have*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{10}$$

Proof. Considering the kth iteration, we see that this iteration has a reference iteration $p(k)$, in turn, the $p(k)$th iteration has a reference iteration $p(p(k)), \ldots,$ up to the point where x_0 is reached by this backward referencing process. Hence, we may construct, for each k, a sequence of iteration indexed by $p_0, p_1, \ldots, p_q$, such that

$$x_0 = x_{p_0},\ x_{p_{j-1}+1} = x_{p(p_j)},\ j = 1, 2, \ldots, q,\ x_{p_q+1} = x_{p(k)}. \tag{11}$$

Note that

$$\begin{aligned} f(x_0) - f(x_{k+1}) = f(x_0) - f(x_{p_0+1}) + \sum_{j=1}^{q}[f(x_{p_{j-1}+1}) - f(x_{p_j+1})] \\ + f(x_{p(k)}) - f(x_{k+1}). \end{aligned}$$

Applying Lemma 3 to each item in the right side of the above equation together with (11), we have

$$f(x_0) - f(x_{k+1}) \geq -\mu\gamma \sum_{j=0}^{k} \lambda_j g_j^T d_j \geq \mu\gamma \sum_{j=0}^{k} \frac{(g_j^T d_j)^2}{\|d_j\|^2}.$$

Since $f(x_k)$ is bounded below on $\mathcal{L}$, we have

$$\sum_{j=0}^{k} \frac{(g_j^T d_j)^2}{\|d_j\|^2} < \infty. \tag{12}$$

On the other hand, from the definition of d_k, we have

$$d_k + \theta_k g_k = \theta_k \beta_k d_{k-1},$$

and squaring both sides of the above equation, we have

$$\|d_k\|^2 = \|\theta_k \beta_k d_{k-1}\|^2 - 2\theta_k g_k^T d_k - \theta_k^2 \|g_k\|^2.$$

Therefore, we obtain that

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^T d_k)^2} &= \frac{\|\theta_k \beta_k d_{k-1}\|^2}{(g_k^T d_k)^2} - 2\theta_k \frac{g_k^T d_k}{(g_k^T d_k)^2} - \theta_k^2 \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\ &= \frac{\|\theta_k \beta_k d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2\theta_k}{g_k^T d_k} - \theta_k^2 \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\ &= \frac{\|\theta_k \beta_k d_{k-1}\|^2}{(g_k^T d_k)^2} - \theta_k \Big(\frac{1}{\|g_k\|} + \frac{\|g_k\|}{g_k^T d_k} \Big)^2 + \theta_k \frac{1}{\|g_k\|^2} \end{aligned}$$

$$\leq \frac{\|\theta_k \beta_k d_{k-1}\|^2}{(-g_k^T d_k)^2} + \frac{\theta_k}{\|g_k\|^2}.$$

From the definition of d_k we have

$$-g_k^T d_k = \theta_k(\|g_k\|^2 - \beta_k d_{k-1}) \geq \theta_k \beta_k \|g_k\| \|d_{k-1}\|,$$

and therefore we have

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{1+\theta_k}{\|g_k\|^2} \leq \frac{1+\theta_{\max}}{\|g_k\|^2},$$

which means

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{\|g_k\|^2}{1+\theta_{\max}}.$$

Hence by (12) we have

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^2}{1+\theta_{\max}} \leq \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty,$$

which implies (10). □

TABLE 1

	$\theta_k = \theta_k^{sg}$		$\theta_k = 1$	
Problem	I_g/I_f	Time	I_g/I_f	Time
HS2	8/34	0.0448	31/282	0.0731
HS26	9/13	0.1060	195/372	0.1073
HS32	11/12	0.0397	486/1792	0.3124
HS42	5/9	0.4472	6/9	0.4589
HS46	15/63	2.2526	171/277	2.806
HS47	31/56	0.0794	146/251	0.7461
HS52	20/27	1.0566	297/1170	1.6368
HS57	6/13	0.1136	109/170	0.0718
HS63	59/110	0.1079	255/311	0.9198
HS76	3/14	0.7355	38/66	2.2139
HS100	14/37	1.9874	251/870	0.2243
HS113	17/30	1.1642	308/829	0.1905
HS213	24/55	1.6886	48/62	0.1448
HS240	12/15	0.3252	277/550	0.1707
HS246	195/2108	0.0378	616/539	7.984
HS256	959/1080	0.4326	*	*
HS257	448/479	0.6675	788/1781	1.2623
HS272	19/27	0.0439	103/709	0.1463
HS273	19/50	0.0434	260/1915	0.3301
HS280	387/432	0.2666	*	*

3. Numerical test

In this section, we test our nonmonotone spectral memory gradient method with MATLAB7.0, the parameters are set as follows: $M = 5$, $\sigma_1 = 0.1$, $\sigma_2 = 1.2$, $\gamma = 10^{-3}$, $\delta = 1$, $\theta_{\min} = 10^{-30}$, $\theta_{\max} = 10^{30}$.

We test 20 examples taken from [7, 11], the numerical results are shown in Table 1. Here HSi denotes the ith example in [7, 11], we report the number of function evaluations (I_f), the number of gradient evaluations (I_g), CUP time (Time), and denote $*$ if the iteration large than 5000.

We compare the method with spectral stepsize ($\theta_k = \theta_k^{sg}$) with the non-spectral stepsize ($\theta_k = 1$), the numerical results show that the performance of spectral method is more efficiency than that of non-spectral method.

Acknowledgments. The authors would like to express their thanks to the anonymous referees for his helpful comments and suggestions.

References

[1] J. Barzilai and J. M. Borwein, *Two-point step size gradient methods*, IMA J. Numer. Anal. **8** (1988), no. 1, 141–148.

[2] E. G. Birgin and J. M. Martinez, *A spectral conjugate gradient method for unconstrained optimization*, Appl. Math. Optim. **43** (2001), no. 2, 117–128.

[3] E. G. Birgin, J. M. Martínez, and M. Raydan, *Nonmonotone spectral projected gradient methods on convex sets*, SIAM J. Optim. **10** (2000), no. 4, 1196–1211.

[4] ———, *Spectral projected gradient methods*, Encyclopedia of Optimization, C. A. Floudas and P. M. Pardalos, (Eds.), to appear.

[5] E. E. Cragg and A. V. Levy, *Study on a supermemory gradient method for the minimization of functions*, J. Optimization Theory Appl. **4** (1969), 191–205.

[6] L. Grippo, F. Lampariello, and S. Lucidi, *A nonmonotone line search technique for Newton's method*, SIAM J. Numer. Anal. **23** (1986), no. 4, 707–716.

[7] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems, 187. Springer-Verlag, Berlin-New York, 1981.

[8] A. Miele and J. W. Cantrell, *Study on a memory gradient method for the minimization of functions*, J. Optimization Theory Appl. **3** (1969), 459–470.

[9] Y. Narushima and H. Yabe, *Global convergence of a memory gradient method for unconstrained optimization*, Comput. Optim. Appl. **35** (2006), no. 3, 325–346.

[10] M. Raydan, *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem*, SIAM J. Optim. **7** (1997), no. 1, 26–33.

[11] K. Schittkowski, *More Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems, 282. Springer-Verlag, Berlin, 1987.

[12] P. L. Toint, *Non-monotone trust-region algorithms for nonlinear optimization subject to convex constraints*, Math. Programming **77** (1997), no. 1, Ser. A, 69–94.

Zhensheng Yu
College of Science
University of Shanghai for Science and Technology
Shanghai, 200093, P. R. China
E-mail address: zhsh-yu@163.com

Jinsong Zang
Modern Teaching Center
University of Shanghai for Science and Technology
Shanghai, 200093, P. R. China
E-mail address: js_zang@163.com

Jingzhao Liu
Editorial Department of Journal of Qufu Normal University
Shandong, 273165, P. R. China
E-mail address: xbljz@mail.dfnu.edu.cn