Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 8, No. 1, pp. 15–25, January 2012

ADAPTIVE NONMONOTONE SPECTRAL RESIDUAL METHOD FOR LARGE-SCALE NONLINEAR SYSTEMS*

Minghou Cheng and Yu-Hong Dai

This paper is to remember Professor Paul Tseng with great sorrow.

Abstract: By combining with the traditional non-monotone linea search, Gruz and Raydan (2003) proposed a spectral residual method for solving large-scale nonlinear systems equations. However, practical performances of the method depend on an important parameter in the non-monotone line search. In this paper, we propose a new and efficient combination of the spectral residual method and the adaptive nonmonotone line search by Dai and Zhang (2001). One advantage of the new algorithm is that, there is no necessity to calculate the Jacobian of the nonlinear system and hence the cost of the algorithm at each iteration is cheap. Numerical results show that the adaptive nonmonotone line search is specially suitable for the spectral residual method. Some convergence analysis is also presented.

Key words: *nonlinear equations, spectral residual method, nonmonotone line search*

Mathematics Subject Classification: *49M37, 65K05, 90C37*

1 Introduction

The purpose of this paper is to propose an adaptive nonmonotone spectral residual method (ANSRM) for solving nonlinear system of equations,

$$F(x) = 0, \tag{1.1}$$

where $F: R^n \to R^n$ is a continuously differentiable mapping. We are interested in large-scale systems where the Jacobian of F is not available or requires a prohibitive amount of storage.

The steepest descent method for solving (1.1) can be dated back to Cauchy [3] in 1847. Its simplicity and ease to use have attracted many researchers' attentions in various research fields. Some recent work along this line can be found in the works by Barzilai and Borwein [1], Birgin *et al.* [2], Dai and Liao [8], Raydan [14, 15], Tseng [16], Tseng and Yun [17], *etc.* A seminal work is due to Barzilai and Borwein [1] for minimizing a smooth n-dimensional function $f(x)$. Assuming the gradient iteration is $x_{k+1} = x_k - \alpha_k g_k$, where $g_k = \nabla f(x_k)$, they proposed the stepsize formula

$$\alpha_k = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}, \tag{1.2}$$

*This work was partly supported by the Chinese NSF grants (no. 10831106), the CAS grant (no. kjcx-yw-s7-03) and the China National Funds for Distinguished Young Scientists (no. 11125107).

where $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$. Despite of its heavy nonmonotonicity, the (unmodified) BB method is proved to be R-superlinearly convergent for two-dimensional quadratics [1] and R-linearly convergent for any dimensional quadratics [8]. In practical computations, the BB method is much superior to the classical steepest descent method.

The BB method has been extended by Cruz and Raydan [6] for solving nonlinear equations (1.1). The new method is called by Spectral Algorithm for Nonlinear Equations (SANE). SANE uses the residual $\pm F(x_k)$ as a search direction in a systematic way. The first trial point at each iteration is $x_k - \alpha_k F(x_k)$, where α_k is some spectral coefficient similarly to (1.2), but with y_{k-1} replaced with $F(x_k) - F(x_{k-1})$. At each iteration, SANE needs to calculate a directional derivative or have its good approximation is necessary since it requires a descent direction of the squared norm of the residual.

To overcome the disadvantage of SANE, Cruz *et al.* [4] recently proposed a Derivative-Free Spectral Algorithm for Nonlinear Equations (DF-SANE) for solving (1.1). The algorithm makes use of the same search direction d_k and the same choice for the initial stepsize σ_k as SANE. Define $f(x) = ||F(x)||^2$, where $\|\cdot\|$ means the two-norm. The algorithm wishes to find a stepsize α_k that satisfies

$$f(x_k + \alpha_k d_k) \le f_{max} + \eta_k - \gamma \alpha_k^2 f(x_k), \tag{1.3}$$

where

$$f_{max} = \max_{0 \le j \le m(k)-1} f(x_{k-j}), \tag{1.4}$$

in which $m(k) = min(k,\, M-1)$, M is some fixed integer, $\eta_k > 0$ satisfies $\sum_k \eta_k \le \eta < \infty$ and γ is a small positive number. The above nonmonotone line search is a combination of the traditional GLL line search [13] and the LF line search [12]. With its use, DF-SANE can avoid the calculations or any approximations of directional derivatives.

However, the numerical performances of DF-SANE heavily depend on the choice of M in the definition of f_{max}. We have observed the influence of M in solving several large-scale test problems and founded that the performance of DF-SANE is sensitive to the choice of M (see Section 2). This motivates us to incorporate the more efficient adaptive nonmonotone line search in [9] in the spectral residual method. The numerical results in Section 4 indeed demonstrated the superiority of the new algorithm for the same set of test problems in [4].

In [9], Dai and Zhang proposed an adaptive nonmonotone line search for large-scale unconstrained optimization problem $\min_{x \in R^n} f(x)$. Numerical results [9] showed that the DZ line search is specially suitable for the BB method. The key point of the adaptive nonmonotone line search is how to update the reference value f_r. Suppose that a descent search direction d_k has been computed at the kth iteration. The DZ line search accepts the first trial stepsize $\alpha_k^{(1)}$ provided it satisfies

$$f(x_k + \alpha_k^{(1)} d_k) \le f_r + \gamma\, \alpha_k^{(1)}\, g_k^T d_k. \tag{1.5}$$

Otherwise, a new stepsize α_{new} is calculated to satisfy the following relatively restrict condition,

$$f(x_k + \alpha_{new} d_k) \le \min\{f_{max}, f_r\} + \gamma\, \alpha_{new}\, g_k^T d_k, \tag{1.6}$$

and set $\alpha_k = \alpha_{new}$. The reference value f_r is updated as follows. Denote by f_{min} the current best function value over all the previous iterations and by l the number of iterations since the value of f_{min} was actually obtained. If l reaches a preset positive integer constant L, they choose the new reference value f_r as follows:

$$f_r = \begin{cases} f_c, & \text{if } \frac{f_{max}-f_{min}}{f_c-f_{min}} > \gamma_1; \\ f_{max}, & \text{otherwise,} \end{cases} \tag{1.7}$$

where f_c is the maximal value of the objective function since the last best function value was found and $\gamma_1 \geq 1$ is a constant. The reference value f_r should also be adjusted if the first trial stepsize is suddenly not accepted by the line search after many successful first trial stepsizes. More exactly, assume that the first trial stepsize at the kth iteration is $\alpha_k^{(1)}$ and that p is the largest integer such that $\{\alpha_{k-i}^{(1)} : i = 1, 2, ..., p\}$ are accepted but $\alpha_{k-p-1}^{(1)}$ not. Let P be some integer and $\gamma_2 \geq 1$ some constant. If $p > P$, we set

$$f_r = \begin{cases} f_{max}, & \text{if } f_{max} > f_k \text{ and } \frac{f_r-f_k}{f_{max}-f_k} \geq \gamma_2; \\ f_r, & \text{otherwise.} \end{cases} \tag{1.8}$$

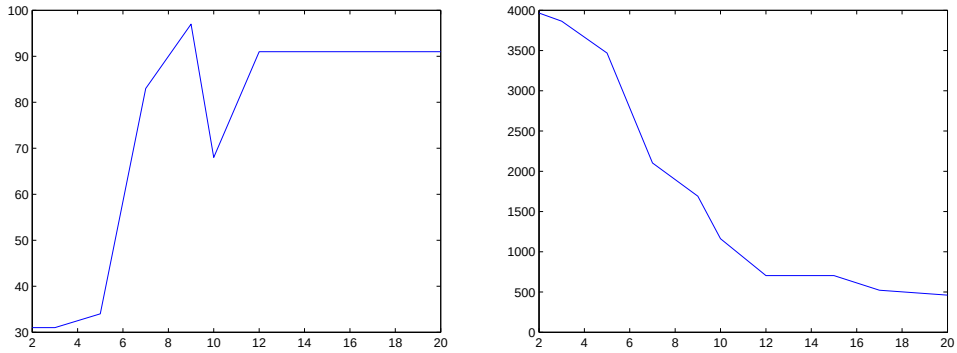
In this work, we will modify the DZ line search so that that they are suitable for the spectral residual method. Then we combine the modified DZ line search and the LF line search in [12] to solve nonlinear systems of equations. It turns out to a robust nonmonotone line search strategy that has the advantages of both schemes.

This paper is organized as follows. In Section 2, we describe the framework of the Adaptive Nonmonotone Spectral Residual Method (ANSRM). In Section 3, we present some convergence analysis. Numerical results with ANSRM are reported in Section 4 and some remarks are made in the last section.

2 Adaptive Nonmonotone Spectral Residual Method

As explained in the previous section, the DF-SANE algorithm makes use of the GLL nonmonotone line search to find a suitable stepsize. This line search avoids the calculations or approximations of directional derivatives and guarantees the global convergence of the algorithm. For quite many problems, however, its numerical efficiency heavily depends on the choice of the parameter M in the line search. We have tested the DF-SANE algorithm with different values of M for solving the problems listed in [5] and observed this phenomenon. Figure 2.1 shows how the number of function evaluations of the DF-SANE algorithm changes with the parameter M for Problem 5 and Problem 43 in [5]. The results for Problem 5 suggest the use of a small value of M, whereas the results for Problem 43 show that a large value of M is preferable. This explains to some extent why $M = 10$ is used for a balance in many numerical algorithms using the GLL line search.

Figure 2.1 Testing DF-SANE with different values of M
(left: Results for Problem 5; right: Results for Problem 43)



For a better replacement, we consider the use of the adaptive nonmonotone line search proposed by Dai and Zhang [9] for unconstrained optimization. By introducing the merit function $f(x) = ||F(x)||^2$, one can directly use the DZ line search for solving (1.1) once a descent direction d_k, namely, $d_k^T \nabla f(x_k) < 0$, is found. In this case, the DZ line search condition is satisfied provided that the stepsize α_k sufficiently close to zero. However, it is possible that $d_k^T \nabla f(x_k) = 0$, in which case the existence of α_k satisfying the DZ line search condition is not guaranteed. In general, if the Jacobian $J(x_k)$ is not calculated, there is some difficulty in finding a search direction d_k satisfying $d_k^T \nabla f(x_k) < 0$.

In order to overcome this difficulty, we recall the LF line search [12]. The LF line search consists in computing a stepsize α_k such that

$$\|F(x_k + \alpha_k d_k)\| \leq (1 + \eta_k)\|F(x_k)\| - \gamma \alpha_k^2 \|d_k\|^2, \tag{2.1}$$

where $\{\eta_k\}$ is a positive sequence and γ is a positive parameter. This line search condition does not require the computation of $J(x_k)$. Moreover, it can be satisfied provided that α_k is sufficiently small, independently of the choice of d_k. However, since η_k is usually very small when k is large, the LF line search generally imposes an almost monotone behavior of the merit function when x_k is close to some solution.

To combine the advantages of the DZ line search (1.5) and (1.6) and the LF line search (2.1), we propose the following adaptive nonmonotone line search for solving (1.1).

Suppose that the initial parameters L, M, P, γ, γ_1, γ_2, $0 < \tau_{min} < \tau_{max} < 1$ are given. The first stepsize $\alpha_k^{(1)}$ is computed at the kth iteration such that $|\alpha_k^{(1)}| \in [\alpha_{min}, \alpha_{max}]$. We describe our adaptive nonmonotone line search as follows.

Algorithm 2.1 (Adaptive nonmonotone line search)**.**

Step 1. *(Possibly reset the reference value).*
(i) *If $l = L$, update f_r by (1.7) and set $l := 0$;*
(ii) *If $p > P$, compute f_r by (1.8).*

Step 2. *(Test the first trial stepsize $\alpha_k^{(1)}$).*
Set $d = -\alpha_k^{(1)} F(x_k)$, $\alpha_+ = 1$, and $\alpha_- = 1$. If

$$f(x_k + \alpha_+ d) \leq f_r + \eta_k - \gamma \alpha_+^2 f(x_k), \tag{2.2}$$

let $d_k = d$, $\alpha_k = \alpha_+$, $p := p + 1$ *and go to Step 4; else if*

$$f(x_k - \alpha_- d) \le f_r + \eta_k - \gamma \alpha_-^2 f(x_k), \tag{2.3}$$

let $d_k = -d$, $\alpha_k = \alpha_-$, $p := p + 1$ *and go to Step 4; else,* $p := 0$.

Step 3. *(Test other trial stepsize till some stepsize is satisfactory).*
(i) $\alpha_{old+} = \alpha_+$, $\alpha_{old-} = \alpha_-$;
(ii) *Compute* $\alpha_{new+} \in [\tau_{min}\alpha_{old+}, \tau_{max}\alpha_{old+}]$ *and* $\alpha_{new-} \in [\tau_{min}\alpha_{old-}, \tau_{max}\alpha_{old-}]$. *If*

$$f(x_k + \alpha_{new+} d) \le \min\{f_{max}, f_r\} + \eta_k - \gamma \alpha_{new+}^2 f(x_k), \tag{2.4}$$

let $\alpha_k = \alpha_{new+}$, $d_k = d$ *and go to Step 4; else if*

$$f(x_k - \alpha_{new-} d) \le \min\{f_{max}, f_r\} + \eta_k - \gamma \alpha_{new-}^2 f(x_k), \tag{2.5}$$

let $\alpha_k = \alpha_{new-}$, $d_k = -d$ *and go to Step 4; else,* $\alpha_{old+} = \alpha_{new+}$ *and* $\alpha_{old-} = \alpha_{new-}$, *repeat (ii).*

Step 4. *(Possibly update the best value found and the candidate value).*
(i) Let $f(x_{k+1}) = f(x_k + \alpha_k d_k)$;
(ii) If $f(x_{k+1}) < f_{min}$, *set* $f_c = f_{min} = f(x_{k+1})$ *and* $l := 0$; *otherwise,* $l := l + 1$;
(iii) If $f(x_{k+1}) > f_c$, *set* $f_c = f(x_{k+1})$;
(iv) Compute f_{max} *by (1.4) with* k *replaced by* $k + 1$.

As in [9], we suggest that the parameters P, M and L in the above algorithm are so chosen that $L \le 5$ and $P \ge 4M \ge 8L$ and the parameters γ_1 and γ_2 in (1.7) and (1.8) are set to M/L and P/M.

Combined the spectral residual method and the above adaptive nonmonotone line search algorithm, we obtain a new algorithm for large-scale nonlinear system of equations. A full description of our adaptive nonmonotone spectral residual method (ANSRM) is given as follows.

Algorithm 2.2 (ANSRM)**.**

Step 1. *(Give the starting point and initialize the parameters).*
(i) *Given* $0 < \alpha_{min} < \alpha_{max} < \infty$, $0 < \sigma_1 < \sigma_2 < 1$ *and* $\varepsilon \ge 0$; *set* $k := 1$;
(ii) *Given positive integers* $P > M > L$ *and constants* $\gamma_1 \ge 1$, $\gamma_2 \ge 1$;
(iii) *Pick up* $x_1 \in R^n$, $\alpha_1^{(1)} \in [\alpha_{min}, \alpha_{max}]$ *and compute* $F(x_1)$;
(iv) Set $l := 0$, $p := 0$ *and* $f_{min} = f_r = f_c := f(x_1)$.

Step 2. *(Test if the stopping condition holds). If* $f(x_k) \le \varepsilon$, *stop.*

Step 3. *Determine the search direction* d_k, *compute a stepsize* α_k *and update* f_r *and* f_{min} *etc. by algorithm 2.1.*

Step 4. *(Update the estimation and compute the residual).*
Compute $x_{k+1} = x_k + \alpha_k d_k$ *and* $F(x_{k+1})$.

Step 5. *(Compute the first trial stepsize* $\alpha_{k+1}^{(1)}$*).*
$s_k = x_{k+1} - x_k$, $y_k = F(x_{k+1}) - F(x_k)$ *and* $\alpha_{k+1}^{(1)} = \max\left\{\alpha_{min}, \min\left\{\frac{s_k^T s_k}{s_k^T y_k}, \alpha_{max}\right\}\right\}$.

Step 6. $k := k + 1$ *and go to Step 2.*

3 Convergence Analysis

In this section, we present a convergence analysis for ANSRM. The following theorem shows that the spectral residual method with the adaptive nonmonotone line search is globally convergent in real computations.

Theorem 3.1. *Suppose that the merit function $f(x) = \|F(x)\|^2$ is twice-continuously differentiable and its level set $\mathcal{L} = \{x \in R^n : f(x) \le f(x_0)+\eta\}$ is bounded, where $\sum_{k=0}^{\infty} \eta_k \le \eta < \infty$. Then for any small constant $c > 0$, Algorithm 2.2 either terminates at a finite iteration j such that $F(x_j) = 0$ or $F(x_j)^T J(x_j)F(x_j) \le c\,\|F(x_j)\|^2$, or generates a sequence $\{x_k\}$ such that*

$$\liminf_{k\to\infty} \|F(x_k)\| = 0. \tag{3.1}$$

Proof. At first, we denote f_r, f_{max}, f_c, l at the kth iteration by f_r^k, f_{max}^k, f_c^k, l^k. Algorithm 2.2 implies that

$$f_k \le f_c^k \le f_{max}^k \quad \text{for all } k. \tag{3.2}$$

From the algorithm, we also know that the whole iteration $\{x_k\}$ remains in the level set $\mathcal{L}$ since $f(x_k) \le f(x_0) + \eta$ for all k, where η is the upper bound of the sum $\sum_{k=0}^{\infty} \eta_k$. Thus, using the fact that $\alpha_k^{(1)} \in [\alpha_{min}, \alpha_{max}]$ and the technique of choosing α_k and (3.2), we can conclude that a stepsize α_k satisfying the line search conditions of Algorithm 2.1 can be found after finite number of trials. Hence there must exist a positive constant c_1 such that $\alpha_k \ge c_1$ for all k (for example, one can see the proof of Theorem 2.1 in [10] for details). It follows from $\alpha_k \ge c_1$ and the conditions of Algorithm 2.1 that

$$f(x_{k+1}) \le f_r^k + \eta_k - \gamma\,\alpha_k^2\, f_k. \tag{3.3}$$

Assuming that Algorithm 2.2 does not stop after finite iterations, we consider the following two cases.

Case 1. $l^k < L$ for all large k. In this case, we know from the algorithm that there exists an infinite subsequence $\{x_{k_i}\}$ such that

$$f(x_{k_{i+1}}) < f(x_{k_i}) + \eta_{k_i},\ k_{i+1} \le k_i + L. \tag{3.4}$$

It is known that in real computations,

$$a < b \quad means \quad a \le b - \epsilon,$$

where $\epsilon > 0$ is the machine precision. Hence (3.4) means $f(x_{k_{i+1}}) \le f(x_{k_i}) + \eta_{k_i} - \epsilon$. This contradicts the fact that $f(x)$ is bounded below on the level set $\mathcal{L}$. So case 1 can not happen in real computations.

Case 2. $l^k = L$ for infinite times. In this case, we define the infinite index set $\mathcal{A} = \{k_i :\ l^{k_i} = L,\ k_i < k_{i+1}\}$. Then for all $k_i < j < k_{i+1}$, f_r^j is determined by the formula (1.8). Since $\gamma_2 \ge 1$, we can then get

$$f_r^j \le f_r^{j-1}, \quad k_i < j < k_{i+1}. \tag{3.5}$$

From (3.2), (3.5) and the definition of set $\mathcal{A}$, we know that

$$f_r^j \le f_r^{k_i} \le f_{max}^{k_i}, \quad k_i < j < k_{i+1}. \tag{3.6}$$

Now assuming the theorem is not true, that is $f_k \geq c_2$ for all k and some constant $c_2 > 0$, we can obtain by (3.3) that

$$f(x_{k+1}) \leq f_r^k + \eta_k - \epsilon_1 \tag{3.7}$$

for all k and $\epsilon_1 = \delta c_1 c_2^2$. Combining (3.6) and (3.7), we get that

$$f_j \leq f_{max}^{k_i} - \epsilon_1, \quad k_i < j \leq k_{i+1}. \tag{3.8}$$

Further, by the definition of f_{max} and $\mathcal{A}$, we know from (3.6) and (3.8) that

$$f_r^{k_{i+1}} \leq f_{max}^{k_{i+1}} \leq f_{max}^{k_i} + \eta_k, \tag{3.9}$$

which implies that $\{f_{max}^{k_i} : k_i \in \mathcal{A}\}$ is an approximately nonincreasing sequence. It follows from (3.8), (3.9) and the approximately nonincreasing property of $\{f_{max}^{k_i} : k_i \in \mathcal{A}\}$ that

$$f_{max}^{k_{i_1}} \leq f_{max}^{k_{i_2}} + \eta_{k_{i_2}} - \epsilon_1, \quad \text{if } k_{i_1} - k_{i_2} > M. \tag{3.10}$$

So there exists a subsequence of $\{f_{max}^{k_i} : k_i \in \mathcal{A}\}$, still denoted by $\{f_{max}^{k_i}\}$, such that

$$f_{max}^{k_{i+1}} \leq f_{max}^{k_i} + \eta_{k_i} - \epsilon_1. \tag{3.11}$$

Since $\mathcal{A}$ is an infinite set and since $\sum_{k\geq 1} \eta_k < +\infty$, we know that the above relation contradicts the lower boundedness of $f(x)$. Therefore the theorem is true. □

4 Numerical Results

We have tested ANSRM and compared it with DF-SANE by using Matlab v7.6 on Core(TM)2 PC with Windows-XP. The initial parameters for ANSRM are $\alpha_{min} = 10^{-10}$, $\alpha_{max} = 10^{10}$, $\sigma_1 = 1$, $\tau_{min} = 0.1$, $\tau_{max} = 0.5$, $\gamma = 10^{-4}$, $M = 10$, $\eta_k = \frac{\|F(x_1)\|}{(1+k)^2}$ for all $k \in \mathbb{N}$. The spectral stepsize was computed by the formula $\alpha_k^{(1)} = \frac{s_k^T s_k}{s_k^T y_k}$, where $s_k = x_{k+1} - x_k$ and $y_k = F(x_{k+1}) - F(x_k)$. However, if $|\alpha_k| \notin [\alpha_{min}, \alpha_{max}]$, we replace the spectral coefficient by

$$\alpha_k^{(1)} = \begin{cases} 1, & \text{if} \quad ||F(x_k)|| > 1; \\ ||F(x_k)||^{-1}, & \text{if} \quad 10^{-5} \leq ||F(x_k)|| \leq 1; \\ 10^5, & \text{if} \quad ||F(x_k)|| < 10^{-5}. \end{cases} \tag{4.1}$$

We also set L, M and P to be 3, 8 and 40, respectively. In addition, in Step 3 of Algorithm 2.1, the new trial stepsize α_{new+} and α_{new-} are obtained by the same rule as the one in [4]. Given $\alpha_+ > 0$, we take $\alpha_{new+} > 0$ as

$$\alpha_{+new} = \begin{cases} \tau_{min}\alpha_+, & \text{if} \alpha_t < \tau_{min}\alpha_+; \\ \tau_{max}\alpha_+, & \text{if} \alpha_t > \tau_{max}\alpha_+; \\ \alpha_t, & \text{otherwise}, \end{cases} \tag{4.2}$$

where

$$\alpha_t = \frac{\alpha_+^2 f(x_k)}{f(x_k + \alpha_+ d) + (2\alpha_+ - 1) f(x_k)}.$$

We use similar formula for choosing α_{new-} as a function of α_-, $f(x_k)$ and $f(x_k - \alpha_- d)$. For both DF-SANE and ANSRM, we stop the process when

$$\frac{||F(x_k)||}{\sqrt{n}} \leq e_a + e_r \frac{||F(x_0)||}{\sqrt{n}}, \tag{4.3}$$

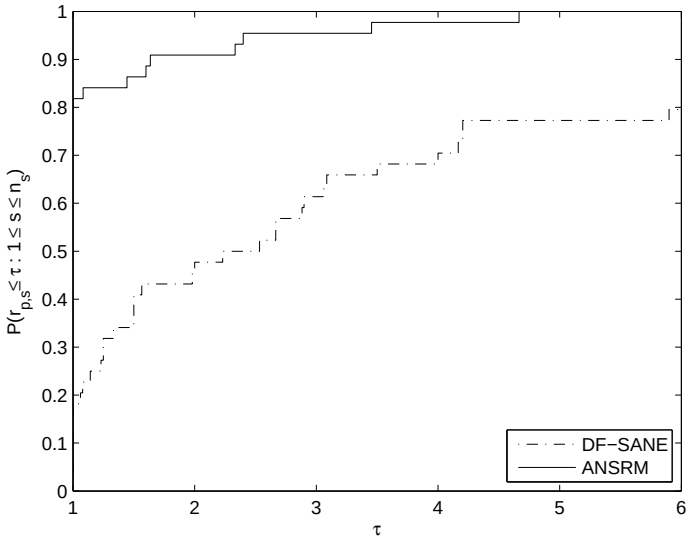
where $e_a = 10^{-5}$ and $e_r = 10^{-4}$.

We ran ANSRM using the same set of large-scale test problems as in [4]. There are 44 test problems, which are fully described in [5]. The first set of 22 problems arises from the discretization of boundary value problems. They have positive definite Jacobians. The second set of 22 problems does not have special characteristics from the point of view of positive definiteness or conditioning. Some of these problems have many solutions.

Table 1: Comparing ANSRM and DF-SANE

Prob(n)	DF-SANE			ANSRM			Prob(n)	DF-SANE			ANSRM		
	It	Fe	Bk	It	Fe	Bk		It	Fe	Bk	It	Fe	Bk
4(99)	99	289	66	12	13	0	30(99)	11	16	2	12	14	0
4(999)	101	325	71	12	13	0	30(9999)	11	16	2	10	12	0
5(9)	42	68	12	15	17	0	33(1000)	37	50	3	6	12	1
5(49)	732	2958	660	20	22	0	33(5000)	4	16	2	3	13	1
7(100)	23	29	2	6	13	0	34(1000)	78	155	26	14	15	0
7(10000)	23	29	2	5	10	0	34(5000)	12	18	1	5	12	1
11(99)	17	49	7	15	16	0	35(1000)	21	27	2	18	25	2
11(399)	17	49	7	16	17	0	35(5000)	38	48	3	43	52	2
12(1000)	30	62	12	7	10	1	36(1000)	28	34	2	27	32	1
12(10000)	23	59	11	7	10	1	36(5000)	26	36	4	18	23	1
16(500)	14	16	1	5	6	0	37(1000)	26	38	5	18	19	0
16(2000)	16	16	0	5	6	0	37(5000)	26	38	5	14	15	0
17(100)	9	11	1	32	38	0	38(1000)	25	30	2	19	20	0
17(1000)	7	9	1	41	42	0	38(5000)	25	30	2	19	20	0
18(50)	19	21	1	5	6	0	39(1000)	14	20	1	15	16	0
18(100)	*	*	*	5	6	0	39(5000)	14	20	1	15	16	0
20(100)	40	42	1	1	4	1	41(500)	7	9	1	12	13	0
20(1000)	44	62	5	1	4	1	41(1000)	3	3	0	6	7	0
24(500)	54	109	18	47	55	2	42(1000)	173	412	85	80	98	5
24(1000)	17	25	3	50	60	3	42(5000)	173	412	85	80	98	5
27(50)	10	10	0	15	16	0	43(100)	86	108	9	32	35	1
27(100)	10	11	0	17	18	0	43(500)	586	1162	193	163	184	8

Figure 4.1. Performance Profiles of ANSRM and DF-SANE



The numerical results are listed in Table 1. It is worth noting that only those problems are listed for which the two algorithms provide different numerical results. In the table, we report the problem number and its dimension (Prob(n)), the number of iterations (It), the

number of function evaluations (Fe) and the number of backtrackings (Bk). Figure 4.1 gives the performance profiles for the number of function evaluation of the two algorithms (see [11] for more details about the performances).

From Table 4.1 and Figure 4.1, we can see that ANSRM performs much better than the DF-SANE algorithm and hence the SANE algorithm (see [4] for comparisons between DF-SANE and SANE). Pariticularly, for Problems 4, 5, 20, 34, 42 and 43, the ANSRM algorithm has considerable savings in both the number of line searches and the number of function evaluations.

We should say that we have also tested ANSRM for different choices of L over $[4, 10]$ and found that its numerical performance is all the same as those with $L = 3$ for all of the test problems.

5 Conclusions and Discussion

In this paper, we have proposed an adaptive nonmonotone spectral residual method (ANSRM) for solving nonlinear system of equations. The important features of ANSRM are that, the reference iteration is chosen adaptively and there is no necessity to calculate the Jacobian of the nonlinear system at each iteration. Our numerical results have demonstrated that it is a better strategy than the traditional GLL nonmonotone line search.

For the choice of initial stepsize $\alpha_{k+1}^{(1)}$ in Step 5 of Algorithm 2.2, we have tested the following three cases

$$\frac{s_k^T s_k}{s_k^T y_k}, \qquad \frac{s_k^T y_k}{y_k^T y_k}, \qquad sign(s_k^T y_k)\sqrt{\frac{s_k^T s_k}{y_k^T y_k}}. \tag{5.1}$$

The first two choices are due to [1] and the third one can be found in for example [7, 18]. Our numerical results show that the first choice is the most efficient one. In addition, although the third choice is not as good as the other two, it gives the best performance for several problems. It still remains under investigation how to choose the initial stepsize efficiently for the gradient method, including ANSRM, in real computations.

References

[1] J. Barzilai and J.M. Borwein, Two-point step size gradient methods, *IMA J. Numer. Anal.* 8 (1988) 141–148.

[2] E.G. Birgin, J.M. Martínez and M. Raydan, Nonmonotone spectral projected gradient methods for convex sets, *SIAM J. Optim.* 10 (2000) 1196–1121.

[3] A. Cauchy, Méthode Générale Pour La Résolution Des Systéms D'equations Simulatanées, *Comp. Rend. Sci. Paris*, 25 (1847) 46–89.

[4] W.L. Cruz, J.M. Martínez and M. Raydan, Spectral residual method without gradient information for solving large-scale nonlinear systems of equations, *Math. Comp.* 75 (2006) 1429–1448.

[5] W.L. Cruz, J.M. Martínez and M. Raydan, Spectral residual method without gradient information for solving large-scale nonlinear systems: Theory and experiments, Technical Report TR-04-08, Dpto. de Computacion, UCV, 2004. Available at *www.kuainasi.ciens.ucv.ve/ccct/mraydan_pub.html*.

[6] W.L. Cruz and M. Raydan, Nonmonotone spectral methods for large-scale nonlinear systems, *Optim. Methods Softw.*, 18 (2003) 583–599.

[7] Y.H. Dai, Alternate Step Gradient Method, *Optimization* 52 (2003) 395–415.

[8] Y.H. Dai and L.Z. Liao, *R*-linear convergence of the Barzilai and Borwein gradient method, *IMA Journal of Numerical Analysis* 22 (2002) 1–10.

[9] Y.H. Dai and H.C. Zhang, Adaptive two-point stepsize gradient algorithm, *Numer. Algorithms* 27 (2001) 377–385.

[10] Y.H. Dai, On the nonmonotone line search,*Journal of Optimization Theory and Applications* 112 (2002) 315–330.

[11] E.D. Dolan and J.J. Moré, Bemchmarking optimization software with performance profiles, *Math. Prog.* 91 (2002) 201–213.

[12] D.H. Li and M. Fukushima, A derivative-free line search and global convergence of Broyden-like method for nonlinear equations, *Optim. Methods Softw.* 13 (2000) 181–201.

[13] L. Grippo, F. Lampariello and S. Lucidi, A nonmonotone line search technique for Newton's method, *SIAM J. Numer. Anal.* 23 (1986) 707–716.

[14] M. Raydan, On the Barzilai and Borwein choice of steplength for the gradient method, *IMA J. Numer. Anal.* 13 (1993) 321–326.

[15] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM J. Optim.* 7 (1997) 26–33.

[16] P. Tseng, Incremental gradient(-projection) method with momentum term and adaptive stepsize rule, *SIAM J. Optim.* 8 (1998) 506–531.

[17] P. Tseng and S. Yun, A coordinate gradient descent method for nonsmooth separable minimization, *Math. Prog.* 117 (2009) 387–423.

[18] R. Varadhan and Ch. Roland, Squared extrapolation methods(SQUAREM): A new class of simple and efficient numerical schemes for accelerating the convergence of the EM algorithm, Department of Biostatistics Working Paper, Johns Hopkins University, Paper 63, 2004.

Manuscript received 30 April 2010
revised 2 February 2011
accepted for publication 17 February

Minghou Cheng
State Key Laboratory of Scientific and Engineering Computing
Institute of Computational Mathematics and Scientific/Engineering Computing
Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing 100190, China
E-mail address: chengmh@lsec.cc.ac.cn

Yu-Hong Dai
State Key Laboratory of Scientific and Engineering Computing
Institute of Computational Mathematics and Scientific/Engineering Computing
Academy of Mathematics and Systems Science, Chinese Academy of Sciences
Beijing 100190, China
E-mail address: dyh@lsec.cc.ac.cn