Journal of Computational and Applied Mathematics 234 (2010) 649–657

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A globally convergent derivative-free method for solving large-scale nonlinear monotone equations[☆]

Qin-Rong Yan[a], Xiao-Zhen Peng[a], Dong-Hui Li[b,*]

[a] Department of Science, Hubei Automotive Industries Institute, Shiyan, Hubei, 442002, China

[b] South China Normal University, Guangzhou, 510631, China

ARTICLE INFO

Article history:
Received 1 June 2008
Received in revised form 26 February 2009

MSC:
65H10
90C25

Keywords:
Monotone equation
Derivative-free method
Global convergence
Projection method

ABSTRACT

In this paper, we propose two derivative-free iterative methods for solving nonlinear monotone equations, which combines two modified HS methods with the projection method in Solodov and Svaiter (1998) [5]. The proposed methods can be applied to solve nonsmooth equations. They are suitable to large-scale equations due to their lower storage requirement. Under mild conditions, we show that the proposed methods are globally convergent. The reported numerical results show that the methods are efficient.

1. Introduction

We consider the nonlinear monotone equation

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

where $F : R^n \to R^n$ is continuous and monotone. By monotonicity, we mean

$$(F(x) - F(y))^T(x - y) \geq 0, \quad \forall x, y \in R^n.$$

We aim at developing conjugate gradient based method for solving large-scale nonlinear equations for which the Jacobian of F is not available. In the solutions of large-scale unconstrained optimization problems, the conjugate gradient methods are particularly welcome efficient due to their simplicity and lower storage. Quite recently, Zhang et al. [1] proposed a three term modified PRP method for unconstrained optimization problems. A good property of the modified PRP method is that it generates descent directions for the objective function. The reported numerical results in [1] show that it is competitive with the CG-DESCENT method [2]. The purpose of this paper is to use the idea of the modified PRP method in [1] to develop two derivative-free methods for solving nonlinear monotone equations.

Monotone equations arise in various applications. One important example is the subproblem in the generalized proximal algorithms with Bregman distances [3]. Some monotone variational inequality problems can also be converted into nonlinear monotone equations [4]. The study in the iterative methods for monotone equations have received much attention in the last decade [5–9]. Using the monotonicity of F, Solodov and Svaiter [5] proposed an inexact Newton method for solving (1.1)

☆ Supported by the NSF of China grant No. 10771057 and the Major Project of Ministry of Education of China grant No. 309023.

* Corresponding author.
E-mail address: dhli@scnu.edu.cn (D.-H. Li).

0377-0427/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2010.01.001

by combining the Newton method with a projection strategy. An attractive property of the method is that the whole sequence of iterates converges to a solution of the system without any regularity assumptions. In addition, the sequence of the distances from the iterates to the solution set of the equations is decreasing. Zhou and Toh [7] extended Solodov and Svaiter's result and obtained the superlinear convergence of a Newton-type methods even if the equation has singular solutions. Zhou and Li [9] applied Solodov and Svaiter's projection strategy to the BFGS methods and limited memory BFGS methods for solving monotone equation. Zhang and Zhou [8] developed a spectral gradient projection method by combining the spectral gradient method [10] with the projection method in [5]. Wang et al. [6] proposed a projection method for a system of nonlinear monotone equations with convex constraints. Based on the technique proposed by Solodov and Svaiter, recently, Li et al. [11] extended the two modified PRP methods in [12,1] to monotone equations. The reported numerical results indicate that the proposed method is promising.

In this paper, we propose two derivative-free iterative methods for solving large-scale nonlinear monotone equations. The proposed methods can be applied to solve nonsmooth equations. Moreover, they are suitable to large-scale equations due to their lower storage requirement. Under some mild conditions, we prove the global convergence of the method.

In the next section, we first propose the methods and show their global convergence. In the last section, we report some preliminary numerical results.

2. Algorithms and their global convergence

Given an initial point x_0, an iterative scheme for problem (1.1) generally generates a sequence of iterates $\{x_k\}$ by

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, \ldots,$$

which employs a line search procedure along the direction d_k to compute the stepsize α_k. Typical line searches include Armijo or Wolfe searches. For monotone equations, however, it is desirable to accelerate the process by using the monotonicity of the equation. Let $z_k = x_k + \alpha_k d_k$. By the monotonicity of F, the hyperplane

$$H_k = \{x \in R^n | (x - z_k)^T F(z_k) = 0\}$$

strictly separates x_k from the solution set of (1.1). Based on this fact, Solodov and Svaiter [5] advised to let the next iterate x_{k+1} be the projection of x_k onto this hyperplane H_k. That is, x_{k+1} is determined by

$$x_{k+1} = x_k - \frac{F(z_k)^T (x_k - z_k)}{\|F(z_k)\|^2} F(z_k).$$

We will propose two kinds of derivative-free methods for (1.1) by using this projection strategy.

In the remainder of this paper, we always assume that F satisfies the following assumptions.

Assumption A. (a) Function F is monotone:

$$(F(x) - F(y))^T (x - y) \geq 0, \quad \forall x, y \in R^n.$$

(b) Function F is Lipschitz continuous, that is, there exists a constant $L > 0$ such that

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad \forall x, y \in R^n.$$

2.1. Algorithm

We first propose a model method that is similar to that in [11], called the Projection Based Method.

Algorithm 2.1 (*Projection Based Method*).

Step 0. Given an initial point $x_0 \in R^n$, and constants $\beta > 0, \sigma > 0, \rho \in (0, 1)$. Let $k := 0$.
Step 1. Stop if $F(x_k) = 0$. Otherwise, determine d_k satisfying

$$F_k^T d_k \leq -\delta \|F_k\|^2, \tag{2.1}$$

where $\delta > 0$ is a constant and F_k is the abbreviation of $F(x_k)$.
Step 2. Let $\alpha_k = \max\{\beta\rho^i : i = 0, 1, \ldots\}$ be determined by the line search rule

$$- F(x_k + \alpha_k d_k)^T d_k \geq \sigma \alpha_k \|F(x_k + \alpha_k d_k)\| \cdot \|d_k\|^2. \tag{2.2}$$

Step 3. Compute $z_k = x_k + \alpha_k d_k$.
Step 4. Stop if $F(z_k) = 0$. Otherwise, let

$$x_{k+1} = x_k - \frac{F(z_k)^T (x_k - z_k)}{\|F(z_k)\|^2} F(z_k). \tag{2.3}$$

Let $k := k + 1$. Go to Step 1.

Remark 2.1. (a) If there exists a real function $f : \Re^n \to \Re$, such that $F(x) = \nabla f(x)$, where $\nabla f(x)$ is the gradient of f, then (2.1) means that d_k is a sufficiently descent direction of f at x_k.
(b) The line search rule (2.2) which is a modification of the derivative-free line search was proposed in [13]. It is not difficult to see that inequality (2.2) holds for all $\alpha_k > 0$ sufficiently small as long as d_k satisfies (2.1). Consequently, it is well defined and can be implemented by some backtracking process.

In this paper, we propose two kinds of d_k that satisfies (2.1). The first one is based on the three term conjugate gradient method proposed in [1]. Specifically, d_k is given by

$$d_k = \begin{cases} -F_0, & k = 0, \\ -F_k + \beta_k^{\mathrm{MHS}} d_{k-1} + \theta_k^M w_{k-1}, & k \geq 1, \end{cases} \tag{2.4}$$

where

$$\beta_k^{\mathrm{MHS}} = \frac{F_k^T w_{k-1}}{w_{k-1}^T d_{k-1}}, \qquad \theta_k^M = -\frac{F_k^T d_{k-1}}{w_{k-1}^T d_{k-1}},$$
$$w_k = y_k + t\|F_k\|\bar{s}_k, \qquad \bar{s}_k = z_k - x_k = \alpha_k d_k,$$
$$y_k = F_{k+1} - F_k, \qquad t = 1 + \|F_k\|^{-1} \max\left(0, -\frac{y_k^T \bar{s}_k}{\|\bar{s}_k\|^2}\right).$$

If in Step 1 of Algorithm 2.1 d_k is determined by (2.4), we call Algorithm 2.1 the MHS based method.

We then propose a two term MHS based method, which is Algorithm 2.1 with d_k determined by

$$d_k = \begin{cases} -F_0, & k = 0, \\ -F_k + \beta_k^{\mathrm{MHS}} \left(I - \dfrac{F_k F_k^T}{\|F_k\|^2}\right) d_{k-1}, & k \geq 1. \end{cases} \tag{2.5}$$

This formula was originated in [12] for unconstrained optimization problems. Here, we extend it to the nonlinear equations.

It is not difficult to know that d_k defined by (2.4) or (2.5) satisfies

$$F_k^T d_k = -\|F_k\|^2. \tag{2.6}$$

The remainder of this section is devoted to the global convergence of MHS based method and the two term MHS based method. We first derive some nice properties of Algorithm 2.1.

Lemma 2.1. *Let sequences $\{x_k\}$ and $\{z_k\}$ be generated by Algorithm 2.1. Then, we have*

$$\alpha_k \geq \min\left\{\beta, \frac{\delta\rho}{L + \sigma\|F(z_k')\|} \frac{\|F_k\|^2}{\|d_k\|^2}\right\}, \tag{2.7}$$

where $z_k' = x_k + \alpha_k' d_k$, $\alpha_k' = \rho^{-1}\alpha_k$.

Proof. If $\alpha_k \neq \beta$, by the line search process, we know that $\alpha_k' = \rho^{-1}\alpha_k$ does not satisfy (2.2). That is,

$$-F(z_k')^T d_k < \sigma\alpha_k' \|F(z_k')\| \cdot \|d_k\|^2.$$

It follows from (2.1) and the Lipschitz continuity of F that

$$\begin{aligned} \delta\|F_k\|^2 &\leq -F_k^T d_k = (F(z_k') - F(x_k))^T d_k - F(z_k')^T d_k \\ &\leq L\alpha_k'\|d_k\|^2 + \sigma\alpha_k'\|F(z_k')\| \cdot \|d_k\|^2 \\ &\leq \rho^{-1}\alpha_k(L + \sigma\|F(z_k')\|)\|d_k\|^2. \end{aligned}$$

Hence, it holds that

$$\alpha_k \geq \frac{\delta\rho}{L + \sigma\|F(z_k')\|} \frac{\|F_k\|^2}{\|d_k\|^2}.$$

This implies (2.7). $\square$

The following lemma shows a nice property of Algorithm 2.1. It can be proved in a way similar to the proof of Lemma 2.1 in [5].

Lemma 2.2. *Suppose that $\bar{x}$ satisfies $F(\bar{x}) = 0$. Let $\{x_k\}$ be generated by Algorithm 2.1. Then, we have*

$$\|x_{k+1} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2 - \|x_{k+1} - x_k\|^2.$$

Particularly, the sequence $\{\|x_k\|\}$ *is bounded and*

$$\sum_{k=0}^{\infty}\|x_{k+1}-x_k\|^2<\infty.$$

Lemma 2.3. *Suppose that* $\{x_k\}$ *is generated by Algorithm 2.1. Let* $s_k=x_{k+1}-x_k$*. Then, we have*

$$\sigma\alpha_k^2\|d_k\|^2\le\|s_k\|\le\alpha_k\|d_k\|. \tag{2.8}$$

Particularly, we have

$$\lim_{k\to\infty}\alpha_k\|d_k\|=0. \tag{2.9}$$

Proof. The first inequality in (2.8) follows from line search rule (2.2) directly, while the second inequality can be obtained by Cauchy inequality. Indeed, we have

$$\begin{aligned}\|x_{k+1}-x_k\| &= \frac{|F(z_k)^T(x_k-z_k)|}{\|F(z_k)\|}\le\frac{\|F(z_k)\|\,\|x_k-z_k\|}{\|F(z_k)\|}\\ &= \|z_k-x_k\|=\alpha_k\|d_k\|. \quad\Box\end{aligned}$$

Lemma 2.4. *Let* $\{x_k\}$ *be generated by Algorithm 2.1,* $\bar{x}$ *satisfy* $F(\bar{x})=0$*,* $z_k'=x_k+\alpha_k'd_k$*,* $\alpha_k'=\rho^{-1}\alpha_k$*. Then, the sequences* $\{\|F_k\|\}$ *and* $\{\|F(z_k')\|\}$ *are bounded. That is, there exists constant* $M\ge0$*, such that*

$$\|F_k\|\le M,\qquad\|F(z_k')\|\le M. \tag{2.10}$$

Proof. From Lemma 2.2, we know $\|x_k-\bar{x}\|\le\|x_0-\bar{x}\|$. It follows from (2.9) that there exists constant $M'>0$ such that $\alpha_k\|d_k\|\le M'$, and that

$$\begin{aligned}\|z_k'-\bar{x}\| &\le \|x_k-\bar{x}\|+\alpha_k'\|d_k\|\le\|x_0-\bar{x}\|+\rho^{-1}\alpha_k\|d_k\|\\ &\le \|x_0-\bar{x}\|+\rho^{-1}M'.\end{aligned}$$

Since $F(x)$ is Lipschitz continuous, we have

$$\|F_k\|=\|F(x_k)-F(\bar{x})\|\le L\|x_k-\bar{x}\|\le L\|x_0-\bar{x}\|$$

and

$$\|F(z_k')\|=\|F(z_k')-F(\bar{x})\|\le L\|z_k'-\bar{x}\|\le L(\|x_0-\bar{x}\|+\rho^{-1}M').$$

Denote $M=\max(L\|x_0-\bar{x}\|,L(\|x_0-\bar{x}\|+\rho^{-1}M'))$. We get (2.10). $\Box$

2.2. Global convergence

In this subsection, we establish the global convergence of the algorithms proposed in the previous subsection.
The following theorem establishes the global convergence of the MHS based method.

Theorem 2.1. *Let* $\{x_k\}$ *be generated by the MHS based method. Then, we have*

$$\liminf_{k\to\infty}\|F_k\|=0. \tag{2.11}$$

Proof. If (2.11) does not hold, there exists $\varepsilon>0$ such that

$$\|F_k\|\ge\varepsilon,\quad\forall k\ge0.$$

From (2.6), we know $\|F_k\|\le\|d_k\|$, which implies

$$\|d_k\|\ge\varepsilon,\quad\forall k\ge0.$$

By the definition of t, it is clear that $t\ge1-\|F_k\|^{-1}\frac{y_k^T\bar{s}_k}{\|\bar{s}_k\|^2}$. So, we get

$$\begin{aligned}w_k^T\bar{s}_k &= y_k^T\bar{s}_k+t\|F_k\|\,\|\bar{s}_k\|^2\\ &\ge y_k^T\bar{s}_k+\left(1-\|F_k\|^{-1}\frac{y_k^T\bar{s}_k}{\|\bar{s}_k\|^2}\right)\|F_k\|\,\|\bar{s}_k\|^2\\ &= \|F_k\|\,\|\bar{s}_k\|^2\ge\varepsilon\|\bar{s}_k\|^2.\end{aligned} \tag{2.12}$$

This means,

$$w_{k-1}^T d_{k-1} = \frac{1}{\alpha_{k-1}} w_{k-1}^T \bar{s}_{k-1} \geq \frac{1}{\alpha_{k-1}} \varepsilon \|\bar{s}_{k-1}\|^2 = \varepsilon \alpha_{k-1} \|d_{k-1}\|^2. \tag{2.13}$$

Moreover, by the definition of w_k, (2.10) and the fact

$$\|y_k\| = \|F(x_{k+1}) - F(x_k)\| \leq L\|s_k\|,$$

we have

$$\begin{aligned}
\|w_k\| &\leq \|y_k\| + t\|F_k\| \, \|\bar{s}_k\| \\
&= \|y_k\| + \left[1 + \|F_k\|^{-1} \max\left(0, -\frac{y_k^T \bar{s}_k}{\|\bar{s}_k\|^2}\right)\right] \|F_k\| \, \|\bar{s}_k\| \\
&\leq \|y_k\| + \left(1 + \|F_k\|^{-1} \frac{|y_k^T \bar{s}_k|}{\|\bar{s}_k\|^2}\right) \|F_k\| \, \|\bar{s}_k\| \\
&\leq 2\|y_k\| + \|F_k\| \, \|\bar{s}_k\| \\
&\leq 2L\|s_k\| + M\alpha_k \|d_k\|.
\end{aligned} \tag{2.14}$$

It follows from (2.4) that for all $k \geq 1$,

$$\begin{aligned}
\|d_k\| &\leq \|F_k\| + \frac{|F_k^T w_{k-1}|}{|w_{k-1}^T d_{k-1}|} \|d_{k-1}\| + \frac{|F_k^T d_{k-1}|}{|w_{k-1}^T d_{k-1}|} \|w_{k-1}\| \\
&\leq \|F_k\| + \frac{2\|F_k\| \, \|w_{k-1}\| \, \|d_{k-1}\|}{|w_{k-1}^T d_{k-1}|}.
\end{aligned} \tag{2.15}$$

Combining (2.10), (2.13), (2.14) and (2.15) with Lemma 2.3, we get

$$\begin{aligned}
\|d_k\| &\leq M + \frac{2M(2L\|s_{k-1}\| + M\alpha_{k-1}\|d_{k-1}\|)}{\varepsilon \alpha_{k-1} \|d_{k-1}\|} \\
&\leq M + \frac{2M}{\varepsilon}(M + 2L).
\end{aligned}$$

Let $\overline{M} = M + \frac{2M}{\varepsilon}(M + 2L)$. We obtain

$$\|d_k\| \leq \overline{M}.$$

This together with Lemmas 2.1 and 2.4 and inequalities $\|d_k\| \geq \varepsilon$, $\|F_k\| \geq \varepsilon$ implies for all k sufficiently large,

$$\begin{aligned}
\alpha_k \|d_k\| &\geq \min\left(\beta, \frac{\delta \rho}{L + \sigma \|F(z_k')\|} \frac{\|F_k\|^2}{\|d_k\|^2}\right) \|d_k\| \\
&\geq \min\left(\beta \varepsilon, \frac{\delta \rho \varepsilon^2}{(L + M\sigma)\overline{M}}\right) > 0.
\end{aligned}$$

The last inequality yields a contradiction with (2.9). Consequently, (2.11) holds. The proof is complete. □

The next theorem establishes the global convergence of the two term MHS based method.

Theorem 2.2. *Let $\{x_k\}$ is generated by the two term MHS based method. Then, Eq. (2.11) holds.*

Proof. Suppose on the contrary that (2.11) does not hold. Then, there is a constant $\varepsilon > 0$ such that $\varepsilon \leq \|F_k\| \leq M$. We also have by (2.5) that

$$\begin{aligned}
\|d_k\| &\leq \|F_k\| + |\beta_k^{\mathrm{MHS}}| \left\| \left(I - \frac{F_k F_k^T}{\|F_k\|^2}\right) d_{k-1} \right\| \\
&\leq \|F_k\| + \frac{|F_k^T w_{k-1}|}{|w_{k-1}^T d_{k-1}|} \|d_{k-1}\| \\
&\leq \|F_k\| + \frac{\|F_k\| \, \|w_{k-1}\| \, \|d_{k-1}\|}{|w_{k-1}^T d_{k-1}|}.
\end{aligned}$$

In a way similar to the proof of Theorem 2.1, we can derive a contradiction. Consequently, (2.11) holds. □

Table 4.1
Test results for Problem 1.

Initial	Dim	MHS			TMHS		
		IT	NF	T	IT	NF	T
x_0	100	9	19	0.00000E+00	9	19	0.00000E+00
x_1	100	8	23	0.00000E+00	8	23	0.00000E+00
x_2	100	11	21	0.00000E+00	10	19	0.00000E+00
x_3	100	10	30	0.00000E+00	10	30	0.00000E+00
x_4	100	9	17	0.00000E+00	7	13	0.00000E+00
x_5	100	4	7	0.00000E+00	2	3	0.00000E+00
x_6	100	9	17	0.00000E+00	7	13	0.00000E+00
x_7	100	11	21	0.00000E+00	10	19	0.00000E+00
x_0	1000	9	19	0.00000E+00	9	19	0.00000E+00
x_1	1000	8	23	0.00000E+00	8	23	0.00000E+00
x_2	1000	11	21	0.15625E−01	11	21	0.00000E+00
x_3	1000	10	30	0.15625E−01	10	30	0.00000E+00
x_4	1000	8	15	0.15625E−01	7	13	0.00000E+00
x_5	1000	4	7	0.00000E+00	3	5	0.15625E−01
x_6	1000	8	15	0.00000E+00	7	13	0.00000E+00
x_7	1000	11	21	0.00000E+00	11	21	0.00000E+00
x_0	3000	9	19	0.15625E−01	9	19	0.15625E−01
x_1	3000	8	23	0.15625E−01	8	23	0.15625E−01
x_2	3000	11	21	0.15625E−01	11	21	0.15625E−01
x_3	3000	10	30	0.31250E−01	10	30	0.15625E−01
x_4	3000	7	13	0.15625E−01	7	13	0.15625E−01
x_5	3000	4	7	0.00000E+00	4	7	0.15625E−01
x_6	3000	7	13	0.00000E+00	7	13	0.15625E−01
x_7	3000	11	21	0.15625E−01	11	21	0.31250E−01

3. Numerical experiments

In this section, we report some numerical results with the proposed method. We test the performance of Algorithm 2.1 on the following four problems with various sizes.

Problem 1. The elements of function F are given by

$$F_i(x) = 2x_i - \sin|x_i|, \quad i = 1, \dots, n.$$

Problem 2. The elements of function F are given by

$$F_i(x) = x_i - \frac{1}{n}x_i^2 + \frac{1}{n}\sum_{i=1}^{n} x_i + i, \quad i = 1, \dots, n.$$

Problem 3. The elements of function F are given by $F_1(x) = \frac{1}{3}x_1^3 + \frac{1}{2}x_2^2$,

$$F_i(x) = -0.5x_i^2 + \frac{i}{3}x_i^3 + \frac{1}{2}x_{i+1}^2, \quad i = 2, \dots, n-1,$$

and $F_n = -\frac{1}{2}x_n^2 + \frac{n}{3}x_n^3$.

Problem 4. The elements of function F are given by $F_1(x) = x_1 - \mathrm{e}^{\cos\left(\frac{x_1+x_2}{n+1}\right)}$,

$$F_i(x) = x_i - \mathrm{e}^{\cos\left(\frac{x_{i-1}+x_i+x_{i+1}}{n+1}\right)}, \quad i = 2, \dots, n-1,$$

and $F_n(x) = x_n - \mathrm{e}^{\cos\left(\frac{x_{n-1}+x_n}{n+1}\right)}$.

Problem 1 comes from [9] with a slight modification. Problems 3 and 4 come from [14]. Problem 2 was constructed by us.

It is not difficult to see that all the functions F in these problems are monotone.

We implemented Algorithm 2.1 with the following parameter. We set $\rho = 0.6$ and $\sigma = 10^{-4}$ in the line search step, i.e. Step 2 of Algorithm 2.1. We choose β in Step 2 of the algorithm to vary with k. Specifically, we let

$$\beta = \frac{s_k^T s_k}{s_k^T y_k},$$

where $s_k = x_{k+1} - x_k$ and $y_k = F(x_{k+1}) - F(x_k)$. Such choice of β was used in [14] and was called the spectral coefficient. From the monotonicity and the Lipschitz continuity of F, it is not difficult to show that

$$0 \le y_k^T s_k \le L s_k^T s_k,$$

Table 4.2
Test results for Problem 2.

Initial	Dim	MHS			TMHS		
		IT	NF	T	IT	NF	T
x_0	100	11	36	0.00000E+00	5	16	0.00000E+00
x_1	100	11	41	0.00000E+00	4	15	0.00000E+00
x_2	100	38	349	0.31250E−01	5	17	0.15625E−01
x_3	100	26	176	0.15625E−01	5	17	0.00000E+00
x_4	100	34	328	0.31250E−01	5	17	0.00000E+00
x_5	100	18	95	0.15625E−01	5	17	0.00000E+00
x_6	100	18	90	0.00000E+00	5	17	0.00000E+00
x_7	100	32	283	0.15625E−01	5	17	0.15625E−01
x_0	1000	18	87	0.57812E+00	14	67	0.45312E+00
x_1	1000	21	115	0.76562E+00	12	54	0.34375E+00
x_2	1000	22	115	0.76562E+00	13	55	0.35938E+00
x_3	1000	19	92	0.60938E+00	14	68	0.45312E+00
x_4	1000	21	113	0.75000E+00	13	57	0.37500E+00
x_5	1000	22	117	0.76562E+00	14	62	0.39062E+00
x_6	1000	22	113	0.73438E+00	13	54	0.35938E+00
x_7	1000	22	110	0.71875E+00	13	55	0.35938E+00
x_0	3000	18	85	0.49531E+01	21	110	0.64062E+01
x_1	3000	22	103	0.60000E+01	23	118	0.68438E+01
x_2	3000	23	108	0.62812E+01	17	78	0.45312E+01
x_3	3000	22	110	0.64062E+01	17	72	0.42031E+01
x_4	3000	17	70	0.40625E+01	15	61	0.35469E+01
x_5	3000	22	104	0.60625E+01	17	75	0.43594E+01
x_6	3000	17	70	0.40781E+01	17	71	0.41250E+01
x_7	3000	23	115	0.67031E+01	17	78	0.45469E+01

Table 4.3
Test results for Problem 3.

Initial	Dim	MHS			TMHS		
		IT	NF	T	IT	NF	T
x_0	100	30	222	0.15625E−01	42	379	0.15625E−01
x_1	100	30	227	0.00000E+00	25	164	0.00000E+00
x_2	100	26	169	0.00000E+00	7	47	0.00000E+00
x_3	100	31	250	0.00000E+00	7	46	0.00000E+00
x_4	100	9	17	0.00000E+00	2	6	0.00000E+00
x_5	100	–	–	–	2	6	0.00000E+00
x_6	100	7	20	0.00000E+00	1	8	0.00000E+00
x_7	100	31	216	0.00000E+00	13	95	0.00000E+00
x_0	1000	97	1057	0.93750E−01	51	360	0.31250E−01
x_1	1000	118	1399	0.12500E+00	57	500	0.46875E−01
x_2	1000	58	514	0.46875E−01	34	217	0.31250E−01
x_3	1000	58	572	0.62500E−01	63	645	0.62500E−01
x_4	1000	8	15	0.00000E+00	1	2	0.00000E+00
x_5	1000	556	6380	0.53125E+00	1	3	0.00000E+00
x_6	1000	8	27	0.00000E+00	3	17	0.00000E+00
x_7	1000	44	343	0.15625E−01	48	383	0.46875E−01
x_0	3000	184	2225	0.54688E+00	98	857	0.25000E+00
x_1	3000	182	2233	0.53125E+00	99	808	0.23438E+00
x_2	3000	98	1006	0.26562E+00	101	900	0.26562E+00
x_3	3000	145	1744	0.42188E+00	110	880	0.23438E+00
x_4	3000	8	15	0.15625E−01	8	15	0.00000E+00
x_5	3000	73	802	0.20312E+00	500	5749	0.15625E+01
x_6	3000	8	29	0.15625E−01	8	29	0.15625E−01
x_7	3000	100	1051	0.28125E+00	83	662	0.18750E+00

where L is Lipschitz constant. If $\beta \notin [\sigma_{\min}, \sigma_{\max}]$, we replace β by

$$\beta = \begin{cases} 1 & \text{if } \|F(x_k)\| > 1, \\ \|F(x_k)\|^{-1} & \text{if } 10^{-5} \le \|F(x_k)\| \le 1, \\ 10^5 & \text{if } \|F(x_k)\| < 10^{-5}, \end{cases}$$

where $\sigma_{\min} = 10^{-10}$ and $\sigma_{\max} = 10^{10}$. We stop the iteration if the inequality

$$\frac{\|F(x_k)\|}{\sqrt{n}} \le e_a + e_r \frac{\|F(x_0)\|}{\sqrt{n}}$$

is satisfied, where $e_a = 10^{-5}$ and $e_r = 10^{-4}$. This stopping criterion comes from [14].

Table 4.4
Test results for Problem 4.

Initial	Dim	MHS			TMHS		
		IT	NF	T	IT	NF	T
x_0	100	9	26	0.00000E+00	15	72	0.15625E−01
x_1	100	9	26	0.00000E+00	13	54	0.00000E+00
x_2	100	10	30	0.00000E+00	9	27	0.00000E+00
x_3	100	11	33	0.00000E+00	9	27	0.15625E−01
x_4	100	11	33	0.00000E+00	9	27	0.00000E+00
x_5	100	11	33	0.00000E+00	9	27	0.00000E+00
x_6	100	11	33	0.00000E+00	9	27	0.00000E+00
x_7	100	10	30	0.00000E+00	9	27	0.00000E+00
x_0	1000	2	5	0.00000E+00	2	5	0.00000E+00
x_1	1000	1	2	0.00000E+00	1	2	0.00000E+00
x_2	1000	10	30	0.15625E−01	10	30	0.15625E−01
x_3	1000	11	33	0.15625E−01	10	30	0.15625E−01
x_4	1000	11	33	0.15625E−01	10	30	0.00000E+00
x_5	1000	11	33	0.15625E−01	10	30	0.15625E−01
x_6	1000	11	33	0.15625E−01	10	30	0.15625E−01
x_7	1000	10	30	0.15625E−01	10	30	0.15625E−01
x_0	3000	2	5	0.00000E+00	2	5	0.15625E−01
x_1	3000	1	2	0.00000E+00	1	2	0.00000E+00
x_2	3000	10	30	0.31250E−01	10	30	0.31250E−01
x_3	3000	11	33	0.46875E−01	11	33	0.46875E−01
x_4	3000	11	33	0.46875E−01	11	33	0.46875E−01
x_5	3000	11	33	0.46875E−01	11	33	0.46875E−01
x_6	3000	11	33	0.46875E−01	11	33	0.31250E−01
x_7	3000	11	33	0.46875E−01	10	30	0.46875E−01

The codes were written in FORTRAN 90 in double precision arithmetic and run on a PC (`CPU 1.6 GHz, 256 MB memory`) with Windows operating system.

The results are listed in Tables 4.1–4.4, where $x_0 = (10, 10, \dots, 10)^T$, $x_1 = (-10, -10, \dots, -10)^T$, $x_2 = (1, 1, \dots, 1)^T$, $x_3 = (-1, -1, \dots, -1)^T$, $x_4 = \left(1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}\right)^T$, $x_5 = (0.1, 0.1, \dots, 0.1)^T$, $x_6 = \left(\frac{1}{n}, \frac{2}{n}, \dots, 1\right)^T$ and $x_7 = \left(1 - \frac{1}{n}, 1 - \frac{2}{n}, \dots, 0\right)^T$. In Tables 4.1–4.4, we report the dimension of the problem (Dim), the number of iterations (IT), the number of function evaluations (including the additional functional evaluations that Algorithm 2.1 uses for approximating initial steplength σ_k) (NF) and the CPU time in seconds (T). We use the symbol '—' to specify either of the following two cases:

(a) the number of iterations is greater than 1000; or
(b) the number of backtracking iterations at some line search step is greater than 50.

In the tables, MHS and TMHS denote the MHS based method and the two term MHS based method, respectively.

Tables 4.1–4.4 show that the proposed algorithm provides an efficient method for large-scale nonlinear systems of equations. We also see from Tables 4.1–4.4 that the TMHS method performs better than the MHS method.

Acknowledgement

The authors would like to thank Dr. Wanyou Cheng for his valuable suggestions in numerical experiments of the paper.

References

[1] L. Zhang, W.J. Zhou, D.H. Li, A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence, IMA Journal of Numerical Analysis 26 (2006).
[2] W.W. Hager, H.C. Zhang, A new conjugate gradient method with guaranteed descent and an efficent line search, SIAM Journal of Optimization 16 (2005) 170–192.
[3] A.N. Iusem, M.V. Solodov, Newton-type methods with generalized distances for constrained optimization, Optimization 41 (1997) 257–278.
[4] Y.B. Zhao, D. Li, Monotonicity of fixed point and normal mappings associated with variational inequality and its application, SIAM Journal on Optimization 11 (2001) 962–973.
[5] M.V. Solodov, B.F. Svaiter, A globally convergent inexact Newton method for systems of monotone equations, in: M. Fukushima, L. Qi (Eds.), Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Kluwer Academic Publishers, 1998, pp. 355–369.
[6] C.W. Wang, Y.J. Wang, C.L. Xu, A projection method for a system of nonlinear monotone equations with convex constraints, Mathematical Methods of Operations Research 66 (2007) 33–46.
[7] G. Zhou, K.C. Toh, Superlinear convergence of a Newton-type algorithm for monotone equations, Journal of Optimization Theory and Applications 125 (2005) 205–221.
[8] L. Zhang, W.J. Zhou, Spectral gradient projection method for solving nonlinear monotone equations, Journal of Computational and Applied Mathematics 196 (2006) 478–484.
[9] W.J. Zhou, D.H. Li, Limited memory BFGS method for nonlinear monotone equations, Journal of Computational Mathematics 25 (2007) 89–96.
[10] J. Barzilai, J.M. Borwein, Two-point step size gradient methods, IMA Journal of Numerical Analysis 8 (1988) 141–148.
[11] D.H. Li, Q.N. Li, A class of derivative-free methods for large-scale nonlinear monotone equations, Working paper, 2007.

[12] W.Y. Cheng, A two term PRP based descent method, Numerical Functional Analysis and Optimization 28 (2007) 1217–1230.
[13] W.J. Zhou, Convergence properties of a quasi-Newton method and its applications, Ph.D. Thesis, College of Mathematics and Econometrics, Hunan University, Changsha, China, 2006.
[14] W. La Cruz, J.M. Martínez, M. Raydan, Spectral residual method without gradient information for solving large-scale nonlinear systems of equations, Mathematics of Computation 75 (2006) 1429–1448.