AMO - Advanced Modeling and Optimization, Volume 14, Number 3, 2012

On the convergence of a derivative-free HS type method for symmetric nonlinear equations

Weijun Zhou[1] and Xinlong Chen
Department of Mathematics, Changsha University of Science and Technology, Changsha 410004, China

Abstract

The Hestenes-Stiefel (HS) method is one of the most efficient nonlinear conjugate gradient methods for nonlinear optimization. In this paper, we propose a derivative-free HS type method for symmetric nonlinear equations without computing its Jacobian, which is suitable for large-scale problems. It is an extension of the three-term HS nonlinear conjugate gradient method for nonlinear optimization presented by Zhang, Zhou and Li [Optim. Methods Softw., 22 (2007), pp. 697-711]. By the use of some approximation norm descent line search, we prove the strong global convergence property of the proposed method. Moreover, R-linear convergence rate is established for the proposed method under some conditions. Some extensions are also discussed.

Keywords. HS method, Symmetric nonlinear equations, global convergence, linear convergence.

AMS subject classification 2010. 65K05, 90C30.

1 Introduction

It is well-known that the Hestenes-Stiefel (HS) method is one of the most efficient conjugate gradient methods for nonlinear optimization. The purpose of this paper is to extend the HS type methods from nonlinear optimization to nonlinear equations.

In this paper, we consider the following symmetric nonlinear equations

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

where $F : R^n \to R^n$ is a continuously differentiable mapping whose Jacobian $J(x) = F'(x)$ is symmetric, i.e., $J(x) = J(x)^T$. This problem covers many practical problems such as the KKT system of unconstrained optimization problem, the saddle problem and the discretized two-point boundary value problem [5, 11]. There are many efficient methods for this problem such as Newton method, Gauss-Newton method and quasi-Newton methods [1, 2, 4, 8, 9, 11, 16, 17, 18].

Recently, Li and Wang [7] extended the MFR conjugate gradient method in [14] for nonconvex optimization to nonlinear equations, which converges globally with a norm descent line search. However, there are no local convergence results for this method.

[1]This work was supported by the NSF foundation (10901026) of China, The Key Project of the Scientific Research Fund of the Hunan Provincial Education Department.
AMO - Advanced Modeling and Optimization. ISSN: 1841-4311

In this paper, we will extend the three-term HS nonlinear conjugate gradient method in [15] for nonlinear optimization to nonlinear equations without computing the Jacobian of the underlying system, which is suitable for large-scale problems. Moreover, we will investigate its global and local convergence properties under some conditions.

In the next section, we give the motivation and the algorithm in detail. In Section 3, we first show the strong global convergence of the proposed method. Then we prove its R-linear convergence under suitable conditions. In Section 4, we extend the proposed method to norm descent case.

2 Motivation and the algorithm

The motivation of the paper is the Gauss-Newton-based BFGS (GN-BFGS) method proposed by Li and Fukushima [5] for symmetric nonlinear equations. At each iteration, the GN-BFGS method [5] produces the search direction d_k by solving the linear equations

$$B_k d = -\tilde{g}_k,$$

where

$$\tilde{g}_k = \alpha_{k-1}^{-1}\big(F(x_k + \alpha_{k-1}F_k) - F_k\big), \tag{2.1}$$

$F_k = F(x_k)$, α_{k-1} is the stepsize given by some line search, and the matrix B_k is updated by the BFGS formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k},$$

where

$$y_k = F(x_k + \delta_k) - F_k, \quad \delta_k = F_{k+1} - F_k, \quad s_k = x_{k+1} - x_k = \alpha_k d_k. \tag{2.2}$$

It is easy to see that if $\|s_k\|$ is small, then

$$B_{k+1}s_k = y_k \approx J_{k+1}^T J_{k+1} s_k.$$

Let f be the norm square merit function defined by

$$f(x) = \frac{1}{2}\|F(x)\|^2. \tag{2.3}$$

Then the problem (1.1) is equivalent to the following global minimization problem:

$$\min f(x), \quad x \in R^n. \tag{2.4}$$

The three-term HS nonlinear conjugate gradient method in [15] for (2.4) generates the search direction by

$$d_k = \begin{cases} -\nabla f(x_k), & \text{if} \quad k = 0, \\ -\nabla f(x_k) + \beta_k^{HS} d_{k-1} - \theta_k z_{k-1}, & \text{if} \quad k \geq 1, \end{cases} \tag{2.5}$$

where $\nabla f(x)$ is the gradient of the function $f(x)$ and

$$\beta_k^{HS} = \frac{\nabla f(x_k)^T z_{k-1}}{d_{k-1}^T z_{k-1}}, \quad \theta_k = \frac{\nabla f(x_k)^T d_{k-1}}{d_{k-1}^T z_{k-1}}, \quad z_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1}). \tag{2.6}$$

An attractive property of (2.5) is that it satisfies $d_k^T \nabla f(x_k) = -\|\nabla f(x_k)\|^2$, which means that d_k provides a sufficient descent direction for f. This property is independent of the objective function convexity and line search used, which is very important for global convergence of this method.

Denote $J_k = J(x_k)$. Then the gradient of the function f defined by (2.3) at x_k is given by

$$\nabla f(x_k) = J_k^T F_k,$$

which is involved in calculation of the Jacobian. This shows that the method (2.5) is not suitable for such problems where the Jacobian is not available or very difficult to compute when it is applied to the special function (2.3).

To avoid computing the Jacobian, we use a similar way to (2.1) to deal with this problem. Let $\{\epsilon_k\}$ and ϵ be a given positive sequence and a positive constant satisfying

$$\sum_{k=0}^{\infty} \epsilon_k \le \epsilon < \infty. \tag{2.7}$$

Then by symmetry of the Jacobian $J(x)$, we have

$$\nabla f(x_k) = J_k^T F_k \approx g_k \triangleq \frac{F(x_k + \epsilon_k F_k) - F_k}{\epsilon_k}. \tag{2.8}$$

Moreover, we note that y_k defined by (2.2) is an approximation to z_k given by (2.6), i.e.,

$$y_k \approx z_k = J_{k+1}^T F_{k+1} - J_k^T F_k.$$

If we replace $\nabla f(x_k)$ and z_{k-1} in (2.5) by g_k and y_{k-1}, respectively, then we get a derivative-free HS type method for (1.1) without computing the Jacobian, that is, at each iteration, the search direction d_k is given by

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} d_{k-1} - \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} y_{k-1}, & \text{if} \quad k \ge 1, \end{cases} \tag{2.9}$$

where g_k and y_k are defined by (2.8) and (2.2), respectively.

In order to guarantee the direction d_k defined by (2.9) is well defined, we use the regularization technique [6] to modify the formula (2.9) as follows

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k^{MHS} d_{k-1} - \theta_k^M \gamma_{k-1}, & \text{if} \quad k \ge 1, \end{cases} \tag{2.10}$$

where

$$\beta_k^{MHS} = \frac{g_k^T \gamma_{k-1}}{d_{k-1}^T \gamma_{k-1}}, \quad \theta_k^M = \frac{g_k^T d_{k-1}}{d_{k-1}^T \gamma_{k-1}}, \tag{2.11}$$

$$\gamma_{k-1} = y_{k-1} + t_k s_{k-1}, \quad t_k = \max\left\{0, -\frac{y_{k-1}^T s_{k-1}}{\|s_{k-1}\|^2}\right\} + \mu, \tag{2.12}$$

y_{k-1} is given by (2.2) and $\mu > 0$ is a given small positive constant.

It is clear from (2.12) that $d_{k-1}^T \gamma_{k-1} \ge \mu \alpha_{k-1} \|d_{k-1}\|^2 > 0$, which implies that the search direction (2.10) is well defined. Moreover, it still satisfies

$$d_k^T g_k = -\|g_k\|^2. \tag{2.13}$$

Note that the search direction d_k is no longer a descent direction of the norm square merit function f in (2.3). Then we adopt the approximate norm descent line search

(2.14) below to globalize this method, which is motivated by that of [5]. The following is the complete derivative-free HS type method for symmetric nonlinear equations (1.1).
Algorithm 2.1

Step 1. Choose a starting point $x_0 \in R^n$, several constants $\mu > 0$, $\sigma \in (0,1)$ and $r \in (0,1)$. Let $k := 0$.

Step 2. Compute d_k by (2.10)-(2.12).

Step 3. Compute $\alpha_k = \max\{1, r^1, r^2, \cdots\}$ with $\alpha = r^i$ satisfying

$$f(x_k + \alpha d_k) - f(x_k) \le \sigma \alpha g_k^T d_k + \epsilon_k f(x_k), \tag{2.14}$$

where $\{\epsilon_k\}$ is a given positive sequence satisfying (2.7).

Step 4. Set $x_{k+1} = x_k + \alpha_k d_k$. Let $k := k+1$ and go to Step 2.

Remark 2.1 The line search (2.14) is well defined since if $\alpha \to 0^+$, the left-hand side of (2.14) tends to zero and the right-hand side goes to the positive term $\epsilon_k f(x_k)$, which implies that the line search (2.14) is satisfied for all sufficiently small $\alpha > 0$.

3 Global and linear convergence properties

In this section, we first investigate global convergence of Algorithm 2.1. Then we discuss its local convergence property.

To begin with, let ϵ be a given constant satisfying (2.7), define the level set

$$\Omega = \{x| \ f(x) \le e^{\epsilon} f(x_0)\}. \tag{3.1}$$

Then we make the following assumptions to ensure global convergence of Algorithm 2.1.
Assumption 3.1

(i) The level set Ω is bounded.

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

$$\|J(x) - J(y)\| \le L\|x - y\|, \quad \forall x, y \in N. \tag{3.2}$$

It is clear that Assumption 3.1 implies that there exist three positive constants L_1, L_2 and M such that

$$\|\nabla f(x) - \nabla f(y)\| \le L_1\|x - y\|, \quad \forall x, y \in N, \tag{3.3}$$

$$\|F(x)\| \le M, \quad \|J(x)\| \le M, \quad \forall x \in N. \tag{3.4}$$

$$\|F(x) - F(y)\| \le L_2\|x - y\|, \quad \forall x, y \in N. \tag{3.5}$$

From (2.8) and (3.4), we have

$$\|g_k\| = \left\| \int_0^1 J(x_k + t\epsilon_k F_k) F_k \right\| \le M^2. \tag{3.6}$$

To prove the global convergence of the proposed method, we present some useful lemmas. By the same argument as Lemma 2.1 in [5], we have the following result.

Lemma 3.1. *Let Assumption 3.1 hold and the sequence* $\{x_k\}$ *be generated by Algorithm 2.1. Then the sequence* $\{f(x_k)\}$ *converges and* $x_k \in \Omega$ *for all* $k \geq 0$.

Lemma 3.2. *Let Assumption 3.1 hold and the sequence* $\{x_k\}$ *be generated by Algorithm 2.1. Then there exists a positive constant* M_1 *such that*

$$d_k^T \gamma_k \geq \mu \|d_k\| \|s_k\|, \quad \|\gamma_k\| \leq M_1 \|s_k\|. \tag{3.7}$$

Proof. The first inequality follows from the definition of (2.12) directly.
From (2.12), (2.2) and (3.5), we have

$$\begin{aligned} \|\gamma_k\| &\leq 2\|y_k\| + \mu\|s_k\| \\ &= 2\|F(x_k + \delta_k) - F_k\| + \mu\|s_k\| \\ &\leq 2L_2\|\delta_k\| + \mu\|s_k\| \\ &= 2L_2\|F_{k+1} - F_k\| + \mu\|s_k\| \\ &\leq (2L_2^2 + \mu)\|s_k\|. \end{aligned}$$

Set $M_1 = (2L_2^2 + \mu)$, then we get the second inequality of (3.7). □

Lemma 3.3. *Let Assumption 3.1 hold and the sequence* $\{x_k\}$ *be generated by Algorithm 2.1. Then we have*

$$\sum_{k=0}^{\infty} \alpha_k \|g_k\|^2 < \infty. \tag{3.8}$$

Proof. It follows from (2.14), (2.13) and (2.7) directly. □

The inequality (3.8) implies that

$$\lim_{k \to \infty} \alpha_k \|g_k\|^2 = 0. \tag{3.9}$$

Lemma 3.4. *Let Assumption 3.1 hold and the sequence* $\{x_k\}$ *be generated by Algorithm 2.1. Then there exists a positive constant* M_2 *such that*

$$\|d_k\| \leq M_2 \|g_k\|. \tag{3.10}$$

Proof. By the definition of the direction (2.10)-(2.12) and (3.7), we have

$$\begin{aligned} \|d_k\| &\leq \|g_k\| + \frac{2\|g_k\|\|\gamma_{k-1}\|\|d_{k-1}\|}{d_{k-1}^T y_{k-1}} \\ &= \Big(1 + \frac{2\|\gamma_{k-1}\|\|d_{k-1}\|}{d_{k-1}^T y_{k-1}}\Big)\|g_k\| \\ &\leq \Big(1 + \frac{2M_1\|s_{k-1}\|\|d_{k-1}\|}{\mu\|s_{k-1}\|\|d_{k-1}\|}\Big)\|g_k\| \\ &= \Big(1 + \frac{2M_1}{\mu}\Big)\|g_k\|. \end{aligned}$$

Set $M_2 = 1 + \frac{2M_1}{\mu}$, then we obtain (3.10). □

Moreover, from (2.13), we have

$$\|g_k\| \leq \|d_k\|, \tag{3.11}$$

which together with (3.10) shows that $\|d_k\|$ is equivalent to $\|g_k\|$.

Lemma 3.5. *Let Assumption 3.1 hold and $\{x_k\}$ be generated by Algorithm 2.1. If $\alpha_k \neq 1$, then we have*

$$\alpha_k \geq m_1 \frac{-g_k^T d_k}{\|d_k\|^2} - m_2 \frac{\epsilon_k \|F_k\|^2}{\|d_k\|} = m_1 \frac{\|g_k\|^2}{\|d_k\|^2} - 2m_2 \frac{\epsilon_k f(x_k)}{\|d_k\|}, \tag{3.12}$$

where m_1 and m_2 are two positive constants.

Proof. If $\alpha_k < 1$, then the line search (2.14) implies that

$$f(x_k + r^{-1}\alpha_k d_k) - f(x_k) > \sigma \alpha_k r^{-1} g_k^T d_k + \epsilon_k f(x_k) \geq \sigma \alpha_k r^{-1} g_k^T d_k.$$

By the mean-value theorem and (3.3), we can easily get

$$f(x_k + r^{-1}\alpha_k d_k) - f(x_k) \leq r^{-1}\alpha_k \nabla f(x_k)^T d_k + L_1 r^{-2} \alpha_k^2 \|d_k\|^2.$$

Then we have

$$\begin{aligned} L_1 r^{-1} \alpha_k \|d_k\|^2 &\geq \sigma g_k^T d_k - \nabla f(x_k)^T d_k \\ &= -(1-\sigma) g_k^T d_k + \big(g_k - \nabla f(x_k)\big)^T d_k \\ &\geq -(1-\sigma) g_k^T d_k - \|g_k - \nabla f(x_k)\| \|d_k\|. \end{aligned}$$

From (2.8), (3.2) and the symmetry of the Jacobian $J(x)$, we have

$$\begin{aligned} \|\nabla f(x_k) - g_k\| &= \left\| J_k F_k - \int_0^1 J(x_k + t\epsilon_k F_k) F_k dt \right\| \\ &\leq \|F_k\| \left\| \int_0^1 \Big(J(x_k + t\epsilon_k F_k) - J_k \Big) dt \right\| \\ &\leq L \epsilon_k \|F_k\|^2. \end{aligned}$$

The above two inequalities yield that

$$\alpha_k \geq \frac{r(1-\sigma)}{L_1} \frac{(-g_k^T d_k)}{\|d_k\|^2} - \frac{rL}{L_1} \frac{\epsilon_k \|F_k\|^2}{\|d_k\|}. \tag{3.13}$$

Set $m_1 = \frac{r(1-\sigma)}{L_1}$ and $m_2 = \frac{rL}{L_1}$, from (2.13), we get (3.12). The proof is then finished. □

Now we give the following global convergence result for Algorithm 2.1.

Theorem 3.1. *Let Assumption 3.1 hold. Then the sequence $\{x_k\}$ be generated by Algorithm 2.1 converges globally in the sense that*

$$\lim_{k \to \infty} \|\nabla f(x_k)\| = 0. \tag{3.14}$$

Proof. We prove the theorem by contradiction. Suppose it is not true, then there exist a positive constant τ and an infinite index set T such that

$$\|\nabla f(x_k)\| \geq \tau, \quad \forall k \in T, \tag{3.15}$$

which implies that

$$\|g_k\| \geq \tau_1, \quad \forall k \in T \tag{3.16}$$

holds for sufficiently large $k \in T$ with some positive constant τ_1.

From (3.12), (3.10), (3.11), (3.16), (2.7) and Lemma 3.1, we deduce that

$$\alpha_k \geq \frac{m_1}{M_2^2} - 2m_2 \frac{\epsilon_k f(x_k)}{\|g_k\|} \geq \frac{m_1}{M_2^2} - \frac{2m_2}{\tau_1} \epsilon_k f(x_k) \geq \frac{m_1}{2M_2^2}, \quad k \in T \tag{3.17}$$

holds for sufficiently large k, which together with (3.9) yields that

$$\lim_{k \in T, k \to \infty} \|g_k\| = 0.$$

This contradicts to (3.16). The proof is then completed. □

Remark 3.1 Theorem 3.1 shows that Algorithm 2.1 is strongly convergent since general conjugate gradient methods for nonlinear optimization [12] or nonlinear equations [7] only have the following weaker global convergence in the sense that

$$\liminf_{k \to \infty} \|\nabla f(x_k)\| = 0.$$

By Theorem 3.1, we have the following result.

Corollary 3.1. *Let Assumption 3.1 hold and the sequence $\{x_k\}$ be generated by Algorithm 2.1. Then every limit point is a stationary point of f. Moreover if $J(x^*)$ is nonsingular at a limit point x^* of the sequence $\{x_k\}$, then x^* is a solution of (1.1).*

Now we begin to discuss the local convergence property of Algorithm 2.1, which requires the following assumptions.

Assumption 3.2

(i) The sequence $\{x_k\}$ converges to x^* and $J(x^*)$ is nonsingular .

(ii) In some neighborhood N_1 of x^*, the Jacobian is Lipschitz continuous, namely, there exists a constant $L_3 > 0$ such that

$$\|J(x) - J(y)\| \leq L_3 \|x - y\|, \quad \forall x, y \in N_1. \tag{3.18}$$

The condition (i) in Assumption 3.2 together with Theorem 3.1 implies that

$$F(x^*) = 0, \quad \nabla^2 f(x^*) = J(x^*)^T J(x^*) \tag{3.19}$$

and the Jacobian is uniformly nonsingular in N_1, that is, there exists a positive constant such that

$$\|J(x)d\| \geq m_3 \|d\|, \quad \forall d \in R^n, \quad x \in N_1. \tag{3.20}$$

This inequality together with (2.8) implies that

$$\|g_k\| = \left\| \int_0^1 J(x_k + t\epsilon_k F_k) dt F_k \right\| \geq m_3 \|F_k\|. \tag{3.21}$$

Theorem 3.2. *Let Assumption 3.2 hold. Then the sequence $\{x_k\}$ be generated by Algorithm 2.1 is R-linearly convergent in the sense that*

$$\|F_k\| \leq \rho^k \|F_0\|, \quad \|x_k - x^*\| \leq m_4 \rho^k$$

hold for constants $\rho \in (0, 1)$ and $m_4 > 0$.

Proof. By (3.12) and (2.14), we have

$$\begin{aligned} f(x_{k+1})-f(x_k) &\le -\sigma m_1\frac{\|g_k\|^4}{\|d_k\|^2}+2m_2\sigma\frac{|g_k^Td_k|}{\|d_k\|}\epsilon_k f(x_k)+\epsilon_k f(x_k)\\ &\le -\frac{\sigma m_1}{M_2^2}\|g_k\|^2+2m_2\sigma\|g_k\|\epsilon_k f(x_k)+\epsilon_k f(x_k)\\ &\le -\frac{\sigma m_1 m_3^2}{M_2^2}\|F_k\|^2+2m_2\sigma\|g_k\|\epsilon_k f(x_k)+\epsilon_k f(x_k)\\ &= \Big(-\frac{2\sigma m_1 m_3^2}{M_2^2}+2m_2\sigma\|g_k\|\epsilon_k+\epsilon_k\Big)f(x_k), \end{aligned}$$

where we use (3.10) and (3.21) in the second inequality and the third inequality, respectively. This inequality together with (2.7) and (3.6) implies that there exists a constant $\rho\in(0,1)$ such that

$$f(x_{k+1})\le\Big(1-\frac{2\sigma m_1 m_3^2}{M_2^2}+2m_2\sigma\|g_k\|\epsilon_k+\epsilon_k\Big)f(x_k)\le\rho^2 f(x_k), \tag{3.22}$$

which means that

$$\|F_{k+1}\|\le\rho\|F_k\|\le\rho^{k+1}\|F_0\|. \tag{3.23}$$

From (3.20), we know that

$$\|F_{k+1}-F(x^*)\|=\Big\|\int_0^1 J\big(x^*+t(x_{k+1}-x^*)\big)dt(x_{k+1}-x^*)\Big\|\ge m_3\|x_{k+1}-x^*\|,$$

which together with (3.23) shows that

$$\|x_k-x^*\|\le m_4\rho^k$$

holds for some positive constant m_4. This finishes the proof. □

4 Extension to norm descent case

In this section, based on the idea of [3, 7], we extend Algorithm 2.1 to norm descent case in the sense that $f(x_{k+1})<f(x_k)$ for all $k\ge0$.

Define

$$g_k(\alpha)\triangleq\frac{F(x_k+\alpha F_k)-F_k}{\alpha}. \tag{4.1}$$

Then by the symmetry of the Jacobian, we have

$$\lim_{\alpha\to0^+}g_k(\alpha)=\nabla f(x_k)=J_k^TF_k,$$

which implies that if α is small, then $g_k(\alpha)$ is a good approximation to $\nabla f(x_k)$.

Now we replace g_k in (2.10) with $g_k(\alpha)$ defined by (4.1), then we obtain a new HS type method as follows

$$d_k(\alpha)=\begin{cases}-g_k(\alpha), & \text{if}\quad k=0,\\ -g_k(\alpha)+\frac{g_k(\alpha)^T\gamma_{k-1}}{d_{k-1}^T\gamma_{k-1}}d_{k-1}-\frac{g_k(\alpha)^Td_{k-1}}{d_{k-1}^T\gamma_{k-1}}\gamma_{k-1}, & \text{if}\quad k\ge1,\end{cases} \tag{4.2}$$

where γ_{k-1} is defined by (2.12). It is easy to get that

$$d_k(\alpha)^T g_k(\alpha) = -\|g_k(\alpha)\|^2. \tag{4.3}$$

Moreover, we have

$$\lim_{\alpha\to 0^+} d_k(\alpha) = \begin{cases} -\nabla f(x_k), & \text{if} \quad k=0, \\ -\nabla f(x_k) + \frac{\nabla f(x_k)^T \gamma_{k-1}}{d_{k-1}^T \gamma_{k-1}} d_{k-1} - \frac{\nabla f(x_k)^T d_{k-1}}{d_{k-1}^T \gamma_{k-1}} \gamma_{k-1}, & \text{if} \quad k \geq 1, \end{cases}$$

which implies that

$$\nabla f(x_k)^T d_k(0) = -\|\nabla f(x_k)\|^2. \tag{4.4}$$

This shows that when α is sufficiently small, $d_k(\alpha)$ will be a descent direction of the function $f(x)$ in (2.3). We compute α satisfying

$$f(x_k + \alpha d_k(\alpha)) - f(x_k) \leq \sigma \alpha g_k(\alpha)^T d_k(\alpha), \tag{4.5}$$

where $\sigma \in (0,1)$ is a constant. The line search (4.5) is different from those of [3, 7]. The relations (4.4) and (4.3) show that (4.5) holds for all sufficiently small $\alpha > 0$.

We give the following procedures which are similar to those of [3] to determine the search direction d_k and the stepsize α_k simultaneously.

Procedure 1.

Let $g_k(\alpha)$ be given by (4.1) and $d_k(\alpha)$ be defined by (4.2). Given $r \in (0,1)$. Let i_k be the smallest nonnegative integer such that (4.5) holds with $\alpha = r^i, i = 0, 1, \cdots$. Let $g_k = g_k(r^{i_k})$ and $d_k = d_k(r^{i_k})$.

Procedure 2.

Let d_k and i_k be generated by Procedure 1. If $i_k = 0$, we choose $\alpha_k = 1$. Otherwise, we let j_k be the largest integer $j_k \in \{0, 1, \cdots, i_k - 1\}$ such that

$$f(x_k + r^{i_k - j_k} d_k) - f(x_k) \leq \sigma r^{i_k - j_k} g_k^T d_k. \tag{4.6}$$

Then we take $\alpha_k = r^{i_k - j_k}$.

The following is a derivative-free HS type method with norm descent.

Algorithm 4.1

Step 1. Choose $x_0 \in R^n$ and two constants $\sigma, r \in (0,1)$. Let $k := 0$.

Step 2. Compute d_k and α_k by Procedures 1 and 2.

Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$. Let $k := k+1$ and go to Step 2.

Using the same argument as that of Section 3, we can prove the global and linear convergence of Algorithm 4.1. Here we list this result but omit its proof.

Theorem 4.1. *Let Assumption 3.1 hold and the sequence* $\{x_k\}$ *be generated by Algorithm 4.1. Then we have* $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$. *Moreover, if Assumption 3.2 holds, then the sequence* $\{x_k\}$ *converges R-linearly.*

5 Conclusions

We have proposed a derivative-free HS type method for symmetric nonlinear equations with strong global and R-linear convergence. We believe that the idea of the paper can be extended to other methods such as [10, 13].

References

[1] W. La Cruz, J. M. Martinez and M. Raydan, Spectral residual method without gradient information for solving large-scale nonlinear systems of equations, Math. Comput., 75 (2006), pp. 1429-1448.

[2] J. E. Dennis, R. B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[3] G. Gu, D. Li, L. Qi and S. Zhou, Descent directions of quasi-Newton method for symmetric nonlinear equations, SIAM J. Numer. Anal., 40 (2002), pp. 1763-1774.

[4] C. T. Kelley, Iterative Methods for Linear and Nonlinear Equations, SIAM, Philadelphia, 1995.

[5] D. Li and M. Fukushima, A globally and superlinearly convergent Gauss-Newton-based BFGS method for symmetric nonlinear equations, SIAM J. Numer. Anal., 37 (1999), pp. 152-172.

[6] D. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, J. Comput. Appl. Math., 129 (2001), pp. 15-35.

[7] D. Li and X. Wang, A modified Fletcher-Reeves-type derivative-free method for symmetric nonlinear equations, Numer. Alge. Cont. Optim., 1 (2011), pp. 71-82.

[8] Q. N. Li and D. Li, A class of derivative-free methods for large-scale nonlinear monotone equations, IMA J. Numer. Anal., doi: 10.1093/imanum/drq015, to appear.

[9] J. M. Martinez, Practical quasi-Newton methods for solving nonlinear systems, J. Comput. Appl. Math., 124 (2000), pp. 97-121.

[10] Y. Narushima, H. Yabe and J. A. Ford, A three-term conjugate gradient method with sucient descent property for unconstrained optimization, SIAM J. Optim., 21 (2011), pp. 212-230.

[11] J. M. Ortega and W. C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, 1970.

[12] W. Sun and Y. Yuan, Optimization Theory and Methods, Springer Science and Business Media, LLC, New York, 2006.

[13] L. Zhang, New versions of the Hestenes-Stiefel nonlinear conjugate gradient method based on the secant condition for optimization, Comput. Appl. Math., 28 (2009), pp. 111-133.

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

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

[16] W. Zhou, A Gauss-Newton-based BFGS method for symmetric nonlinear least squares problems, Pac. J. Optim., to appear.

[17] W. Zhou and X. Chen, Global convergence of a new hybrid Gauss-Newton structured BFGS methods for nonlinear least squares problems, SIAM J. Optim., 20 (2010), pp. 2422-2441.

[18] W. Zhou and D. Li, A globally convergent BFGS method for nonlinear monotone equations without any merit functions, Math. Comput., 77 (2008), pp. 2231-2240.