CONVERGENCE PROPERTIES OF THE BFGS ALGORITHM *

Yu-Hong Dai†

Abstract

The BFGS method is one of the most famous quasi-Newton algorithms for unconstrained optimization. In 1984, Powell presented an example of a function of two variables that shows that the PRP conjugate gradient method and the BFGS quasi-Newton method may cycle round eight nonstationary points if each line search picks a local minimum that provides a reduction in the objective function. In this paper, a new technique of choosing parameters is introduced and an example with only six cyclic points is provided. It is also noted by the examples that, the BFGS method with Wolfe line searches need not converge for nonconvex objective functions.

Keywords. unconstrained optimization, conjugate gradient method, quasi-Newton method, Wolfe line search, nonconvex, global convergence.

1. The BFGS algorithm

The BFGS algorithm is one of the most efficient quasi-Newton methods for unconstrained optimization,

$$\min f(x), \quad x \in \mathcal{R}^n. \tag{1.1}$$

The algorithm was proposed by Broyden [2], Fletcher [5], Goldfarb [7] and Shanno [18] individually and can be stated as follows.

Algorithm 1.1 *(The Broyden-Fletcher-Goldfarb-Shanno algorithm)*

Step 0 *Given* $x_1 \in \mathcal{R}^n$; $B_1 \in \mathcal{R}^{n\times n}$ *positive definite;*
Compute $g_1 = \nabla f(x_1)$. *If* $g_1 = 0$, *stop; otherwise, set* $k := 1$.

Step 1 *Set* $d_k = -B_k^{-1} g_k$.

Step 2 *Carry out a line search along* d_k, *getting* $\alpha_k > 0$,
$x_{k+1} = x_k + \alpha_k d_k$, *and* $g_{k+1} = \nabla f(x_{k+1})$;
If $g_{k+1} = 0$, *stop.*

*This work was supported by the Chinese NSF grants 19801033 and 10171104 and a Youth Innovation Fund of Chinese Academy of Science.

†State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, P. O. Box 2719, Beijing 100080, P. R. China. E-mail: *dyh@lsec.cc.ac.cn*

Step 3 Set

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}, \tag{1.2}$$

where

$$\begin{aligned} s_k &= \alpha_k d_k, & (1.3) \\ y_k &= g_{k+1} - g_k. & (1.4) \end{aligned}$$

Step 4 $k := k+1$; *go to Step 1.*

The line search in Step 2 requires the steplength α_k to meet certain conditions. If exact line search is used, α_k satisfies

$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k). \tag{1.5}$$

In the implementations of the BFGS algorithm, one normally requires that the steplength α_k satisfies the Wolfe conditions ([19]):

$$\begin{aligned} f(x_k + \alpha_k d_k) - f(x_k) &\leq \delta_1 \alpha_k d_k^T g_k, & (1.6) \\ d_k^T \nabla f(x_k + \alpha_k d_k) &\geq \delta_2 d_k^T g_k, & (1.7) \end{aligned}$$

where $\delta_1 \leq \delta_2$ are constants in $(0, 1)$. For convenience, we call the line search that satisfies the Wolfe conditions (1.6)-(1.7) as Wolfe line search.

Another famous quasi-Newton method is the DFP method, which was discovered by Davidon [3] and modified by Fletcher and Powell [6]. Broyden [2] proposed a family of quasi-Newton methods:

$$B_{k+1}(\theta) = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \theta (s_k^T B_k s_k) v_k v_k^T, \tag{1.8}$$

where $\theta \in \mathcal{R}^1$ is a scalar and $v_k = \frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k}$. The choice $\theta = 0$ gives rise to the BFGS update, whereas $\theta = 1$ defines the DFP method.

For uniformly convex functions, Powell [12] showed that the DFP algorithm with exact line searches stops at the unique minimum or generate a sequence that converges to the minimum. Dixon [4] found that all methods in the Broyden family with exact line searches produce the same iterations for general functions. For inexact line searches, Powell [14] first proved the global convergence of the BFGS algorithm with Wolfe line searches for convex functions. His result was extended by Byrd, Nocedal and Yuan [1] to all methods in the restricted Broyden family with $\theta \in [0, 1)$. However, the following questions remain open for many years (for example, see Nocedal [9] and Yuan [20]): (i) *does the DFP method with Wolfe line searches converge for convex functions?* and (ii) *does the BFGS method with Wolfe line searches converge for nonconvex functions?*

In this paper, we will consider the $n = 2, m = 8$ example in [15] for the PRP conjugate gradient method [10, 11]. The two-dimensional example shows that the PRP method may cycle round eight nonstationary points if each line search picks a local minimum that provides a reduction in the objective function. By introducing a new technique of choosing parameters, we will present a new example for the PRP method (see §2). The example has only six cyclic points.

Since in the case that $g_{k+1}^T d_k = 0$ for all k, the BFGS method can produce the same iterations as the PRP method does for two-dimensional functions, it can be shown by the examples that the BFGS method with Wolfe line searches need not converge for nonconvex objective functions (see §3). Thus a negative answer is given to question (ii). Some discussions are made in the last section.

2. A counter-example with six cyclic points

The PRP method uses the negative gradient as its initial search direction. For $k \geq 1$, the method defines d_{k+1} as follows:

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T y_k}{\|g_k\|_2^2} d_k \,. \tag{2.1}$$

Powell [15] constructed a two-dimensional example, showing that the PRP method with the line search (2.2) may cycle round eight nonstationary points.

$$\alpha_k \textit{ is a local minimum of } \Phi_k(\alpha) \textit{ and such that } \Phi_k(\alpha_k) < \Phi_k(0), \tag{2.2}$$

where $\Phi_k(\alpha)$ is the line search function

$$\Phi_k(\alpha) = f(x_k + \alpha d_k), \quad \text{where } \alpha > 0. \tag{2.3}$$

However, examples with fewer cyclic points seem not possible from the practice in [15]. In this section, we will introduce a new technique of choosing parameters and provide an example of only six cyclic points.

Assume that $n = 2$. Similar to [15], our example will be so constructed that all the iterations generated by the PRP method converge to the horizontal axis in $\mathcal{R}^2$. For m even, we consider the steps $\{s_k\}$ in the form

$$s_{mj+i} = a_i \begin{pmatrix} 1 \\ b_i \phi^{2j} \end{pmatrix}, \quad s_{mj+\frac{m}{2}+i} = a_i \begin{pmatrix} -1 \\ b_i \phi^{2j+1} \end{pmatrix}, \quad i = 1, \ldots, \frac{m}{2}, \tag{2.4}$$

where ϕ, $\{a_i\}$, $\{b_i\}$ are parameters to be determined, satisfying $\phi \in (0,1)$ and $a_i > 0\,(i = 1, \ldots, \frac{m}{2})$. To be such that

$$g_{k+1}^T d_k = 0, \quad \text{for all } k, \tag{2.5}$$

we assume that the gradients $\{g_k\}$ have the form

$$\begin{cases} g_{mj+1} = c_1 \begin{pmatrix} b_{\frac{m}{2}} \phi^{2j-1} \\ 1 \end{pmatrix}; & g_{mj+i} = c_i \begin{pmatrix} -b_{i-1} \phi^{2j} \\ 1 \end{pmatrix}, i = 2, \ldots, \frac{m}{2}, \\ g_{mj+\frac{m}{2}+1} = c_1 \begin{pmatrix} -b_{\frac{m}{2}} \phi^{2j} \\ 1 \end{pmatrix}; & g_{mj+\frac{m}{2}+i} = c_i \begin{pmatrix} b_{i-1} \phi^{2j+1} \\ 1 \end{pmatrix}, i = 2, \ldots, \frac{m}{2}, \end{cases} \tag{2.6}$$

where $\{c_i\}$ are also parameters to be determined. In this section, we are interesting in the case that $m = 6$.

By relations (2.1) and (2.5), we know that the PRP method satisfies the conjugacy condition

$$d_{k+1}^T y_k = 0 \tag{2.7}$$

and the descent condition

$$d_{k+1}^T g_{k+1} < 0. \tag{2.8}$$

The above conditions require that $g_{6j+i}^T s_{6j+i} = g_{6j+i-1}^T s_{6j+i} < 0$, yielding

$$\begin{cases} c_2(b_2 - b_1) = c_1(b_2 + b_3\phi^{-1}) < 0, \\ c_3(b_3 - b_2) = c_2(b_3 - b_1) < 0, \\ c_1(b_1\phi + b_3) = c_3(b_1\phi + b_2) < 0. \end{cases} \tag{2.9}$$

Denoting $b_0 = -b_3\phi^{-1}$ and $b_4 = -b_1\phi$, we can draw the following conditions on $\{b_i\}$ from (2.9):

$$\begin{cases} (b_2 - b_1)(b_3 - b_2)(b_4 - b_3) = (b_2 - b_0)(b_3 - b_1)(b_4 - b_2), \\ (b_3 - b_4)(b_2 - b_0) > 0, (b_2 - b_1)(b_3 - b_1) > 0, (b_3 - b_2)(b_2 - b_4) > 0. \end{cases} \tag{2.10}$$

Defining $\varphi_i = b_i - b_{i-1}$, the above relations are equivalent to

$$\begin{cases} \varphi_2\varphi_3\varphi_4 = (\varphi_1 + \varphi_2)(\varphi_2 + \varphi_3)(\varphi_3 + \varphi_4), \\ \varphi_4(\varphi_1 + \varphi_2) < 0, \varphi_2(\varphi_2 + \varphi_3) > 0, \varphi_3(\varphi_3 + \varphi_4) < 0. \end{cases} \tag{2.11}$$

Further, letting $t_i = \varphi_{i+1}/\varphi_i$ and noting that $\varphi_4/\varphi_1 = -\phi$, we can obtain

$$\begin{cases} t_1t_2t_3 = (1 + t_1)(1 + t_2)(1 + t_3) = -\phi, \\ t_1 > -1, t_2 > -1, t_3 < -1. \end{cases} \tag{2.12}$$

The first line in (2.12) is equivalent to

$$-t_1t_2t_3 = \frac{t_1t_2(1 + t_1)(1 + t_2)}{1 + t_1 + t_2} = \phi. \tag{2.13}$$

Thus for any $\phi \in (0, 1)$ and $t_3 < -1$, we may solve t_1 and t_2 from (2.13). If the solved t_1 and t_2 are such that $t_1 > -1$ and $t_2 > -1$, then we can further consider the choices of $\{a_i\}$. In our real construction, we pick $t_3 = -2$. This with (2.13) indicates that

$$t_1t_2 = 1 + t_1 + t_2. \tag{2.14}$$

Further, we find that the following values of $\{t_i\}$ and ϕ satisfy (2.13) and allow suitable $\{a_i; i = 1, 2, 3\}$:

$$t_1 = -\frac{3}{4}, \quad t_2 = -\frac{1}{7}, \quad t_3 = -2, \quad \phi = \frac{3}{14}. \tag{2.15}$$

Now, by the definitions of φ_i and t_i, we can express $\sum_{i=2}^4 \varphi_i$ in two ways:

$$\begin{aligned} \sum_{i=2}^4 \varphi_i &\overset{(1)}{=} b_4 - b_1 = -b_1(1 + \phi) \\ &\overset{(2)}{=} \varphi_2(1 + t_2 + t_2t_3) = (b_2 - b_1)(1 + t_2 + t_2t_3). \end{aligned} \tag{2.16}$$

We then get that

$$b_2 = \left[1 - \frac{1 + \phi}{1 + t_2 + t_2t_3}\right] b_1. \tag{2.17}$$

Further, we have

$$b_3 = b_2 + \varphi_3 = b_2 + t_2\varphi_2 = (1 + t_2)b_2 - t_2b_1. \tag{2.18}$$

Thus, letting $b_1 = 1$, we have from this, (2.17) and (2.18) that

$$b_1 = 1, \quad b_2 = -\frac{1}{16}, \quad b_3 = \frac{5}{56}. \tag{2.19}$$

Letting $c_2 = 1$, we obtain from (2.9) that

$$c_1 = -3, \quad c_2 = 1, \quad c_3 = -6. \tag{2.20}$$

As will be shown, the parameters chosen above allow the function value to be monotonically decreased. Denote f^* to be the limit of $f(x_k)$. Since all the iterations are required to converge to the horizontal axis, and for each value of the first variable, the dependence of $f(x)$ on the second variable is linear, we have that

$$f(x_k) - f^* = (x_k)_2(g_k)_2, \quad \text{for all } k \geq 1, \tag{2.21}$$

where $(v)_i$ means the i-th component of vector v. Given the limit $\hat{x}_1 = \lim_{j\to\infty} x_{6j+1}$, we can compute $\{x_{6j+i}; i = 1, \ldots, 4\}$ in the following way:

$$\begin{cases} x_{6j+1} = \hat{x}_1 - \sum_{k=j}^{\infty} \sum_{i=1}^{6} s_{6k+i}; \\ x_{6j+i} = x_{6j+i-1} + s_{6j+i-1}, \; i = 2, 3, 4. \end{cases} \tag{2.22}$$

As a result, the second components of $\{x_{6j+i}; i = 1, \ldots, 4\}$ can be expressed as follows:

$$\left(x_{6j+i}\right)_2 = -h_i(1-\phi)^{-1}\phi^{2j}, \quad i = 1, \ldots, 4, \tag{2.23}$$

where

$$\begin{cases} h_1 = a_1 b_1 + a_2 b_2 + a_3 b_3, \\ h_2 = a_1 b_1 \phi + a_2 b_2 + a_3 b_3, \\ h_3 = a_1 b_1 \phi + a_2 b_2 \phi + a_3 b_3, \\ h_4 = h_1 \phi. \end{cases} \tag{2.24}$$

Using the relations (2.21) and (2.23) and noting that the structure of this example has some symmetry, we know that the monotonicity of $f(x_k)$ requires $\{a_i\}$ to meet

$$-c_1 h_1 > -c_2 h_2 > -c_3 h_3 > -c_1 h_4. \tag{2.25}$$

This relation can be satisfied if we choose

$$a_1 = 14, \quad a_2 = 160, \quad a_3 = 1. \tag{2.26}$$

In this case, the four terms in (2.25) have the values

$$\frac{687}{56}, \quad \frac{387}{56}, \quad \frac{159}{28}, \quad \text{and} \quad \frac{2061}{784},$$

respectively. So (2.25) is satisfied. Further, if we let $(x_1)_1 = -87.5$, then $\{(x_{6j+i})_1; i = 1, \cdots, 6\}$ have the values of -87.5, -73.5, 86.5, 87.5, 73.5, and -86.5, which are all different.

Finally, we discuss how to construct a smooth function $f(x) \in \mathcal{R}^2$ that satisfies the gradient conditions (2.6). At first, for given real numbers p_1, $p_2 (\neq 0)$, p_3, p_4 and any $j \geq 1$, we see that the following function

$$\Psi(u_1, u_2) = \left[p_4 + p_2^{-1} p_3 (u_1 - p_1)\right] u_2 \tag{2.27}$$

is such that

$$\nabla\Psi\begin{pmatrix} p_1 \\ p_2\phi^j \end{pmatrix} = \begin{pmatrix} p_3\phi^j \\ p_4 \end{pmatrix}. \tag{2.28}$$

Note that $\{x_{6j+i}; i=1,\ldots,6\}$ are as follows:

$$\begin{pmatrix} -87.5 \\ -\frac{229}{44}\phi^{2j} \end{pmatrix}, \begin{pmatrix} -73.5 \\ \frac{387}{44}\phi^{2j} \end{pmatrix}, \begin{pmatrix} 86.5 \\ -\frac{53}{44}\phi^{2j} \end{pmatrix}, \begin{pmatrix} 87.5 \\ -\frac{229}{44}\phi^{2j+1} \end{pmatrix}, \begin{pmatrix} 73.5 \\ \frac{387}{44}\phi^{2j+1} \end{pmatrix}, \begin{pmatrix} -86.5 \\ -\frac{53}{44}\phi^{2j+1} \end{pmatrix}.$$

Letting $\mathcal{B}_i = \{u_1; |u_1 - (x_{6j+i})_1| \le 0.1\}$, it is easy to find one-dimensional C^∞ functions ξ and γ such that their values at the intervals $\{\mathcal{B}_i; i=1,\ldots,6\}$ are

$$\frac{8251}{458},\ -\frac{2847}{387},\ -\frac{6981}{212},\ \frac{8251}{458},\ -\frac{2847}{387},\ -\frac{6981}{212}$$

and

$$\frac{55}{229},\ -\frac{44}{387},\ \frac{33}{106},\ -\frac{55}{229},\ \frac{44}{387},\ -\frac{33}{106},$$

respectively. Then we can test that the function

$$f(u_1,u_2) = [\xi(u_1) + \gamma(u_1)u_1]u_2 \tag{2.29}$$

is a C^∞ function in $\mathcal{R}^2$ and satisfies the gradient conditions (2.6). One deficiency of the function (2.29) is that, the point x_{6j+i+1} may not be a local minimum of $\Phi_{6j+i}(\alpha)$ (see (2.3) for the definition of Φ). For example, x_{6j+2}. For this, we can further choose a one-dimensional C^∞ function τ such that for $i=1,\ldots,6$, its value at $\mathcal{B}_i$ is equal to $(x_{6j+i})_1$. Then the C^∞ function

$$f(u_1,u_2) = [\xi(u_1) + \gamma(u_1)u_1 + M(u_1 - \tau(u_1))^2]u_2 \tag{2.30}$$

with $M > 0$ sufficiently large can guarantee each x_{6j+i+1} is a local minimum of $\Phi_{6j+i}(\alpha)$. This completes the construction of our new example.

Thus by introducing the quantities φ_i and t_i, we have obtained a new example. The example shows that the PRP method with the line search (2.2) may cycle round six nonstationary point. One advantage of this example over the one in [15] is that, it has only six cyclic points, whereas the latter has eight.

It is easy to see that the above example applies to the BFGS method if the choice of B_1 is such that $B_1 s_1 = -l g_1$, where l is any positive number. If one changes the definition of f in a small neighborhood of x_1 to meet the necessary initial conditions, the example is also efficient for the BFGS method with any positive definite matrix B_1 or the PRP method with $d_1 = -g_1$.

3. Nonconvergence of the BFGS algorithm for nonconvex functions

Generally, the line search (2.2) needs not satisfy the Wolfe conditions (1.6)–(1.7). For example, consider the function

$$f(x) = \cos x, \quad x \in \mathcal{R}^1. \tag{3.1}$$

Assume that $x_k = 0$ and $d_k = 1$. For any nonnegative integer i, $\alpha = (2i+1)\pi$ is a local minimum of $\Phi_k(\alpha)$. Then (1.6) is false if i is large. For the line search in the example of §2, however, we can directly test that the Wolfe conditions (1.6)-(1.7) hold (see Theorem 3.1). Thus the example in §2 also shows that the BFGS algorithm with Wolfe line searches need not converge for nonconvex objective functions.

Theorem 3.1 *Consider the BFGS algorithm with the Wolfe line search (1.6)-(1.7), where* $\delta_1 \le \frac{69}{7480}$ *and* $\delta_2 \in (\delta_1, 1)$. *Then for any* $n \ge 2$, *there exists a starting point* x_1 *and a* C^∞ *function* f *in* $\mathcal{R}^n$ *such that, the sequence* $\{\|g_k\|_2 : k = 1, 2, \ldots\}$ *generated by the algorithm is bounded away from zero.*

Proof Consider the example in §2. For any starting matrix B_1, we may slightly modify the example such that it satisfies the necessary initial conditions. By (2.21), (2.23) and (2.6), we see that

$$f(x_{6j+i}) = f^* - c_i h_i (1-\phi)^{-1} \phi^{2j}, \quad i = 1, \cdots, 4. \tag{3.2}$$

Still denote $b_0 = -b_3\phi^{-1}$, $b_4 = -b_1\phi$ and let $a_4 = a_1$, $c_4 = c_1$. We have by (2.4) and (2.6) that

$$g_{6j+i}^T s_{6j+i} = a_i c_i (b_i - b_{i-1}) \phi^{2j}, \quad i = 1, \cdots, 4. \tag{3.3}$$

Combining (3.2) and (3.3) and noting the symmetry of the example, we know that the first Wolfe condition (1.6) holds with any constant δ_1 satisfying

$$\begin{aligned}
\delta_1 &\le \min\left\{ \frac{f(x_{6j+i+1}) - f(x_{6j+i})}{g_{6j+i}^T s_{6j+i}} : i = 1, 2, 3 \right\} \\
&= \frac{1}{1-\phi} \min\left\{ \frac{c_i h_i - c_{i+1} h_{i+1}}{a_i c_i (b_i - b_{i-1})} : i = 1, 2, 3 \right\} \\
&= \frac{69}{7480}.
\end{aligned} \tag{3.4}$$

In addition, relations (2.5) and (2.8) imply that the second Wolfe condition (1.7) holds for $\delta_2 \in (\delta_1, 1)$. Thus the example in §2 shows that the BFGS algorithm with Wolfe line searches needs not converge for two-dimensional functions.

In the case when $n \ge 3$, we only need to consider the function

$$\hat{f}(x) = \hat{f}(u_1, u_2, \ldots, u_n) = f(u_1, u_2), \tag{3.5}$$

where f is the function in the example of §2. This completes our proof. $\square$

The parameter δ_1 in the above theorem is required to be not greater than $\frac{69}{7480} \approx 0.0092$. If we consider Powell's example with eight cyclic points, then Theorem 3.1 can be extended to $\delta_1 \le \frac{1}{84} \approx 0.0119$.

4. Some discussions

In this paper, it has been shown by one of Powell's examples in [15] and a new example with six cyclic points that, the BFGS algorithm with Wolfe line searches needs not converge for nonconvex objective functions. This result also applies to the HS conjugate gradient method [8], the Broyden positive family (1.8) with $\theta \ge 0$, and the limited-memory quasi-Newton methods, since all these methods satisfy both the conjugacy condition (2.7) and the descent condition (2.8) if $g_{k+1}^T d_k = 0$ for all k.

To my knowledge, the parameters δ_1 and δ_2 in (1.6)-(1.7) are often set to 0.01 (or a smaller value) and 0.9 respectively in the implementations of the BFGS algorithm. According to the remark after it, Theorem 3.1 can be extended to the case that $\delta_1 \leq \frac{1}{84}$. Since $\frac{1}{84} > 0.01$, one would be satisfied with this result for the BFGS algorithm. As Professor J. C. Gilbert discussed with me, however, we wonder whether Theorem 3.1 holds for any $\delta_1 < 1$ in theory.

Using the same technique as in §2, we can show that there do not exist examples of four cyclic points having similar structures. This means that, the number of cyclic points, six, can not be decreased if we assume m to be even. In fact, if $m = 4$, we have by (2.4) and (2.6) that

$$\begin{cases} c_2(b_2 - b_1) = c_1(b_2 + b_2\phi^{-1}) < 0, \\ c_1(b_2 + b_1\phi) = c_2(b_1 + b_1\phi) < 0, \end{cases} \tag{4.1}$$

where $\phi \in (0,1)$. Denote $b_0 = -b_2\phi^{-1}$, $b_3 = -b_1\phi$, $\varphi_i = b_i - b_{i-1} (i = 1,2,3)$ and $t_i = \varphi_{i+1}/\varphi_i (i = 1,2)$. Similar to (2.10), (2.11), (2.12), we can obtain

$$\begin{cases} t_1 t_2 = (1 + t_1)(1 + t_2) = -\phi, \\ t_1 > -1, \ t_2 < -1. \end{cases} \tag{4.2}$$

The above imply that $t_2 = -(1 + t_1)$ and $\phi = t_1(1 + t_1)$. Since $\phi \in (0,1)$, we can then get that $t_1 > 0$. Further, letting $b_1 = 1$, we can similar to (2.16) obtain that $b_2 = (1 + t_1)^2/t_1$. Since b_1, b_2 and ϕ are all positive, we know by $c_1(b_2 + b_1\phi) < 0$ that $c_1 < 0$. Letting $c_1 = -t_1$, we can get by (4.1) that $c_2 = -(1 + t_1)$. Similar to (2.21)–(2.25), it is easy to see that the condition $f(x_{4j+1}) > f(x_{4j+2})$ requires

$$-c_1(a_1 b_1 + a_2 b_2) > -c_2(a_1 b_1 \phi + a_2 b_2). \tag{4.3}$$

Substituting the expressions of ϕ, c_1 and c_2 with t_1, (4.3) is equivalent to

$$-(2 + t_1)t_1^2 a_1 b_1 - a_2 b_2 > 0. \tag{4.4}$$

This is not possible since t_1, a_1, a_2, b_1 and b_2 are all positive. The contradiction shows the nonexistence of examples of four cyclic points.

Under the assumption that $x_k \to \bar{x}$, Powell [13] showed that the BFGS algorithm with exact line searches converges globally for general functions when there are only two variables. This result was extended by Pu and Yu [17] to the case that $n \geq 2$. Therefore an interesting question may be, if $x_k \to \bar{x}$, is the BFGS algorithm with Wolfe line searches globally convergent for general functions? Another question is, does there exist an inexact line search that ensures the global convergence of the BFGS method for general functions?

Recently, Powell [16] showed that, if the line search always finds the first local minimum of $\Phi_k(\alpha)$ in (2.3), the BFGS method is globally convergent for two-dimensional twice-continuously differentiable functions with bounded level sets. Powell (private communications) and the author are trying to construct a three-dimensional example showing that the BFGS algorithm with the above line search needs not converge.

Acknowledgements. The author is much indebted to Professors Y. Yuan, J. C. Gilbert, and M. J. D. Powell, who once discussed with him about the idea of this paper and gave many valuable suggestions and comments. Thanks are also due to the two anonymous referees, whose comments and suggestions greatly improved this paper.

References

[1] Byrd R H, Nocedal J, Yuan Y. Global convergence of a class of quasi-Newton methods on convex problems. SIAM J. Numer. Anal. 1987, 24: 1171-1190.

[2] Broyden C G. The convergence of a class of double rank minimization algorithms: 2. the new algorithm. J. Inst. Math. Appl. 1970, 6: 222-231.

[3] Davidon W C. Variable metric methods for minimization. Report ANL 5990 (revised), Argonne National Laboratory, Illinois, USA, 1959 (see also: SIAM J. Optim. 1991, 1: 1-17).

[4] Dixon L C W. Variable metric algorithms: Necessary and sufficient conditions for identical behavior of nonquadratic functions. J. Optim. Theory Appl. 1972, 10: 34-40.

[5] Fletcher R. A new approach to variable metric algorithms. Computer J. 1970, 13: 317-322.

[6] Fletcher R, Powell M J D. A rapidly convergent descent method for minimization. Comput. J. 1963, 6: 163-168.

[7] Goldfarb D. A family of variable metric methods derived by variational means. Math. Comp. 1970, 24: 23-26.

[8] Hestenes M R, Stiefel E. Method of conjugate gradient for solving linear system. J. Res. Nat. Bur. Stand. 1952, 49: 409-436.

[9] Nocedal J. Theory of algorithms for unconstrained optimization. Acta Numerica, 1992, 1: 199-242.

[10] Polak E, Ribière G. Note sur la convergence de methodes de directions conjuguées. Rev. française Informat. Recherche Opérationelle, 1969, 16: 35-43.

[11] Polyak B T. Conjugate gradient method in extremal problems. USSR Comp. Math. and Math. Phys. 1969, 9: 94-112.

[12] Powell M J D. On the convergence of the variable metric algorithm. J. Inst. Maths. Appl. 1971, 7: 21-36.

[13] Powell M J D. Quadratic termination properties of minimization algorithm, Part I and Part II. J. Inst. Maths. Appl. 1972, 10: 332-357.

[14] Powell M J D. Some global convergence properties of a variable metric algorithm for minimization without exact line searches, in: Cottle R W, Lemke C E (Eds.), Nonlinear Programming, SIAM-AMS Proceedings vol. IX, SIAM, Philadelphia, PA, 1976, 53-72.

[15] Powell M J D. Nonconvex minimization calculations and the conjugate gradient method. in: Griffiths D F (Ed.), Numerical Analysis, Lecture Notes in Mathematics vol. 1066, Springer-Verlag, Berlin, 1984, 122-141.

[16] Powell M J D. On the convergence of the DFP algorithm for unconstrained optimization when there are only two variables. Math. Program. Ser. B, 2000 (published online).

[17] Pu D, Yu W. On the convergence property of DFP algorithm. Annals of Operations Research, 1990, 24: 175-184.

[18] Shanno D F. Conditioning of quasi-Newton methods for function minimization. Math. Comp. 1970, 24: 647-650.

[19] Wolfe P. Convergence conditions for ascent methods. SIAM Review, 1969, 11: 226-235.

[20] Yuan Y. Numerical methods for nonlinear programming. Shanghai Scientific & Technical Publishers, 1993 (in Chinese).