Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT
ELSEVIER
Journal of Computational and Applied Mathematics 154 (2003) 319–339
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

The revised DFP algorithm without exact line search[☆]

Dingguo Pu[a,*], Weiwen Tian[b]

[a] *Department of Applied Mathematics, Tongji University, 1239 Siping Road, Shanghai, 200092, China*
[b] *Department of Mathematics, Shanghai University, 99 Shangda Road, Shanghai 200436, China*

Received 15 May 2001; received in revised form 12 October 2002

Abstract

In this paper, we discuss the convergence of the DFP algorithm with revised search direction. Under some inexact line searches, we prove that the algorithm is globally convergent for continuously differentiable functions and the rate of convergence of the algorithm is one-step superlinear and n-step second order for uniformly convex objective functions.

From the proof of this paper, we obtain the superlinear and n-step second-order convergence of the DFP algorithm for uniformly convex objective functions.

MSC: 90C30

Keywords: DFP algorithm; Line search; Convergence; Convergence rate

1. Introduction

We know that in order to obtain a superlinearly convergent method, it is necessary to approximate the Newton step asymptotically—this is the principle of Dennis and Moré [7]. How can we do this without actually evaluating the Hessian matrix by any approximate to the Hessian matrix at every iteration? The answer was discovered by Davidon [5] and was subsequently developed and popularized by Fletcher and Powell [10]. It consists of starting with any approximation to the Hessian matrix, and at each iteration, updating this matrix by incorporating the curvature of the problem measured along the step. If this update is done appropriately, one obtains some remarkably robust and efficient algorithms, called Quasi–Newton methods or variable metric algorithms. They revolutionized nonlinear optimization by providing an alternative to Newton's method which is too costly for many applications.

☆ This research is supported by the Ministry of Education of P.R.C.
* Corresponding author.

0377-0427/03/$ - see front matter
doi:10.1016/S0377-0427(02)00856-7

One, maybe the most important, class of variable metric algorithms is Broyden algorithms [3]. With exact line search, Dixon [8] proved that all Broyden algorithms produce the same iterations for general functions. Powell [16] proved that the rate of convergence of these algorithms is one-step Q-superlinear for the uniformly convex object functions, and Pu [24] extend this result for LC^1 objective function. Pu and Yu [28] proved that if the points which are given by these algorithms are convergent, they are globally convergent, for continuously differentiable functions.

Without exact line search several results have been obtained. A global convergence result for the BFGS algorithm is obtained by Powell [17]. He demonstrated that if the objective function f is convex, then the BFGS algorithm gives $\liminf \|\nabla f(x_k)\| = 0$ under given conditions on the line search, and if in addition the sequence $\{x_k\}$ converges to a solution point at which the Hessian matrix is positive definite, then the rate of convergence is Q-superlinear.

This analysis has been extended by Byrd et al. [4] to the restricted Broyden algorithms. They proved the global and Q-superlinear convergence on convex problems for all the restricted Broyden algorithms except for the DFP algorithm, i.e., for $\phi \in (0,1]$ in the Broyden update class (the ϕ in the Broyden update is shown in (6)). Pu [23,25] proved the global convergence of the DFP algorithm for the uniformly convex object function under some modified Wolfe conditions.

Other variable metric algorithms have also been proposed. For example, the Huang's updating formula is characterized by three independent parameters. For the relationship among Huang's updates, Oren's updates [14] and the Broyden algorithms see [32,33].

For the choice of the parameter ϕ in the Broyden update formula, some optimal conditions are suggested in some methods. For example, Davidon [6] proposed a method in which B_{k+1} (B_k and B_{k+1} are denoted in (6)) is chosen to be the member of the Broyden class that minimizes the condition number of $B_k^{-1}B_{k+1}$, subject to preserving positive definiteness. Other work in this area includes [1,13,11], and so on. Besides Zhang and Tewarson [37] performed numerical tests with negative values of ϕ.

One can also attempt to improve variable metric methods by introducing automatic scaling strategies to adjust the size of matrix B_k. An idea proposed by Oren and Luenberger [15] consists of multiplying B_k by a scaling length θ before the update takes place. For example, for BFGS methods, the update would be of the form

$$B_{k+1} = \theta \left[B_k - \frac{B_k s_k s_k^{\mathrm{T}} B_k}{s_k^{\mathrm{T}} B_k s_k} \right] + \frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k}, \tag{1}$$

where g_k is the gradient of $f(x)$ at x_k, $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$.

Another strategy has been proposed by Powell [21], and further developed by Lalee and Nocedal [12] and Siegel [30,31]. Powell's idea is to work with the lengthization

$$H_k = (B_k)^{-1} = z_k z_k^{\mathrm{T}} \tag{2}$$

of the inverse Hessian approximation H_k.

There are many theoretical and computational results on rank-one updating formulas as well as rank-two updating formulas proposed (for example, see [34]).

The Broyden algorithms are also applied to the methods for solving the constrained nonlinear optimization problems, for example, see [18,2,29].

However, there are several unsolved theoretical problems for the Broyden algorithms. We cannot prove the convergence of the Broyden algorithms for nonconvex functions, some computational results show that the points given by the Broyden algorithms may not converge to the optimum if objective functions are not convex. We do not know that, whether or not, the DFP algorithm is convergent if the line search satisfies the Wolfe conditions too (see [9]).

To overcome the shortcoming that the Broyden algorithms may not converge for general functions, Pu and Tian proposed ([22,26]) a class of modified Broyden algorithms in which the updating formula is rank three, and proved the convergence and the one-step superlinear convergence of these algorithms. They advanced above algorithms and proposed a new class of variable metric algorithms in which the Broyden update is used, but the line searches directions are revised properly (see [27]). They call them the Broyden algorithms with revised search direction, or revised Broyden algorithms, and proved that these algorithms are convergent for continuously differentiable objective functions, and superlinear and n-step second-order convergent for the uniformly convex objective functions under exact line search.

In this paper, we discuss the revised DFP algorithm under inexact line search. We prove that the algorithm is convergent for the continuously differentiable objective functions. Also the new algorithm is superlinear and n-step second-order convergent for uniformly convex functions when the line search is inexact, but satisfies some search conditions. We list the convergent and superlinearly convergent results, but do not give the detail proof of superlinear convergence for other revised Broyden algorithms. We also list the n-step second-order convergence results of revised Broyden algorithms without the detail proof.

The revised Broyden algorithms are iterative. Given a starting point x_1 and an initial positive definite matrix B_1, they generate a sequence of points $\{x_k\}$ and a sequence of matrices of $\{B_k\}$ which are given by the following equations ((3) and (6)):

$$x_{k+1} = x_k + s_k = x_k + \alpha_k d_k, \tag{3}$$

where $\alpha_k > 0$ is the step factor and d_k is the search direction satisfying

$$-d_k = H_k g_k + \|Q_k H_k g_k\| R_k g_k, \tag{4}$$

where g_k is the gradient of $f(x)$ at x_k and H_k is the inverse of B_k.

$\{Q_k\}$ and $\{R_k\}$ are two sequences of positive definite or positive semi-definite matrices which are uniformly bounded. All eigenvalues of these matrices are included in $[q,r]$, $0 \leqslant q \leqslant r$, i.e., for all k and $x \in R^n$, $x \neq 0$

$$q\|x\|^2 \leqslant x^{\mathrm{T}} Q_k x \leqslant r\|x\|^2; \quad q\|x\|^2 \leqslant x^{\mathrm{T}} R_k x \leqslant r\|x\|^2. \tag{5}$$

If $g_k = 0$, the algorithms terminate, otherwise let

$$B_{k+1} = B_k - \frac{B_k s_k s_k^{\mathrm{T}} B_k}{s_k^{\mathrm{T}} B_k s_k} + \frac{y_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \phi(s_k^{\mathrm{T}} B_k s_k) v_k v_k^{\mathrm{T}}, \tag{6}$$

where $y_k = g_{k+1} - g_k$, $v_k = y_k(s_k^{\mathrm{T}} y_k)^{-1} - B_k s_k (s_k^{\mathrm{T}} B_k s_k)^{-1}$ and $\phi \in [0,1]$. In the above algorithms if $R_k \equiv 0$, we get the Broyden algorithms, and if $\phi = 0$ we call it revised BFGS algorithm or RBFGS algorithm, and if $\phi = 1$ we call it revised DFP algorithm or RDFP algorithm. In this paper, we discuss the convergence of algorithms for $q > 0$.

The matrix H_{k+1} denotes the inverse of B_{k+1}, the recurrence formula of H_{k+1} is

$$H_{k+1} = H_k - \frac{H_k y_k y_k^{\mathrm{T}} H_k}{y_k^{\mathrm{T}} H_k y_k} + \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \frac{\rho \mu_k \mu_k^{\mathrm{T}}}{y_k^{\mathrm{T}} H_k y_k}, \tag{7}$$

where

$$\mu_k = H_k y_k - \frac{y_k^{\mathrm{T}} H_k y_k}{s_k^{\mathrm{T}} y_k} s_k, \tag{8}$$

where $\rho \in [0,1]$, ρ and ϕ satisfying

$$\phi = \frac{(1-\rho)(s_k^{\mathrm{T}} y_k)^2}{(1-\rho)(s_k^{\mathrm{T}} y_k)^2 + \rho y_k^{\mathrm{T}} H_k y_k s_k^{\mathrm{T}} B_k s_k} \tag{9}$$

or

$$\rho = \frac{(1-\phi)(s_k^{\mathrm{T}} y_k)^2}{(1-\phi)(s_k^{\mathrm{T}} y_k)^2 + \phi y_k^{\mathrm{T}} H_k y_k s_k^{\mathrm{T}} B_k s_k}. \tag{10}$$

We may obtain the quasi-Newton formula

$$H_{k+1} y_k = s_k \tag{11}$$

in the Broyden algorithms.

In this paper, the line search is inexact, and in order to guarantee descentness of the objective function values and the convergence of the algorithms, we must give some conditions for determining α_k. We use the Wolfe conditions or modified Wolfe conditions as follows:

$$f(x_k) - f(x_{k+1}) \geqslant \zeta_k(-g_k^{\mathrm{T}} s_k) \tag{12}$$

and

$$|g_{k+1}^{\mathrm{T}} s_k| \leqslant \theta_k(-g_k^{\mathrm{T}} s_k). \tag{13}$$

Let ζ_0 and θ_0 be two constants satisfying $0 < \zeta_0 \leqslant \theta_0 < 1/2$, we discuss the following cases:

Case 1 (*Wolfe condition*): $\zeta_k = \zeta_0$ and $\theta_k = \theta_0$ are two constants.

Case 2 (*modified Wolfe condition*):

$$\zeta_k = \zeta_0 \min\{1, \alpha_k^{-1}\}, \quad \theta_k = \theta_0 \min\{1, \alpha_k^{-1}\} \tag{14}$$

and $-s_k^{\mathrm{T}} g(x_k + \lambda s_k) > 0$ for all $\lambda \in (0,1]$.

Case 3 (*another modified Wolfe condition*):

$$\zeta_k = \zeta_0 \min\{1, \|g_k\|^{-1}\}, \quad \theta_k = \theta_0 \min\{1, \|g_k\|^{-1}\}. \tag{15}$$

The above three cases will be called the line search condition 1, 2 or 3, respectively. We always try $\alpha_k = 1$ first in choosing the step factor.

From the Broyden algorithms we know that if B_k is a positive definite matrix and line research satisfies one of above cases, then $s_k^{\mathrm{T}} y_k > 0$ and B_{k+1} is positive definite. Using the mathematical induction, it is easy to imply that B_k and H_k are positive definite matrices if H_1 and B_1 are so.

If no ambiguities are arisen, we may drop the subscript of the characters, for example, g, x, R denote g_k, x_k, R_k, and use subscript * to denote the amounts obtained by the next iteration, i.e., g_*, x_*, R_* denote g_{k+1}, x_{k+1}, R_{k+1}, respectively.

For simplicity, we let

$$U_k = \frac{-g_k^{\mathrm{T}} H_k y_k}{y_k^{\mathrm{T}} H_k y_k}, \quad V_k = \frac{y_k^{\mathrm{T}} H_k y_k}{s_k^{\mathrm{T}} y_k}, \quad W_k = \frac{-g_k^{\mathrm{T}} d_k}{y_k^{\mathrm{T}} d_k} = \frac{-g_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k},$$

$$\begin{aligned} Z_k &= H_k g_k + \frac{-g_k^{\mathrm{T}} H_k y_k}{s_k^{\mathrm{T}} y_k} s_k \\ &= \frac{\|Q_k H_k g_k\| y_k^{\mathrm{T}} R_k g_k}{s_k^{\mathrm{T}} y_k} s_k - \|Q_k H_k g_k\| R_k g_k. \end{aligned} \tag{16}$$

The paper is outlined as follows:

Section 2 gives several convergence results without the convexity assumption. Section 3 gives some results for convex objective functions. In Sections 4, we prove that the revised DFP algorithm is linearly convergent. In Section 5 we prove that our algorithm is one-step superlinearly convergent and in Section 6 we show that the algorithm has a quadratical convergence rate under some conditions on line search and give some numerical results.

Throughout this paper the vector norms are Euclidian.

2. Several convergence results without convexity assumption

In this section, we assume:

1. $f(x) \in C^{1,1}$, i.e., there exists an $L > 1$ such that for any $x, y \in R^n$,

$$\|g(x) - g(y)\| \leqslant L\|x - y\|. \tag{17}$$

2. For any $x_1 \in R^n$, the level set $S(x_1) = \{x \mid f(x) \leqslant f(x_1)\}$ is bounded.
3. Let $\bar{x}$ be the minimum point of f, then $f(x)$ and x are replaced by $f(x - \bar{x}) - f(\bar{x})$ and $x - \bar{x}$, respectively. So, we may assume for simplicity

$$f(0) = \min f(x) = 0.$$

We get the following by the properties of R and Q:

$$(1 + r^2\|g\|)\|Hg\| \geqslant \|d\| \geqslant (1 - r^2\|g\|)\|Hg\|. \tag{18}$$

If $2r^2\|g_k\| < 1$ for sufficiently large k, then there exists a constant $c_0 > 0$ such that for all k,

$$\|d\|(1 - c_0\|x\|) \leqslant \|Hg\| \leqslant \|d\|(1 + c_0\|x\|). \tag{19}$$

The following holds for all k:

$$-g^{\mathrm{T}}s = \alpha[g^{\mathrm{T}}Hg + \|QHg\|g^{\mathrm{T}}Rg] \geqslant \frac{q^2\|g\|^2\|s\|}{1+r^2\|g\|}. \tag{20}$$

Assumption 1 and (13) imply

$$-(1-\theta_0)g^{\mathrm{T}}s \leqslant s^{\mathrm{T}}y \leqslant L\|s\|^2. \tag{21}$$

From (20) and (21) we obtain

$$\|s\| \geqslant \frac{-(1-\theta_0)g^{\mathrm{T}}s}{L\|s\|} \geqslant \frac{q^2(1-\theta_0)\|g\|^2}{L(1+r^2\|g\|)} \tag{22}$$

and

$$(-g^{\mathrm{T}}s) \geqslant \frac{(1-\theta_0)q^4\|g\|^4}{L(1+r^2\|g\|)^2} \geqslant \frac{(1-\theta_0)\|g\|^2}{4L}\min\left\{q^4\|g\|^2; \frac{q^4}{r^4}\right\}. \tag{23}$$

Then the following theorem can be given.

Theorem 2.1. *The algorithms are globally convergent under the line search condition* 1:

$$\lim_{k\to\infty} g_k = 0. \tag{24}$$

Proof. Suppose the theorem is not true, then there exists an $\varepsilon > 0$ such that $\|g_k\| \geqslant \varepsilon > 0$ for infinitely many k. The $f(x_k)$ is bounded below because the level set $S(x_1)$ is bounded. This implies

$$\lim_{k\to\infty}\{f(x_k) - f(x_{k+1})\} = 0. \tag{25}$$

But (12) and (23) imply that, for those k with $\|g_k\| \geqslant \varepsilon$,

$$f(x_k) - f(x_{k+1}) \geqslant \zeta_0(-g_k^{\mathrm{T}}s_k) \geqslant \frac{(1-\theta_0)\varepsilon^2\zeta_0}{4L}\min\left\{q^4\varepsilon^2; \frac{q^4}{r^4}\right\} > 0. \tag{26}$$

The contradiction between (25) and (26) leads to the theorem. □

Remark. Under line search condition 2 or 3, Theorem 2.1 still holds.

Except there is an extra statement, in the remainder part of this paper we discuss the revised DFP algorithm, i.e., $\phi = 1$ or $\rho = 0$.

By taking the trace of both sides of (6), we get

$$\mathrm{tr}(B_*) = \mathrm{tr}(B) + \frac{\|y\|^2}{s^{\mathrm{T}}y} + \frac{\|y\|^2 s^{\mathrm{T}}Bs}{(s^{\mathrm{T}}y)^2} - \frac{2y^{\mathrm{T}}Bs}{s^{\mathrm{T}}y}. \tag{27}$$

By taking the trace of both sides of (7), we obtain

$$\mathrm{tr}(H_*) = \mathrm{tr}(H) - \frac{\|Hy\|^2}{y^{\mathrm{T}}Hy} + \frac{\|s\|^2}{s^{\mathrm{T}}y}. \tag{28}$$

Multiplying both sides of (7) by g_*, we get

$$H_*g_* = Hg_* - \frac{g_*^{\mathrm{T}}Hy}{y^{\mathrm{T}}Hy}Hy + \frac{g_*^{\mathrm{T}}s}{s^{\mathrm{T}}y}s = \frac{-g^{\mathrm{T}}Hy}{y^{\mathrm{T}}Hy}Hy + Hg + \frac{g_*^{\mathrm{T}}s}{s^{\mathrm{T}}y}s$$
$$= U\mu + Z + (1-W)s, \tag{29}$$

where U and W are defined in (16). Then we get

$$\mu = U^{-1}[H_*g_* - Z - (1-W)s] \tag{30}$$

and

$$Hy = \mu + Vs = U^{-1}[H_*g_* - Z - (1-W)s] + Vs. \tag{31}$$

Let $c_1 = [(1 + r^2 \sup\{\|g_k\|\})]^{-1}q^2$ and β_k denote the angle between g_k and H_kg_k, then from (20) we know that for all k,

$$\frac{M\|s\|}{(1-\theta)\|g\|} \geqslant \frac{y^{\mathrm{T}}s}{(1-\theta)\|g\|\|s\|} \geqslant \cos\beta = \frac{-g^{\mathrm{T}}s}{\|g\|\|s\|} \geqslant c_1\|g\|. \tag{32}$$

3. Some results for the uniformly convex objective functions

In this section, we assume:

1. The objective function $f(x)$ is uniformly convex and there exist M and m, $M \geqslant m > 0$, such that, for all $x, y \in R^n$,

$$m\|x\|^2 \leqslant x^{\mathrm{T}}G(y)x \leqslant M\|x\|^2, \tag{33}$$

where $G(y)$ is the Hessian of $f(x)$ at y.

2. $G(x)$ satisfies the Lipschitz condition, i.e., there exists an $L > 1$ such that, for all $x, y \in R^n$,

$$\|G(x) - G(y)\| \leqslant L\|x - y\|. \tag{34}$$

For simplicity, we assume

3. $f(0) = \min f(x) = 0$ and $G(0) = I_{n\times n}$, i.e., the nth order identity matrix.

Assumption 3 is equivalent to in having a linear affine transformation for the objective function which does not affect the results in the paper.

By Byrd et al. (1987) (cf. p. 1175), there exists a $c_2 > 0$ such that for all k,

$$f(x_{k+1}) \leqslant (1 - c_2\cos^2\beta_k)f(x_k), \tag{35}$$

where $\cos\beta_k$ is the same as in (32). Since

$$\frac{m}{2}\|x\|^2 \leqslant f(x) = x^{\mathrm{T}}\left(\int_0^1\int_0^u G(tx)\,\mathrm{d}t\,\mathrm{d}u\right)x \leqslant \frac{M}{2}\|x\|^2 \tag{36}$$

and $\{f(x_k)\}$ is a monotonically nonincreasing sequence of k, we get, for all k and $i>0$,

$$M\|x_k\|^2 \geqslant m\|x_{k+i}\|^2. \tag{37}$$

Let

$$G_k = \int_0^1 G(x_k + ts_k)\,\mathrm{d}t \tag{38}$$

and $c_3 = L\sqrt{M/m}(1+1/m)$, then (34) and (37) imply, for all k,

$$\|I-G\| = \|G(0)-G\| \leqslant L\sqrt{\frac{M}{m}}\|x\| \leqslant c_3\|x\|, \tag{39}$$

where the subscript k of G_k is dropped. Since $y = Gs$, and $\|y\|^2 - s^{\mathrm{T}}y = s^{\mathrm{T}}(G)^{1/2}(G-I)(G)^{1/2}s$, we get

$$\max\{m; 1-c_3\|x\|\} \leqslant \frac{\|y\|^2}{s^{\mathrm{T}}y} \leqslant \min\{M; 1+c_3\|x\|\}. \tag{40}$$

For the same reason, let $(G)^{-1}$ denote the inverse of G, we get

$$\|I-G^{-1}\| \leqslant \|G^{-1}\|\|I-G\| \leqslant L\sqrt{\frac{M}{m}}\frac{1}{m}\|x\| \leqslant c_3\|x\| \tag{41}$$

and the following holds for all k:

$$\max\left\{\frac{1}{M}; 1-c_3\|x\|\right\} \leqslant \frac{\|s\|^2}{s^{\mathrm{T}}y} \leqslant \min\left\{\frac{1}{m}; 1+c_3\|x\|\right\}. \tag{42}$$

The Quasi–Newton $H_*y = s$ and (39) imply that $g_*^{\mathrm{T}}s = g_*^{\mathrm{T}}H_*y$ and

$$|g_*^{\mathrm{T}}H_*s - g_*^{\mathrm{T}}s| = |g_*^{\mathrm{T}}H_*(I-G)s| \leqslant L\sqrt{\frac{M}{m}}\|x\|\|H_*g_*\|\|s\|. \tag{43}$$

So, by (42) and (43)

$$\begin{aligned}|g_*^{\mathrm{T}}H_*s - (1-W)\|s\|^2| &\leqslant \left|g_*^{\mathrm{T}}H_*s - \frac{g_*^{\mathrm{T}}H_*s\|s\|^2}{s^{\mathrm{T}}y}\right| + \left|\frac{g_*^{\mathrm{T}}H_*s\|s\|^2}{s^{\mathrm{T}}y} - \frac{g_*^{\mathrm{T}}s\|s\|^2}{s^{\mathrm{T}}y}\right| \\ &\leqslant 2c_3\|x\|\|H_*g_*\|\|s\|. \end{aligned} \tag{44}$$

Eqs. (16) and (19) imply that there is a constant $c_4>0$ such that, for all k,

$$\|Z\| \leqslant c_4\|d\|\|x\|. \tag{45}$$

Eqs. (30), (44) and (45) imply

$$|s^{\mathrm{T}}\mu| = |U^{-1}(g_*^{\mathrm{T}}H_*s - s^{\mathrm{T}}Z - (1-W)\|s\|^2)|$$
$$\leqslant U^{-1}\|x\|\|s\|(2c_3\|H_*g_*\| + c_4\|d\|). \tag{46}$$

By (31) and (46) we obtain

$$\frac{\|Hy\|^2}{y^{\mathrm{T}}Hy} = \frac{V\|s\|^2}{s^{\mathrm{T}}y} + \frac{\|\mu\|^2}{y^{\mathrm{T}}Hy} + \frac{2s^{\mathrm{T}}\mu}{s^{\mathrm{T}}y}$$
$$\geqslant V(1-2c_3\|x\|) + \frac{\|\mu\|^2}{y^{\mathrm{T}}Hy} - \frac{2\|x\|\|s\|(2c_3\|H_*g_*\| + c_4\|d\|)}{Us^{\mathrm{T}}y}. \tag{47}$$

Eq. (45) implies $2|Z^{\mathrm{T}}H_*g_*| \leqslant c_4\|x\|(\|H_*g_*\|^2 + \|d\|^2)$, (30) and (46) imply that there exists a $c_5 > 0$ such that, for all k,

$$\|\mu\|^2 = U^{-2}\|H_*g_* - Z - (1-W)s\|^2$$
$$\geqslant U^{-2}\{\|H_*g_*\|^2 - 2|Z^{\mathrm{T}}H_*g_*| - (1-W)^2\|s\|^2\} - 2|U^{-1}(1-W)s^{\mathrm{T}}\mu|$$
$$\geqslant U^{-2}[\|H_*g_*\|^2(1-c_5\|x\|) - c_5\|x\|\|d\|^2 - (1-W)g_*^{\mathrm{T}}s(1+c_5\|x\|)]. \tag{48}$$

We discuss the relation among $y^{\mathrm{T}}Hy$, $-g^{\mathrm{T}}Hy$, $d^{\mathrm{T}}y$ and $g^{\mathrm{T}}d$. Eq. (4) means

$$-g^{\mathrm{T}}d - r^2\|Hg\|\|g\|^2 \leqslant g^{\mathrm{T}}Hg$$
$$= -g^{\mathrm{T}}d - \|QHg\|g^{\mathrm{T}}Rg \leqslant -g^{\mathrm{T}}d - q^2\|Hg\|\|g\|^2 \tag{49}$$

and

$$d^{\mathrm{T}}y - r^2\|g\|\|Hg\|\|y\| \leqslant -g^{\mathrm{T}}Hy$$
$$= d^{\mathrm{T}}y + g^{\mathrm{T}}Ry\|QHg\| \leqslant d^{\mathrm{T}}y + r^2\|g\|\|Hg\|\|y\|. \tag{50}$$

On the other hand, $-g^{\mathrm{T}}dy^{\mathrm{T}}Hy \geqslant y^{\mathrm{T}}Hyg^{\mathrm{T}}Hg \geqslant (-g^{\mathrm{T}}Hy)^2$ implies $y^{\mathrm{T}}Hy \geqslant (-g^{\mathrm{T}}Hy)^2/(-g^{\mathrm{T}}d)$.

So, by (18) and (19), there exists a constant $c_6 > 0$ such that the following (51)–(53) hold for sufficiently large k:

$$d^{\mathrm{T}}y(1-c_6\|x\|) \leqslant -g^{\mathrm{T}}Hy \leqslant d^{\mathrm{T}}y(1+c_6\|x\|), \tag{51}$$

$$y^{\mathrm{T}}Hy \geqslant (1-\theta_0)(1-c_6\|x\|)(-g^{\mathrm{T}}Hy) \tag{52}$$

and

$$y^{\mathrm{T}}Hy \geqslant (1-\theta_0)^2(1-2c_6\|x\|)y^{\mathrm{T}}d. \tag{53}$$

Without loss of generality, we may assume that (51)–(53) hold for all k. Substituting (48) and (53) into (47), there exists a $c_7 > 0$ such that, for all k,

$$\frac{\|Hy\|^2}{y^THy} \geqslant V(1 - c_7\|x\|) + \frac{U^{-2}[\|H_* g_*\|^2(1 - c_7\|x\|)}{y^THy} - \frac{(1-W)g_*^T s(1 + c_7\|x\|)]}{y^THy} - c_7\|x\|. \tag{54}$$

Because of

$$g_k = \int_0^1 G(tx_k)\,\mathrm{d}t\, x_k, \tag{55}$$

Eq. (33) implies that the following holds,

$$m\|x\| \leqslant \|g\| \leqslant M\|x\|. \tag{56}$$

By (37) we obtain the following result for all k and $i > 0$,

$$m^3\|g_{k+i}\|^2 \leqslant M^3\|g_k\|^2. \tag{57}$$

From (35) we know $f(x_{k+1}) \leqslant \prod_{j=1}^{k}(1 - c_2\cos^2\beta)f(x_1)$, and (32) indicates

$$\sum_{k=1}^{\infty}\|g_k\|^2 < +\infty, \quad \sum_{k=1}^{\infty}\|x_k\|^2 < +\infty. \tag{58}$$

Lemma 3.1. *There exists a sequence of monotonically nonincreasing positive numbers* $\{b_k\}$ *such that, for all* k

$$\|x_k\| \leqslant b_k \leqslant c_3\|x_k\|. \tag{59}$$

Proof. Let $k = 1$ and $b_k = c_3\|x_k\|$, (37) implies that $b_k \geqslant \|x_{k+i}\|$, for all $i > 0$. We choose $b_{k+1} = \min\{b_k; c_3\|x_{k+1}\|\}$, then we can obtain $b_2, b_3, \ldots$, recursively. Clearly, Lemma 3.1 holds. □

Lemma 3.2. *Let* $\{D_k\}$ *be a sequence of positive numbers, and let* t_1 *be a positive number. If there exists a positive number* $t_2 > 0$ *such that the following holds for all* k:

$$\sum_{j=1}^{k} D_j(1 - t_1\|x_j\|) \leqslant t_2 k, \tag{60}$$

then there exists a positive number t_3 *such that the following holds for all* k:

$$\sum_{j=1}^{k} D_j\|x_j\| \leqslant t_3\|x_j\|. \tag{61}$$

Proof. Because $x_j \to 0$ as $j \to \infty$, we know that there exists a constant t_4 such that, for all k,

$$\sum_{j=1}^{k} D_j/t_4 = \sum_{j=1}^{k} E_j \leqslant k, \tag{62}$$

where $E_j = D_j/t_4$.

We first prove by mathematical induction that for any sequence of positive numbers $\{E_j\}$, if (62) holds, then the following holds for all k,

$$\sum_{j=1}^{k} E_j b_j \leqslant \sum_{j=1}^{k} b_j, \tag{63}$$

where b_j is monotonically nonincreasing, and defined in Lemma 3.1. Clearly, the result holds for $k = 1$. Assume it is true for k. If $E_{k+1} \leqslant 1$, then the above result holds for $k+1$. If $E_{k+1} > 1$, then let $F_j = E_j$, $j = 1, 2, \ldots, k-1$, $F_k = E_k - 1 + E_{k+1}$ and $F_{k+1} = 1$, (62) holds for $\{F_k\}$. So, the assumption of mathematical induction implies

$$\sum_{j=1}^{k+1} b_j \geqslant \sum_{j=1}^{k} F_j b_j + b_{k+1}$$
$$= \sum_{j=1}^{k-1} E_j b_j + (E_k - 1 + E_{k+1}) b_k + b_{k+1} \geqslant \sum_{j=1}^{k+1} E_j b_j. \tag{64}$$

The result is true for $k+1$. So, we get, for all k, that

$$\sum_{j=1}^{k} D_j \|x_j\| \leqslant \sum_{j=1}^{k} D_j b_j \leqslant \sum_{j=1}^{k} t_4 b_j \leqslant \sum_{j=1}^{k} t_4 c_3 \|x_j\|. \tag{65}$$

The lemma is proved. □

By Lemma 3.2, we obtain the following conclusion immediately.

Corollary 3.1. *Let $\{D_k\}$ be a sequence of positive numbers, and let t_1, t_2, t_3, t_4 and t_5 be positive numbers. If the following equation holds for all k:*

$$t_1 + \sum_{j=1}^{k} D_j (1 - t_2 \|x_j\|) \leqslant t_3 + t_4 k + \sum_{j=1}^{k} t_5 \|x_j\|, \tag{66}$$

then there exists a positive number $t_6 > 0$ such that for all k,

$$t_1 + \sum_{j=1}^{k} D_j \leqslant t_4 k + \sum_{j=1}^{k} t_6 \|x_j\|. \qquad \square \tag{67}$$

4. The linear convergence of RDFP algorithm

In this section, we assume assumptions 1–3 in Section 3 hold. Under the line search condition 2, we discuss the linear convergence for the RDFP algorithm. This result is also true under the line search condition 3, and the proof of the linear convergence for the RDFP algorithm under the line search condition 3 is almost the same as that under the line search condition 2.

Lemma 4.1. *There exists a* $c_8 > 0$ *such that for all* k,

$$\mathrm{tr}(B_{k+1}) \leqslant c_8 k \,\mathrm{Exp}\left\{ c_8 \sum_{j=1}^{k} \|x_j\| \right\}. \tag{68}$$

Proof. Eq. (39) implies

$$2y^{\mathrm{T}}Bs - \frac{\|y\|^2 s^{\mathrm{T}}Bs}{s^{\mathrm{T}}y} \geqslant s^{\mathrm{T}}Bs - 3\|I-G\|\|Bs\|\|s\| \geqslant s^{\mathrm{T}}Bs - \frac{3c_3\|x\|\|Bs\|^2 s^{\mathrm{T}}y}{s^{\mathrm{T}}Bs}. \tag{69}$$

Without loss of generality we may assume $1 > c_3\|x\|$ for all k. Then (27) implies

$$\begin{aligned}\mathrm{tr}(B_*) &\leqslant \mathrm{tr}(B) + \frac{3c_3\|x\|\|Bs\|^2}{s^{\mathrm{T}}Bs} + M \\ &\leqslant \mathrm{tr}(B)(1+3c_3\|x\|) + M.\end{aligned} \tag{70}$$

Clearly, there exists a constant $c_8 > 0$ such that

$$\begin{aligned}\mathrm{tr}(B_{k+1}) &\leqslant \mathrm{tr}(B_k)(1+3c_3\|x_k\|) + M \\ &\leqslant \sum_{j=2}^{k}\left[M \prod_{i=j}^{k}(1+3c_3\|x_j\|)\right] + M + \mathrm{tr}(B_1)\prod_{i=j}^{k}(1+\|x_j\|) \\ &\leqslant (1+M)(1+\mathrm{tr}(B_1))k\prod_{j=1}^{k}[1+3c_3\|x_j\|] \\ &\leqslant c_8 k\,\mathrm{Exp}\left\{ c_8\sum_{j=1}^{k}\|x_j\|\right\},\end{aligned} \tag{71}$$

which completes the proof of this lemma. □

Lemma 4.2. *There exists a constant* $c_9 > 0$ *such that, for all* k,

$$\mathrm{tr}(H_{k+1}) + \sum_{j=1}^{k}\left[V_j\left(1 - \frac{\theta_0^2}{(1-\theta_0)^2}\right) + \frac{V_j\|H_{j+1}g_{j+1}\|^2}{\|H_j g_j\|^2}\right] \leqslant k + \sum_{j=1}^{k} c_9\|x_j\|. \tag{72}$$

Proof. Definition (16) of U and (51) imply that when $1-2c_6\|x\|>0$,

$$\frac{1}{U^2y^THy}=\frac{y^THy}{(-g^THy)^2}\leqslant\frac{y^THy(1+2c_6\|x\|)^2}{(d^Ty)^2}$$
$$=\frac{\alpha V(1+2c_6\|x\|)^2}{d^Ty}\leqslant\frac{V[(1+2c_6\|x\|)(1+c_0\|x\|)]^2}{m\|Hg\|^2}. \tag{73}$$

By (73), we get the following equation under the line search condition 2:

$$\frac{(1-W)g_*^Ts}{y^THy}=\frac{U^2y^THy(g_*^Ts)^2}{(-g^THy)^2s^Ty}$$
$$\leqslant\frac{V\theta_0^2(-g^Ts)^2(1+2c_6\|x\|^2)}{(s^Ty)^3}\leqslant\frac{V\theta_0^2(1+2c_6\|x\|)^2}{(1-\theta_0)^2}. \tag{74}$$

On the other hand, (19) and (15) imply

$$\frac{1}{U^2y^THy}=\frac{y^THy}{(-g^THy)^2}\geqslant\frac{Vm(1-c_6\|x\|)^2(1-c_0\|x\|)^2}{\|Hg\|^2}. \tag{75}$$

Substituting (74) and (75) into (54), we obtain, for sufficiently large k, that

$$\frac{\|Hy\|^2}{y^THy}\geqslant\left[V\left(1-\frac{\theta_0^2}{(1-\theta_0)^2}-c_{10}\|x\|\right)\right]+\left[\frac{V\|H_*g_*\|^2}{\|Hg\|^2}(1-c_{10}\|x\|)\right], \tag{76}$$

where $c_{10}>0$ is a constant. Eqs. (76) and (28) imply

$$\mathrm{tr}(H_*)+\left[V\left(1-\frac{\theta_0^2}{(1-\theta_0)^2}-c_{10}\|x\|\right)\right]$$
$$+\left[\frac{V\|H_*g_*\|^2}{\|Hg\|^2}(1-c_{10}\|x\|)\right]\leqslant\mathrm{tr}(H)+1+c_{10}\|x\|. \tag{77}$$

We may assume that (77) holds for all k. Adding both sides of (77) over $j=1,2,\ldots,k$ we get

$$\mathrm{tr}(H_{k+1})+\sum_{j=1}^{k}V_j\left(1-\frac{\theta_0^2}{(1-\theta_0)^2}-c_{10}\|x_j\|\right)+\sum_{j=1}^{k}\left[\frac{V_j\|H_{j+1}g_{j+1}\|^2}{\|H_jg_j\|^2}(1-c_{10}\|x_j\|)\right]$$
$$\leqslant\mathrm{tr}(H_1)+k+\sum_{j=1}^{k}c_9\|x_j\|. \tag{78}$$

The Corollary 3.1 implies this lemma. □

The recurrence formula of the RDFP algorithm is the same as that of the DFP algorithm. So the determinants of the matrices $\{B_k\}$ satisfy the following recurrence relation for the RDFP algorithm (cf. [16]):

$$\det(B_{k+1}) = \det(B_k)V_k = \det(B_1)\prod_{j=1}^{k} V_j. \tag{79}$$

Lemma 4.2 implies that there exists a constant $c_{11} > 0$ such that for all k,

$$\operatorname{tr}(H_k) \leqslant c_{11}k \tag{80}$$

Theorem 4.1. *There exists a constant* δ, $0 < \delta < 1$, *such that for sufficiently large* k,

$$f(x_{k+1}) \leqslant \delta^k f(x_1). \tag{81}$$

Proof. Eqs. (72), (68), (79), (80) and $\|x_k\| \to 0$ imply that given any constant t, $t \in (0,1)$, there exists a positive integer number K_t such that the following equations hold for all $k \geqslant K_t$:

$$\begin{aligned}\sum_{j=1}^{k} V_j &\geqslant k\left[\prod_{j=1}^{k} V_j\right]^{1/k} = k\left[\frac{\det(B_{k+1})}{\det(B_1)}\right]^{1/k} \\ &\geqslant k\left[\frac{1}{\det(B_1)(c_{11}k)^n}\right]^{1/k} \geqslant kt,\end{aligned} \tag{82}$$

$$\left[\frac{1}{\|H_1 g_1\|^2 \operatorname{tr}^2(B_{k+1})}\right]^{1/k} \geqslant \left(\frac{1}{2(\|H_1 g_1\|^2)(c_{18}k)^2}\right)^{1/k}\left(\operatorname{Exp}\left\{-\frac{2c_{18}}{k}\sum_{j=1}^{k}\|x_j\|\right\}\right) \geqslant t, \tag{83}$$

and

$$\frac{1}{k}\sum_{j=1}^{k} c_9\|x_j\| \leqslant 1 - t. \tag{84}$$

Combining (82) and (83) we obtain

$$\begin{aligned}\sum_{j=1}^{k}\frac{V_j\|H_{j+1}g_{j+1}\|^2}{\|H_j g_j\|^2} &\geqslant k\left[\prod_{j=1}^{k}\frac{V_j\|H_{j+1}g_{j+1}\|^2}{\|H_j g_j\|^2}\right]^{1/k} \\ &= k\left[\frac{\|H_{k+1}g_{k+1}\|^2}{\|H_1 g_1\|^2}\right]^{1/k}\prod_{j=1}^{k}[V_j]^{1/k} \geqslant kt^2\|g_{k+1}\|^{2/k}.\end{aligned} \tag{85}$$

Substituting (82), (84) and (85) into (72), we obtain, for all $k \geqslant K_t$,

$$\operatorname{tr}(H_{k+1}) + k\{t[1 - [\theta_0/(1-\theta_0)]^2 + t^2\|g_{k+1}\|^2\}$$

$$\leqslant \mathrm{tr}(H_{k+1}) + \sum_{j=1}^{k} \left[V_j \left(1 - \frac{\theta_0^2}{(1-\theta_0)^2} \right) + \frac{V_j \|H_{j+1} g_{j+1}\|^2}{\|H_j g_j\|^2} \right]$$

$$\leqslant k + \sum_{j=1}^{k} c_9 \|x_j\| \leqslant k + k(1-t) \tag{86}$$

or

$$\|g_{k+1}\| \leqslant \left[\frac{2 - 2t + t[\theta_0/(1-\theta_0)]^2}{t^2} \right]^{2/k}. \tag{87}$$

Clearly, this theorem holds for $\theta_0 < 1/2$ and t can be any number in $(0, 1)$. □

5. The one-step superlinear convergence of the algorithms

In this section, we assume assumptions 1–3 in Section 3 hold. We discuss the RDFP algorithm under the line search condition 2 or 3. The algorithm presented in this paper has been proved to have linear convergence rate. The Theorem 4.1 implies

$$\sum_{j=1}^{\infty} \|x_j\| < \infty, \quad \sum_{j=1}^{\infty} \|g_j\| < \infty. \tag{88}$$

Similar to the proof in [17], (88) may imply that our algorithm has one-step superlinear convergence rate. But we would rather use another way which is somewhat different from Powell's method to get some interesting results.

Lemma 5.1. *There exists a constant $c_{12} > 0$ such that, for all k,*

$$\mathrm{tr}(B_{k+1}) + \sum_{j=1}^{k} \frac{s_j^{\mathrm{T}} B_j s_j}{s_j^{\mathrm{T}} y_j} \leqslant k + c_{12}. \tag{89}$$

Proof. Lemma 4.1, (28) and (88) imply that there exists a constant $c_{13} > 0$ such that, for all k,

$$\mathrm{tr}(B_k) \leqslant c_{13} k, \quad \mathrm{tr}(H_k) \leqslant c_{13} k, \tag{90}$$

and for sufficiently large k, $\mathrm{tr}(B)\|x\| \leqslant \|x\|^{1/2}$. Substituting (69) into (27), we get, for all k, that

$$\mathrm{tr}(B_{k+1}) + \frac{s_k^{\mathrm{T}} B_k s_k}{s_k^{\mathrm{T}} y_k}(1 - 3c_3\|x_k\|) \leqslant \mathrm{tr}(B_k) + \frac{c_3 \|x_k\| \|y_k\| \|B_k s_k\|}{s_k^{\mathrm{T}} B_k s_k} + 1 + c_3\|x_k\|$$

$$\leqslant \mathrm{tr}(B_k) + c_{13} M \|x_k\| k + 1 + 3c_3\|x_k\|. \tag{91}$$

Adding both sides of (91) over $j = 1, 2, \ldots, k$, we get

$$\mathrm{tr}(B_{k+1}) + \sum_{j=1}^{k} \frac{s_j^{\mathrm{T}} B_j s_j}{s_j^{\mathrm{T}} y_j}(1 - c_3\|x_j\|) \leqslant \mathrm{tr}(B_1) + k + \sum_{j=1}^{k} [Mc_{13}\|x_j\| j + c_3\|x_j\|]. \tag{92}$$

By (92) and (88), it is clear that this lemma holds. □

Lemma 5.2. *There exists a* $c_{14} > 0$ *such that, for all* k,

$$\mathrm{tr}(H_{k+1}) + \sum_{j=1}^{k} V_j \leqslant k + c_{14}. \tag{93}$$

Proof. Eq. (90) implies that there exists a constant $c_{15} > 0$ such that, for all k,

$$\sum_{j=1}^{\infty} \frac{2\|x_j\|\|s_j\|(2c_3\|H_{j+1}g_{j+1}\| + c_4\|d_j\|)}{U_j s_j^{\mathrm{T}} y_j} = c_{15}. \tag{94}$$

Substituting (47) and (94) into (28) and then adding both sides over $j = 1, 2, 3, \ldots, k$, we get

$$\mathrm{tr}(H_{k+1}) + \sum_{j=1}^{k} V_j(1 - c_3\|x_j\|) - c_{15} \leqslant \mathrm{tr}(H_{k+1}) + \sum_{j=1}^{k} \frac{\|H_j y_j\|^2}{y_j^{\mathrm{T}} H_j y_j}$$

$$= \mathrm{tr}(H_1) + \frac{\|s_j\|^2}{s_j^{\mathrm{T}} y_j} \leqslant \mathrm{tr}(H_1) + k + \sum_{j=1}^{k} c_3\|x_j\|. \tag{95}$$

Now it is easy to see that the lemma holds. □

Theorem 5.1. *The algorithm presented in this paper is one-step superlinearly convergent for uniformly objective functions, i.e.*,

$$\lim_{k\to\infty} \|g_{k+1}\|/\|g_k\| = 0. \tag{96}$$

Proof. Adding both sides of (89) and (93), respectively, we get

$$\mathrm{tr}(B_{k+1} + H_{k+1}) + 2\sum_{j=1}^{k} \left[\frac{(y_j^{\mathrm{T}} H_j y_j)^{1/2} (s_j^{\mathrm{T}} B_j s_j)^{1/2}}{s_j^{\mathrm{T}} y_j} - 1 \right]$$

$$+ \frac{[(y_j^{\mathrm{T}} H_j y_j)^{1/2} - (s_j^{\mathrm{T}} B_j s_j)^{1/2}]^2}{s_j^{\mathrm{T}} y_j} \leqslant c_{12} + c_{14}. \tag{97}$$

As $y_j^{\mathrm{T}} H_j y_j s_j^{\mathrm{T}} B_j s_j \geqslant (s_j^{\mathrm{T}} y_j)^2$, we have

$$\mathrm{tr}(H_{k+1} + B_{k+1}) \leqslant c_{12} + c_{14}, \tag{98}$$

$$\lim_{k\to\infty} \frac{y_k^{\mathrm{T}} H_k y_k s_k^{\mathrm{T}} B_k s_k}{(s_k^{\mathrm{T}} y_k)^2} = 1 \tag{99}$$

and

$$\lim_{k\to\infty} \left| \frac{y_k^{\mathrm{T}} H_k y_k}{s_k^{\mathrm{T}} y_k} - \frac{s_k^{\mathrm{T}} B_k s_k}{s_k^{\mathrm{T}} y_k} \right| = 0. \tag{100}$$

Substituting (99) into (100), we get

$$\lim_{k\to\infty} \left| \frac{s_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} B_k s_k} - \frac{s_k^{\mathrm{T}} B_k s_k}{s_k^{\mathrm{T}} y_k} \right| = 0 \tag{101}$$

and

$$\begin{aligned} \lim_{k\to\infty} \frac{s_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} B_k s_k} &= \lim_{k\to\infty} \frac{s_k^{\mathrm{T}} B_k s_k}{s_k^{\mathrm{T}} y_k} \\ &= \lim_{k\to\infty} \frac{s_k^{\mathrm{T}} y_k}{y_k^{\mathrm{T}} H_k y_k} = \lim_{k\to\infty} \frac{y_k^{\mathrm{T}} H_k y_k}{s_k^{\mathrm{T}} y_k} = 1. \end{aligned} \tag{102}$$

From (102) we get

$$\lim_{k\to\infty} \frac{-\alpha_k g_k^{\mathrm{T}} d_k}{y_k^{\mathrm{T}} d_k} = \lim_{k\to\infty} \frac{-g_k^{\mathrm{T}} d_k}{\|d_k\|^2} = 1, \tag{103}$$

and for sufficiently large k,

$$1 < 2\alpha_k < 4. \tag{104}$$

We get by (103),

$$\begin{aligned} g^{\mathrm{T}}(x_k + d_k) d_k - g_k^{\mathrm{T}} d_k &= d_k^{\mathrm{T}} \left(\int_0^1 G(x_k + t d_k)\, \mathrm{d}t \right) d_k \\ &= \|d_k\|^2 + \mathrm{o}(\|d_k\|^2). \end{aligned} \tag{105}$$

Eqs. (103) and (105) imply

$$d_k^{\mathrm{T}} g(x_k + d_k) = \mathrm{o}(\|d_k\|^2) = \mathrm{o}(-g_k^{\mathrm{T}} d_k) \tag{106}$$

and

$$\lim_{k\to\infty} \frac{g(x_k + d_k)^{\mathrm{T}} d_k}{g_k^{\mathrm{T}} d_k} = 0. \tag{107}$$

Therefore, for sufficiently large k,

$$f(x_k) - f(x_k + d_k) = d_k^{\mathrm{T}} \left(\int_0^1 \int_u^1 G(x_k + t_k d_k)\, \mathrm{d}t\, \mathrm{d}u \right) d_k - d_k^{\mathrm{T}} g(x_k + d_k)$$

$$\geqslant \frac{\|d_k\|^2 - L\|d_k\|^3}{2} - d_k^{\mathrm{T}} g(x_k + d_k) \geqslant \zeta_0 g_k^{\mathrm{T}} H_k g_k. \tag{108}$$

Eqs. (106) and (108) show that $\alpha_k = 1$ must satisfy (12) and (13) for sufficiently large k. So we can take $\alpha \equiv 1$ for sufficiently k. Eqs. (102) and (106) imply

$$\lim_{k\to\infty} \frac{y_k^{\mathrm{T}} H_k y_k}{g_k^{\mathrm{T}} d_k} - 1 = \lim_{k\to\infty} \frac{g_{k+1}^{\mathrm{T}} H_k g_{k+1}}{g_k^{\mathrm{T}} d_k} = 0 \tag{109}$$

and

$$\lim_{k\to\infty} \frac{\|g_{k+1}\|}{\|g_k\|} = 0. \tag{110}$$

This completes the proof of Theorem 5.1. □

From the proof of Theorem 5.1 we may obtain conclusions below, (1) $\alpha \equiv 1$, and

$$\sum_{j=1}^{\infty} \frac{g_{j+1}^{\mathrm{T}} H_j g_{j+1}}{g_j^{\mathrm{T}} H_j g_j} < \infty, \tag{111}$$

(by (89) and (93)). (2) There exist H and B satisfying

$$\lim_{k\to\infty} H_k = H, \qquad \lim_{k\to\infty} B_k = B. \tag{112}$$

The following theorem follows from the proof of Theorem 5.1.

Theorem 5.2. *If $x^k \to x^*$ then under the line condition* 2 *the DFP algorithm is one-step superlinearly convergent for uniformly objective functions*, *i.e.*,

$$\lim_{k\to\infty} \|g_{k+1}\|/\|g_k\| = 0. \tag{113}$$

Remark 5.1. Theorem 5.1 holds for all revised Broyden algorithms under the line condition 1 or 2 or 3.

We list the n-step quadratic convergence of the algorithm without detailed proof.

Theorem 5.3. *If the line search satisfies the line search condition* 3, *then the algorithm presented in this paper is n-step quadratically convergent*, *i.e.*,

$$\|x_k\|^2 = \mathrm{O}(\|x_{k+n}\|).$$

Remark 5.2. Theorem 5.3 holds for all revised Broyden algorithms.

6. Discussion

In Sections 2–5, we have shown that the revised DFP algorithm proposed in this paper have good convergence properties, that is, the algorithms guarantee one-step superlinear convergence and n-step quadratical convergence for uniformly convex objective functions. Furthermore, they are globally convergent for the continuously differentiable functions. So, we use them not only to solve unconstrained nonlinear problems, but also to solve constrained nonlinear optimization problems. For example, we change constrained nonlinear optimization problems into unconstrained nonlinear optimization problems which are equivalent to the prime original problems by multiplier methods or penalty function methods. Generally, the objective functions obtained in the unconstrained nonlinear optimization problems may not be convex. So, in this case the revised DFP algorithm are usually more efficient than the DFP algorithm.

We have done some computational experiments for the DFP algorithm and the revised DFP algorithm under both the Wolfe conditions and the modified Wolfe condition. The testing results show that, for uniform convex functions, the two classes of algorithms are same effective under both the Wolfe conditions and the modified Wolfe conditions, and for nonconvex functions the revised DFP algorithm has better stability than the DFP algorithm. Here we compare the performance of the BFGS algorithm with revised BFGS algorithm under the Wolfe conditions for blow function.

Function. Let

$$\Phi(x) = \sum_{j=0}^{100} [1 - e^{-jh} r(jh, x)]^2, \quad x \in R^5. \tag{114}$$

where $h = 0.05$ and $r(t,x)$ has the value

$$r(t,x) = \frac{x_1 + x_2 t + x_3 t^2}{1 + (x_4 + x_5 t)}. \tag{115}$$

The objective function itself is the expression

$$f(x) = \Phi(Dx), \tag{116}$$

where D is a 5×5 positive diagonal matrix. We choose the starting point

$$x_0 = (15d_{11}^{-1}, 10d_9^{-1}, 5d_{33}^{-1}, 6d_{44}^{-1}, -d_{55}^{-1})^{\mathrm{T}}, \tag{117}$$

where $d_{ii}/d_{i+1,i+1}=$ constant. We set $\theta = 0.7$ and $B_0 = I$, and the stopping condition is the inequality

$$|f(x_k) - f(x_*)| < 10^{-10}, \tag{118}$$

where the optimal function value of the problem is to ten decimal places,

$$f(x_*) = 3.085557482 \times 10^{-3}. \tag{119}$$

The computing results are listed in Table 1.

Table 1

Function	x_0	BFGS		RBFGS	
		IN	FN	IN	FN
1	$15 \times 1, \ldots, -1 \times 1$	42	58	35	54
1	$15 \times 1, \ldots, -1 \times 10^4$	68	102	57	96
1	$15 \times 1, \ldots, -1 \times 10^8$	72	136	64	123
1	$15 \times 1, \ldots, -1 \times 10^{12}$	81	150	69	125
1	$15 \times 10^{-4}, \ldots, -1 \times 1$	91	111	63	99
1	$15 \times 10^{-8}, \ldots, -1 \times 1$	67	121	67	103
1	$15 \times 10^{-12}, \ldots, -1 \times 1$	F	F	70	120

7. Uncited references

[19,20,35,36,38]

References

[1] M. Al-Baali, Variational quasi-Newton methods for unconstrained optimization, J. Optim. Theory Appl. 77 (1993) 127–143.
[2] P. Armand, J.C. Gilbert, S. Jan-Jǵou, A feasible BFGS interior point algorithm for solving convex minimization problem, SIAM J. Optim. 11 (2000) 199–222.
[3] C.G. Broyden, A class of algorithms for solving nonlinear simultaneous equations, Math. Comput. 19 (1965) 577–593.
[4] R.H. Byrd, J. Nocedal, Y. Yuan, Global convergence of a class of quasi-Newton methods on convex problems, SIAM J. Numer. Anal. 24 (1987) 1171–1190.
[5] W.C. Davidon, Variable metric algorithms for minimization, Argonne National Laboratory Report, 1959.
[6] W.C. Davidon, Optimally conditioned optimization algorithms without line searches, Math. Programming 9 (1975) 1–30.
[7] J.E. Dennis Jr., J. Jorge, Moré, quasi-Newton method, motivation and theory, SIAM Rev. 19 (1977) 46–89.
[8] L.C.W. Dixon, Variable metric algorithms: necessary and sufficient conditions for identical behavior of nonquadratical functions, J. Optim. Theory Appl. 10 (1972) 34–40.
[9] R. Fletcher, Practical Methods of Optimization 1, Unconstrained Optimization, Wiley, New York, 1987.
[10] R. Fletcher, M.J.D. Powell, A rapidly convergent descent method for minimization, Comput. J. 6 (1963) 163–168.
[11] Y.F. Hu, C. Storey, Family of optimally conditioned quasi-Newton updates for unconstrained optimization, J. Optim. Theory Appl. 83 (1994) 421–431.
[12] M. Lalee, J. Nocedal, Automatic column scaling strategies for quasi-Newton methods, SIAM J. Optim. 3 (1993) 637–653.
[13] J.L. Nazareth, R.B. Mifflin, The least prior deviation quasi-Newton update, Math. Programming 65 (1994) 247–261.
[14] S.S. Oren, Self-scaling variable metric algorithms, for unconstrained minimization, Ph.D. Thesis, Department of Engineering Economic Systems, Stanford University, 1972.
[15] S.S. Oren, D.G. Luenberger, Self-scaling variable metric algorithms, J. Optim. Theory Appl. 37 (1974) 137–147.
[16] M.J.D. Powell, On the convergence of the variable metric algorithm, Inst. Math. Appl. 7 (1971) 21–36.
[17] M.J.D. Powell, Some global convergence properties of the variable metric algorithm for minimization without exact line searchers, in: R.W. Cottle, C.E. Lemke (Eds.), SIAM-AMS Proceedings, Nonlinear Programming, Vol. 6, SIAM Publications, Philadelphia, PA, 1976.
[18] M.J.D. Powell, Variable metric methods for constrained optimization, in: A. Bachem, M. Grotschel, B. Korte (Eds.), Mathematical Programming: The State the Art, Bonn, 1982, Springer, Berlin, New York, 1983, pp. 288–311.

[19] M.J.D. Powell, Nonconvex minimization calculations and the conjugate gradient method, in: D.F. Griffiths (Ed.), Numerical Analysis, Lecture Notes in Mathematics, Vol. 1066, Springer, Berlin, 1984, pp. 122–141.
[20] M.J.D. Powell, How bad are the BFGS and DFP methods when the objective function is quadratic? Math. Programming 34 (1986) 34–47.
[21] M.J.D. Powell, Update conjugate directions by the BFGS formula, Math. Programming 38 (1987) 29–46.
[22] D. Pu, Modified DFP algorithm, J. Acta Math. Appl. Sinica 13 (1990) 118–122.
[23] D. Pu, A class of DFP algorithm without exact linear search, Asia-Pacific J. Oper. Res. 9 (1992) 207–220.
[24] D. Pu, The convergence of the Broyden algorithms for LC gradient function, J. Acta Math. Appl. Sinica (English Ser.) 16 (2000) 313–319.
[25] D. Pu, The convergence of DFP algorithm without exact linear search, J. Optim. Theory Appl. 112 (2002) 187–211.
[26] D. Pu, W. Tian, A class of modified Broyden algorithms, J. Comp. Math. 8 (1994) 366–379.
[27] D. Pu, W. Tian, A class of Broyden algorithms with revised search direction, Asia-Pacific J. Oper. Res. 14 (1997) 93–109.
[28] D. Pu, W. Yu, On the convergence property of the DFP algorithm, J. Ann. Oper. Res. 24 (1990) 175–184.
[29] H. Qi, L. Qi, A new QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization, SIAM J. Optim. 11 (2000) 113–132.
[30] D. Siegel, Updating of conjugate direction matrices using members of Broyden's family, Math. Programming 60 (1993) 167–185.
[31] D. Siegel, Modifying the BFGS update by a new column scaling technique, Math. Programming 66 (1994) 45–78.
[32] E. Spedicato, A variable-metric method for function minimization derived from invariancy to nonlinear scaling, J. Optim. Appl. 20 (1976) 315–329.
[33] E. Spedicato, On a conjecture of Dixon and other topic in variable metric methods, Math. Programming 15 (1978) 123–129.
[34] E. Spedicato, A class of rank-one positive definite quasi-Newton updates for unconstrained minimization, Math. Oper. Statist. Ser. Optim. 14 (1982) 61–70.
[35] P. Wolfe, Convergence conditions for ascent methods, SIAM Rev. 11 (1969) 226–235.
[36] P. Wolfe, Convergence conditions for ascent methods II: some corrections, SIAM Rev. 13 (1971) 185–188.
[37] Y. Zhang, R.P. Tewarson, Quasi-Newton algorithms with updates from the pre-convex part of Broyden's family, IMA J. Numer. Anal. 8 (1988) 487–509.
[38] G. Zoutendijk, Nonlinear programming, computational methods, in: J. Abadie (Ed.), Integer and Nonlinear Programming, North-Holland, Amsterdam, 1970, pp. 37–86.