Journal of Computational Mathematics, Vol.23, No.4, 2005, 357–372.

CONVERGENCE PROPERTIES OF MULTI-DIRECTIONAL PARALLEL ALGORITHMS FOR UNCONSTRAINED MINIMIZATION [*1)]

Cheng-xian Xu Yue-ting Yang
(*Faculty of Sciences, Xi'an Jiaotong University, Xi'an 710049, China*)

Abstract

Convergence properties of a class of multi-directional parallel quasi-Newton algorithms for the solution of unconstrained minimization problems are studied in this paper. At each iteration these algorithms generate several different quasi-Newton directions, and then apply line searches to determine step lengths along each direction, simultaneously. The next iterate is obtained among these trail points by choosing the lowest point in the sense of function reductions. Different quasi-Newton updating formulas from the Broyden family are used to generate a main sequence of Hessian matrix approximations. Based on the BFGS and the modified BFGS updating formulas, the global and superlinear convergence results are proved. It is observed that all the quasi-Newton directions asymptotically approach the Newton direction in both direction and length when the iterate sequence converges to a local minimum of the objective function, and hence the result of superlinear convergence follows.

Mathematics subject classification: 65K05.
Key words: Unconstrained minimization, Multi-directional parallel quasi-Newton method, Global convergece, Superlinear convergence.

1. Introduction

This paper concerns with quasi-Newton methods for unconstrained nonlinear minimization

$$\min \quad f(x), \tag{1.1}$$

where $f:\ R^n \to R$ is assumed to be twice continuously differentiable. Starting from an initial point x_1 and an initial symmetric positive definite matrix B_1, a quasi-Newton method generates sequences $\{x_k\}$ and $\{B_k\}$ by the iteration

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

and an updating formula for B_k, where α_k is a step length and d_k is a descent search direction that is generated by solving the following system of equations

$$B_k d_k = -g_k,$$

$g_k = \nabla f(x_k)$ is the gradient of $f(x)$ at x_k. B_k is an $n \times n$ symmetric matrix that approximates the Hessian $G(x) = \nabla^2 f(x)$ of $f(x)$ at x_k, and satisfies the so-called quasi-Newton equation

$$B_k s_{k-1} = y_{k-1} \tag{1.3}$$

with $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$.

* Received November 12, 2003; final revised January 11, 2005.
1) This work is supported by National Science Foundation of China: 10231060.

Various updating formulae that satisfy equation (1.3) exist, and one of the most widely used class of updates was the Broyden family (see [3])

$$B_{k+1}(\phi) = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \phi(s_k^T B_k s_k) u_k u_k^T, \tag{1.4}$$

where ϕ is a scale parameter and

$$u_k = \frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k}.$$

The computational characters and convergence properties of quasi-Newton methods in Broyden family have been widely studied (see [6], [7], [9], [10], [11], [12], [13], [14], [16], [24], [25]).

One of the most widely used quasi-Newton update is the BFGS update

$$B_{k+1}^{BFGS} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}, \tag{1.5}$$

that is obtained by setting $\phi = 0$ in (1.4), denoted by $B_{k+1}(0)$. Broyden, Dennis and More in [5] proved that the BFGS method with unit step length for all k is superlinearly convergent provided that the initial point x_1 and the initial Hessian approximation B_1 are sufficiently accurate. Powell in [24] proved the global convergence for the BFGS method when it is applied to convex functions and the step length α_k satisfies the Wolfe conditions for all k. Furthermore, if the function $f(x)$ is strictly convex and the step length $\alpha_k = 1$ is taken whenever it satisfies the Wolfe conditions, the result of Broyden, Dennis and More in [5] applies, i.e., the convergence rate is superlinear. These convergence properties of the BFGS method have been extended to the convex Broyden class, except for the DFP method, $(0 \le \phi < 1)$ by Ritter [26], Byrd, Nocedal and Yuan [7], and to the preconvex Broyden class $(\phi_{k0} < \phi < 1)$ by Byrd, Liu and Nocedal [8] where

$$\phi_{k0} = (s_k^T y_k)^2 / [(s_k^T y_k)^2 - s_k^T B_k s_k y_k^T B_k^{-1} y_k] < 0. \tag{1.6}$$

The easiest update in (1.4) is the symmetric rank one (**SR1**) update

$$B_{k+1}^{SR1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k},$$

that is obtained by setting

$$\phi = s_k^T y_k / (y_k - B_k s_k)^T s_k \stackrel{\triangle}{=} \hat{\phi}_k \tag{1.7}$$

in (1.4), denoted by $B_{k+1}(\hat{\phi}_k)$. The drawback of the SR1 update is that the matrix B_{k+1}^{SR1} may not be positive definite or it may even not well defined when the denominator approaches zero. However, some recent works on the SR1 method have sparked renewed interesting in this updating formula (see [9], [21], [18] and [1]). It is proved in [9] that the sequence $\{B_k\}$ generated by the SR1 update converges to the actual Hessian $G(x^*)$ at the solution x^*, provided that the search directions $\{d_k\}$ are uniformly linearly independent, and that the denominators in the SR1 update are always sufficiently different from zero, and that the iterates $\{x_k\}$ converges to x^*. Moreover, numerical tests (see [9] and [29]) show that in comparison with the BFGS update, the SR1 update generates more accurate Hessian approximations. Khalfan, Byrd and Schnabel in [18] provided a proof of $(n+1)$-step super-linear convergence result for the SR1 method under an assumption that the updating matrices $\{B_k\}$ are positive definite for all k and bounded asymptotically.

Based on the idea of obtaining more accurate Hessian approximation in the direction s_{k-1} through using more available function value information in updating formulae, Zhang and Xu in [32] proposed a modification to quasi-Newton equation (1.3)

$$B_k s_{k-1} = \left(1 + \frac{\theta_{k-1}}{s_{k-1}^T y_{k-1}}\right) y_{k-1} \stackrel{def}{=} \hat{y}_{k-1},$$

where

$$\theta_{k-1} = 6(f_{k-1} - f_k) + 3(g_{k-1} + g_k)^T s_{k-1}. \tag{1.8}$$

The resulting modified BFGS update has the form

$$B_{k+1}^{MBFGS} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\hat{y}_k \hat{y}_k^T}{s_k^T \hat{y}_k}, \tag{1.9}$$

and is denoted by $B_{k+1}(\theta_k)$, which belongs to the class of non-quasi-Newton updates derived by Yuan and Byrd (see [31]). And a similar update as (1.9) was first derived by Yuan (see [30]). When inverse updating formula is used, the corresponding formula is just the modified inverse BFGS updating suggested by Biggs (see [2])

$$H_{k+1}^{MBFGS} = H_k + \frac{1}{s_k^T y_k}[(\frac{1}{\pi_k} + \frac{y_k^T H_k y_k}{s_k^T y_k}) s_k s_k^T - s_k y_k^T H_k - H_k y_k s_k^T],$$

where $\pi_k = 1/\theta_k$. It is proved that with a safeguard strategy to keep positive definite updating, the modified BFGS method maintains the global convergence property (see [19]), and the super-linear convergence property (see [32]) of the BFGS method. Comparisons between the BFGS method and the modified BFGS method show a favorite to the modified BFGS method either on computational costs or on the Hessian approximation accuracy (see [32]).

It has been observed (see [20], [22] and [21]) that in practical applications, a particular quasi-Newton method may be "good" in solving certain types of minimization problems, but its efficiency degenerates when it is applied to solve other categories of problems. For instance, when the SR1 method (without any modification) is applied to solve a practical problem, two possible cases are likely to occur. If it solves the problem, its efficiency is usually better than other quasi-Newton methods, such as the BFGS method. However, it may fail to solve the problem due to the problems mentioned above. Any subsequent modifications made to the SR1 update may force it to solve the problem, but its efficiency degenerates.

Based on these observations, multi-directional parallel quasi-Newton (PQN) algorithms are proposed by Phua in [20] to explore different quasi-Newton directions simultaneously. Phua, Fan and Zeng in [22] proposed multi-step, multi-directional PQN algorithms for solving large-scale nonlinear optimization problems. When three parallel processors are used for computing parallel search directions, a reduction of 200% or more in terms of the number of iterations and function/gradient evaluations has been achieved by these PQN algorithms over a wide range of 63 test problems (see [22]). Self-scaling PQN methods are also considered by Phua, Fan and Zeng in [23] based on a new class of three parameter quasi-Newton updates. They reported that the average speedup factors obtained by these new self-scaling PQN algorithms over the conventional QN methods are more than 300%, both in terms of the total number of iterations and the total number of function/gradient evaluations, when three parallel processors are used to compute search directions, simultaneously. However, whether the multi-directional parallel quasi-Newton algorithms converge globally is an open question. It is the purpose of this paper that we explore the global and super-linear convergence properties of these multi-directional parallel quasi-Newton algorithms. The paper is organized as follows. In the next section we describe the multi-directional parallel quasi-Newton algorithms. The global convergence properties of these algorithms will be proved in sections 3 and 4. In section 5, we study the super-linear convergence property of these algorithms. Conclusions are presented in section 6. Throughout the paper, $\|\cdot\|$ denotes the Euclidean vector norm or its induced matrix norm.

2. The Algorithms

In this section we describe the multi-directional parallel quasi-Newton algorithms for the solution of problem (1.1). Assume that p processors are available for calculating p search directions simultaneously, and the sequel $\mathcal{N}$ is used to denote the set of integers $\mathcal{N} = \{1, 2, \cdots, p\}$.

Algorithm 2.1.

Step 1 Initialization : Give $\epsilon > 0$, initial point x_1, and $B_1(= I)$; Calculate $f_1 = f(x_1)$ and $g_1 = g(x_1)$ and set $k = 1$;

Step 2 Convergence test : If $\|g_k\| \leq \epsilon$ then terminate else go to step 3;

Step 3 Calculate parallel search directions : Calculate

$$d_{ki} = -B_k(\phi_{k-1,i})^{-1} g_k, \quad i = 1, 2, \cdots, p; \tag{2.1}$$

Step 4 Perform parallel line searches : Along each of these p directions calculate step lengths α_{ki} in parallel to satisfy the Wolfe descent conditions

$$f(x_k + \alpha_{ki} d_{ki}) \leq f(x_k) + \rho \alpha_{ki} g_k^T d_{ki}, \tag{2.2}$$

$$g(x_k + \alpha_{ki} d_{ki})^T d_{ki} \geq \sigma g_k^T d_{ki}, \tag{2.3}$$

with $\rho \in (0, 1/2)$ and $\sigma \in (\rho, 1)$;

Step 5 Calculate the new point : Determine a subset $I_k \subset \mathcal{N}$ and calculate

$$f(x_k + \alpha_{kl} d_{kl}) = \min \{ f(x_k + \alpha_{ki} d_{ki}) \mid i \in I_k \},$$

and set $x_{k+1} = x_k + \alpha_{kl} d_{kl}$;

Step 6 Calculate new Hessian approximations :
Calculate $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$; Choose one value of ϕ, ϕ_{kj} say, and update B_k to obtain $B_{k+1}(\phi_{kj})$; Set $k := k + 1$ and go to step 2;

The value of $\phi_{k-1,i}$ in step 3 can be chosen as 0, ϕ_{k0} given by (1.6), and $\hat{\phi}_k$ given by (1.7) in equation (1.4); or θ_{k-1} in (1.8) and (1.9).

Methods are available to calculate the parallel search directions effectively. It follows from (1.4), (1.9) and the formula

$$(A + uu^T)^{-1} = A^{-1} - \frac{A^{-1} u u^T A^{-1}}{1 + u^T A^{-1} u} \quad (1 + u^T A^{-1} u \neq 0) \tag{2.4}$$

that once a search direction, say d_{kj}, is calculated from the matrix $B_k(\phi_{k-1j})$ that is updated in step 6, the other search directions are just combinations of the direction d_{kj} and a vector v_k, and only difference is the combination coefficients. For instance, if the BFGS update (1.5) is used to get the matrix $B_{k+1}(0)$ in step 6, then $d_{k+1,j} = -B_{k+1}(0)^{-1} g_{k+1}$ is the BFGS direction. The SR1 direction is

$$d_{k+1,i} = d_{k+1,j} + \frac{\hat{\phi}_k s_k^T B_k s_k u_k^T B_{k+1}(0)^{-1} g_{k+1}}{1 + \hat{\phi}_k s_k^T B_k s_k u_k^T B_{k+1}(0)^{-1} u_k} B_{k+1}(0)^{-1} u_k,$$

and the modified BFGS direction is

$$d_{k+1,i} = d_{k+1,j} + \frac{\theta_k y_k^T B_{k+1}(0)^{-1} g_{k+1}}{(s_k^T y_k)^2 + \theta_k y_k^T B_{k+1}(0)^{-1} y_k} B_{k+1}(0)^{-1} y_k.$$

For other choices of the updating formula in the Broyden class, the search directions can be similarly calculated.

Other line search conditions such as the Goldstein condition and the Curry-Altman condition can also be used in step 4. For convergence analysis of quasi-Newton algorithms, the Wolfe conditions are usually considered. It is known that when d_{ki} is a descent direction, intervals of acceptable α values for (2.2) and (2.3) exist (see [14]). If a step length satisfying (2.2) and (2.3) does not exist for some ascent direction d_{ki}, α_{ki} is set to zero.

The subset I_k in step 5 can be determined according to different criterions. In this paper we consider the following criteria for the selection of the subset I_k. Let $B_k(\phi_{k-1,j})$ be the updating formula selected in step 6, and d_{kj} be the search direction generated from (2.1) with $B_k(\phi_{k-1,j})$. The subset I_k is determined as follows

$$I_{k1} = \{\ i \mid \frac{{d_{ki}}^T B_k(\phi_{k-1,i}) d_{ki}}{\|B_k(\phi_{k-1,i}) d_{ki}\| \|d_{ki}\|} \geq \beta \frac{{d_{ki}}^T B_k(\phi_{k-1,j}) d_{ki}}{\|B_k(\phi_{k-1,j}) d_{ki}\| \|d_{ki}\|},\ i \in \mathcal{N}\}, \tag{2.5}$$

$$I_{k2} = \{\ i \mid \frac{{d_{kj}}^T B_k(\phi_{k-1,j}) d_{kj}}{\|B_k(\phi_{k-1,j}) d_{kj}\| \|d_{kj}\|} \geq \beta \frac{{d_{ki}}^T B_k(\phi_{k-1,j}) d_{ki}}{\|B_k(\phi_{k-1,j}) d_{ki}\| \|d_{ki}\|},\ i \in \mathcal{N}\}, \tag{2.6}$$

where $\beta \in (0, 1)$ is a small positive number. Both sets I_{k1} and I_{k2} are not empty, since there is at least one element j in each set. It can be expected that for small enough β, both sets I_{k1} and I_{k2} contains all or most indices in $\mathcal{N}$.

Any updating formula can be selected in step 6, for example, $B_k(\phi_{k-1,l})$ where the index l is determined in step 5 (see [20]). Alternatively, the same updating formula that generates positive definite matrices can be employed in step 6, for example the BFGS update or the modified BFGS update. As for the modified BFGS update (1.9), the following result holds.

Lemma 2.1. *Assume that $f(x)$ is twice continuously differentiable. If the sequence $\{x_k\}$ converges to a local minimizer x^* of $f(x)$ with $g(x^*) = 0$ and $G(x^*)$ positive definite, then*

$$\lim_{k \to \infty} \frac{\theta_k}{s_k^T y_k} = 0.$$

Proof. It follows from Taylor expansions of $f(x)$ and $g(x)$ at the point x_k that

$$\theta_k = O(\|s_k\|^3).$$

Since $G(x^*)$ is positive definite and $\{x_k\}$ converges to x^*, there exist $m_1 > 0$ and an integer K such that

$$s_k^T y_k = s_k^T \int_0^1 G(x_k + t s_k) dt s_k \geq m_1 \|s_k\|^2$$

holds for all $k \geq K$. Then the conclusion of the lemma follows from the convergence of the sequence $\{x_k\}$ to x^*.

The lemma shows that when the iteration sequence $\{x_k\}$ converges to a strong local minimizer x^*, the condition

$$s_k^T \hat{y}_k = (1 + \frac{\theta_k}{s_k^T y_k}) s_k^T y_k > 0$$

which is required to keep positive definite updates, will be satisfied for sufficiently large k. However, the case $s_k^T \hat{y}_k \leq 0$ can occur whether line searches are exact or inexact when x_k is remote from x^*, though it is rare in practical calculations (see [32]). Strategies are required to maintain the positive definite updates for the modified BFGS update. Here we put a restriction on the value of θ_k to ensure positive definite updates, that is,

$$\theta_k = \max\{(\omega - 1) s_k^T y_k,\ 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k\}, \tag{2.7}$$

where $\omega \in (0, 1)$. This strategy ensure $s_k^T \hat{y}_k \geq \omega s_k^T y_k$ and θ_k takes the value in (1.8) for sufficiently large k.

In the next two sections, we will consider convergence properties of the following particular multi-directional parallel quasi-Newton algorithms:

Algorithm B1.

$I_k = I_{k1}$ in step 5 and $\phi_{kj} = 0$, i.e., the BFGS update, for all k in step 6.

Algorithm B2.

$I_k = I_{k2}$ in step 5, and the BFGS update, for all k in step 6.

Algorithm C1.
$I_k = I_{k1}$ in step 5, and $\phi_{kj} = \theta_k$, i.e., the modified BFGS update, for all k in step 6.
Algorithm C2.
$I_k = I_{k2}$ in step 5, and the modified BFGS update, for all k in step 6.

3. Convergence of Algorithms B1 and B2

In this section we will study the global convergence properties of Algorithms B1 and B2 using the way that is presented by Byrd and Nocedal in [6]. It is assumed that the function $f(x)$ is twice continuously differentiable.
Assumption 3.1. *The level set $D = \{ x \in R^n \ : \ f(x) \leq f(x_1) \}$ is convex, and there exist positive constants $M \geq m > 0$ such that*

$$m\|z\|^2 \leq z^T G(x) z \leq M\|z\|^2$$

holds for all $z \in R^n$ and all $x \in D$.

Assumption 3.1 implies that f has a unique minimizer x^* in D, and that the following inequalities

$$m\|s_k\|^2 \leq s_k^T y_k \leq M\|s_k\|^2, \tag{3.1}$$

$$\frac{\|y_k\|^2}{s_k^T y_k} \leq M, \tag{3.2}$$

$$\frac{1}{2} m\|x_k - x^*\|^2 \leq f(x_k) - f(x^*) \leq \frac{1}{m}\|g_k\|^2 \tag{3.3}$$

hold for all $x_k,\ x_{k+1} \in D$, $x_k \neq x_{k+1}$. Under assumption 3.1, the Wolfe line search conditions (2.2) and (2.3) imply that there is a constant $\eta > 0$ such that

$$f(x_k + \alpha_{ki} d_{ki}) - f(x_k) \leq -\eta \frac{(g_k^T d_{ki})^2}{\|d_{ki}\|^2} \tag{3.4}$$

holds for all descent direction d_{ki} (see [28]). Thus, we have

$$f(x_k) - f(x_{k+1}) \geq \eta \frac{(s_k^T g_k)^2}{\|s_k\|^2 \|g_k\|^2} \|g_k\|^2.$$

Let

$$\cos\gamma_k = \frac{-g_k^T s_k}{\|g_k\| \|s_k\|},$$

then using (3.3) we obtain

$$f(x_{k+1}) - f(x^*) \leq (1 - \eta m \cos^2\gamma_k)[f(x_k) - f(x^*)],$$

where γ_k is the angle between the steepest descent direction $-g_k$ and the step s_k, and the decreasing property of the sequence $\{f(x_k)\}$ implies $0 \leq 1 - \eta m \cos^2\gamma_k \leq 1$. Therefore, if there is a subsequence of the iterates $\{x_k\}$ for which $\cos\gamma_k$ are bounded away from zero, called *good iterates*, then the conclusion that the sequence $\{x_k\}$ converges to x^* follows.

In the general BFGS method, the step s_k is generated in the BFGS search direction $d_k = -B_k^{-1} g_k$ $(s_k = \alpha_k d_k)$, and

$$\cos\gamma_k = \frac{-g_k^T s_k}{\|g_k\| \|s_k\|} = \frac{s_k^T B_k s_k}{\|B_k s_k\| \|s_k\|}.$$

Byrd and Nocedal in Theorem 2.1 of [6] employ the function

$$\psi(B_k) = \mathrm{tr}(B_k) - \ln\,\det(B_k),$$

to get the result that the most of the iterates in the general BFGS method are good iterates, where $\mathrm{tr}(B_k)$ and $\det(B_k)$ are the trace and determinant of the matrix B_k. Moreover

$$\psi(B_k) = \sum_{i=1}^{n}[\lambda_i(B_k) - \ln(\lambda_i(B_k))], \tag{3.5}$$

where $\lambda_i(B_k)$ are eigenvalues of the matrix B_k. So $\psi(B_k) > 0$ if B_k is positive definite. For the BFGS updating formula, we have

$$\psi(B_{k+1}) \quad = \quad \psi(B_k) + \frac{\|y_k\|^2}{s_k^T y_k} - 1 - \ln\frac{s_k^T y_k}{s_k^T s_k} + \ln r_k^2 + \varphi(t_k), \tag{3.6}$$

$$\tag{3.7}$$

where

$$\varphi(t) = 1 - t + \ln t, \qquad t_k = \frac{q_k}{r_k^2},$$

$$r_k = \frac{s_k^T B_k s_k}{\|B_k s_k\|\|s_k\|}, \quad q_k = \frac{s_k^T B_k s_k}{s_k^T s_k}.$$

Note that the function $\varphi(t)$ is non-positive for all $t > 0$ and $\varphi(t) \to 0$ as $t \to 1$.

In fact, Under conditions (3.1) and (3.2), Theorem 2.1 in [6] also holds for the sequences $\{x_k\}$ and $\{B_k\}$ generated in algorithms B1 and B2.

Lemma 3.1. *Let assumption 3.1 hold, B_1 be symmetric positive definite and $g_k \neq 0$ for all $k \geq 1$. Then the sequence $\{B_k\}$ generated in algorithms B1 and B2 is well-defined, and for any $\delta \in (0,1)$ there exist positive constants c_1, c_2, and $c_3 > 0$ such that for any $k > 1$, the following inequalities*

$$\frac{s_j^T B_j s_j}{\|B_j s_j\|\|s_j\|} \geq c_1, \tag{3.8}$$

$$c_2 \leq \frac{s_j^T B_j s_j}{s_j^T s_j} \leq c_3, \tag{3.9}$$

$$c_2 \leq \frac{\|B_j s_j\|}{\|s_j\|} \leq c_3/c_1 \tag{3.10}$$

hold for at least $[\delta k]$ values of $j \in [1, k]$.

Note that $r_j = s_j^T B_j s_j/(\|B_j s_j\|\|s_j\|)$ is no longer the cosine of the angle between the steepest descent direction $-g_j$ and the step s_j, since the step s_j may be generated from a search direction different from the BFGS search direction.

With Lemma 3.1 the convergence results for algorithms B1 and B2 follow.

Theorem 3.2. *Let x_1 be a starting point for which assumption 3.1 is satisfied. Then for any symmetric positive definite matrix B_1, the sequence $\{x_k\}$ generated in algorithm B1 converges to x^* at a linear rate. Moreover,*

$$\sum_{k=1}^{\infty} \|x_k - x^*\| < \infty, \tag{3.11}$$

and there is a constant $0 \leq r_1 < 1$ such that

$$f(x_{k+1}) - f(x^*) \leq r_1^k [f(x_1) - f(x^*)] \tag{3.12}$$

holds for all $k \geq 1$.

Proof. Let J be the set of indices for which (3.8), (3.9) and (3.10) are satisfied. Then for iteration $j \in J$, from (3.4), (3.8) and the definition (2.5) of the subset I_{k1}, we have

$$\begin{aligned}
f(x_j) - f(x_{j+1}) &= f(x_j) - f(x_j + \alpha_{jl} d_{jl}) \\
&\geq \eta \frac{(g_j^T d_{jl})^2}{\|g_j\|^2 \|d_{jl}\|^2} \|g_j\|^2 \\
&= \eta \frac{({d_{jl}}^T B_j(\phi_{j-1,l}) d_{jl})^2}{\|B_j(\phi_{j-1,l}) d_{jl}\|^2 \|d_{jl}\|^2} \|g_j\|^2 \\
&\geq \eta\beta \frac{({d_{jl}}^T B_j d_{jl})^2}{\|B_j d_{jl}\|^2 \|d_{jl}\|^2} \|g_j\|^2 \\
&= \eta\beta \frac{(s_j^T B_j s_j)^2}{\|B_j s_j\|^2 \|s_j\|^2} \|g_j\|^2 \\
&\geq \eta\beta c_1^2 \|g_j\|^2 = \xi \|g_j\|^2, \qquad (3.13)
\end{aligned}$$

where $s_j = \alpha_{jl} d_{jl}$ is determined in step 5 of the algorithm, and $B_j(\phi_{j-1,l})$ is the matrix that generates the direction d_{jl}. From (3.13) and (3.3) we obtain that for all $j \in J$

$$f(x_{j+1}) - f(x^*) \leq r_1^{\frac{1}{\delta}} [f(x_j) - f(x^*)],$$

where $r_1^{\frac{1}{\delta}} = (1 - \xi m) \geq 0$, because $\{f(x_j)\}$ is a decreasing sequence. Since there exist at least $[\delta k]$ indices in $J \cap [1, k]$, it follows from the decreasing of the sequence $\{f(x_j)\}$ that

$$f(x_{k+1}) - f(x^*) \leq r_1^k [f(x_1) - f(x^*)]$$

holds for all $k \geq 1$. This gives (3.12). Then from (3.3) we have

$$\begin{aligned}
\sum_{k=1}^{\infty} \|x_k - x^*\| &\leq (\frac{2}{m})^{\frac{1}{2}} \sum_{k=1}^{\infty} [f(x_k) - f(x^*)]^{\frac{1}{2}} \\
&\leq [\frac{2(f(x_1) - f(x^*))}{m}]^{\frac{1}{2}} \sum_{k=0}^{\infty} (r_1^{\frac{1}{2}})^k < \infty,
\end{aligned}$$

which gives (3.11) and shows the convergence of the sequence $\{x_k\}$ to the point x^*. The linear convergence rate comes from (3.12).

As for the convergence of Algorithm B2, we have the following result.

Theorem 3.3. *Let x_1 be a starting point for which assumption 3.1 is satisfied. Then for any symmetric positive definite matrix B_1, the sequence $\{x_k\}$ generated in algorithm B2 converges to x^* at a linear rate. Moreover, the formula (3.11) holds and (3.12) with $0 \leq r_1 < 1$ is satisfied for all $k \geq 1$.*

Proof. Let J be the set defined in the proof of Theorem 3.2, and $\{\hat{x}_k\}$ be an auxiliary sequence that is generated from the BFGS direction in algorithm B2, that is,

$$\hat{x}_{k+1} = x_k + \alpha_{k1} d_{k1},$$

where $d_{k1} = -B_k^{-1} g_k$ denotes the BFGS search direction in algorithm B2. Then it follows from

$f(x_{j+1}) - f(x_j) \le f(\hat{x}_{j+1}) - f(x_j)$, (2.6), (3.4) and (3.8) that for iteration $j \in J$ we have

$$\begin{aligned}
f(x_j) - f(x_{j+1}) &\ge f(x_j) - f(\hat{x}_{j+1}) = f(x_j) - f(x_j + \alpha_{j1} d_{j1}) \\
&\ge \eta \frac{(g_j^T d_{j1})^2}{\|d_{j1}\|^2 \|g_j\|^2} \|g_j\|^2 \\
&= \eta \frac{(d_{j1}^T B_j d_{j1})^2}{\|B_j d_{j1}\|^2 \|d_{j1}\|^2} \|g_j\|^2 \\
&\ge \eta\beta \frac{(d_{jl}^T B_j d_{jl})^2}{\|B_j d_{jl}\|^2 \|d_{jl}\|^2} \|g_j\|^2 \\
&= \eta\beta \frac{(s_j^T B_j s_j)^2}{\|B_j s_j\|^2 \|s_j\|^2} \|g_j\|^2 \\
&\ge \eta\beta c_1^2 \|g_j\|^2.
\end{aligned}$$

The reset of the proof is the same as those in Theorem 3.2.

4. Convergence of Algorithms C1 and C2

In this section we will present the global convergence results for algorithms C1 and C2. The processes of proving these results are the same as those in section 3. All we need is to show that the results of lemma 3.1 still hold when the vector $(1 + \frac{\theta_k}{s_k^T y_k}) y_k$ with safeguard strategy (2.7) is used to replace the vector y_k in the BFGS updating formula. Since Lemma 3.1 can be proved under inequalities (3.1) and (3.2), we give similar inequalities for the vector $\hat{y}_k = (1 + \frac{\theta_k}{s_k^T y_k}) y_k$.

Lemma 4.1. *Under assumption 3.1 and strategy (2.7), there exist positive constants $\hat{M} \ge \hat{m} > 0$ such that*

$$\begin{aligned}
&\hat{m}\|s_k\|^2 \le s_k^T \hat{y}_k \le \hat{M}\|s_k\|^2, \\
&\frac{\|\hat{y}_k\|^2}{s_k^T \hat{y}_k} \le \hat{M},
\end{aligned}$$

hold for all $x_k,\ x_{k+1} \in D$ and $x_k \ne x_{k+1}$.

Proof. Using the Taylor expansions of $f(x)$ and $g(x)$ at the point x_k

$$\begin{aligned}
&f(x_{k+1}) = f(x_k) + g_k^T s_k + \frac{1}{2} s_k^T G(x_k + \zeta s_k) s_k, \quad \zeta \in (0,1), \\
&g_{k+1}^T s_k = g_k^T s_k + s_k^T \int_0^1 G(x_k + t s_k) dt s_k,
\end{aligned}$$

we obtain

$$\theta_k \ge 3[s_k^T \int_0^1 G(x_k + t s_k) dt s_k - s_k^T G(x_k + \zeta s_k) s_k]. \tag{4.1}$$

The definition (2.7) of θ_k implies

$$1 + \frac{\theta_k}{s_k^T y_k} \ge \omega. \tag{4.2}$$

The definition of $\hat{y}_k$ gives

$$\frac{s_k^T \hat{y}_k}{s_k^T s_k} = \frac{(1 + \frac{\theta_k}{s_k^T y_k}) s_k^T y_k}{s_k^T s_k}, \quad \frac{\|\hat{y}_k\|^2}{s_k^T \hat{y}_k} = \frac{(1 + \frac{\theta_k}{s_k^T y_k}) \|y_k\|^2}{s_k^T y_k}.$$

Then from (3.1), (3.2), (4.1) and (4.2) we obtain

$$m\omega \leq \frac{s_k^T \hat{y}_k}{s_k^T s_k} \leq M(1 + 6\frac{M}{m}),$$
$$\frac{\|\hat{y}_k\|^2}{s_k^T \hat{y}_k} \leq M(1 + 6\frac{M}{m}).$$

Therefore the conclusion of the lemma hold with $\hat{m} = m\omega$ and $\hat{M} = M(1 + 6M/m)$.

With Lemma 4.1, global convergence results for algorithms C1 and C2 immediately follow. We give these results, but ignore their proofs.

Lemma 4.2. *Let B_1 be symmetric positive definite and $g_k \neq 0$ for all $k \geq 1$. Then the sequence $\{B_k = B_k(\theta_k)\}$ generated in algorithms C1 and C2 with strategy (2.7) is well-defined, and for any $\delta \in (0, 1)$ there exist positive constants $\hat{c}_1, \hat{c}_2$, and $\hat{c}_3 > 0$ such that for any $k > 1$, the following inequalities*

$$\frac{s_j^T B_j s_j}{\|B_j s_j\| \|s_j\|} \geq \hat{c}_1,$$
$$\hat{c}_2 \leq \frac{s_j^T B_j s_j}{s_j^T s_j} \leq \hat{c}_3,$$
$$\hat{c}_2 \leq \frac{\|B_j s_j\|}{\|s_j\|} \leq \hat{c}_3/\hat{c}_1$$

hold for at least $[\delta k]$ values of $j \in [1, k]$.

Theorem 4.3. *Let x_1 be an initial point for which assumption 3.1 is satisfied. Then for any symmetric positive definite matrix B_1, the sequence $\{x_k\}$ generated by algorithm C1 (or C2) with strategy (2.7) converges to x^* at a linear rate. Moreover, the formula (3.11) holds and there is a constant $0 \leq \hat{r}_1 < 1$ such that (3.12) with $\hat{r}_1$ replacing r_1 holds for all $k \geq 1$.*

5. The Super-linear Convergence Property

In this section we study the super-linear convergence property of the multi-directional parallel quasi-Newton algorithms. The super-linear convergence result of quasi-Newton methods is usually proved by showing that the search directions approach the Newton directions in both direction and length so that the step length of one is eventually taken for all iterates. This is characterized by the following theorem (see Theorem 6.4 of [13]).

Theorem 5.1. *Let $f(x)$ be twice continuously differentiable in an open set E and $\{x_k\} \subset E$ be a sequence that is generated by an iteration in form (1.2) with d_k a descent direction of $f(x)$ at x_k, and the step length α_k satisfies conditions (2.2) and (2.3). If the sequence $\{x_k\}$ converges to a point $x^* \in E$ at which $G(x^*)$ is positive definite, and if*

$$\lim_{k \to \infty} \frac{\|g_k + G_k d_k\|}{\|d_k\|} = 0, \tag{5.1}$$

then there is an integer K such that $\alpha_k = 1$ is acceptable to conditions (2.2) and (2.3) for all $k \geq K$. Moreover, $g(x^*) = 0$, and $\{x_k\}$ converges superlinearly to x^* if $\alpha_k = 1$ is taken for all suficiently large k.

For a general quasi-Newton method with a positive definite update, condition (5.1) is equivalent to the condition

$$\lim_{k \to \infty} \frac{\|(B_k - G_k)s_k\|}{\|s_k\|} = 0. \tag{5.2}$$

since $B_k d_k = -g_k$ and $s_k = \alpha_k d_k$. Moreover, if the Hessian matrix $G(x)$ of the function $f(x)$ is Lipschitz continuous at x^*, i.e., there exists a positive constant L such that

$$\|G(x) - G(x^*)\| \leq L\|x - x^*\| \tag{5.3}$$

holds for all x in a neighborhood of x^*, then following the convergence of the sequence $\{x_k\}$ to x^*, condition (5.2) is further equivalent to the condition

$$\lim_{k\to\infty} \frac{\|[B_k - G(x^*)]s_k\|}{\|s_k\|} = 0. \tag{5.4}$$

For the multi-directional parallel quasi-Newton methods, we define the sequence $\{B_k\}$ generated in step 6 of the algorithms as the main matrix sequence. Since the step $s_k = \alpha_{kl} d_{kl}$ may not be generated by the matrix B_k in the main sequence $\{B_k\}$, $B_k d_{kl} = -g_k$ may not hold. We consider an auxiliary sequence of matrices $\{\hat{B}_k\}$ which generates the direction d_{kl} in step 5 of the algorithms, that is,

$$\hat{B}_k = B_k(\phi_{k-1,l}), \quad \text{and} \quad \hat{B}_k d_{kl} = -g_k.$$

Then it is clear that if

$$\lim_{k\to\infty} \frac{\|[\hat{B}_k - G(x^*)]s_k\|}{\|s_k\|} = 0, \tag{5.5}$$

the super-linear convergence result follows if the initial step length is one in all line searches and is accepted whenever it satisfies Wolfe conditions.

Based on these analysis, the following further assumption on the function $f(x)$ and restrictions on the algorithms are made.

Assumption 5.1. *The Hessian $G(x)$ of $f(x)$ is Lipschitz continuous at $x^* \in D$, that is, condition (5.3) holds with a constant L.*

Restriction 5.1. *The values of ϕ_{ki}, $i = 1, 2, \cdots, p$ in step 3 are chosen for only positive definite updates, that is, $\phi_{ki} > \phi_{k0}$ (see section 1 for the expression of ϕ_{k0}). Since for a super-linearly convergent sequence $\{x_k\}$ with the sequence $\{\|B_k^{-1}\|\}$ bounded, $\phi_{k0} \to -\infty$ (see [8]), we give the values of ϕ_{ki} the following restriction*

$$\phi_{ki} \in [\nu_k, \varrho_1], \qquad \text{with} \qquad \nu_k = \max\{\phi_{k0} + \varrho_2, -\varrho_3\},$$

where $\varrho_1 \geq 1$, ϱ_2 and ϱ_3 are some positive constants. The modified BFGS update with strategy (2.7) is also included in step 3. As for the SR1 update, since $\hat{\phi}_k > \phi_{k0}$ is not guaranteed, we place the following modification to this update

$$\hat{\phi}_k = \max\{\frac{s_k^T y_k}{s_k^T(y_k - B_k s_k)}, \nu_k\}$$

so that $B_{k+1}(\hat{\phi}_k)$ keeps positive definite update.

Restriction 5.2. *The initial step length is one in all line searches and is accepted whenever it satisfies conditions (2.2) and (2.3).*

Using Theorem 3.2 of [6], we directly obtain the following result for the main sequences $\{B_k\}$ and $\{x_k\}$ generated in the multi-directional parallel quasi-Newton Algorithm B1 (or B2).

Theorem 5.2. *Let assumption 3.1 hold and B_1 be symmetric positive definite. Consider the main sequences $\{B_k\}$ and $\{x_k\}$ generated in algorithm B1 (or B2). If there is a positive sequence $\{\epsilon_k\}$ with*

$$\sum_{k=1}^{\infty} \epsilon_k < \infty \tag{5.6}$$

such that

$$\frac{\|y_k - G(x^*)s_k\|}{\|s_k\|} \leq \epsilon_k \tag{5.7}$$

holds for all $k \geq 1$, then

$$\lim_{k\to\infty} \frac{\|[B_k - G(x^*)]s_k\|}{\|s_k\|} = 0, \tag{5.8}$$

and the sequences $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded.

Proof. Define

$$\bar{s}_k = G^{*\frac{1}{2}} s_k, \qquad \bar{y}_k = G^{*-\frac{1}{2}} y_k, \qquad \bar{B}_k = G^{*-\frac{1}{2}} B_k G^{*-\frac{1}{2}},$$
$$\bar{r}_k = \frac{\bar{s}_k^T \bar{B}_k \bar{s}_k}{\|\bar{B}_k \bar{s}_k\|\|\bar{s}_k\|}, \qquad \bar{q}_k = \frac{\bar{s}_k^T \bar{B}_k \bar{s}_k}{\bar{s}_k^T \bar{s}_k}.$$

Then from (1.4) we have

$$\bar{B}_{k+1}(\phi) = \bar{B}_k - \frac{\bar{B}_k \bar{s}_k \bar{s}_k^T \bar{B}_k}{\bar{s}_k^T \bar{B}_k \bar{s}_k} + \frac{\bar{y}_k \bar{y}_k^T}{\bar{s}_k^T \bar{y}_k} + \phi(\bar{s}_k^T \bar{B}_k \bar{s}_k) \bar{u}_k \bar{u}_k^T,$$

$$\bar{u}_k = \frac{\bar{y}_k}{\bar{s}_k^T \bar{y}_k} - \frac{\bar{B}_k \bar{s}_k}{\bar{s}_k^T \bar{B}_k \bar{s}_k}.$$

For the BFGS update ($\phi = 0$) it follows from (3.6) that

$$\psi(\bar{B}_{k+1}) \quad = \quad \psi(\bar{B}_k) + \frac{\|\bar{y}_k\|^2}{\bar{s}_k^T \bar{y}_k} - 1 - \ln \frac{\bar{s}_k^T \bar{y}_k}{\bar{s}_k^T \bar{s}_k} + \ln \bar{r}_k^2 + \varphi(\bar{t}_k), \tag{5.9}$$

where $\bar{t}_k = \bar{q}_k / \bar{r}_k^2$. Using condition (5.7) we can obtain

$$\frac{\bar{s}_k^T \bar{y}_k}{\bar{s}_k^T \bar{s}_k} \geq 1 - c_4 \epsilon_k, \tag{5.10}$$

$$\frac{\|\bar{y}_k\|^2}{\bar{s}_k^T \bar{y}_k} \leq 1 + c_5 \epsilon_k, \tag{5.11}$$

where $c_4 = \|G^{*-\frac{1}{2}}\|^2$ and $c_5 > c_4$ are constants. It follows from these two inequalities and (5.9) that

$$0 \quad < \quad \psi(\bar{B}_{k+1}) \leq \psi(\bar{B}_1) + c_6 + \sum_{j=1}^{k} [3c_5 \epsilon_j + \ln \bar{r}_j^2 + \varphi(\bar{t}_j)],$$

where $c_6 > 0$ is a positive constant. Then condition (5.6) and non-positive properties of $\ln \bar{r}_j^2$ and $\varphi(\bar{t}_j)$ imply the boundedness of $\{\psi(\bar{B}_k)\}$, and

$$\ln \bar{r}_j^2 \to 0, \qquad \varphi(\bar{t}_j) \to 0. \tag{5.12}$$

The boundedness of $\{\psi(\bar{B}_k)\}$ and (3.5) imply that both sequences $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded. It follows from (5.12) and the properties of the function $\varphi(t) = 1 - t + \ln t,\ t > 0$ that

$$\lim_{k\to\infty} \bar{r}_k = \lim_{k\to\infty} \bar{q}_k = 1. \tag{5.13}$$

Since

$$\frac{\|G^{*-\frac{1}{2}}(B_k - G^*)s_k\|^2}{\|G^{*\frac{1}{2}} s_k\|^2} = \frac{\|\bar{B}_k \bar{s}_k\|^2 - 2\bar{s}_k^T \bar{B}_k \bar{s}_k + \bar{s}_k^T \bar{s}_k}{\bar{s}_k^T \bar{s}_k} = \frac{\bar{q}_k^2}{\bar{r}_k^2} - 2\bar{q}_k + 1,$$

conclusion (5.8) immediately follows from (5.13).

Now we present the super-linear convergence results for both algorithms B1 and B2.

Theorem 5.3. *Let assumptions 3.1 and 5.1 hold with a given initial point* x_1 *and the sequence* $\{x_k\}$ *be generated by algorithm B1 (or B2) with restrictions 5.1 and 5.2, and a given symmetric positive definite matrix* B_1. *Then the sequence* $\{x_k\}$ *converges super-linearly to* x^*, *and the sequences* $\{\|\hat{B}_k\|\}$ *and* $\{\|\hat{B}_k^{-1}\|\}$ *are bounded.*

Proof. We only need to show that limit (5.5) hold for both algorithms B1 and B2. It is clear that the global convergence results of Theorems 3.2 and 3.3 hold for both algorithms. Since the Hessian $G(x)$ is Lipschitz continuous at x^*, we have

$$\begin{aligned}\frac{\|y_k - G(x^*)s_k\|}{\|s_k\|} &= \frac{\|[\int_0^1 G(x_k + ts_k)dt - G(x^*)]s_k\|}{\|s_k\|} \\ &\leq L\max\{\|x_{k+1} - x^*\|,\ \|x_k - x^*\|\}.\end{aligned}$$

Let $\epsilon_k = L\max\{\|x_{k+1} - x^*\|,\ \|x_k - x^*\|\}$. Then it follows from (3.11) that conditions (5.6) and (5.7) in Theorem 5.2 hold, and hence conclusion (5.8) follows for the main matrix sequence $\{B_k\}$.

Since

$$\frac{\|[\hat{B}_k - G(x^*)]s_k\|}{\|s_k\|} \leq \frac{\|[B_k - G(x^*)]s_k\|}{\|s_k\|} + \frac{\|[\hat{B}_k - B_k]s_k\|}{\|s_k\|}, \tag{5.14}$$

consider the second term in the right-hand side. If $\hat{B}_k$ is obtained from B_{k-1} with the modified BFGS updating formula, then

$$\frac{\|[\hat{B}_k - B_k]s_k\|}{\|s_k\|} \leq \frac{|\theta_{k-1}|}{s_{k-1}^T y_{k-1}} \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}}.$$

It follows from Lemma 2.1, (3.2) and the convergence of the sequence $\{x_k\}$ to x^* that there is a positive constant c_7 such that

$$\frac{\|[\hat{B}_k - B_k]s_k\|}{\|s_k\|} \leq c_7\epsilon_k \tag{5.15}$$

hold. If $\hat{B}_k$ is obtained from an updating formula in the Broyden family with $\phi_{k-1,l} \in [\nu_k, \varrho_1]$, then

$$\frac{\|[\hat{B}_k - B_k]s_k\|}{\|s_k\|} \leq |\phi_{k-1,l}(s_{k-1}^T B_{k-1}s_{k-1})|\|u_{k-1}\|^2. \tag{5.16}$$

From the definition of $\bar{u}_k$ we have

$$\|\bar{u}_k\|^2 = \frac{1}{\|\bar{s}_k\|^2}[\frac{\|\bar{y}_k\|^2}{\bar{s}_k^T\bar{y}_k}\frac{\|\bar{s}_k\|^2}{\bar{s}_k^T\bar{y}_k} - 2\frac{1}{\bar{q}_k}\frac{\bar{s}_k^T\bar{B}_k\bar{y}_k}{\bar{s}_k^T\bar{y}_k} + \frac{1}{\bar{r}_k^2}].$$

Since

$$y_k = G(x^*)s_k + \int_0^1 [G(x_k + ts_k) - G(x^*)]\mathrm{d}ts_k,$$

we have

$$\bar{y}_k = \bar{s}_k + \bar{E}_k\bar{s}_k,$$

where

$$\|\bar{E}_k\| = \|G(x^*)^{-\frac{1}{2}}\int_0^1 [G(x_k + ts_k) - G(x^*)]\mathrm{d}tG(x^*)^{-\frac{1}{2}}\| \leq c_4^2\epsilon_k.$$

Thus we have

$$\frac{\bar{s}_k^T\bar{B}_k\bar{y}_k}{\bar{s}_k^T\bar{y}_k} = \frac{\bar{s}_k^T\bar{s}_k}{\bar{s}_k^T\bar{y}_k}\frac{\bar{s}_k^T\bar{B}_k\bar{s}_k + \bar{s}_k^T\bar{B}_k\bar{E}_k\bar{s}_k}{\bar{s}_k^T\bar{s}_k}.$$

Then from (5.10), (5.11), and the boundedness of the sequence $\{\|B_k\|\}$, there is a constant $c_8 > 0$ such that

$$\|\bar{u}_k\|^2 \leq \frac{1}{\|\bar{s}_k\|^2}(1 - \frac{1}{\bar{r}_k^2} + c_8\epsilon_k). \tag{5.17}$$

Then from (5.14), (5.15), (5.16) and (5.17) we obtain

$$\begin{aligned}\frac{\|[\hat{B}_k - G(x^*)]s_k\|}{\|s_k\|} \quad \leq \quad & \frac{\|[B_k - G(x^*)]s_k\|}{\|s_k\|} + \max\{c_7\epsilon_{k-1}, \\ & |\phi_{k-1,l}|\|G(x^*)^{\frac{1}{2}}\|^2 \frac{\bar{s}_{k-1}^T\bar{B}_{k-1}\bar{s}_{k-1}}{\bar{s}_{k-1}^T\bar{s}_{k-1}}[1 - \frac{1}{\bar{r}_{k-1}^2} + c_8\epsilon_{k-1}]\}\end{aligned}$$

It then follows from (5.8), (5.13), the boundedness of the value $\phi_{k-1,l}$ and the sequence $\{\|B_k\|\}$ that limit (5.5) holds for the matrix sequence $\{\hat{B}_k\}$, and hence the superlinear convergence result holds for both algorithms.

Finally, since

$$\begin{aligned}\|\hat{B}_k\| \quad \leq \quad & \|B_k\| + \max\{\frac{|\theta_{k-1}|\|y_{k-1}\|^2}{(s_{k-1}^T y_{k-1})^2}, |\phi_{k-1,l}|s_{k-1}^T B_{k-1}s_{k-1}\|u_{k-1}\|^2\} \\ \leq \quad & \|B_k\| + \max\{c_7\epsilon_{k-1}, |\phi_{k-1,l}|\|G(x^*)^{\frac{1}{2}}\|^2 \frac{\bar{s}_{k-1}^T\bar{B}_{k-1}\bar{s}_{k-1}}{\bar{s}_{k-1}^T\bar{s}_{k-1}}(1 - \frac{1}{\bar{r}_{k-1}^2} + c_8\epsilon_{k-1})\},\end{aligned}$$

the boundedness of the sequence $\{\|\hat{B}_k\|\}$ follows from the boundedness of the sequence $\{\|B_k\|\}$ and the convergence of $\bar{r}_k$ to one and ϵ_k to zero. The boundedness of the sequence $\{\|\hat{B}_k^{-1}\|\}$ can be similarly proved using equation (2.4), the boundedness of the sequence $\{\|B_k^{-1}\|\}$ and the convergence of the sequences $\{\bar{r}_k\}$ and $\{\epsilon_k\}$ to one and zero, respectively.

As for algorithms C1 and C2, it follows from (3.1), (4.1) and (5.3) that

$$\begin{aligned}\frac{\|\hat{y}_k - G(x^*)s_k\|}{\|s_k\|} \quad \leq \quad & \frac{\|y_k - G(x^*)s_k\|}{\|s_k\|} + \frac{|\theta_k|\|y_k\|}{s_k^T y_k\|s_k\|} \\ \leq \quad & (1 + \frac{6M}{m})L\max\{\|x_{k+1} - x^*\|,\ \|x_k - x^*\|\}.\end{aligned}$$

Thus, conditions (5.6) and (5.7) hold with $\epsilon_k = (1+6M/m)L\max\{\|x_{k+1}-x^*\|, \|x_k-x^*\|\}$, which derives the limit (5.8) for the main sequence $\{B_k\}$ in algorithms C1 and C2, i.e., the results of Theorem 5.2 also hold for both algorithms C1 and C2. Then the superlinear convergence results for both these algorithms can be obtained using the same dementration as that used in the proof of Theorem 5.3.

Theorem 5.4. *Let assumptions 3.1 and 5.1 hold with a given initial point x_1 and the sequences $\{x_k\}$, $\{B_k\}$ and $\{\hat{B}_k\}$ be generated by algorithm either C1 or C2 with strategy (2.7), restrictions 5.1 and 5.2, and a symmetric positive definite matrix B_1. Then limits (5.4) and (5.5) hold, and the sequence $\{x_k\}$ converges superlinearly to the unique local minimizer x^* of $f(x)$ in D, and the sequences $\{\|\hat{B}_k\|\}$ and $\{\|\hat{B}_k^{-1}\|\}$ are bounded.*

Theorems 5.3 and 5.4 indicate that when the sequence $\{x_k\}$ converges to x^*, all the search directions generated in step 3 by different quasi-Newton updates with $\phi_{ki} \in [\nu_k, \varrho_1]$ approach simultaneously the Newton direction in both direction and length. This coincides with the numerical performance of these algorithms. Though these algorithms often take either the BFGS steps or the SR1 steps for final iterations, the differences between the steps obtained in different search directions are very small at these final iterations.

6. Conclusions

A class of multi-directional parallel quasi-Newton algorithms for the solution of unconstrained minimization problems is presented in the paper. At each iteration, these algorithms generate several parallel quasi-Newton directions, and then apply line searches along each direction, simultaneously. The next iterate is obtained as the "best" point from the results of parallel searches. Different quasi-Newton updating formulas in Broyden family can be used to generate a main sequence of Hessian matrix approximations. Based on the BFGS and the modified BFGS updating formulae, the global and superlinear convergence properties are proved. It can be observed from the proof of these convergence results that all the quasi-Newton directions approach the Newton direction in both direction and length when the sequence $\{x_k\}$ converges to the local minimum x^* of the function $f(x)$.

Numerical results (see [22]) for a broad class of middle and large scale test problems show that these algorithms are efficient and robust in solving practical optimization problems, particularly for large scale problems, and that a benefit of 300% reduction, or more, in terms of the number of iterations and function/gradient evaluations can be obtained over the serial BFGS method. It has been observed that compared with the serial BFGS method, a spedup factor of up to 28 times can be achieved by these algorithms in solving certain large scale optimization problems.

References

[1] M. Al-Baali, Variational quasi-Newton methods for unconstrained optimization, *Journal of Optimization Theory and Application*, **77** (1993), 127-143.

[2] M.C. Biggs, A note on minimization algorithms which make use of non-quadratic properties of the objective function, *J. Inst. Maths. Applics.*, **12** (1973), 337-338.

[3] C.G. Broyden, Quasi-Newton methods and their application to function minimization, *Math. Comp.*, **21** (1967), 368-381.

[4] C.G. Broyden, The convergence of a class double-rank minimization algorithms, *J. Inst. Maths. Applics.*, **6** (1970), 76-90.

[5] C.G. Broyden, J. E. Dennis and J. J. Moré, On the local and superlinear convergence of quasi-Newton methods, *J. Inst. Maths. Applics.*, **12** (1973), 223-245.

[6] R.H. Byrd and J. Nocedal, A tool for the analysis of quasi-Newton methods with applications to unconstrained minimization, *SIAM J. Numer. Anal.*, **26** (1989), 727-739.

[7] R.H. Byrd, J. Nocedal and Yuan Yaxian, Global convergence of a class of quasi-Newton methods on convex problems, *SIAM J. Numer. Anal.*, **24** (1987), 1171-1190.

[8] R.H. Byrd, D. Liu and J. Nocedal, On the behavior of Broyden's class of quasi-Newton methods, *SIAM J. Optimization*, **2** (1992), 533-557.

[9] A.R. Conn, N. Gould and Ph.L. Toint, Convergence of quasi-Newton matrices generated by the symmetric rank one update, *Math. Prog.*, **50** (1991), 177-196.

[10] W.C. Davidon, Variable metric methods for minimization, AEC Res. and Dev. Report ANL-5990, 1959.

[11] Nai-yang Deng and Zheng-feng Li, Convergence properties of a modified BFGS algorithm for minimization with Armijio-Goldstein step lengths, *J. Compt. Math.*, **17** (1999), 645-652.

[12] J.E. Dennis and J.J. Moré, A characterization of superlinear convergence and its application to quasi-Newton methods, *Mathematics of Computation*, **28** (1974), 549-560.

[13] J.E. Dennis and J.J. Moré, Quasi-Newton methods, motivation and theory, *SIAM Review*, **19** (1977), 46-89.

[14] R. Fletcher, *Practical Methods of Optimization V.1, Unconstrained Optimization*, John Wily & Sons New York, 1980.

[15] R. Fletcher, A new approach to variable metric algorithms, *Computer J.*, **13** (1970), 392-399.

[16] R. Fletcher and M.J.D. Powell, A rapidly convergent descent method for minimization, *Computer J.*, **6** (1963), 163-168.

[17] D. Goldfarb, A family of variable metric methods derived by variation means, *Math. Comput.*, **24** (1970), 23-26.

[18] F. Khalfan, R.H. Byrd and R.B. Schnabel, A theoretical and experimental study of the symmetric rank-one update, *SIAM J. Optimization*, **3** (1993), 1-24.

[19] W.D. Lin and C.X. Xu, Global convergence properties of convex Broyden quasi-Newton methods based on the new quasi-Newton equations, Presented at the International Conference on Nonlinear Programming and Variational Inequalities, City University of Hong Kong, December, 1998.

[20] P.K.H. Phua, Multi-directional parallel algorithms for unconstrained optimization, *Optimization*, **38** (1996), 107-125.

[21] P.K.H. Phua and S.B.W. Chew, Symmetric rank-one update and quasi-Newton methods, In: *Optimization Techniques and Applications*, Proceedings of the International Conference on Optimization Techniques and Applications, P. K. H. Phua *et al.*, eds., World Scientific, Singapore, (1992), 52-63.

[22] P.K.H. Phua, W.G. Fan and Y.L. Zeng, Parallel algorithms for large-scale nonlinear optimization, *Int. Trans. Opl. Res.*, **5** (1998), 67-77.

[23] P.K.H. Phua, W.G. Fan and Y.L. Zeng, Self-scaling parallel quasi-Newton methods, In: *Optimization Techniques and Applications*, Proceedings of the International Conference on Optimization Techniques and Applications,

[24] M.J.D. Powell, Some global convergence properties of a variable metric algorithm for minimization without exact line searches, In: *Nonlinear Programming*, SIAM-AMS Proceedings V.IX (R.W. Cotte and C.E. Lemke eds), SIAM, Philadephia, 1976.

[25] D.G. Pu, S.H. Gui and W.W. Tian, A class of revised Broyden algorithms without exact line search, *J. Comput. Math.*, **22** (2004), 11-20.

[26] K. Ritter, Global and superlinear convergence of a class of variable metric methods, *Math. Prog. Study*, **14** (1981), 178-205.

[27] D.F. Shanno, Conditioning of quasi-Newton methods for function minimization, *Math. Comput.*, **24** (1970), 647-656.

[28] J. Werner, Uber die global konvergenz von variable-metric verfahren mit nichtexakte schrittweitenbestimmung, *Numer. Math.*, **31** (1978), 321-334.

[29] C.X. Xu and J.Z. Zhang, A survey of Quasi-Newton Equations and Quasi-Newton Methods for Optimization, *Annals of Operations Research*, **103** (2001), 213-234.

[30] Y. Yuan, A modified BFGS algorithm for unconstrained optimization, *IMA J. Numerical Analysis*, **11** (1991), 325-332.

[31] Y. Yuan and R.H. Byrd, Non-quasi-Newton updates for unconstrained optimization, *J. Comput. Math*, **13** (1995), 95-107.

[32] J.Z. Zhang and C.X. Xu, Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations, *Journal of Computational and Applied Mathematics*, **137** (2001), 269-278.