<table>
<thead>
<tr>
<th>Title</th>
<th>FACTORIZED QUASI-NEWTON METHODS FOR NONLINEAR LEAST SQUARES PROBLEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Takahashi, Toshihiko; Yabe, Hiroshi</td>
</tr>
<tr>
<td>Citation</td>
<td>数理解析研究所講究録 (1988), 676: 170-194</td>
</tr>
<tr>
<td>Issue Date</td>
<td>1988-12</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/2433/100963">http://hdl.handle.net/2433/100963</a></td>
</tr>
<tr>
<td>Type</td>
<td>Departmental Bulletin Paper</td>
</tr>
<tr>
<td>Textversion</td>
<td>publisher</td>
</tr>
</tbody>
</table>
FACTORIZED QUASI-NEWTON METHODS FOR NONLINEAR LEAST SQUARES PROBLEMS
（非線形二乗問題に対する分解型ニュートン法）

鹿島建設（株）・情報システム部 高橋俊彦（Toshihiko Takahashi）
東京理科大学・工学部 矢部 博（Hiroshi Yabe）

1. Introduction

We consider the following nonlinear least squares problem:

\begin{equation}
\text{Minimize } F(x) = \frac{1}{2} \sum_{j=1}^{m} (f_j(x))^2, \quad x \in \mathbb{R}^n, \quad m \geq n,
\end{equation}

where each \( f_j : \mathbb{R}^n \to \mathbb{R} \) is twice continuously differentiable. This problem is extremely important in many fields of mathematical programming applications, e.g. maximum likelihood estimations, nonlinear data fitting or parameter estimation, respectively.

Most iterative methods for the above problem are variants of Newton’s method. At the \( k \)-th iteration of Newton’s method, the search direction \( d_k \) is computed by

\begin{equation}
\nabla^2 F(x_k) d_k = -\nabla F(x_k),
\end{equation}

and the new point is generated by

\begin{equation}
x_{k+1} = x_k + d_k.
\end{equation}

Here \( x_k \) is the current estimate of the minimum point \( x^* \), and \( \nabla F, \nabla^2 F \) are the gradient vector and the Hessian matrix of \( F \), respectively, and are given by

\begin{align}
\nabla F(x) &= J(x)^T f(x), \\
\nabla^2 F(x) &= J(x)^T J(x) + \sum_{j=1}^{m} f_j(x) \nabla^2 f_j(x),
\end{align}

where

\begin{equation}
f(x) = (f_1(x), \ldots, f_m(x))^T,
\end{equation}

and \( J \) is the \( m \times n \) Jacobian matrix of \( f \), and the symbol "^T" denotes the transpose of a vector or a matrix.

Since the cost of providing the complete Hessian matrix is often expensive, some methods have been derived which use only the first derivative information. For example, the Gauss-Newton method and the Levenberg-Marquardt method are well known. These methods neglect the second part of the Hessian matrix of \( F \), so they can be expected to perform well when the residuals at \( x^* \) are small or each \( f_j \) is close to linear. However,
they can be much less efficient when the neglected part of $\nabla^2 F(x)$ is not small compared with $J(x)^T J(x)$ in the sense of Meyer's result [9].

On the other hand, quasi-Newton approximations to the second part of the Hessian matrix have been considered [7]. Recently, two robust algorithms have been proposed by Bartholomew-Biggs[1] and Dennis, Gay and Welsch [8]. These methods are shown in Section 2. Our approach is based on the idea of structured quasi-Newton updating which utilizes the structure of the Hessian matrix of $F$. The main purpose of this paper is to obtain descent search directions for the objective function, which may enable us to establish global convergence property under suitable conditions. Subsequently, to accomplish the above desirable property, we propose factorized versions of the structured quasi-Newton methods and derive various types of factorized quasi-Newton updating formulae in Section 3. Consequently, in Section 4, we prove the local and q-superlinear convergence of our algorithms. Finally, some computational experiments are given in order to show that our methods are comparable to other effective methods.

Throughout this paper, the norm $\| \cdot \|$ denotes the 2-norm for vectors and matrices. For any matrix $Q$ and a nonsingular $M$, $\| Q \|_F$ and $\| Q \|_{F,M}$ denote the Frobenius and the weighted Frobenius norms of $Q$, respectively, and are defined by

$$\| Q \|_F = \left( \text{Trace}(QQ^T) \right)^{1/2} \quad \text{and} \quad \| Q \|_{F,M} = \| M^T Q M \|_F.$$  

For a symmetric positive definite matrix $Q$, $Q^{1/2}$ denotes the symmetric matrix which satisfies $(Q^{1/2})^2 = Q$.


A straightforward application of general quasi-Newton methods to the nonlinear least squares problem is not desirable, because these methods approximate all of the Hessian matrix. Since the nonlinear least squares algorithms usually calculate the Jacobian matrix $J(x)$ analytically or numerically, the portion $J(x)^T J(x)$ of $\nabla^2 F(x)$ is always readily available, so we only have to approximate the second part of $\nabla^2 F(x)$. Therefore, for the nonlinear least squares problem, it has been considered that the search direction can be computed by

$$\left( J_k^T J_k + A_k \right) d_k = -J_k^T f_k,$$

where $f_k = f(x_k)$, $J_k = J(x_k)$, and the matrix $A_k$ is the $k$-th approximation to
the second part of the Hessian matrix of $F$ [7]. The matrix $A_k$ is updated such that the new matrix $A_{k+1}$ satisfies the secant condition

\begin{equation}
A_{k+1} s_k = u_k, \quad u_k = y_k - J_{k+1} f_{k+1} s_k
\end{equation}

or

\begin{equation}
A_{k+1} s_k = v_k, \quad v_k = (J_{k+1} - J_k)^T f_{k+1},
\end{equation}

where

\begin{equation}
s_k = x_{k+1} - x_k, \quad y_k = \nabla F_{k+1} - \nabla F_k, \quad \nabla F_k = \nabla F(x_k).
\end{equation}

The first is proposed by Broyden and Dennis (BD) [4], and the second is proposed by Bartholomew-Biggs and Dennis, Cay and Welsch (DGW) [8]. We call these strategies structured quasi-Newton methods.

Structured quasi-Newton updates are usually of rank one or of rank two. Broyden and Dennis gave the following update:

(i) the BD update

\begin{equation}
A_{k+1} = A_k + ((u_k - A_k s_k)s_k^T + s_k(s_k - A_k s_k)^T) / s_k^T s_k
\end{equation}

\[-((u_k - A_k s_k)s_k^T / (s_k^T s_k)^2) s_k^T s_k.

Recently, by using sizing techniques, Bartholomew-Biggs and Dennis et al. have proposed the robust algorithms for the both cases of large and small residual problems. When the residuals are large, their algorithms perform as well as the Broyden and Dennis method does. On the other hand, for the very small residual problems, their algorithms perform almost as well as the Gauss-Newton method does. Their updates are as follows:

(ii) the Biggs update

\begin{equation}
A_{k+1} = \beta_k A_k + (v_k - \beta_k A_k s_k)(v_k - \beta_k A_k s_k)^T / (v_k - \beta_k A_k s_k)^T s_k,
\end{equation}

\begin{equation}
\beta_k = 1 f_{k+1}^T f_k / f_k^T f_k 1,
\end{equation}

(iii) the DGW update

\begin{equation}
A_{k+1} = \beta_k A_k + ((v_k - \beta_k A_k s_k)y_k^T + y_k(v_k - \beta_k A_k s_k)^T) / s_k^T y_k
\end{equation}

\[-((s_k^T (v_k - \beta_k A_k s_k) / (s_k^T y_k)^2) y_k y_k^T).

\begin{equation}
\beta_k = \min(1 s_k^T y_k / s_k^T A_k s_k, 1),
\end{equation}

where $\beta_k$ is a sizing factor.

3. Factorized Versions of Structured Quasi-Newton Methods

In order to obtain a descent search direction, it is desirable that the
coefficient matrix in (2.1) is positive definite. However, it is not clear
how to construct updating formulae of $A_k$ such that the matrix $J_{k}^T J_k + A_k$ is
positive definite. To overcome this difficulty, several strategies have
been proposed, for example, the modified Cholesky decomposition of the matrix
$J_{k}^T J_k + A_k$, the Levenberg-Marquardt modification (the model/trust region
strategy) [8] or switching to the Gauss-Newton method [2].

In this section, a direct approach is proposed in order to maintain
positive definiteness of the coefficient matrix in (2.1). We try to compute
the search direction by solving the linear system of equations

\[(3.1) \quad (L_k + J_k)^T (L_k + J_k) d_k = -J_k^T f_k,\]

where the matrix $L_k$ is an $m \times n$ correction matrix to the Jacobian matrix
such that $L_k^T L_k + L_k^T J_k + J_k L_k$ is the $k$-th approximation to the second part of
the Hessian matrix of $F$ [10]. Since the coefficient matrix is expressed by
the factorized form, the search direction may be expected to be a descent
direction for $F$.

Now we construct updating formulae of the matrix $L_k$. The secant
condition (2.2) or (2.3) for $A_{k+1}$ can be reduced to the following secant
condition for $L_{k+1}$

\[(3.2) \quad (L_{k+1} + J_{k+1})^T (L_{k+1} + J_{k+1}) s_k = z_k,\]

where

\[(3.3) \quad z_k = y_k\]

or

\[(3.4) \quad z_k = v_k + J_{k+1}^T J_{k+1} s_k,\]

and the vectors $v_k$, $s_k$ and $y_k$ are given in (2.3) and (2.4), respectively.
It is easily shown that, for nonzero $s_k$ and $z_k$, the matrix equation (3.2) is
consistent if and only if

\[(3.5) \quad L_{k+1}^T h = z_k - J_{k+1}^T h \quad \text{and} \quad L_{k+1} s_k = h - J_{k+1} s_k\]

for some $m$-dimensional vector $h$. Further, the matrix equations (3.5) have a
common solution $L_{k+1}$ if and only if each equation separately has a solution
and $h^T h = s_k^T z_k$.

In the sequent two subsections, we find a rectangular matrix $L_{k+1}$
which satisfies the equations (3.5) under the assumption of $s_k^T z_k > 0$. 

-4-
3.1 Least-change secant updates of \( L_k \)

Since the equations (3.5) may not uniquely determine the solution matrix \( L_{k+1} \), we use the least-change secant update technique following to Dennis and Schnabel [8].

For any unknown \( m \)-dimensional vector \( h \) such that \( h^T h = s_k^T z_k \), minimizing the Frobenius norm

\[
\| L_{k+1}^T - L_k^T \|_F
\]

with respect to \( L_{k+1} \), subject to

\[
L_{k+1}^T h = z_k - J_{k+1}^T h,
\]

we have

\[
L_{k+1} = L_k + h(z_k - (L_k + J_{k+1})^T h)/h^T h.
\]

By substituting the above for the other condition in (3.5) and using \( h^T h = s_k^T z_k \), the vector \( h \) can be determined by the form

\[
h = (s_k^T z_k / s_k^T s_k)^{1/2}(L_k + J_{k+1}) s_k,
\]

where

\[
B_k^# = (L_k + J_{k+1})^T (L_k + J_{k+1}).
\]

Thus we have the rank one update of \( L_k \) as follows:

\[
L_{k+1} = L_k + ((L_k + J_{k+1}) s_k / s_k^T B_k^# s_k) (s_k^T s_k / s_k^T s_k)^{1/2} z_k - B_k^# s_k^T.
\]

Setting

\[
B_{k+1} = (L_{k+1} + J_{k+1})^T (L_{k+1} + J_{k+1}),
\]

we have

\[
B_{k+1} = B_k^# - B_k^# s_k s_k^T B_k^# / s_k^T B_k^# s_k + z_k z_k^T / s_k^T s_k,
\]

which is the analogy of the BFGS update. Note that (3.11) differs from the standard BFGS update in that the structure of the Hessian matrix (1.5) is included in \( z_k \) and \( B_k^# \).

Next, we consider an analogy of the DFP update. For any unknown \( m \)-dimensional vector \( h \), which satisfies \( h^T h = s_k^T z_k \), and the nonsingular matrices \( W_L \in \mathbb{R}^{m \times m} \), \( W_R \in \mathbb{R}^{n \times n} \), minimizing the Frobenius norm

\[
\| W_L (L_{k+1} - L_k) W_R \|_F
\]

with respect to \( L_{k+1} \), subject to
(3.13) \[ L_{k+1} s_k = h - J_{k+1} s_k, \]
we have
\[ L_{k+1} = L_k + (h - (l_k + J_{k+1}) s_k) s_k^T w / s_k^T w s_k, \]
where \( w = w_R s_k^T w_R^{-1}. \) If the symmetric positive definite matrix \( w \) is chosen such that \( w s_k = z_k, \) by substituting the above for the other condition in (3.5), we have
\[ (l_k + J_{k+1})^T h = c z_k, \quad c = s_k^T (l_k + J_{k+1}) h / s_k^T z_k. \]
If the matrix \( l_k + J_{k+1} \) is of full rank, then
\[ h = c ((l_k + J_{k+1})^T)^+ z_k, \]
where \( ((l_k + J_{k+1})^T)^+ \) is the Moore-Penrose generalized inverse of \( (l_k + J_{k+1})^T. \)

Setting \( h^T s_k = z_k, \) we have the rank one update of \( l_k \) as follows:
(3.14) \[ L_{k+1} = L_k + (l_k + J_{k+1})(s_k^T z_k / z_k (b_k^S-k^S) z_k^1/2 (b_k^S-k^S)^{-1} z_k - s_k (z_k / s_k z_k^T)). \]
Further, we have
(3.15) \[ b_{k+1} = b_k^S + (1 + s_k^T b_k^S s_k / s_k^T z_k) z_k z_k^T / s_k z_k - (b_k^S s_k^T z_k + z_k s_k^T b_k^S) / s_k^T z_k, \]
which is the analogy of the DFP update.

3.2 Least-change secant updates of \( l_k + J_k \)

In order to obtain the updating formulae which satisfy the secant condition (3.2), we minimized the norm (3.6) or (3.12) in the previous subsection. Instead of this strategy, we can minimize the norms
(3.16) \[ \| (l_k + J_k + J_{k+1})^T - (l_k + J_k)^T \|_F, \]
and
(3.17) \[ \| w_{l}(l_k + J_k) - (l_k + J_k) w_{l} \|_F. \]

Then, by the same way as the subsection 3.1, we have the following updates:
(i) the update corresponding to (3.9)
(3.18) \[ L_{k+1} = L_k + J_k - J_{k+1} + (l_k + J_k) s_k / s_k^T b_k s_k ((s_k^T b_k s_k / s_k^T z_k)^1/2 z_k - b_k s_k)^T, \]
(3.19) \[ b_{k+1} = b_k - b_k^S s_k^T b_k / s_k^T b_k s_k + z_k z_k^T / s_k^T z_k, \]
(ii) the update corresponding to (3.14)
(3.20) \[ L_{k+1} = L_k + J_k - J_{k+1} + (l_k + J_k) ((s_k^T z_k / z_k^T b_k^S-k^S z_k^1/2 b_k^S-k^S)^{-1} z_k - s_k (z_k / s_k z_k^T))^T, \]
(3.21) \[ b_{k+1} = b_k + (1 + s_k^T b_k s_k / s_k^T z_k) z_k z_k^T / s_k z_k^T - (b_k^S s_k^T z_k + z_k s_k^T b_k^S) / s_k^T z_k. \]
where $B_k = (L_k + J_k)^T(L_k + J_k)$. For the case of $z_k = y_k$, we have the standard BFGS and DFP updates, which are identical with the results of Dennis and Schnabel [6]. For the case of $z_k = v_k + J_{k+1}^T J_{k+1} s_k$, we have another updating formulae.

For (3.19) and (3.21), the inverse updating formulae can be obtained by letting $H_k = B_k^{-1}$. Then we have the inverse updates as follows:

(i) the inverse update corresponding to (3.19)

$$H_{k+1} = H_k + (1 + z_k^T H_k z_k / s_k^T z_k) s_k s_k^T H_k / s_k^T z_k,$$

(ii) the inverse update corresponding to (3.21)

$$H_{k+1} = H_k - s_k s_k^T / z_k^T H_k z_k + s_k s_k^T / z_k^T z_k.$$

In this case, the search direction can be calculated by

$$d_k = -H_k f_k,$$

without solving the linear system of equations.

3.3. Sizing of the updating matrix

We know that, for zero residual problems, the matrices $A_k$ and $L_k^T L_k + L_k^T J_k + J_k L_k$ should ideally converge to zero. If the matrices do not at least become small in those cases, then structured quasi-Newton methods cannot be hoped to compete with the Gauss-Newton method. Noting that the quasi-Newton updates do not generate the zero matrix, some remedies must be employed.

Among them, the sizing of the updating matrices which has been introduced by Bartholomew-Biggs [1] and Dennis et al. [8] seems most promising. The Biggs' sizing factor (2.7) is based on the idea such that if $f_{k+1} = f_k f_k$ for some $\beta_k A_k = \sum_{i=1}^m f_k^2 f_{k+1}^2$ and each $f_i$ is quadratic, then $\sum_{i=1}^m f_{k+1}^2 f_i^2 = \beta_k A_k$, where $f_{k+1}^2 f_i$ and $f_{k+1}^2 f_i^2$ denote $f_i(x_{k+i})$ and $\nabla^2 f_i(x_{k+i})$, respectively. Dennis et al. proposed the strategy such that the spectrum of the sized matrix $\beta_k A_k$ overlaps that of the second part of the Hessian matrix in the direction of $s_k$. They obtained the factor (2.9) by using the relation

$$l[s_k, \sum_{i=1}^m f_{k+1}^2 f_i^2 s_k / s_k^T ] s_k^T (\beta_k A_k) s_k / s_k^T s_k^{-1} l \approx l(s_k^T v_k) / s_k^T (\beta_k A_k) s_k l = 1,$$
where the vector \( v_k \) is defined in (2.3). The structured quasi-Newton methods with the sizing factors (2.7) and (2.9) are reasonable in the sense that if the function \( f_{k+1} \) becomes zero, then \( v_k = 0 \) and \( \beta_k = 0 \), so the new matrix \( A_{k+1} \) also becomes zero. This fact is based on using the condition (2.3).

Now we can use the above mentioned factors for our factorized versions. Then we have the following updates:

(i) the sized BFGS-type update

\[
L_{k+1} = \beta_k L_k + ((\beta_k L_k + J_{k+1})s_k / s_k^T B_k^# s_k) \left( (s_k^T B_k^# s_k / s_k^T z_k)^{1/2} z_k - B_k^# s_k \right)^T,
\]

(ii) the sized DFP-type update

\[
L_{k+1} = \beta_k L_k + ((\beta_k L_k + J_{k+1}) (s_k^T z_k / z_k^T (B_k^#)^{-1} z_k)^{1/2} (B_k^#)^{-1} z_k - s_k) (z_k / s_k^T z_k)^T,
\]

where \( z_k \) is given by (3.4), the factor \( \beta_k \) is the Biggs' sizing factor (2.7) and the matrix \( B_k^# \) is rewritten as

\[
B_k^# = (B_k^# L_k + J_{k+1})^T (B_k^# L_k + J_{k+1}).
\]

It is reasonable to use the above because if the function \( f_{k+1} \) becomes zero, then the new matrix \( L_{k+1} \) becomes zero, so we have the Gauss-Newton direction at the \((k+1)\)-th iteration. Since the DGW's sizing factor (2.9) contains the matrix \( A_k \), we can not use it directly. However, for the factorized version, the strategy similar to the DGW's one can be considered. The factor \( \beta_k \) should be chosen such that the matrix

\[
(\beta_k L_k)^T (\beta_k L_k) + (\beta_k L_k)^T J_{k+1} + J_{k+1}^T (\beta_k L_k)
\]

has the same spectrum as that of the second part of the Hessian matrix in the direction of \( s_k \). So we have the following relation

\[
s_k^T v_k / s_k^T \left[ (\beta_k L_k)^T (\beta_k L_k) + (\beta_k L_k)^T J_{k+1} + J_{k+1}^T (\beta_k L_k) \right] s_k = 1,
\]

which yields

\[
\beta_k = \{- (L_k s_k)^T J_{k+1} s_k + \text{sgn}(L_k s_k)^T J_{k+1} s_k \} \frac{\xi_k}{1/2} / \| L_k s_k \|_2^2,
\]

where

\[
\xi_k = \left( (L_k s_k)^T J_{k+1} s_k \right)^2 + \| L_k s_k \|_2^2 \| s_k^T v_k \|
\]

and the symbol \( \text{sgn}(\zeta) \) denotes the sign of \( \zeta \). In practice, it is also reasonable to use the sizing factor

\[
\beta_k = \min \{ 1 - (L_k s_k)^T J_{k+1} s_k + \text{sgn}(L_k s_k)^T J_{k+1} s_k \xi_k \frac{1/2}{1/2} / \| L_k s_k \|_2^2, 1 \}
\]
in the above sense. Though the idea seems interesting, its computational cost is more expensive than that of (2.7).

3.4 New Algorithms

Now we present the two kinds of structured quasi-Newton methods. First, we show the algorithm described in the subsection 3.1 as follows:
(FACNLS Algorithm)

Starting with a point \( x_1 \in \mathbb{R}^n \) and an \( m \times n \) matrix \( L_1 \), the algorithm proceeds, for \( k = 1, 2, \ldots \), as follows:

Step 1. Having \( x_k \) and \( L_k \), find the search direction \( d_k \) by solving the linear system of equations

\[
(L_k + J_k)^T(L_k + J_k) d_k = -J_k^T f_k.
\]

Step 2. Choose a steplength \( \alpha_k \) by a suitable line search algorithm.

Step 3. Set \( x_{k+1} = x_k + \alpha_k d_k \).

Step 4. If the new point satisfies the convergence criterion, then stop; otherwise, go to Step 5.

Step 5. Construct \( L_{k+1} \) by using the updating formula (3.26) or (3.27).

Next we present the algorithm described in the subsection 3.2:
(INVNLS Algorithm)

Starting with a point \( x_1 \in \mathbb{R}^n \) and an \( n \times n \) symmetric positive definite matrix \( H_1 \), the algorithm proceeds, for \( k = 1, 2, \ldots \), as follows:

Step 1. Having \( x_k \) and \( H_k \), calculate the search direction \( d_k \) by

\[
d_k = -H_k^{1/2} J_k^T f_k.
\]

Step 2. Choose a steplength \( \alpha_k \) by a suitable line search algorithm.

Step 3. Set \( x_{k+1} = x_k + \alpha_k d_k \).

Step 4. If the new point satisfies the convergence criterion, then stop; otherwise, go to Step 5.

Step 5. Construct \( H_{k+1} \) by using the updating formula (3.22) or (3.23).

If \( z_k \) is given by (3.4), the information of the second part of the Hessian matrix \( \nabla^2 F(x) \) is contained only in the secant condition. Further, in the case where \( f_{k+1} = 0 \), we have \( H_{k+1} s_k = (J_{k+1}^T J_{k+1})^{-1} s_k \).
4. Local and q-Superlinear Convergence of FACNLS Algorithm

We prove the local and q-superlinear convergence of FACNLS algorithms. Our proof is based on the bounded deterioration theorem by Broyden et al. [3]. Let \( D \) be the open convex subset of \( \mathbb{R}^n \) which contains the minimum point \( x^* \). We make the following two assumptions throughout this section:

(A1) There exist positive constants \( \xi_1, \xi_2 \) and \( p \) such that

\[
\| \nabla^2 F(u) - \nabla^2 F(x^*) \| \leq \xi_1 \| u - x^* \|^p \quad \text{for any } u \text{ in } D,
\]

and

\[
\| J(u_1) - J(u_2) \| \leq \xi_2 \| u_1 - u_2 \|^p \quad \text{for any } u_1 \text{ and } u_2 \text{ in } D.
\]

(A2) \( \nabla^2 F \) is symmetric positive definite at \( x^* \).

By using the above assumptions, we have

\[
\| \nabla F(u_1) - \nabla F(u_2) - \nabla^2 F(x^*)(u_1 - u_2) \| \leq \xi_1 \max(\| u_1 - x^* \|, \| u_2 - x^* \|)^p \| u_1 - u_2 \|
\]

for any \( u_1 \) and \( u_2 \) in \( D \), and

\[
\| J(u) \| \leq \xi_2 \| u - x^* \|^p + \| J(x^*) \| \quad \text{for any } u \text{ in } D.
\]

At first, we show the local and q-superlinear convergence of our algorithm with the DFP-type update (3.14). The following is the key lemma, which is shown by Broyden, Dennis and More [3, Lemma 5.2].

**Lemma 1.** Let \( M \) be an \( n \times n \) nonsingular symmetric matrix which satisfies

\[
\| Mc - M^{-1}a \| \leq \gamma \| M^{-1}a \|
\]

for some \( \gamma \in [0, 1/3] \) and vectors \( a \) and \( c \) in \( \mathbb{R}^n \) with \( c^Ta \neq 0 \). Let \( X \) be an \( n \times n \) symmetric matrix, \( b \) any vector in \( \mathbb{R}^n \) and define \( Y \) by

\[
Y = X + \{ (b - Xa)c^T + c(b - Xa)^T \}/c^Ta - \{ a^T(b - Xa)/(c^Ta)^2 \} c c^T.
\]

Then, for any \( n \times n \) symmetric \( G \),

\[
\| Y - G \|_{F, M} \leq [(1 - \alpha \theta^2)^{1/2} + 5/2 \| Mc - M^{-1}a \| / \{ (1 - \gamma) \| M^{-1}a \| \}] \| X - G \|_{F, M} + 2(1 + 2n^{1/2}) \| M \|_F \| b - Ga \| / \| M^{-1}a \|,
\]

where \( \alpha = (1 - 2\gamma)/(1 - \gamma^2) \in [3/8, 1] \), \( \theta = 0 \) if \( X = G \), and

\[
\theta = \| M(X - G)a \| / (\| X - G \|_{F, M} \| M^{-1}a \|) \quad \text{if } X \neq G.
\]

Let \( M \) be \( \nabla^2 F(x^*)^{-1/2} \). By the equivalence of norms for any \( n \times n \) matrix
there exists a positive constant $\eta$ such that
\[(4.5) \quad \|C\| \leq \eta \|C\|_{F,M}.
\]
For each iteration, set
\[(4.6) \quad \sigma_k = \max(\|x_k - x^*\|, \|x_{k+1} - x^*\|).\]

**Theorem 2.** Suppose that the assumptions (A1) and (A2) are satisfied. Let the matrix $L_k$ be updated by the DFP-type formula (3.14), where $z_k$ is given by (3.3) or (3.4). Let the sequence $\{x_k\}$ be generated by
\[(4.7) \quad x_{k+1} = x_k - ((L_k + J_k)^T(L_k + J_k))^{-1}J_k^Tf_k.\]

Then, for any $r \epsilon (0, 1)$, there exist positive constants $\varepsilon(r)$ and $\delta(r)$ such that if $\|x_1 - x^*\| \leq \varepsilon(r)$ and $\|L_1 + J_1\|^T(L_1 + J_1) - \nabla^2F(x^*)\|_{F,M} \leq \delta(r)$, the sequence $\{x_k\}$ generated by (4.7) is well defined and converges q-linearly to $x^*$ with
\[(4.8) \quad \|x_{k+1} - x^*\| \leq r \|x_k - x^*\|, \quad k \geq 1.
\]
Further, $\{\|L_k + J_k\|^T(L_k + J_k)\}$, $\{\|(L_k + J_k)^T(L_k + J_k))^{-1}\|\}$, $\{\|L_k - J_k+1\|^T(L_k + J_k+1)\}$ and $\{\|(L_k + J_k+1)^T(L_k + J_k+1))^{-1}\|\}$ are uniformly bounded.

**Proof.** For given $r \in (0, 1)$, choose $\delta$ such that
\[(4.9) \quad 0 < \delta \leq r/(2\eta \zeta_3),\]
and choose $\varepsilon$ so small that
\[(4.10) \quad \varepsilon \leq \frac{1}{\eta}, \quad (4.11) \quad \varepsilon \eta \leq \frac{r}{\zeta_3 - 2\eta \delta} \frac{\varepsilon}{\xi_1}, \quad (4.12) \quad 2^{p+1}(1+r) \eta \|\nabla^2F(x^*)\|^q(\zeta_1^{1/2} + 2 \zeta_2) \leq 9/10, \quad (4.13) \quad \varepsilon \eta \leq \frac{1}{3 \|\nabla^2F(x^*)\|^{1/2} \|\nabla^2F(x^*)\|^2} \zeta_4, \]
and
\[(4.14) \quad (2\mu_1 \delta + \mu_2) \varepsilon \eta \leq \frac{1}{(1 - \eta^p)} \delta \]
with
\[(4.15) \quad \zeta_1 = 2\eta \delta + \|\nabla^2F(x^*)\|, \quad (4.16) \quad \zeta_2 = \xi_2 + \|J(x^*)\|.\]
(4.17) \( \xi_3 = (1+r) \parallel \nabla^2 F(x^\#) \parallel^{-1} \parallel \),

(4.18) \( \xi_4 = \xi_1 + 2^p (p+2) \xi_2 \xi_2 / (p+1) \),

(4.19) \( \mu_1 = (15/4) \xi_4 \parallel M \parallel^2 \),

(4.20) \( \mu_2 = (n \parallel M \parallel) \xi_2 \xi_2 \xi_2^p + (1+\mu_1)(\xi_1^{1/2} + 2 \xi_2) + 2(1+2n^{1/2}) \xi_4 \parallel M \parallel F \parallel M \parallel \).

Set

(4.21) \( N_1 = \{ x \in \mathbb{R}^n \mid \parallel x - x^\# \parallel \leq \varepsilon \} \)

and

(4.22) \( N_2 = \{ B \in \mathbb{R}^{n \times n} \mid \parallel B - \nabla^2 F(x^\#) \parallel F, M \leq 2 \delta \} \).

Now we prove, by using the mathematical induction, that the following expressions hold for all \( k \geq 1 \):

(E1;k) \( \parallel B_k \parallel \leq \xi_1 \) and \( B_k \in N_2 \),

(E2;k) \( \parallel L_k \parallel \leq \xi_1^{1/2} + \xi_2 \),

(E3;k) \( \parallel B_k^{-1} \parallel \leq \xi_3 \),

(E4;k) \( \parallel x_{k+1} - x^\# \parallel \leq \varepsilon \parallel x_k - x^\# \parallel \) and \( x_{k+1} \in N_1 \),

(E5;k) \( \parallel B_k^p \parallel \leq 2^p + n^2 \xi_2 \xi_2 \varepsilon^p \parallel M \parallel^2 (\xi_1^{1/2} + 2 \xi_2) + \xi_1 \),

(E6;k) \( \parallel (B_k^p)^{-1} \parallel \leq 10 \xi_3 \),

(E7;k) The matrix \( B_{k+1} \) in (3.15) is well defined and

\[ \parallel B_{k+1} - \nabla^2 F(x^\#) \parallel F, M \leq \left( (1-3 \theta_k^2 / 8) / 2 + \mu_1 \sigma_k^p \right) \parallel B_k - \nabla^2 F(x^\#) \parallel F, M + \mu_2 \sigma_k^p \],

\[ \leq (1+\mu_1 \sigma_k^p) \parallel B_k - \nabla^2 F(x^\#) \parallel F, M + \mu_2 \sigma_k^p \],

where \( \sigma_k \), \( \mu_1 \) and \( \mu_2 \) are defined in (4.6), (4.19) and (4.20), respectively, and \( \theta_k \) is given by

(4.23) \( \theta_k = 0 \) if \( B_k^p = \nabla^2 F(x^\#) \)

and

(4.24) \( \theta_k = \parallel M(B_k^p - \nabla^2 F(x^\#))s_k \parallel / (\parallel B_k^p - \nabla^2 F(x^\#) \parallel F, M \parallel M^{-1} s_k \parallel) \), otherwise.

First, we consider the case of \( k = 1 \).

(E1;1) Since

(4.25) \( \parallel B_1 - \nabla^2 F(x^\#) \parallel F, M \leq \delta \leq 2 \delta \),
it is clear that $B_1 \in N_2$ and

\begin{equation}
\|B_1\| \leq \eta \|B_1 - \nabla^2 F(x^*)\|_{F,M} + \|\nabla^2 F(x^*)\| \leq 2\eta \delta + \|\nabla^2 F(x^*)\|.
\end{equation}

(E2;1) By (4.4) and (4.26), we have

\begin{equation}
\|L_1\| \leq \|L_1 + J(x_1)\| + \|J(x_1)\| \leq \|B_1\|^{1/2} + \|J(x_1)\|
\end{equation}

\begin{equation}
(4.27)
\|\nabla^2 F(x^*)^{-1} B_1 - \nabla^2 F(x^*) \| \leq 2\eta \delta + \|\nabla^2 F(x^*)\|^{1/2} + \|\nabla^2 F(x^*)\|^{1/2} + \|\nabla^2 F(x^*)\|
\end{equation}

(E3;1) It follows from (4.9) and (4.25) that

\begin{equation}
\|\nabla^2 F(x^*)^{-1} B_1 - \nabla^2 F(x^*) \| \leq 2\eta \delta + \|\nabla^2 F(x^*)\|^{1/2} + \|\nabla^2 F(x^*)\|^{1/2} + \|\nabla^2 F(x^*)\|
\end{equation}

\begin{equation}
\leq r/(1+r) < 1.
\end{equation}

By Banach Perturbation Lemma, $B_1$ is nonsingular, so $B_1$ is positive definite, and we have

\begin{equation}
\|B_1^{-1}\| \leq (1+r)\|\nabla^2 F(x^*)^{-1}\|.
\end{equation}

(E4;1) By (4.3), (4.11), (4.25) and (4.28), we have

\begin{equation}
\|x_2 - x^*\| = \|x_1 - B_1^{-1}\nabla F(x_1)\| - x^*\|
\end{equation}

\begin{equation}
\leq \|B_1^{-1}\| \|\nabla F(x_1) - \nabla F(x^*) - \nabla^2 F(x^*)(x_1 - x^*)\|
\end{equation}

\begin{equation}
+ \|B_1^{-1}\| \|B_1 - \nabla^2 F(x^*)\| \|x_1 - x^*\|
\end{equation}

\begin{equation}
\leq (1+r)(2\eta \delta + \|\nabla^2 F(x^*)\|^{1/2} + \|\nabla^2 F(x^*)\|^{1/2} + \|\nabla^2 F(x^*)\|)
\end{equation}

\begin{equation}
\leq \|x_1 - x^*\| \leq \varepsilon.
\end{equation}

Thus $x_2 \in N_1$.

(E5;1) Since $J(x_2)$ is available, the matrix $B_1^*$ is well defined. Thus we have

\begin{equation}
\|B_1^* - B_1\| \leq \|\sum_{1 \leq j \leq 2} (J_2 - J_1)^T (L_1 + J_1)\|_{F,M} + \|\sum_{1 \leq j \leq 2} (J_2 - J_1)^T (L_1 + J_1)\|_{F,M}
\end{equation}

\begin{equation}
\leq 2\|\sum_{1 \leq j \leq 2} (L_1 + J_1)\|_{F,M} + \|\sum_{1 \leq j \leq 2} (J_2 - J_1)\|_{F,M} + \|\sum_{1 \leq j \leq 2} (J_2 - J_1)\|_{F,M} + \|\sum_{1 \leq j \leq 2} (J_2 - J_1)\|_{F,M}
\end{equation}

\begin{equation}
(4.29) \leq 2^{p+1}n^2 \|M\| \|\sum_{1 \leq j \leq 2} (J_2 - J_1)\|_{F,M} \leq 2\|\nabla^2 F(x^*)\|^{1/2} + 2\|\nabla^2 F(x^*)\|^{1/2} + 2\|\nabla^2 F(x^*)\|
\end{equation}

Therefore, we have

\begin{equation}
\|B_1^*\| \leq \|B_1^* - B_1\| + \|B_1\|\leq 2^{p+1}n^2 \|M\| \|\nabla^2 F(x^*)\|^{1/2} + 2\|\nabla^2 F(x^*)\|^{1/2} + 2\|\nabla^2 F(x^*)\|\varepsilon + 2\|\nabla^2 F(x^*)\|.\end{equation}
(E6;1) By (4.12), (4.28) and (4.29), we have
\[
\|B_1^{-1}\| \|B_1^\# - B_1\| \leq 2^{p+1}n^2 \|M\| \|\xi_2\| \{(2\eta \delta + \|\nabla^2 F(x^\#)\|)^{1/2}
+ 2(\xi_2 \varepsilon + \|J(x^\#)\|)\|\nabla^2 F(x^\#)^{-1}\|/(1+r)\varepsilon\}
\leq 9/10 < 1.
\]
Thus, by Banach Perturbation Lemma, the matrix \(B_1^\#\) is nonsingular, so \(B_1^\#\) is positive definite, and we have
\[
\|B_1^\#\|^{-1} \leq 10/(1+r)\|\nabla^2 F(x^\#)^{-1}\|.
\]

(E7;1) At first, we show that there holds for \(\xi_4\) in (4.18)
\[
(4.30) \quad \|z_1 - \nabla^2 F(x^\#)s_1\| \leq \xi_4 \sigma_1 \|s_1\|
\]
where \(z_1\) is defined by (3.3) or (3.4).

If \(z_1 = y_1\), then it follows directly from (4.3) that
\[
\|y_1 - \nabla^2 F(x^\#)s_1\| \leq \xi_1 \sigma_1 \|s_1\|
\]
Consider the case of \(z_1 = v_1 + J_2^T J_2 s_1\). Since
\[
f_2 - f_1 - J_2 s_1 = \int_0^1 (x_1 + ts_1) s_1 dt - J_2 s_1,
\]
we have, by (4.2),
\[
\|z_1 - \nabla^2 F(x^\#)s_1\| \leq \|y_1 - \nabla^2 F(x^\#)s_1\| + \|J_1^T\| \|f_2 - f_1 - J_2 s_1\|
+ \|J_2 - J_1\| \|J_2\| \|s_1\|
\leq \{(\xi_1 + (p+2) \xi_2^2 \varepsilon P + \|J(x^\#)\|/(p+1)) \sigma_1 \|s_1\|
\]
Next, we show that \(s_1^T z_1 > 0\) if \(s_1 \neq 0\). This can be shown by a similar way to the proof of Lemma 4.2(a) in [3]. Since
\[
(4.31) \quad \|M z_1 - M^{-1} s_1\| \leq \|M\| \|z_1 - \nabla^2 F(x^\#)s_1\|
\leq \xi_4 \sigma_1 \|M\| \|s_1\| \leq (1/3) \|M^{-1} s_1\|
\]
and
\[
s_1^T z_1 = (M^{-1} s_1)^T (M z_1 - M^{-1} s_1) + (M^{-1} s_1)^T (M^{-1} s_1),
\]
we have
\[
\|s_1^T z_1\| \leq \|M^{-1} s_1\| \|M z_1 - M^{-1} s_1\| \leq (1/3) \|M^{-1} s_1\|^2.
\]
Thus
\[
(2/3) \|M^{-1} s_1\|^2 \leq s_1^T z_1 \leq (4/3) \|M^{-1} s_1\|^2.
\]
which suggests that \( s_1 \neq 0 \) implies \( s_1 z_1 > 0 \).

Since the matrix \( B_1^\# \) is symmetric positive definite and \( s_1 z_1 > 0 \), \( L_2 \) is well defined, and is given by
\[
L_2 = L_1 + (L_1 + J_2) (s_1 z_1 / z_1 B_1^\# - 1)^{1/2} (b_1^\# - 1 z_1 - s_1) (z_1 / s_1 z_1)^T.
\]
Setting \( B_1 = (L_1 + J_1)^T (L_1 + J_1) \), \( B_2 = (L_2 + J_2)^T (L_2 + J_2) \) and \( B_1^\# = (L_1 + J_2)^T (L_1 + J_2) \) gives
\[
B_2 = B_1 + (1 + s_1 B_1^\# s_1 / s_1 z_1) z_1 z_1 / s_1 z_1 - (b_1^\# s_1 z_1 + z_1 s_1 B_1^\#) / s_1 z_1.
\]

Note that the above corresponds to the DFP update. Let
\[
X = B_1^\#, \quad Y = B_2, \quad G = \nabla^2 f(x^\#), \quad \gamma = 1 / 3, \quad a = s_1, \quad b = z_1, \quad c = z_1.
\]
Then using Lemma 1 and (4.31), we have the bounded deterioration property for \( B_1^\# \) and \( B_2 \) such that
\[
(4.32) \quad \| B_2 - \nabla^2 f(x^\#) \|_F, M \leq \{ (1 - 3 \theta^2 / 8)^{1/2} + \tau_1 \sigma^p \} \| B_1^\# - \nabla^2 f(x^\#) \|_F, M + \tau_2 \sigma^p,
\]
where
\[
\tau_1 = (15/4) \| M \|_2 \zeta_4, \quad \tau_2 = 2(1 + 2n^{1/2}) \| M \|_F \zeta_4,
\]
and \( \theta_1 \) is given by (4.23) or (4.24). Moreover, by noting that
\[
\| B_1^\# - \nabla^2 f(x^\#) \|_F, M \leq \| B_1^\# - B_1 \|_F, M + \| B_1 - \nabla^2 f(x^\#) \|_F, M
\]
and using (4.32), we obtain the bounded deterioration property for \( B_1 \) and \( B_2 \).

Assume that the expressions (E1;\( k \)) through (E7;\( k \)) hold for \( k = 1, \ldots, m-1 \). Then we have
\[
\| B_{k+1} - \nabla^2 f(x^\#) \|_F, M \leq \| B_k - \nabla^2 f(x^\#) \|_F, M
\]
\[
\leq (\mu_1 \| B_k - \nabla^2 f(x^\#) \|_F, M + \mu_2 \sigma^p_k) \sigma^p_k \leq (2 \mu_1 \delta + \mu_2) \sigma^p_k
\]
for \( k = 1, \ldots, m-1 \), and by summing both sides from \( k = 1 \) to \( m-1 \), it follows from (4.14) that
\[
\| B_m - \nabla^2 f(x^\#) \|_F, M \leq \| B_1 - \nabla^2 f(x^\#) \|_F, M + (2 \mu_1 \delta + \mu_2) \| x_1 - x^\# \|_1 \sum_{j=1}^{m-2} (r^p_j)^j
\]
\[
\leq \delta + (2 \mu_1 \delta + \mu_2) \varepsilon^p(1 - r^p) \leq 2 \delta,
\]
which implies (E1;\( m \)). We can prove (E2;\( m \)) through (E7;\( m \)) by the same way as the case of \( k = 1 \).

Therefore, this concludes the induction, which completes the proof. ■
Further, by combining Theorem 3.4 in [5] and (E7;k),
\begin{equation}
\lim_{k \to \infty} \| \nabla^2 F(x^*)^{-1/2} B_k \nabla^2 F(x^*)^{-1/2} I \|_F
\end{equation}
exists and we have
\begin{equation}
\lim_{k \to \infty} \| (B_k - \nabla^2 F(x^*)) s_k \| / \| s_k \| = 0
\end{equation}
Thus, by Theorem 2.2 in [5], we obtain the following theorem.

**Theorem 3.** Suppose that all conditions of Theorem 2 hold. Then the sequence \( \{x_k\} \) converges \( q \)-superlinearly to \( x^* \), that is,
\begin{equation}
\lim_{k \to \infty} \| x_{k+1} - x^* \| / \| x_k - x^* \| = 0.
\end{equation}

Next consider the BFGS-type update (3.9) of \( L_k \). Let \( H_k = B_k^{-1} \) and
\( H_k^* = (B_k^*)^{-1} \). Then the relation between \( H_k^* \) and \( H_{k+1}^* \) can be given by
\begin{equation}
H_{k+1} = H_k^* - (1 + z_T^k H_k^* z_k / z_T^k s_k s_k^T / z_T^k s_k \cdot z_T^k s_k) H_k^* + (H_k^* z_k z_k^T s_k / z_T^k s_k + s_k^T H_k^* z_k) / z_T^k s_k,
\end{equation}
which is the form obtained by performing the interchange \( B_k^* \leftrightarrow H_k^* \) and
\( s_k \leftrightarrow z_k \) in (3.15). So we can prove the local and \( q \)-superlinear convergence of our algorithm with (3.9) by the same means as the above.

Let \( M \) be \( \nabla^2 F(x^*)^{-1/2} \). By the equivalence of norms for any \( n \times n \) matrix \( C \), there exist positive constants \( \eta \) and \( \eta' \) such that
\begin{equation}
(1 / \eta') \| C \|_{F,M} \leq \| C \| \leq \eta \| C \|_{F,M}.
\end{equation}

**Theorem 4.** Suppose that the assumptions (A1) and (A2) are satisfied. Let the matrix \( L_k \) be updated by the BFGS-type formula (3.9), where \( z_k \) is given by (3.3) or (3.4). Let the sequence \( \{x_k\} \) be generated by
\begin{equation}
x_{k+1} = x_k - (L_k + J_k)^T(L_k + J_k)^{-1} J_k f_k.
\end{equation}
Then, for any \( r \in (0, 1) \), there exist positive constants \( \varepsilon(r) \) and \( \delta(r) \) such that if \( \| x_1 - x^* \| \leq \varepsilon(r) \) and \( \| (L_1 + J_1)^T(L_1 + J_1)^{-1} - \nabla^2 F(x^*)^{-1} \|_{F,M} \leq \delta(r) \), the sequence \( \{x_k\} \) generated by (4.38) is well defined and converges \( q \)-linearly to \( x^* \) with
\begin{equation}
\| x_{k+1} - x^* \| \leq r \| x_k - x^* \|, \quad k \geq 1.
\end{equation}
Further, \{\| (L_k + L_k')^T (L_k + L_k') \| \}, \{\| (L_k + L_k')^T (L_k + L_k')^{-1} \| \},
\{\| (L_k + L_k+1)^T (L_k + L_k+1) \| \} \text{ and } \{\| (L_k + L_k+1)^T (L_k + L_k+1)^{-1} \| \} \text{ are uniformly bounded.}

Proof. For given \( r \in (0, 1) \), choose \( \delta \) such that
(4.40) \quad 0 < \delta \leq r/(2 \pi \xi_3^*),
and choose \( \varepsilon \) so small that
(4.41) \quad \varepsilon \leq 1,
(4.42) \quad \xi_1^* \xi_1 \varepsilon \leq r^2/(1+r),
(4.43) \quad 2^{p+1} \xi_2^* \xi_1^* (2 \xi_2^* + \xi_3^* 1/2) \varepsilon \leq 9/10,
(4.44) \quad \varepsilon^p \leq 1/(4 \| \nabla^2 F(x^*)^{-1} \| M^2 M^{-1} \| \xi_4^* \|
and
(4.45) \quad (2 \mu_1^* \delta + \mu_2^*) \varepsilon \leq (1 - r^p) \leq \delta
with
(4.46) \quad \xi_1^* = 2 \eta \delta + \| \nabla^2 F(x^*)^{-1} \| ,
(4.47) \quad \xi_2^* = \xi_2 + \| J(x^*) \| ,
(4.48) \quad \xi_3^* = (1+r) \| \nabla^2 F(x^*) \| ,
(4.49) \quad \xi_4^* = \xi_1^* + 2^p (p+2) \xi_2^* \xi_2^* / (p+1),
(4.50) \quad \mu_1^* = 5 \xi_4^* \| \nabla^2 F(x^*)^{-1} \| ,
(4.51) \quad \mu_2^* = 10 \xi_2^* 2^{p+1} \eta \{ (1+\mu_1^*) \xi_1^* (2 \xi_2^* + (\xi_3^* 1/2)
+ (8/15)(1 + 2n^{1/2}) \mu_1^* \| M \| M^{-1} \| .

Set
(4.52) \quad N_1^* = \{ x \in \mathbb{R}^n \mid \| x - x^* \| \leq \varepsilon \}
and
(4.53) \quad N_2^* = \{ h \in \mathbb{R}^n \times \mathbb{R}^n \mid \| h - \nabla^2 F(x^*)^{-1} \| \leq 2 \delta \}.

Now we prove, by using the mathematical induction, that the following expressions hold for all \( k \geq 1 \):
(EL1;k)’ \quad \| H_k \| \leq \xi_1^* \text{ and } H_k \in N_2^*.
(EL2;k)’ \quad \| B_k \| \leq \xi_3^*.
\( (E3;k)' \quad \| L_k \| \leq \xi'_2 + (\xi'_3)^{1/2} \),

\( (E4;k)' \quad \| x_{k+1} - x^\# \| \leq r \| x_k - x^\# \| \) and \( x_{k+1} \in N_1^\# \),

\( (E5;k)' \quad \| \sigma_k^p \| \leq 2^{p+1} \xi_2 (2 \xi'_2 + (\xi'_3)^{1/2}) + \xi'_3 \),

\( (E6;k)' \quad \| H_k^\# \| \leq 10 \xi'_1 \),

\( (E7;k)' \quad H_{k+1} \) is well defined and

\[
\| H_{k+1} - \nabla^2 F(x^\#)^{-1} \|_{F,M} \leq \{(1 - 3 \theta_k^2 / 8)^{1/2} + \mu_1 \sigma_k^p \} \| H_k - \nabla^2 F(x^\#)^{-1} \|_{F,M} + \mu_2 \sigma_k^p,
\]

\[
\leq (1 + \mu_1 \sigma_k^p) \| H_k - \nabla^2 F(x^\#)^{-1} \|_{F,M} + \mu_2 \sigma_k^p,
\]

where \( \sigma_k \), \( \mu_1 \) and \( \mu_2 \) are defined in (4.6), (4.50) and (4.51), respectively, and \( \theta_k \) is given by

\( (4.54) \quad \theta_k = 0 \quad \text{if} \quad H_k^\# = \nabla^2 F(x^\#)^{-1} \)

and

\( (4.55) \quad \theta_k = \| M (H_k^\# - \nabla^2 F(x^\#)^{-1}) z_k \| / (\| H_k^\# - \nabla^2 F(x^\#)^{-1} \|_{F,M} \| M^{-1} z_k \|), \) otherwise.

First, we consider the case of \( k = 1 \).

\( (E1;1)' \quad \text{Since the proof is very similar to that of Theorem 2, it is omitted.} \)

\( (E2;1)' \quad \text{It follows from (4.40) that} \)

\[
\| \nabla^2 F(x^\#) \| \| H_1^\# - \nabla^2 F(x^\#)^{-1} \| \leq 2 \eta \delta \| \nabla^2 F(x^\#) \| \leq r/(1+r) < 1.
\]

By Banach Perturbation Lemma, we have

\[
\| H_1 \| = \| H_1^\# \| \leq (1+r) \| \nabla^2 F(x^\#) \|.
\]

Since the proofs of \( (E3;1)' \) through \( (E6;1)' \) are very similar to those of Theorem 2, they are omitted.

\( (E7;1)' \quad \text{By the same way as the proof of Theorem 2, we can show that} \)

\( (4.56) \quad \| z_1 - \nabla^2 F(x^\#) s_1 \| \leq \xi'_4 \sigma_1^p \| s_1 \| \)

for \( \xi'_4 \) in (4.49), where \( z_1 \) is defined by (3.3) or (3.4), and that \( s_1^Ts_1 > 0 \) if \( s_1 \neq 0 \). Using Lemma 1, we have the bounded deterioration property for \( H_1^\# \)

and \( H_2 \) such that

\( (4.57) \quad \| H_2 - \nabla^2 F(x^\#)^{-1} \|_{F,M} \leq \{(1 - 3 \theta_1^2 / 8)^{1/2}
\]
\[ \tau_1^* = 5 \| M \| \| M^{-1} \| \| \nabla^2 F(x^*)^{-1} \| \xi_4, \]
\[ \tau_2^* = (8/3)(1 + 2n^{1/2}) \| M^{-1} \| 2 \| M \| \| M \| \| \nabla^2 F(x^*)^{-1} \| \xi_4, \]
and \( \theta_1 \) is given by (4.54) or (4.55).

Moreover, by (E1;1)' and (E6;1)', we have
\[ \| H_1^* - H_1 \| _F, M \leq \| H_1^* (B_1 - B_1) H_1 \| _F, M \leq \eta \| H_1^* \| \| H_1 \| \| B_1 - B_1 \| \]
\[ \leq 10 \eta (\xi_1^*)^2 2^{p+1} \xi_2 (2 \xi_2 + (\xi_3)^{1/2}) \sigma_1 \]
and
\[ \| H_1^* - \nabla^2 F(x^*)^{-1} \| _F, M \leq \| H_1^* - H_1 \| _F, M + \| H_1 - \nabla^2 F(x^*)^{-1} \| _F, M. \]

Therefore, we obtain the bounded deterioration property for \( H_1 \) and \( H_2 \).

Since the remainder of the proof is very similar to that of Theorem 2, it is omitted. \( \blacksquare \)

Moreover, by a similar way to the proof of Theorem 3.4 in [5],
\[ (4.58) \quad \lim_{k \to \infty} \| \nabla^2 F(x^*)^{-1/2} H_k \| \nabla^2 F(x^*)^{-1/2} - I \|_F \]
exists and we have
\[ (4.59) \quad \lim_{k \to \infty} \| (H_k - \nabla^2 F(x^*)^{-1}) z_k \| / \| z_k \| = 0. \]

Finally, we show that the above is at least the sufficient condition for the superlinear convergence of the algorithm.

Theorem 5. Suppose that the same assumptions of Theorem 4 hold. Let the sequence \( \{x_k\} \) be generated by
\[ (4.60) \quad x_{k+1} = x_k - H_k \nabla F(x_k). \]
Then \( \{x_k\} \) converges q-superlinearly to \( x^* \).

Proof. For the case of \( z_k = y_k \), it is proved in [5, p.559] that (4.59) implies the q-superlinear convergence. So we only consider the DGW secant condition (3.4). Since
\[ z_k = y_k - J_k^T (f_{k+1} - f_k - J_k S_k) + (J_{k+1} - J_k)^T J_{k+1} S_k, \]
we have
\[ \| H_k \nabla F(x_{k+1}) \| \leq \| (H_k - \nabla^2 F(x^*))^{-1} z_k \| + \| \nabla^2 F(x^*)^{-1} \| \| y_k - \nabla^2 F(x^*) s_k \| \\
+ \| H_k - \nabla^2 F(x^*)^{-1} \| \| J_k \| \| f_{k+1} - f_k - J_k s_k \| \\
+ \| H_k - \nabla^2 F(x^*)^{-1} \| \| J_{k+1} - J_k \| \| J_{k+1} \| \| s_k \|. \]

Using (4.2), (4.4), (4.56), (4.58), (4.59) and (E1;k)' , a similar way to [5, p.559] yields

(4.61) \[ \lim_{k \to \infty} \| \nabla F(x_{k+1}) \| / \| x_{k+1} - x_k \| = 0 \]

and

(4.62) \[ \lim_{k \to \infty} \| x_{k+1} - x^* \| / \| x_k - x^* \| = 0. \]

Thus the proof is complete.

5. Computational Experiments

Computational experiments were performed to compare the factorized versions proposed in this paper with the Gauss-Newton method and the structured quasi-Newton methods from the viewpoint of the number of iterations and the number of the objective function evaluations.

The numerical calculations were carried out in double precision arithmetic on a NEC PC-9801VX personal computer, and the program is coded in FORTRAN 77. For all the methods, the initial matrices \( A_1 \) and \( L_1 \) are set to zero matrices, and \( H_1 \) is set to the unit matrix. The iterative process is terminated

(1) if \( \| f(x_k) \|_\infty \leq \max(TOL1, \varepsilon) \),

or

(2) if \( \| e_j^T f(x_{k+1}) \| \leq \max(TOL2, \varepsilon) \| f(x_{k+1}) \| \| J(x_{k+1}) e_j \| \\
\quad \text{for } j=1, \ldots, n \) and \( \| x_{k+1} - x_k \|_\infty \leq \max(TOL3, \varepsilon) \max( \| x_{k+1} \|_\infty, 1.0) \),

where \( e_j \) denotes the \( j \)-th column of the unit matrix,

or

(3) if the number of iterations exceeds the prescribed limit (ITMAX),

or

(4) if the number of function evaluations exceeds the prescribed limit (NFEMAX),

where \( \| \cdot \|_\infty \) denotes the maximum norm and \( \varepsilon \) is a machine epsilon.

Further, the Jacobian matrix is evaluated by the forward difference approximation and the bisection line search method with Armijo's rule.
\[ F(x_k + \alpha_k d_k) \leq F(x_k) + 0.1 \alpha_k \nabla F(x_k)^T d_k \]

is employed.

In the experiments, we set TOL1 = TOL2 = TOL3 = 10^{-4}, ITMAX = 500 and
NFEMAX = 2000. Since the sized BFGS-type update (3.27) includes the inverse
matrix of \( B_k \), we used only the sized BFGS-type update (3.26) in the FACNLS
methods. The test functions to be minimized are listed as follows [2]:

Problem 1. (Powell)
\[ F = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4, \]
Starting point=(3, -1, 0, 1), Minimum point=(0, 0, 0, 0), Value = 0,
Problem 2. (Freudenstein and Roth)
\[ F = (-13 + x_1 - 2x_2 + 5x_2^2 - x_2^3)^2 + (-29 + x_1 - 14x_2 + x_2^2 + x_2^3)^2, \]
Starting point=(15, -2) or (6,6), Local minimum point = (11.4128, -0.89681), Value = 48.98425; Global minimum point = (5,4), Value = 0,
Problem 3. (Kowalik)
\[ F = \sum_{j=1}^{11} (a_j - x_1(u_j^2 + x_2u_j)/(u_j^2 + x_3u_j + x_4))^2, \]
The data \( a_j \) and \( u_j \) are given in [2],
Starting point = (0.25, 0.39, 0.415, 0.39),
Minimum point = (0.19281,0.19128,0.12306,0.13606), Value=3.075 \times 10^{-4}
Problem 4. (Jennrich)
\[ F = \sum_{j=1}^{10} (a_j - (\exp(jx_1) + \exp(jx_2))^2, \text{ where } a_j = 2 + 2j, \]
Starting point=(0.3,0.4), Minimum point=(0.25783,0.25783), Value=124.36
Problem 5. (Osborne)
\[ F = \sum_{j=1}^{33} (a_j - (x_1 + x_2\exp(-x_4t_j) + x_3\exp(-x_5t_j))^2, \]
The data \( a_j \) are given in [2] and \( t_j = 10(j-1), \)
Starting point = (0.5, 1.5, -1, 0.01, 0.02),
Minimum point = (0.3754, 1.9358, -1.4647, 0.01287, 0.02212),
Minimum Value = 0.546 \times 10^{-4}

The computational results are summarized in Tables 1 through 6. In each table, we use the following symbols:
GN : the Gauss-Newton method,
Biggs: the Bartholomew-Biggs update (2.6),
DGW : the Dennis, Gay and Welsch update (2.8),
FAC0 : the FACNLS algorithm with (3.3) and (3.9),
FAC1 : the FACNLS algorithm with (3.4) and (3.9),
FAC2 : the FACNLS algorithm with (2.7), (3.4) and (3.26),
FAC3 : the FACNLS algorithm with (3.33), (3.4) and (3.26),
INV0 : the INVNLS algorithm with (3.4) and (3.22),
INV1 : the INVNLS algorithm with (3.4) and (3.23),
BFGS : the standard BFGS method,
DFP : the standard DFP method,
IT : the number of iterations for convergence,
EV : the number of the objective function evaluations,
FV : the obtained final function value,
# : the method had failed to converge in the specified number of
the objective function evaluations.

Table 1. Results for Problem 1

<table>
<thead>
<tr>
<th>IT</th>
<th>EV</th>
<th>FV</th>
</tr>
</thead>
<tbody>
<tr>
<td>GN</td>
<td>9</td>
<td>50 2.3 x 10^{-9}</td>
</tr>
<tr>
<td>Biggs</td>
<td>14</td>
<td>75 7.0 x 10^{-9}</td>
</tr>
<tr>
<td>DGW</td>
<td>14</td>
<td>75 6.6 x 10^{-9}</td>
</tr>
<tr>
<td>FAC0</td>
<td>20</td>
<td>105 6.2 x 10^{-9}</td>
</tr>
<tr>
<td>FAC1</td>
<td>14</td>
<td>75 4.8 x 10^{-9}</td>
</tr>
<tr>
<td>FAC2</td>
<td>14</td>
<td>75 5.4 x 10^{-9}</td>
</tr>
<tr>
<td>FAC3</td>
<td>14</td>
<td>75 5.1 x 10^{-9}</td>
</tr>
<tr>
<td>INV0</td>
<td>27</td>
<td>158 3.5 x 10^{-9}</td>
</tr>
<tr>
<td>INV1</td>
<td>77</td>
<td>419 1.5 x 10^{-8}</td>
</tr>
<tr>
<td>BFGS</td>
<td>32</td>
<td>184 1.0 x 10^{-8}</td>
</tr>
<tr>
<td>DFP</td>
<td>97</td>
<td>513 7.6 x 10^{-9}</td>
</tr>
</tbody>
</table>

Table 2. Results for Problem 2 with (15, -2)

<table>
<thead>
<tr>
<th>IT</th>
<th>EV</th>
<th>FV</th>
</tr>
</thead>
<tbody>
<tr>
<td>GN</td>
<td>105</td>
<td>#</td>
</tr>
<tr>
<td>Biggs</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>DGW</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>FAC0</td>
<td>9</td>
<td>33</td>
</tr>
<tr>
<td>FAC1</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td>FAC2</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td>FAC3</td>
<td>11</td>
<td>102</td>
</tr>
<tr>
<td>INV0</td>
<td>8</td>
<td>36</td>
</tr>
<tr>
<td>INV1</td>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>BFGS</td>
<td>9</td>
<td>39</td>
</tr>
<tr>
<td>DFP</td>
<td>10</td>
<td>42</td>
</tr>
</tbody>
</table>
Table 3. Results for Problem 2 with (6,6)

<table>
<thead>
<tr>
<th>IT</th>
<th>EV</th>
<th>FV</th>
</tr>
</thead>
<tbody>
<tr>
<td>GN</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$2.8 \times 10^{-17}$</td>
</tr>
<tr>
<td>Biggs</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$8.7 \times 10^{-15}$</td>
</tr>
<tr>
<td>DGW</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$8.1 \times 10^{-15}$</td>
</tr>
<tr>
<td>FAC0</td>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$8.5 \times 10^{-13}$</td>
</tr>
<tr>
<td>FAC1</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$6.5 \times 10^{-9}$</td>
</tr>
<tr>
<td>FAC2</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$6.2 \times 10^{-15}$</td>
</tr>
<tr>
<td>FAC3</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$6.2 \times 10^{-15}$</td>
</tr>
<tr>
<td>INV0</td>
<td>92</td>
<td>$\ast$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>730.41</td>
</tr>
<tr>
<td>INV1</td>
<td>92</td>
<td>$\ast$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>730.41</td>
</tr>
<tr>
<td>BFGS</td>
<td>13</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48.98 ($)</td>
</tr>
<tr>
<td>DFP</td>
<td>44</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48.98 ($)</td>
</tr>
</tbody>
</table>

($)$The local minimum is obtained.

Table 4. Results for Problem 3

<table>
<thead>
<tr>
<th>IT</th>
<th>EV</th>
<th>FV</th>
</tr>
</thead>
<tbody>
<tr>
<td>GN</td>
<td>19</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3.075 \times 10^{-4}$</td>
</tr>
<tr>
<td>Biggs</td>
<td>10</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3.075 \times 10^{-4}$</td>
</tr>
<tr>
<td>DGW</td>
<td>12</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3.075 \times 10^{-4}$</td>
</tr>
<tr>
<td>FAC0</td>
<td>14</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3.075 \times 10^{-4}$</td>
</tr>
<tr>
<td>FAC1</td>
<td>11</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3.075 \times 10^{-4}$</td>
</tr>
<tr>
<td>FAC2</td>
<td>10</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3.075 \times 10^{-4}$</td>
</tr>
<tr>
<td>FAC3</td>
<td>11</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3.075 \times 10^{-4}$</td>
</tr>
<tr>
<td>INV0</td>
<td>29</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3.075 \times 10^{-4}$</td>
</tr>
<tr>
<td>INV1</td>
<td>399</td>
<td>$\ast$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.711 $\times 10^{-4}$</td>
</tr>
<tr>
<td>BFGS</td>
<td>32</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$3.075 \times 10^{-4}$</td>
</tr>
<tr>
<td>DFP</td>
<td>399</td>
<td>$\ast$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.772 $\times 10^{-4}$</td>
</tr>
</tbody>
</table>

Table 5. Results for Problem 4

<table>
<thead>
<tr>
<th>IT</th>
<th>EV</th>
<th>FV</th>
</tr>
</thead>
<tbody>
<tr>
<td>GN</td>
<td>138</td>
<td>$\ast$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$2.7 \times 10^{-9}$</td>
</tr>
<tr>
<td>Biggs</td>
<td>9</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124.36</td>
</tr>
<tr>
<td>DGW</td>
<td>9</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124.36</td>
</tr>
<tr>
<td>FAC0</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124.36</td>
</tr>
<tr>
<td>FAC1</td>
<td>9</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124.36</td>
</tr>
<tr>
<td>FAC2</td>
<td>10</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124.36</td>
</tr>
<tr>
<td>FAC3</td>
<td>15</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124.36</td>
</tr>
<tr>
<td>INV0</td>
<td>11</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124.36</td>
</tr>
<tr>
<td>INV1</td>
<td>11</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124.36</td>
</tr>
<tr>
<td>BFGS</td>
<td>13</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124.36</td>
</tr>
<tr>
<td>DFP</td>
<td>30</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td></td>
<td>124.36</td>
</tr>
</tbody>
</table>

Table 6. Results for Problem 5

<table>
<thead>
<tr>
<th>IT</th>
<th>EV</th>
<th>FV</th>
</tr>
</thead>
<tbody>
<tr>
<td>GN</td>
<td>6</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$5.465 \times 10^{-5}$</td>
</tr>
<tr>
<td>Biggs</td>
<td>27</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$5.465 \times 10^{-5}$</td>
</tr>
<tr>
<td>DGW</td>
<td>21</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$5.465 \times 10^{-5}$</td>
</tr>
<tr>
<td>FAC0</td>
<td>43</td>
<td>274</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$5.465 \times 10^{-5}$</td>
</tr>
<tr>
<td>FAC1</td>
<td>26</td>
<td>190</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$5.465 \times 10^{-5}$</td>
</tr>
<tr>
<td>FAC2</td>
<td>18</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$5.465 \times 10^{-5}$</td>
</tr>
<tr>
<td>FAC3</td>
<td>16</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$5.465 \times 10^{-5}$</td>
</tr>
<tr>
<td>INV0</td>
<td>93</td>
<td>$\ast$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.560 $\times 10^{-5}$</td>
</tr>
<tr>
<td>INV1</td>
<td>329</td>
<td>$\ast$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.580 $\times 10^{-5}$</td>
</tr>
<tr>
<td>BFGS</td>
<td>51</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$5.465 \times 10^{-5}$</td>
</tr>
<tr>
<td>DFP</td>
<td>172</td>
<td>$\ast$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.019 $\times 10^{-5}$</td>
</tr>
</tbody>
</table>
From these tables, we can see that the Gauss-Newton method performed very well for the zero or small residual problems (Tables 1 and 6), but did not necessarily well for the large residual problems (Tables 2 and 5). For all the problems, the FACNLS methods except FAC0, the structured quasi-Newton methods with the Biggs and the DGW updates performed well and were numerically stable. These numerical results suggest that the FACNLS methods are comparable with the structured quasi-Newton methods with the Biggs and the DGW updates. However, the INVNLs methods, the standard BFGS and the standard DFP methods did not perform well compared with the other methods.

6. Concluding Remarks

This paper has been concerned with the iterative methods based on the structured quasi-Newton methods for nonlinear least squares problems. Our idea is to compute the search direction by solving the linear system of equations (3.1). This enables us to obtain descent search directions for the objective function.

We proposed the FACNLS and the INVNLs algorithms. However, since the information of the second part \( \sum_{j=1}^{m} f_j(x)\nabla^2 f_j(x) \) in (1.5) is contained only in the secant condition, the INVNLs methods did not perform well compared with the other structured quasi-Newton methods in our experiments. We recommend the FACNLS algorithm with the sized BFGS-type update (3.26) for practical computations. In addition, the FAC1 seems numerically stable though it does not employ a sizing technique.

References


