Journal of Computational Mathematics, Vol.14, No.2, 1996, 143-158.

A CLASS OF FACTORIZED QUASI-NEWTON METHODS FOR NONLINEAR LEAST SQUARES PROBLEMS*

C.X. Xu X.F. Ma M.Y. Kong
(*Department of Mathematics, Xi'an Jiaotong University, Xi'an, China*)

Abstract

This paper gives a class of descent methods for nonlinear least squares solution. A class of updating formulae is obtained by using generalized inverse matrices. These formulae generate an approximation to the second part of the Hessian matrix of the objective function, and are updated in such a way that the resulting approximation to the whole Hessian matrix is the convex class of Broyden-like updating formulae. It is proved that the proposed updating formulae are invariant under linear transformation and that the class of factorized quasi-Newton methods are locally and superlinearly convergent. Numerical results are presented and show that the proposed methods are promising.

1. Introduction

This paper deals with the problem of minimizing a sum of squares of nonlinear functions

$$f(x) = \frac{1}{2}\sum_{i=1}^{m}(r_i(x))^2 = \frac{1}{2}r(x)^T r(x) \tag{1}$$

where $r_i(x), i = 1, 2, \cdots, m$ are twice continuously differentiable, $m \geq n$, $r(x) = (r_1(x), r_2(x), \cdots, r_m(x))^T$ and "T" denotes transpose. Nonlinear least squares problem is a kind of important optimization problems and is appeared in many fields such as scientific experiments, maximum likelihood estimation, solution of nonlinear equations, pattern recognition and etc. The derivatives of the function $f(x)$ are given by

$$g(x) = \bigtriangledown f(x) = A(x)^T r(x) \tag{2}$$

$$G(x) = \bigtriangledown^2 f(x) = A(x)^T A(x) + \sum_{i=1}^{m} r_i(x) \bigtriangledown^2 r_i(x) \tag{3}$$

where $A \in R^{m\times n}$ is the Jacobian matrix of $r(x)$ and its elements are $a_{ij} = \partial r_i(x)/\partial x_j$, $i = 1, 2, \cdots, m, j = 1, 2, \cdots, n$.

Various iterative methods for problem (1) are available and can be divided into two kinds, trust region methods and descent methods. Trust region methods are globally

* Received August 23, 1994.

convergent, but complicated in implementation. In this paper we consider Newton-like descent methods. Suppose that $x^{(k)}$ is a current estimation of the minimum point x^*. A descent direction $d^{(k)}$ is assigned to $x^{(k)}$ by solving a system

$$B_k d^{(k)} = -g^{(k)} \tag{4}$$

and a new estimate point is generated by

$$x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} \tag{5}$$

where $g^{(k)} = g(x^{(k)}), B_k$ is a symmetric positive definite approximation to the Hessian matrix $G_k = G(x^{(k)}), \alpha_k$ is a step length determined by line search. An ideal choice for the step length is

$$\alpha_k = \arg\min_{\alpha>0} f(x^{(k)} + \alpha d^{(k)}). \tag{6}$$

Since the ideal choice of the step length is impracticable and unnecessary, inexact line searches are usually carried out to give a step length satisfying

$$f(x^{(k)} + \alpha_k d^{(k)}) \leq f(x^{(k)}) + \rho \alpha_k {g^{(k)}}^T d^{(k)} \tag{7}$$

$$\mid g(x^{(k)} + \alpha_k d^{(k)})^T d^{(k)} \mid \leq -\sigma {g^{(k)}}^T d^{(k)}. \tag{8}$$

With $\rho \in (0, \frac{1}{2})$ and $\sigma \in (\rho, 1)$, an interval of acceptable α values always exists and an efficient line search strategy to find such a step length can be found in [2].

Different choices for B_k in (4) generates different descent methods. For example, the Gauss-Newton method with $B_k = A_k^T A_K$ and the quasi-Newton methods with B_k being obtained from qausi-Newton updating formulae are well known. Since B_k in the Gauss-Newton method is obtained by neglecting the second part of G_k, the method is expected to perform well when residuals at x^* are small enough or the function $r_i(x), i = 1, 2, \cdots, m$ are close to linear. The quasi-Newton methods such as the BFGS method and the DFP method approximate the whole Hessian matrix by using quasi-Newton update formulae and information obtained from first derivative values. However the quasi-Newton methods do not take account of the special structure of the problem.

Another kind of descent methods for nonlinear least squares is the hybrid method between the Gauss-Newton and the quasi-Newton method[1,10]. Depending upon the outcome of a certain test, the method chooses B_k to be either the Gauss-Newton matrix or the result of applying an updating formula to B_{k-1}. Numerical experiments[10] show that hybrid methods match or improve on the better of the Gauss-Newton and the quasi-Newton methods for every test problem and therefore give reliable, superlinearly convergent methods that contain the best features of both the Gauss-Newton and the quasi-Newton methods.

Since the Jacobian matrix $A(x)$ is usually calculated analytically or numerically in nonlinear lesat squares algorithms, the first portion $A(x)^T A(x)$ of $G(x)$ is always readily available. It is only necessary to approximate the second part of $G(x)$. If we use S_k to denote an approximation to this part, the search direction $d^{(k)}$ can be calculated by

$$(A_k^T A_k + S_k)d^{(k)} = -A_k^T r^{(k)} = -g^{(k)} \tag{9}$$

Updating formulae to generate S_k have been proposed[3,5] and formulated methods are called structured quasi-Newton methods. However, the matrix $A_k^T A_k + S_k$ may be indefinite and it is not clear how to construct updating formula for S_k such that $A_k^T A_k + S_k$ is positive definite.

Yabe and Takahashi[13] proposed a factorized quasi-Newton method, in which the direction d_k is computed by

$$B_k d^{(k)} = (A_k + L_k)(A_k + L_k)^T d^{(k)} = -g^{(k)} \tag{10}$$

and L_k is generated by updating formulae such that $L_k^T L_k + L_k^T A_k + A_k^T L_k$ is an approximation to the second portion of G_k. Variatonal method is used to give updating formulae for L_k and the resulting matrix $B_k = (A_k + L_k)^T(A_k + L_k)$ is either a BFGS-like or a DFP-like updating. The positive definiteness of the matrix B_k is guaranteed when the matrix $A_k + L_k$ is full rank.

In this paper we propose a class of factorized quasi-Newton updating formulae for nonlinear least squares solution using generalized inverse of matrix. The BFGS-like and the DFP-like updating formulae proposed by Yabe and Takahashi[13] are special cases of the class. The derivation of these updating formulae is given in section 2.

In section 3, it is proved that the factorized quasi-Newton updating formulae are also invariant under linear transformation.

Convergence properties of the factorized quasi-Newton methods are discussed in section 4. It is shown that these methods are locally and superlinearly convergent.

Numerical experiments and comparison are presented in section 5. The comparison shows that the sized factorized quasi-Newton method is as efficient and robust as the hybrid methods in [10].

At the rest of this paper, we make the following assumptions for nonlinear least squares problems (1):

(A1): $G(x)$ and $A(x)$ are local Lipschitz continuous at a local solution x^* of the problem (1), that is, there exist a neighbourhood $N(x^*, \epsilon)$ of x^* and constants $L_G \geq 0$ and $L_A \geq 0$ such that

$$\|G(x) - G(x^*)\| \leq L_G \|x - x^*\|, \forall x \in N(x^*, \epsilon) \tag{11}$$

$$\|A(x) - A(x^*)\| \leq L_A \|x - x^*\|, \forall x \in N(x^*, \epsilon) \tag{12}$$

where

$$N(x^*, \epsilon) = \{x \mid \|x - x^*\| \leq \epsilon\}$$

(A2): $G(x^*)$ is positive definite, that is, there exist constants $M \geq m > 0$ such that

$$m \leq \|G(x^*)\| \leq M \tag{13}$$

In this paper $\| \bullet \|$ denotes the 2-norm for vectors or matrices, while $\| \bullet \|_F$ denotes the Frobenius norm of a matrix. Then there exist constants $\eta \geq \tau > 0$ such that

$$\tau \| \bullet \|_F \leq \| \bullet \| \leq \eta \| \bullet \|_F. \tag{14}$$

2. Factorized Quasi-Newton Updating Formulae

In quasi-Newton methods, the matrix B_{k+1} is required to satisfy the *quasi-Newton equation*[8]

$$B_{k+1}\delta^{(k)} = \gamma^{(k)} \tag{15}$$

where

$$\delta^{(k)} = x^{(k+1)} - x^{(k)} \tag{16}$$

$$\gamma^{(k)} = g^{(k+1)} - g^{(k)}. \tag{17}$$

For nonlinear least squares problems, when B_k is choosen to be the form $B_k = (A_k + L_k)^T(A_k + L_k)$, the equation (15) becomes

$$(A_{k+1} + L_{k+1})^T(A_{k+1} + L_{k+1})\delta^{(k)} = \gamma^{(k)}. \tag{18}$$

Another possible choice for $\gamma^{(k)}$ is to use the special structure of problem (1) and to define

$$\gamma^{(k)} = (A_{k+1} - A_k)^T r^{(k+1)} + A_{k+1}^T A_{k+1}\delta^{(k)}. \tag{19}$$

Numerical experiments[10] show that quasi-Newton methods with $\gamma^{(k)}$ in (19) gives better results. It comes from the knowledge of matrix theory[4] that the solution of equation (18) exists if and only if there is an m-vector h such that each of equations

$$(A_{k+1} + L_{k+1})^T h = \gamma^{(k)} \tag{20}$$

$$(A_{k+1} + L_{k+1})\delta^{(k)} = h$$

is consistent and

$$h^T h = {\delta^{(k)}}^T \gamma^{(k)} \tag{21}$$

For convenience, we introduce the following notations

$$\begin{aligned} \widehat{L}_{k+1} &= A_{k+1} + L_{k+1} \\ L_k^{\#} &= A_{k+1} + L_k \\ B_k^{\#} &= {L_k^{\#}}^T L_k^{\#} \end{aligned}$$

The system (20) can be rearranged into

$$\begin{aligned} {\delta^{(k)}}^T \widehat{L}_{k+1}^T &= h^T \\ \widehat{L}_{k+1}^T h &= \gamma^{(k)}. \end{aligned} \tag{22}$$

Suppose $B_k^{\#}$ is nonsingular. Then a possible choice for h is

$$h = aL_k^{\#}\delta^{(k)} + bL_k^{\#}B_k^{\#^{-1}}\gamma^{(k)} \tag{23}$$

where a and b are constants. Then each equation in (22) is consistent and a general solution of (22) can be expressed as[4]

$$\widehat{L}_{k+1} = U^{(1)}V^T + \gamma^{(k)}V^{(1)} - U^{(1)}U\gamma^{(k)}V^{(1)} + \overline{Y} = X_0 + \overline{Y} \tag{24}$$

where $U = \delta^{(k)^T}, V = h, U^{(1)}$ and $V^{(1)}$ are the generalized $\{1\}$-inverses of matrices U and V, respectively, $\overline{Y}$ is any matrix satisfying

$$U\overline{Y} = 0, \ \ \overline{Y}V = 0. \tag{25}$$

Combining (21) and (23) gives an equation for constants a and b

$$a^2\delta^{(k)^T}B_k^{\#}\delta^{(k)} + 2ab\delta^{(k)^T}\gamma^{(k)} + b^2\gamma^{(k)^T}B_k^{\#^{-1}}\gamma^{(k)} = \delta^{(k)^T}\gamma^{(k)}. \tag{26}$$

It can be verified that matrices

$$h^{(1)} = \frac{h^T}{h^Th} = \frac{1}{\delta^{(k)^T}\gamma^{(k)}}(a\delta^{(k)^T}L_k^{\#^T} + b\gamma^{(k)^T}B_k^{\#^{-1}}L_k^{\#^T}) \tag{27}$$

$$(\delta^{(k)^T})^{(1)} = \frac{1}{\delta^{(k)^T}\delta^{(k)}}\delta^{(k)} \tag{28}$$

are the $\{1\}$-inverses of matrices U and V, respectively. Using (26)-(28) and after some manipulation, we have

$$X_0 = a\frac{\gamma^{(k)}\delta^{(k)^T}L_k^{\#^T}}{\delta^{(k)^T}\gamma^{(k)}} + b\frac{\gamma^{(k)}\gamma^{(k)^T}B_k^{\#^{-1}}L_k^{\#^T}}{\delta^{(k)^T}\gamma^{(k)}}. \tag{29}$$

One possible choice for $\overline{Y}$ is

$$\overline{Y} = cL_k^{\#^T}\left(I - \frac{L_k^{\#}\delta^{(k)}\delta^{(k)^T}L_k^{\#^T}}{\delta^{(k)^T}B_k^{\#}\delta^{(k)}}\right) + d\left(I - \frac{\gamma^{(k)}\delta^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}}\right)L_k^{\#^T} \tag{30}$$

It is obvious that $U\overline{Y} = \delta^{(k)^T}\overline{Y} = 0$ is satisfied for any constants c and d. Furthermore if c and d are choosen to satisfy

$$\frac{ad}{\delta^{(k)^T}\gamma^{(k)}} = \frac{bc}{\delta^{(k)^T}B_k^{\#}\delta^{(k)}} \tag{31}$$

then $\overline{Y}V = \overline{Y}h = 0$ is also satisfied. If a constraint

$$c + d = 1 \tag{32}$$

is imposed, then (26), (31) and (32) form a system of 3 linear equations with 4 unknowns, that is, there is one degree of freedom left in the choice of values of parameters a, b, c and d.

Substituting (29), (30) into (24), and using (32) generates a class of updating formulae for L_k:

$$L_{k+1} = L_k + (a-d)\frac{L_k^{\#}\delta^{(k)}\gamma^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}} + b\frac{L_k^{\#}B_k^{\#^{-1}}\gamma^{(k)}\gamma^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}} - c\frac{L_k^{\#}\delta^{(k)}\delta^{(k)^T}B_k^{\#}}{\delta^{(k)^T}B_k^{\#}\delta^{(k)}} \tag{33}$$

The resulting updating formulae for $B_{k+1} = (A_{k+1} + L_{k+1})^T(A_{k+1} + L_{k+1})$ are

$$\begin{aligned}B_{k+1} =& B_k^{\#} - (c^2+2cd)\frac{B_k^{\#}\delta^{(k)}\delta^{(k)^T}B_k^{\#}}{\delta^{(k)^T}B_k^{\#}\delta^{(k)}} - d^2\frac{\gamma^{(k)}\delta^{(k)^T}B_k^{\#} + B_k^{\#}\delta^{(k)}\gamma^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}}\\ &+\left(a^2\frac{\delta^{(k)^T}B_k^{\#}\delta^{(k)}}{\left(\delta^{(k)^T}\gamma^{(k)}\right)^2} + 2ab\frac{1}{\delta^{(k)^T}\gamma^{(k)}} + b^2\frac{\gamma^{(k)^T}B_k^{\#^{-1}}\gamma^{(k)}}{\left(\delta^{(k)^T}\gamma^{(k)}\right)^2}\right.\\ &\left.+d^2\frac{\delta^{(k)^T}B_k^{\#}\delta^{(k)}}{\left(\delta^{(k)^T}\gamma^{(k)}\right)^2}\right)\gamma^{(k)}\gamma^{(k)^T}\\ =&\alpha B_{k+1}^{BFGS} + \beta B_{k+1}^{DFP}\end{aligned} \tag{34}$$

where $\alpha = c^2 + 2cd, \beta = d^2, \alpha + \beta = (c+d)^2 = 1$

$$B_{k+1}^{BFGS} = B_k^{\#} - \frac{B_k^{\#}\delta^{(k)}\delta^{(k)^T}B_k^{\#}}{\delta^{(k)^T}B_k^{\#}\delta^{(k)}} + \frac{\gamma^{(k)}\gamma^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}} \tag{35}$$

$$B_{k+1}^{DFP} = B_k^{\#} - \frac{B_k^{\#}\delta^{(k)}\gamma^{(k)^T} + \gamma^{(k)}\delta^{(k)^T}B_k^{\#}}{\delta^{(k)^T}\gamma^{(k)}} + \left(1 + \frac{\delta^{(k)^T}B_k^{\#}\delta^{(k)}}{\delta^{(k)^T}\gamma^{(k)}}\right)\frac{\gamma^{(k)}\gamma^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}}. \tag{36}$$

The formulae (35) and (36) are the BFGS-like and the DFP-like updating formulae proposed by Yabe and Takahashi[13]. It can be seen from (35) and (36) that if $B_k^{\#}$ is positive definite and $\delta^{(k)^T}\gamma^{(k)} > 0$, then the resulting matrix B_{k+1} in (34) is also positive definite[8].

For choices of values of parameters, we give two special cases. When $a = 0$ is chosen, equations (26), (31) and (32) generate

$$b = \left(\delta^{(k)^T}\gamma^{(k)}/\gamma^{(k)^T}B_k^{\#^{-1}}\gamma^{(k)}\right)^{\frac{1}{2}}, c = 0, \text{and } d = 1.$$

The updating formula for L_k is

$$L_{k+1} = L_k + \frac{L_k^{\#}B_k^{\#^{-1}}\gamma^{(k)}\gamma^{(k)^T}}{\left(\delta^{(k)^T}\gamma^{(k)}.\gamma^{(k)^T}B_k^{\#^{-1}}\gamma^{(k)}\right)^{\frac{1}{2}}} - \frac{L_k^{\#}\delta^{(k)}\gamma^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}} \tag{37}$$

and the corresponding updating formula of B_k is (36). When $b=0$, the values of a,c and d are

$$a=\left(\delta^{(k)^T}\gamma^{(k)}/\delta^{(k)^T}B_k^{\#}\delta^{(k)}\right)^{\frac{1}{2}},\ d=0,\ \text{and } c=1$$

The updating formula for L_k is

$$L_{k+1}=L_k+\frac{L_k^{\#}\delta^{(k)}\gamma^{(k)^T}}{\left(\delta^{(k)^T}\gamma^{(k)}.\delta^{(k)^T}B_k^{\#}\delta^{(k)}\right)^{\frac{1}{2}}}-\frac{L_k^{\#}\delta^{(k)}\delta^{(k)^T}B_k^{\#}}{\delta^{(k)^T}B_k^{\#}\delta^{(k)}} \tag{38}$$

and the corresponding updating formula of B_k is (35). It can be seen from (34)-(36) that when values of parameters a,b,c and d are determined from equations (26), (31) and (32), the matrix lk is updated in such a way that the resulting matrix $B_k=(A_k+L_k)^T(A_k+L_k)$ consists of a convex class of Broyden-like updating formulae and the BFGS-like and the DFP-like updating formulae are two extremes of this convex class.

For small residual problems, the second part of $G(x^{(k)})$ converges to zero as $x^{(k)}$ approaches x^*. In this case, it is necessary for $S_k={L_k}^TL_k+{L_k}^TA_k+{A_k}^TL_k$ to converge to zero, so that the method can be expected to be comparable with the Gauss-Newton method. Since the quasi-Newton updates do not generate zero matrix, sizing strategy can be employed to force the updated matrix converging to zero. Various sizing strategies are available[3,13]. Among them we prefer Biggs sizing factor[3]

$$\beta_k=\min\{r^{(k+1)^T}r^{(k)}/r^{(k)^T}r^{(k)},1\} \tag{39}$$

which is computationally simple and efficient.

With the sizing strategy, the updating formulae (33) and (34) can be reexpressed as

$$\begin{aligned}L_{k+1}=&\beta_kL_k+(a-d)\frac{\widehat{L}_k^{\#}\delta^{(k)}\gamma^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}}+b\frac{\widehat{L}_k^{\#}\widehat{B}_k^{\#-1}\gamma^{(k)}\gamma^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}}\\&-c\frac{\widehat{L}_k^{\#}\delta^{(k)}\delta^{(k)^T}\widehat{B}_k^{\#}}{\delta^{(k)^T}\widehat{B}_k^{\#}\delta^{(k)}}\end{aligned} \tag{40}$$

$$B_{k+1}=\alpha\widehat{B}_{k+1}^{BFGS}+\beta\widehat{B}_{k+1}^{DFP} \tag{41}$$

$$\widehat{B}_{k+1}^{BFGS}=\widehat{B}_k^{\#}-\frac{\widehat{B}_k^{\#}\delta^{(k)}\delta^{(k)^T}\widehat{B}_k^{\#}}{\delta^{(k)^T}\widehat{B}_k^{\#}\delta^{(k)}}+\frac{\gamma^{(k)}\gamma^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}} \tag{42}$$

$$\widehat{B}_{k+1}^{DFP}=\widehat{B}_k^{\#}-\frac{\widehat{B}_k^{\#}\delta^{(k)}\gamma^{(k)^T}+\gamma^{(k)}\delta^{(k)^T}\widehat{B}_k^{\#}}{\delta^{(k)^T}\gamma^{(k)}}+\left(1+\frac{\delta^{(k)^T}\widehat{B}_k^{\#}\delta^{(k)}}{\delta^{(k)^T}\gamma^{(k)}}\right)\frac{\gamma^{(k)}\gamma^{(k)^T}}{\delta^{(k)^T}\gamma^{(k)}} \tag{43}$$

where

$$\begin{aligned}\widehat{L}_k^{\#}=&\beta_kL_k+A_{k+1}\\\widehat{B}_k^{\#}=&\widehat{L}_k^{\#T}\widehat{L}_k^{\#}\end{aligned}$$

3. Invariance

It is well known that quasi-Newton methods with fixed step length α_k or with α_k determined by (7) and (8) are invariant[8] under general linear transformation. When using an invariant method, the performance is not easily upset by a problem in which G is ill-conditioned, because one can implicitly transform to $G = I$ without changing the method. In this section we discuss the invariance of the factorized quasi-Newton methods for nonlinear least squares.

Let T be an $n \times n$ nonsingular matrix, b an n-vector. Consider the linear transformation

$$y = Tx + b. \tag{44}$$

Then $f(x)$ can be regarded as being computed either from x or from y. The derivatives of f with respect to x and the derivatives of f with respect to y have following relations:

$$g_x = T^T g_y \tag{45}$$
$$A_x^T = T^T A_y^T \tag{46}$$
$$G_x = T^T G_y T. \tag{47}$$

With these relations, we have

$$\delta_y^{(k)} = T\delta_x^{(k)} \tag{48}$$
$$\gamma_x^{(k)} = T^T \gamma_y^{(k)}. \tag{49}$$

Theorem 3.1. *Updating formulae* (33) *and* (34) *are invariant under linear transformation* (44).

Proof: At first, we prove the conclusion holds for the BFGS-like updating formula, that is, (38) and (35).

Let

$$\widetilde{L}_1^T = (T^{-1})^T L_1^T \tag{50}$$

Then

$$\widetilde{B}_1 = (\widetilde{A}_1 + \widetilde{L}_1)^T(\widetilde{A}_1 + \widetilde{L}_1) = (T^{-1})^T B_1 T^{-1} \tag{51}$$
$$\widetilde{L}_1^{\#T} = (\widetilde{A}_2 + \widetilde{L}_1)^T = (T^{-1})^T L_1^{\#T} \tag{52}$$

where "∼" is used to denote quantities in y-space. Now we prove, by induction, that following relations hold for all $k \geq 1$.

$$\begin{aligned}
\widetilde{B}_k^{\#} &= (T^{-1})^T B_k^{\#} T^{-1} \\
\widetilde{L}_{k+1}^T &= (T^{-1})^T L_{k+1}^T \\
\widetilde{B}_{k+1} &= (T^{-1})^T B_{k+1} T^{-1} \\
\widetilde{L}_{k+1}^{\#\,T} &= (T^{-1})^T L_{k+1}^{\#\,T}
\end{aligned} \tag{53}$$

For $k = 1$ using (52), we have

$$\widetilde{B}_1^{\#} = \widetilde{L}_1^{\#T}\widetilde{L}_1^{\#} = (T^{-1})^T L_1^{\#T} L_1^{\#} T^{-1} = (T^{-1})^T B_1^{\#} T^{-1} \tag{54}$$

while using (48), (49), (50), (52) and (54) we obtain

$$\begin{aligned}\widetilde{L}_2^T =& \widetilde{L}_1^T + \frac{\widetilde{\gamma}^{(1)}\widetilde{\delta}^{(1)T}\widetilde{L}_1^{\#T}}{\left(\widetilde{\delta}^{(1)T}\widetilde{\gamma}^{(1)}.\widetilde{\delta}^{(1)T}\widetilde{B}_1^{\#}\widetilde{\delta}^{(1)}\right)^{\frac{1}{2}}} - \frac{\widetilde{B}_1^{\#}\widetilde{\delta}^{(1)}\widetilde{\delta}^{(1)T}\widetilde{L}_1^{\#T}}{\widetilde{\delta}^{(1)T}\widetilde{B}_1^{\#}\widetilde{\delta}^{(1)}} \\ =& (T^{-1})\left(L_1^T + \frac{\gamma^{(1)}\delta^{(1)T}L_1^{\#T}}{\left(\delta^{(1)T}\gamma^{(1)}.\delta^{(1)T}B_1^{\#}\delta^{(1)}\right)^{\frac{1}{2}}} - \frac{B_1^{\#}\delta^{(1)}\delta^{(1)T}L_1^{\#T}}{\delta^{(1)T}B_1^{\#}\delta^{(1)}}\right) \\ =& (T^{-1})^T L_2^T.\end{aligned} \tag{55}$$

Then using (46) and (55) generates

$$\begin{aligned}\widetilde{B}_2 =& (\widetilde{A}_2 + \widetilde{L}_2)^T(\widetilde{A}_2 + \widetilde{L}_2) = (T^{-1})^T(A_2 + L_2)^T(A_2 + L_2)T^{-1} \\ =& (T^{-1})^T B_2 T^{-1} \\ \widetilde{L}_2^{\#T} =& (\widetilde{A}_3 + \widetilde{L}_2)^T = (T^{-1})^T(A_3 + L_2)^T = (T^{-1})^T L_2^{\#}\end{aligned}$$

which completes the proof that (53) holds for $k = 1$. In the same way, we can prove that (53) hold for $k = l > 1$ with an inductive assumption being made.

In a similar way, we can prove the conclusion holds for the DFP-like updating formula, that is, (37) and (36). Then the expression in (33) and (34) show that the conclusion of theorem holds and the proof is completed.

Theorem 3.1 indicates that the factorized quasi-Newton methods with step length α_k determined by (7) and (8) or with fixed step length α_k are invariant. So without loss of generality, we assume, at rest of the paper, that

$$G(x^*) = I. \tag{56}$$

4. Convergence Properties

In this section we study the convergence properties of the factorized quasi-Newton methods. As regards global convergence, the method could be equivalent to the famous BFGS method for which global convergence is still an open question. So local convergence properties of the methods are concerned, and superlinear convergence rate is proved under assumptions **(A1)** and **(A2)**. The continuity of $A(x)$ and $G(x)$ implies that there is a neighbourhood, $N(x^*, \epsilon)$ say, of x^* and constants c_1 and β such that

$$\|A(x)\| \le c_1, \forall x \in N(x^*, \epsilon) \tag{57}$$

$$\|G(x)\| \ge \beta, \forall x \in N(x^*, \epsilon) \tag{58}$$

and for any $x^{(k)}, x^{(k+1)} \in N(x^*, \epsilon)$

$$\delta^{(k)T}\gamma^{(k)} = \int_0^1 \delta^{(k)T} G(x^{(k)} + t\delta^{(k)})\delta^{(k)} dt \geq \beta\|\delta^{(k)}\|^2. \tag{59}$$

Thus if $\delta^{(k)} \neq 0$, then $\delta^{(k)T}\gamma^{(k)} > 0$. In this case, as long as the matrix $B_k^{\#}$ is positive definite, the matrix B_{k+1} obtained in (34) is also positive definite[8].

Following lemma is basic for our convergence results.

Lemma 4.1. *Suppose that the conditions* **(A1)** *and* **(A2)** *are satisfied. Let* $\theta \in (0,1)$. *There exist* $0 < \epsilon(\theta) \leq \epsilon$ *and* $\delta(\theta) > 0$ *such that when the point* $x^{(k)}$ *and the matrix* L_k *satisfy*

$$\|x^{(k)} - x^*\| \leq \epsilon(\theta) \tag{60}$$

$$\|(A_k + L_k)^T(A_k + L_k) - I\|_F \leq 2\delta(\theta) \tag{61}$$

then $B_k = (A_k + L_k)^T(A_k + L_k)$ *is positive definite and the point*

$$x^{(k+1)} = x^{(k)} - B_k^{-1}A_k^T r^{(k)} \tag{62}$$

is well-defined for $A_k^T r^{(k)} \neq 0$, *and*

$$\|x^{(k+1)} - x^*\| \leq \theta\|x^{(k)} - x^*\| \tag{63}$$

Furthermore, the matrix $B_k^{\#}$ *is also positive definite and*

$$\begin{aligned}\|(A_{k+1}+&L_{k+1})^T(A_{k+1} + L_{k+1}) - I\|_F \\ &\leq(1 + \mu\|x^{(k)} - x^*\|)\|B_k - I\|_F + \mu\|x^{(k)} - x^*\|\end{aligned} \tag{64}$$

where L_{k+1} *is obtained from* L_k *by using updating formula* (34) *with parameters* $c+d = 1$ *and* μ *is a positive constant.*

Proof: For given $\theta \in (0,1)$, choose $\epsilon(\theta)$ and $\delta(\theta)$ small enough such that

$$2\eta\delta(\theta) \leq \theta/(1+\theta) < 1 \tag{65}$$

$$(1+\theta)(L_G\epsilon(\theta) + 2\eta\delta(\theta)) \leq \theta \tag{66}$$

$$2(1+\theta)^2((2\eta\delta(\theta)+1)^{\frac{1}{2}} + 2c_1)L_A\epsilon(\theta) \leq \omega < 1 \tag{67}$$

Since $\|I\|\|B_k - I\| \leq 2\eta\delta \leq \theta/(1+\theta) < 1$ the pertubation lemma[12] implies B_k is nonsingular and

$$\|B_k^{-1}\| \leq 1 + \theta = c_2. \tag{68}$$

Hence B_k is positive definite and the iterate (62) is well-defined when $A_k^T r^{(k)} \neq 0$.

Using $g^* = 0$ and $G^* = I$ we have, with (66)

$$\begin{aligned}\|x^{(k+1)} - x^*\| &= \|x^{(k)} - B_k^{-1}g^{(k)} - x^*\| \\ &\leq \|B_k^{-1}\| \left[\|g^{(k)} - g^* - (x^{(k)} - x^*)\| + \|B_k - I\|\|x^{(k)} - x^*\|\right] \\ &\leq c_2(L_G\epsilon(\theta) + 2\eta\delta(\theta))\|x^{(k)} - x^*\| \\ &\leq \theta\|x^{(k)} - x^*\|.\end{aligned}$$

Since

$$\|B_k\| \le \|B_k - I\| + \|I\| \le 2\eta\delta(\theta) + 1 = c_3 \tag{69}$$
$$\|L_k\| \le \|L_k + A_k\| + \|A_k\| \le c_3^{\frac{1}{2}} + c_1 = c_4 \tag{70}$$

we obtain

$$\begin{aligned}\|B_k^{\#} - B_k\| \le & 2\|L_k\|\|A_{k+1} - A_k\| + (\|A_{k+1}\| + \|A_k\|)\|A_{k+1} - A_k\| \\ \le & 2(c_4 + c_1)L_A\|x^{(k+1)} - x^{(k)}\| \\ \le & 2(c_4 + c_1)L_A\epsilon(\theta)(1+\theta) \end{aligned}\tag{71}$$

and

$$\|B_k^{-1}\|\|B_k^{\#} - B_k\| \le 2(1+\theta)^2(c_4 + c_1)L_A\epsilon(\theta) \le \omega. \tag{72}$$

Then the pertubation lemma again gives that $B_k^{\#}$ is nonsingular, hence positive definite, and

$$\|B_k^{\#-1}\| \le \|B_k^{-1}\|/(1-\omega) \tag{73}$$
$$\|B_k^{\#}\| \le \|B_k^{\#} - B_k\| + \|B_k\| \le 2(1+\theta)^2(c_4 + c_1)L_A\epsilon(\theta) + c_3 = c_5. \tag{74}$$

Now we denote the updating formula (34) by $B_{k+1} = \mathrm{upd}(B_k^{\#}, \delta^{(k)}, \gamma^{(k)})$ and set a matrix $B'_{k+1} = \mathrm{upd}(B_k^{\#}, \delta^{(k)}, \delta^{(k)})$. Then following relations[11] hold among matrices B_{k+1}, B'_{k+1} and $B_k^{\#}$,

$$\|B_{k+1} - B'_{k+1}\|_F \le (\|B_k^{\#} - I\|_F + 2)O(\|x^{(k)} - x^*\|) \tag{75}$$
$$\|B'_{k+1} - I\|_F \le \|B_k^{\#} - I\|_F. \tag{76}$$

With these two relations and (71), we have

$$\begin{aligned}\|B_{k+1} - I\|_F \le & \|B_{k+1} - B'_{k+1}\|_F + \|B'_{k+1} - I\|_F \\ \le & (1 + O(\|x^{(k)} - x^*\|))\|B_k^{\#} - I\|_F + O(\|x^{(k)} - x^*\|) \\ \le & (\|B_k^{\#} - B_k\|_F + \|B_k - I\|_F)(1 + O(\|x^{(k)} - x^*\|)) + O(\|x^{(k)} - x^*\|) \\ \le & (1 + O(\|x^{(k)} - x^*\|))\|B_k - I\|_F + O(\|x^{(k)} - x^*\|) \\ \le & (1 + \mu\|x^{(k)} - x^*\|)\|B_k - I\|_F + \mu\|x^{(k)} - x^*\| \end{aligned}\tag{77}$$

where μ is large enough such that $O(\|x^{(k)} - x^*\|) \le \mu\|x^{(k)} - x^*\|$.

With Lemma 4.1, the convergence result of the factorized quasi-Newton methods can be described as follows.

Theorem 4.2. *Suppose that the assumptions of Lemma* 4.1 *are satisfied, and* $\epsilon(\theta)$ *and* $\delta(\theta)$ *are small enough such that besides* (65)-(67),

$$(2\mu\delta(\theta) + \mu)\epsilon(\theta) \le \delta(\theta)(1-\theta)$$

is satisfied. Then for any initial point

$$x^{(1)} \in N_1 = \{x \mid \|x - x^*\| \le \epsilon(\theta)\} \tag{78}$$

and any initial matrix

$$L_1 \in N_2 = \{L \mid \|(A_1 + L)^T (A_1 + L) - I\|_F \le \delta(\theta)\} \tag{79}$$

the sequence $\{x^{(k)}\}$ *generated by* (62) *is well defined and linearly convergent to* x^* *at a rate*

$$\|x^{(k+1)} - x^*\| \le \theta \|x^{(k)} - x^*\| \quad k = 1, 2, \cdots \tag{80}$$

where updating formula (33) *is used to generate* $\{L_k\}$. *Furthermore, sequences* $\{\|B_k\|\}$, $\{\|B_k^{-1}\|\}$, $\{\|B_k^{\#}\|\}$ *and* $\{\|B_k^{\#^{-1}}\|\}$ *are uniformly bounded.*

Proof: By Lemma 4.1, we only need to prove that (60) and (61) hold at each iteration. This can be done by induction.

By assumption made in theorem, (60) and (61) hold for $k = 1$. We prove that (60) and (61) also hold for $k = 2$. By Lemma 4.1, (61) is obviously satisfied for $k = 2$. From (79), (64) and (77), we have

$$\begin{aligned} \|B_2 - I\|_F &\le \|B_1 - I\|_F + (2\mu\delta(\theta) + \mu)\|x^{(1)} - x^*\| \\ &\le \delta(\theta) + (2\mu\delta(\theta) + \mu)\epsilon(\theta) \le 2\delta(\theta) \end{aligned}$$

Hence (60) and (61) hold for $k = 2$.

Now, assume that (60) and (61) hold for $k = 1, 2, \cdots, l$ we prove (60) and (61) also hold for $k = l + 1$. By Lemma 4.1, (60) is obviously satisfied. By inductive assumption, (64) holds for $k = 1, 2, \cdots, l$. Using (64), (77) and (79), we obtain

$$\|B_{l+1} - I\|_F - \|B_l - I\|_F \le (2\mu\eta\delta(\theta) + \mu)\theta^{l-1}\|x^{(1)} - x^*\|$$

. Summing both sides from $k = 1$ to $l - 1$ generates

$$\begin{aligned} \|B_{l+1} - I\|_F &\le \|B_1 - I\|_F + (2\mu\delta(\theta) + \mu)\|x^{(1)} - x^*\| \sum_{k=1}^{l-1} \theta^k \\ &\le \delta(\theta) + (2\mu\delta(\theta) + \mu)\epsilon(\theta)/(1 - \theta) \le 2\delta(\theta). \end{aligned}$$

Hence both (60) and (61) hold for $k = l + 1$, which completes the inductive proof. Then Lemma 4.1 shows that the sequence $\{x^{(k)}\}$ is linearly convergent and (68), (69), (73) and (74) tell that sequences $\{\|B_k\|\}, \{\|B_k^{-1}\|\}, \{\|B_k^{\#}\|\}$ and $\{\|B_k^{\#^{-1}}\|\}$ are uniformly bounded.

Finally, by a similar way to the proof of Proposition 4 in [11] we can obtain

$$\lim_{k \to \infty} \frac{\|(B_k - I)\delta^{(k)}\|}{\|\delta^{(k)}\|} = 0.$$

Then by Theorem 2.2 in [7], following superlinear convergence result is obtained.

Theorem 4.3 *Suppose that all the conditions of Theorem* 4.2 *hold. Then the sequence generated by* (62) *converges superlinearly to* x^* *that is*

$$\lim_{k\to\infty} \frac{\|x^{(k+1)} - x^*\|}{\|x^{(k)} - x^*\|} = 0.$$

5. Numerical Results

Numerical experiments have been performed on a personal computer, and 27 test problems are used to compare the factorized quasi-Newton methods with various available methods. The information about these test problems are given in Table 1. The test problems are ordered in optimal values of functions, with "Z", "S" and "L" being used to denote zero, small and large residual problems, respectively. The first column of the table gives names of tested problems where Osborn 1 stands for the first Osborn test problem, Freudstein(1) and Freudstein(2) denote the Freudstein test problem with different initial points, Chebyquad[6] and Watson[6] denote Chebyquad and Watson test problems with 6 variables, while Signo.[6-2] stands for the Signomial test problem[1] with $m = 6$ and $n = 2$, and so on.

Following methods are compared on the test problems:

GN: Gauss-Newton method

BFGS: normal BFGS method

FX: hybrid method[10] between GN and BFGS

F-BFGS: factorized BFGS method (with $c = 1, d = 0$)

F-Broyden: factorized Broyden method (with $c = d = \frac{1}{2}$)

S-F-BFGS: sized factorized BFGS method

S-F-Broyden: sized factorized Broyden method

In the implementation of these methods, search direction $d^{(k)}$ is computed from equation (1.4) with different choices of B_k. In GN, F-BFGS, F-Broyden, S-F-BFGS and S-F-Broyden methods, this is done by forming LL^T factors of the matrices ${A_k}^T A_k$ and $(A_k + L_k)^T(A_k + L_k)$ using QR factorization technique. In normal BFGS method LDL^T factors of the matrix B_k is updated by calling the subroutine MC11A[9]. The line search method[2] is employed to determine a step length. This line search method is specially designed for nonlinear least squares solution. Line searches are terminated when α_k satisfies both the conditions (7) and (8) with parameter values $\rho = 0.01$ and $\sigma = 0.1$. The iteration is terminated when

(1) $g^{(k)} = 0$ or

(2) $f^{(k)} - f^{(k+1)} \leq 10^{-8} \max(1, f^{(k)})$ or

(3) The prreset maximum number (100) of iteration is approached

Table 1. Test Problems

Problem	m	n	$x_{(0)}^T$	Optimal value	Type
Woods	7	4	(-3,-1,-3,-1)	0	Z
Eugvall	5	3	(1,2,0)	0	Z
Helix	3	3	(-1,0.001,0.001)	0	Z
Box	10	3	(0,10,20)	0	Z
Beale	3	2	(0.1,0.1)	0	Z
Freudstein(2)	2	2	(6,6)	0	Z
Rosenbrok	2	2	(-1.2,1)	0	Z
Singular	4	4	(3,-1,0,1)	0	Z
Chebyquad[6]	6	6	$(1,2,3,\cdots,6)/7$	0	Z
Chebyquad[9]	9	9	$(1,2,3,\cdots,9)/10$	0	Z
Osborn 1	33	5	(.5,1.5,-1,.01,.02)	.5464804E-4	S
Kowa.&Osb.	11	4	(.25,.39,.415,.39)	.3075055E-3	S
Watson[6]	31	6	$(0,0,\cdots,0)$	.2287659E-2	S
Chebyquad[8]	8	8	$(1,2,3,\cdots,8)/9$	.3516872E-2	S
Chebyquad[10]	10	10	$(1,2,3,\cdots,10)/11$	.4772715E-2	S
Bard	15	3	(1,1,1)	.8214878E-2	S
Madsen	3	2	(3,1)	.773199	S
Freudstein(1)	2	2	(15,-2)	.4898425E+2	L
Meyer(2)	16	3	(0.005,6140,340)	.8793119E+2	L
Jennrich	10	2	(0.3,0.4)	.1243622E+3	L
Modified Box	10	3	(0,10,20)	.3073099E+3	L
Signo.[6-2]	6	2	Random generate	.894283E+4	L
Signo.[10-2]	10	2	Random generate	.1218381E+5	L
Signo.[18-6]	18	6	Random generate	.4540915E+5	L
Signo.[12-4]	12	4	Random generate	.7746263E+5	L
Brown	20	4	(25,5,-5,-1)	.8582217E+5	L
Signo.[20-4]	20	4	Random generate	.158455E+6	L

A FORTRAN 77 program with single precision was coded to carry out these experiments. Numerical results are given in Table 2, where F denotes failure of convergence at a local solution. Each entry in the table contains the number of iterations, function evaluations and gradient evaluations required to terminate the iteration. From the table, we can see that the factorized quasi-Newton methods are as robust as the BFGS and the hybrid methods, that is, the factorized quasi-Newton methods solve almost all the test problems. Especially, the sized factorized Broyden method is comparable with the hybrid method FX, which is known to be the currently most preferred nonlinear least squares method[10].

Table 2. Numerical results

Problem	GN	BFGS	FX	F-BFGS	F-Broyden	S-F-BFGS	S-F-Broyden
Woods	43 83 45	F	F	F	43 86 45	40 77 42	41 79 42
Eugvall	9 17 10	12 27 14	9 17 10	16 32 17	14 31 18	12 20 13	12 20 14
Helix	5 8 6	7 11 8	5 8 6	5 10 6	6 11 7	6 10 7	6 10 7
Box	5 9 7	8 12 9	5 9 7	5 7 6	5 7 6	5 7 6	5 7 6
Beale	5 10 7	7 21 10	7 20 11	8 19 10	8 21 10	6 17 8	6 17 8
Freudstein(2)	F	7 11 7	6 13 6	6 9 6	5 9 5	6 9 6	6 10 6
Rosenbrok	18 27 18	15 26 16	16 28 16	15 26 16	15 30 16	14 25 15	14 25 15
Singular	4 8 7	11 24 17	4 8 7	9 14 11	10 20 14	10 16 13	10 16 13
Chebyquad[6]	5 16 7	8 19 10	5 16 7	6 17 8	6 16 8	5 15 7	5 15 7
Chebyquad[9]	F	14 26 18	6 11 8	15 47 26	10 42 11	8 15 9	9 16 10
Osborn 1	9 15 9	23 44 27	10 19 12	23 42 30	F	21 44 28	19 41 29
Kowa.&Osb.	6 19 9	10 22 12	6 17 8	6 17 8	6 17 8	6 17 8	6 17 8
Watson[6]	6 8 6	31 83 35	6 8 7	12 20 12	10 14 10	8 12 9	7 9 7
Chebyquad[8]	F	10 23 13	13 35 16	15 39 18	12 30 15	13 34 17	12 28 14
Chebyquad[10]	F	14 30 20	10 25 12	36 75 38	13 31 16	26 57 29	13 28 15
Bard	5 7 7	8 15 9	5 7 7	6 9 7	6 9 7	6 8 7	6 8 7
Madsen	5 10 5	6 9 6	5 16 6	5 8 5	5 7 5	5 11 5	4 12 4
Freudstein(1)	5 6 5	6 9 6	5 6 5	5 8 6	5 8 6	5 7 6	5 7 6
Meyer(2)	4 11 4	6 19 6	4 11 4	6 15 6	9 59 5	19 99 21	6 60 8
Jennrich	F	8 24 10	6 16 6	8 23 8	9 25 9	7 20 9	8 15 8
Modified Box	6 29 10	13 45 10	9 46 11	13 39 18	13 37 16	12 44 17	12 46 17
Signo.[6-2]	3 8 3	4 12 4	3 8 3	4 8 4	5 17 5	4 10 4	4 12 4
Signo.[10-2]	8 20 8	9 24 9	8 22 8	7 20 9	8 19 10	8 14 9	10 24 11
Signo.[18-6]	26 53 29	25 54 26	17 39 19	14 35 15	14 37 16	14 39 15	15 31 15
Signo.[12-4]	18 51 18	14 38 16	9 18 9	11 42 13	13 34 14	12 36 14	12 39 12
Brown	F	10 20 11	10 22 11	11 22 12	12 22 12	11 20 11	12 20 12
Signo.[20-4]	21 47 23	29 71 34	11 25 12	21 55 24	27 76 32	16 41 16	16 39 18

6. Conclusion

Efficient and robust descent methods for nonlinear least squares problems are considered in this paper. The search direction $d^{(k)}$ is computed using (4). A class of updating formulae for generating L_k is derived such that ${L_k}^T L_k + {L_k}^T A_k + {A_k}^T L_k$ approximates the second part of $G(x^{(k)})$. Generalized inverse method is used and L_k is updated in such a way that the resulting updating formula for B_k is the convex class of the Broyden-like updating formula. Sizing technique is employed to enforce the approximation to approach zero for zero and small residual problems. It is shown that the factorized quasi-Newton methods with fixed step length or with a step length determined by (7) and (8) are still invariant. When the condition of full rank on matrix $A_k + L_k$ is imposed (Lemma 4.1), the matrix B_k is positive definite and the search direction $d^{(k)}$ is descent. Local superlinear convergence property is analyzed. Numerical experiments show that the sized factorized Broyden method is as efficient as the hybrid method FX, that is, the method matches and improves the performance of the better of the GN and the BFGS methods in almost all test problems.

After we finished this paper, it is pointed out to us by Prof. Yuan[15] that similar resuts have been obtained by Yabe and Yamaki[14].

References

[1] M. Al-Baali and R. Fletcher, Variational methods for nonlinear least squares, *Journal of the Operation Research Society*, **36**:5 (1985), 405-421.

[2] M. Al-Baali and R. Fletcher, An efficient line search for nonlinear least squares, *Journal of Optimization Theory and Applications*, **48**:3 (1986), 359-377.

[3] M.C. Bartholomew-Biggs, The estimation of the Hessian matrix in nonlinear least squares problems with non-zero residuals, *Mathematical Programming*, 12 (1977), 67-80.

[4] A. Ben-Israel and T.N.E. Greville, *Generalized Inverse: Theory and Application*, John Wiley & Sons, 1974.

[5] J.E. Dennis, Jr., D.M. Gay and R.E. Welsch, An adaptive nonlinear least-squares algorithm, *ACM Transactions on Mathematical Software*, **7**:3 (1981), 348-368.

[6] J.E. Dennis, Jr., H.J. Martinez and R.A. Tapia, Convergence theory for structured BFGS secant method with an application to nonlinear least squares, *Journal of Optimization Theory and Applications*, **61**:2 (1989), 161-178.

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

[8] R. Fletcher, Practical Methods of Optimization, Vol. 1, Unconstrained Optimization, Chichester: Wiley, 1980.

[9] R. Fletcher and M.J.D. Powell, On the modification of LDL^T factorizations, *Mathematics of Computation*, 28 (1974), 1067-1087.

[10] R. Fletcher and C. Xu, Hybrid methods for nonlinear least squares, *IMA Journal of Numerical Analysis*, 7 (1987), 371-389.

[11] A. Griewank and Ph.L. Toint, Local convergence analysis for partitioned quasi-Newton updates, *Numerische Mathematik*, 39 (1982), 429-448.

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

[13] H. Yabe and T. Takahashi, Factorized quasi-Newton methods for nonlinear least squares problems, *Mathematical Programming*, 51 (1991), 75-100.

[14] H. Yabe and N. Yamaki, Convergence of a factorized Broyden-like family for nonlinear least squares problems, in: K. Tanabe ed., Optimazation: Modeling and Algorithms 4, The institute of statistical mathematics, cooperative research report 53, Tokyo, Japan, 1994.

[15] Y. Yuan, private communication, April, 1995.