数理解析研究所講究録
第 717 巻 1990 年 144-158

Numerical Comparison among Structured Quasi-Newton Methods for Nonlinear Least Squares Problems
(非線形最小２乗問題に対するSongbai and Zhihong法の変種について)

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

1. Introduction

This paper is concerned with the problem of minimizing a sum of squared nonlinear functions

(1.1) $F(x) = (1/2)\ \Sigma_{i=1}^{m}(f_i(x))^2, \qquad m \geqq n$

where $f_i : R^n \rightarrow R$ are twice continuously differentiable for $i = 1,\ldots,m$.

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 solving

(1.2) $\nabla^2 F(x_k)\ d_k = -\nabla F(x_k)$

and the new point is generated by

(1.3) $x_{k+1} = x_k + d_k$.

Here x_k is the current estimate of the minimum point x^* and ∇F, $\nabla^2 F$ are given by

(1.4) $\nabla F(x) = J(x)^T f(x)$,

(1.5) $\nabla^2 F(x) = J(x)^T J(x) + \Sigma_{i=1}^{m} f_i(x) \nabla^2 f_i(x)$,

where

(1.6) $f(x) = (f_1(x),\ \ldots\ ,\ f_m(x))^T$

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

For the problem (1.1), quasi-Newton approximations to only the second part of the Hessian matrix have been developed [3],[4]. Recently, two robust algorithms have been proposed by Bartholomew-Biggs[1] and Dennis, Gay and Welsch[5].

In Section 3, we present factorized versions of structured quasi-Newton methods, and derive a BFGS-like and a DFP-like update, which were first given by Yabe and Takahashi[9]. On the other hand, Songbai and Zhihong[7] have been studying factorized versions of structured quasi-Newton methods independently of us. In Section 4, the Songbai and Zhihong method is presented. In Section 5, we establish sizing techniques. Further, a factorized algorithm is given in Section 6. Finally, some computational experiments are described in order to investigate the effectiveness of several structured quasi-Newton methods. Throughout this paper, $\| \cdot \|$ denotes the 2-norm for vectors or matrices and $\| \cdot \|_F$ denotes the Frobenius norm.

2. Structured Quasi-Newton Methods for Nonlinear Least Squares Problems

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 solving

(2.1) $\qquad (J_k^T J_k + A_k) d_k = -J_k^T f_k,$

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 [4]. The matrix A_k is updated such that the new matrix A_{k+1} satisfies the secant condition

(2.2) $\qquad A_{k+1} s_k = y_k - J_{k+1}^T J_{k+1} s_k$

or

(2.3) $\qquad A_{k+1} s_k = v_k, \quad v_k = (J_{k+1} - J_k)^T f_{k+1},$

where

(2.4) $\qquad s_k = x_{k+1} - x_k, \quad y_k = \nabla F_{k+1} - \nabla F_k.$

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. Their updates are as follows:

(i)the Biggs update

(2.5) $\qquad A_{k+1} = \beta_k A_k + (v_k - \beta_k A_k s_k)(v_k - \beta_k A_k s_k)^T / (v_k - \beta_k A_k s_k)^T s_k,$

(2.6) $\qquad \beta_k = f_{k+1}^T f_k \,/\, f_k^T f_k,$

(ii)the DGW update

(2.7) $\qquad A_{k+1} = \beta_k A_k + ((v_k - \beta_k A_k s_k) y_k^T + y_k (v_k - \beta_k A_k s_k)^T) / s_k^T y_k$
$\qquad\qquad - \{s_k^T (v_k - \beta_k A_k s_k) / (s_k^T y_k)^2\} y_k y_k^T,$

(2.8) $\qquad \beta_k = \min(| s_k^T v_k / s_k^T A_k s_k |, 1),$

where β_k is a sizing factor.

3. Factorized Versions of Structured Quasi-Newton Methods

For general quasi-Newton methods, the hereditary positive definiteness property is desirable because a descent search direction for objective function is obtained. On the other hand, for structured quasi-Newton updates, it is not clear how to construct updating formulae for 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)[5] and switching to the Gauss-Newton method.

In [9],we proposed a direct approach which maintains positive definiteness of the coefficient matrix in (2.1). We compute the search direction by solving the linear system of equations

(3.1) $(L_k + J_k)^T(L_k + J_k)\, d_k = -J_k{}^T f_k$,
where the matrix L_k is an $m \times n$ correction matrix to the Jacobian matrix such that $L_k{}^T L_k + L_k{}^T J_k + J_k{}^T L_k$ is the k-th approximation to the second part of the Hessian matrix of F. Since the coefficient matrix is expressed by the factorized form, the search direction may be expected to be a descent direction for F. Successful updates for L_k would lead to simplified line search algorithms in contrast to the more complex trust region algorithms.

The secant condition for L_{k+1} is as follows:

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

where

(3.3) $z_k = y_k$

or

(3.4) $z_k = v_k + J_{k+1}{}^T J_{k+1} s_k$,

and the vectors v_k, s_k and y_k are given in (2.3) and (2.4), respectively. Then we have two types of updates as follows:

(i)the BFGS-like update

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

(ii)the DFP-like update

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

where

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

The local and q-superlinear convergence of these methods are shown by the following theorem[8].

Theorem. Let D be the open convex subset of R^n which contains the minimum point x^*. Suppose that there exist positive constants ξ_1, ξ_2 and p such that

(3.8) $\| \nabla^2 F(u) - \nabla^2 F(x^*) \| \leq \xi_1 \| u - x^* \|^p$ for any u in D,

(3.9) $\| J(u_1) - J(u_2) \| \leq \xi_2 \| u_1 - u_2 \|^p$ for any u_1 and u_2 in D,

and that $\nabla^2 F$ is symmetric positive definite at x^*. Let the matrix L_k be updated by the BFGS-like formula (3.5) or the DFP-like formula (3.6), where z_k is given by (3.3) or (3.4). Then the sequence $\{x_k\}$ generated by

(3.10) $x_{k+1} = x_k - ((L_k + J_k)^T(L_k + J_k))^{-1} J_k{}^T f_k$.

converges to the minimum point x^* locally and q-superlinearly.

4. Songbai and Zhihong method

Independently of us, Songbai and Zhihong [7] have been studying factorized versions of structured quasi-Newton methods. They proposed the approximation of f(x) around x_k as follows;

(4.1) $m(x_k + d) = f_k + (J_k + L_k)d$.

If $L_k = 0$, then the above is reduced to the Gauss-Newton model. In which case, the model function to be minimized is $(1/2)\| m(x_k + d)\|^2$, and the search direction d_k is given by solving the normal equation

(4.2) $(J_k + L_k)^T(J_k + L_k)d = -(J_k + L_k)^T f_k$.

Since this does not correspond to the Newton equation (1.2),they imposed the condition $L_k{}^T f_k = 0$ on a matrix L_k, in addition to the secant condition. So the conditions which the matrix L_{k+1} should satisfy are as follows:

(4.3) $(L_{k+1}+J_{k+1})^T(L_{k+1}+J_{k+1})\, s_k = z_k$

and

(4.4) $L_{k+1}{}^T f_{k+1} = 0$

where

(4.5) $z_k = (J_{k+1} - J_k)^T f_{k+1} + J_{k+1}{}^T J_{k+1} s_k$.

It is easily shown that, for nonzero s_k and z_k, the matrix equation (4.3) is consistent if and only if

(4.6) (a) $L_{k+1}{}^T h = z_k - J_{k+1}{}^T h$ and (b) $L_{k+1} s_k = h - J_{k+1} s_k$

for some m-dimensional vector h. Further, the matrix equations (4.6) have a common solution L_{k+1} if and only if each equation separately has a solution and $h^T h = s_k{}^T z_k$ [2, Chapter 2]. So the purpose is to find a rectangular matrix L_{k+1} which satisfies the equations (4.4) and (4.6) under the assumption of $s_k{}^T z_k > 0$.

In the following, we construct a updating formula which corresponds to the BFGS update, by a slight different way from Songbai and Zhihong[7]. Now we drop the suffix k and replace the suffix (k+1) by '+' for simplicity of notation. Assume that $f_+ \neq 0$. For a matrix M,let R(M) denote a space spanned by column vectors of M. Then we can consider the following two cases:

(Case1)

When h is contained in $R(f_+)$, h is represented by

(4.7) $h = \pm(s^T z)^{1/2} f_+ / \| f_+ \|$.

If $z - J_+{}^T h \neq 0$, then the matrix equations (4.4) and (4.6.a) are inconsistent. Otherwise, since (4.6.a) is equivalent to (4.4), the conditions are reduced to the expressions

(4.8) $L_+ s = \pm(s^T z)^{1/2} f_+ / \| f_+ \| - J_+ s$ and $L_+{}^T f_+ = 0$.

(Case2)

When h is not contained in $R(f_+)$, we can consider a least change secant update following to Songbai and Zhihong[7]. For any unknown m-dimensional vector h such that $h^T h = s^T z$, minimizing the Frobenius norm $\| L_+{}^T - L^T \|_F$ with respect to L_+, subject to $L_+{}^T h = z - J_+{}^T h$ and $L_+{}^T f_+ = 0$, we have a unique solution

(4.9) $L_+ = PL + Ph(z - M^T h)^T / \| Ph \|^2$, where

(4.10) $M = PL + J_+$ and $P = I - f_+ f_+^T / f_+^T f_+$.

By substituting the above for the other condition (4.6.b), we have

(4.11) $(1-(z-M^T h)^T s/\|Ph\|^2)h = Ms - \{(z-M^T h)^T s \cdot f_+^T h/(\|Ph\|^2 \|f_+\|^2)\} f_+$.

Then we can further consider two cases:

(Case2-1)

When Ms is not contained in $R(f_+)$, $1-(z-M^T h)^T s/\|Ph\|^2 \neq 0$. In fact, for an m-dimensional vector h such that $(z-M^T h)^T s = \|Ph\|^2$, the left-hand side of (4.11) becomes zero, on the other hand, the right-hand side of (4.11) becomes $Ms-(f_+^T h/\|f_+\|^2)f_+ \neq 0$, which is a contradiction. Consequently, h can be represented by the form

(4.12) $h = \tau_1 Ms + \tau_2 f_+$, $\tau_1 \neq 0$.

Substituting the above for the expression (4.11), we have

$$\{1-(s^T z - \tau_1 s^T M^T Ms - \tau_2 s^T J_+^T f_+)/(\tau_1^2 \|PMs\|^2)\}(\tau_1 Ms + \tau_2 f_+) = Ms - \{(s^T z - \tau_1 s^T M^T Ms - \tau_2 s^T J_+^T f_+)\cdot(\tau_1 s^T J_+^T f_+ + \tau_2 \|f_+\|^2)/(\tau_1^2 \|PMs\|^2 \|f_+\|^2)\} f_+.$$

By arranging the coefficients of the vectors Ms and f_+, and using the linear independence of Ms and f_+,

(4.13.a) $(s^T z - \tau_1 s^T M^T Ms - \tau_2 s^T J_+^T f_+)/(\tau_1 \|PMs\|^2) = \tau_1 - 1$,

(4.13.b) $(s^T z - \tau_1 s^T M^T Ms - \tau_2 s^T J_+^T f_+) f_+^T J_+ s/(\tau_1 \|PMs\|^2 \|f_+\|^2) = -\tau_2$.

Then we have

$$\tau_2 = (1-\tau_1) f_+^T J_+ s/\|f_+\|^2.$$

Substituting the above for the expression (4.13.a) and setting $\tau = \tau_1$, we have the quadratic equation of τ

$$\|PMs\|^2 \tau^2 + (\|Ms\|^2 - \|PMs\|^2 - (f_+^T J_+ s)^2/\|f_+\|^2)\tau + ((f_+^T J_+ s)^2/\|f_+\|^2 - s^T z) = 0,$$

which yields

(4.14) $\|PMs\|^2 \tau^2 = s^T z - (f_+^T J_+ s)^2/\|f_+\|^2$.

If $s^T z - (f_+^T J_+ s)^2/\|f_+\|^2 \geqq 0$, then the above can be solved and we obtain

(4.15) $h = \tau Ms + (1-\tau)(f_+^T J_+ s/\|f_+\|^2) f_+$,

which corresponds to the Songbai and Zhihong updating formula.

(Case2-2)

When Ms is contained in $R(f_+)$, Ms is formed by $Ms = (f_+^T Ms/\|f_+\|^2) f_+$. Thus, it follows from (4.11) that

(4.16) $\{(h^T f_+)(Ms-h)^T f_+\} h = \{(f_+^T Ms)\|Ph\|^2 - (z-M^T h)^T s \cdot (f_+^T h)\} f_+$.

Since h is independent of f_+, the coefficients of the both sides should be zero. Therefore, we can choose a vector such that

(4.17) $(Ms-h)^T f_+ = 0$, i.e., $f_+^T h = f_+^T Ms$.

Then, noting the condition $h^T h = s^T z$, we have a general solution of (4.17)

(4.18) $h = Ms + (s^T z - \|Ms\|^2)^{1/2} Pu/\|Pu\|$, where P is an orthogonal projection matrix (4.10) onto the null space of f_+ and u is an n-dimensional arbitrary vector which is not contained in $R(f_+)$.

Finally, summarizing (Case2-1) and (Case2-2), we obtain the following updating formula:

(4.19) $L_{k+1} = P_k L_k + P_k h_k (z_k - M_k^T h_k)^T / \| P_k h_k \|^2$,

(4.20) $h_k = (f_{k+1}^T J_{k+1} s_k / \| f_{k+1} \|^2) f_{k+1} + \rho_k P_k w_k / \| P_k w_k \|$,

(4.21) $P_k = I - f_{k+1} f_{k+1}^T / \| f_{k+1} \|^2$,

(4.22) $M_k = P_k L_k + J_{k+1}$,

(4.23) $\rho_k^2 = s_k^T z_k - (f_{k+1}^T J_{k+1} s_k)^2 / \| f_{k+1} \|^2$,

where

(4.24) $w_k = M_k s_k$ if $M_k s_k$ is not contained in $R(f_{k+1})$, i.e., $\| P_k M_k s_k \| \neq 0$,
otherwise, w_k is chosen to be a linear independent vector to f_{k+1}.

Note that the update with $w_k = M_k s_k$ coincides with the Songbai and Zhihong update by choosing a positive ρ_k.

5. Sizing techniques of the updating matrix

We know that,for zero residual problems,the matrices A_k and $L_k^T L_k + L_k^T J_k + J_k^T L_k$ 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. Since the quasi-Newton updates do not generate the zero matrix, some remedies must be employed. For example, Songbai and Zhihong proposed the switching to the Gauss-Newton method. Among remedies, the sizing of the updating matrices which has been introduced by Bartholomew-Biggs[1] or Dennis et al.[5] seems most promising.

The structured quasi-Newton methods with the sizing factors (2.6) and (2.8) may be 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 derived by using the secant condition (2.3).

Now we can apply the above mentioned techniques to our factorized versions and the Songbai-Zhihong update. Then we have the following updates:

(i)the sized BFGS-like update

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

(ii)the sized DFP-like update

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

(iii)the sized Songbai-Zhihong update

(5.3) $L_{k+1} = \beta_k P_k L_k + P_k h_k (z_k - M_k^T h_k)^T / \| P_k h_k \|^2$,

(5.4) $h_k = (f_{k+1}^T J_{k+1} s_k (\| f_{k+1} \|^2)^+) f_{k+1} + \rho_k P_k M_k s_k / \| P_k M_k s_k \|$,

(5.5) $P_k = I - (\| f_{k+1} \|^2)^+ f_{k+1} f_{k+1}^T$, (5.6) $M_k = \beta_k P_k L_k + J_{k+1}$,

(5.7) $\rho_k = (s_k^T z_k - (f_{k+1}^T J_{k+1} s_k)^2 (\| f_{k+1} \|^2)^+)^{1/2}$,

where z_k is given by (3.4), q^+ denotes the Moore-Penrose generalized inverse of q, β_k is a suitable sizing factor and the matrix $B_k^\#$ is rewritten as

(5.8) $B_k^\# = (\beta_k L_k + J_{k+1})^T(\beta_k L_k + J_{k+1})$.

Note that we can apply the Biggs' sizing parameter (2.6) to the factorized quasi-Newton updates. On the other hand, since the DGW's sizing factor (2.8) contains the matrix A_k, we can not employ it directly. However, for the factorized version, a strategy similar to the DGW's one can be considered. The factor β_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

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

which yields

(5.9) $\beta_k' = \{-(L_k s_k)^T J_{k+1} s_k + \mathrm{sgn}((L_k s_k)^T J_{k+1} s_k) \phi_k^{1/2}\} / \| L_k s_k \|^2$,

where $\phi_k = ((L_k s_k)^T J_{k+1} s_k)^2 \pm \| L_k s_k \|^2 (s_k^T v_k)$ and the symbol $\mathrm{sgn}(\zeta)$ denotes the sign of ζ.

Now we can obtain the following two strategies (a) and (b) by investigating the signs of $s_k^T v_k$ and ϕ_k:

(a) Set $\phi_k = ((L_k s_k)^T J_{k+1} s_k)^2 + \| L_k s_k \|^2 | s_k^T v_k |$. For β_k' in (5.9), we choose

(a-1) $\beta_k = \min\{ | \beta_k' |, 1 \}$;

or

(a-2)
$$\beta_k = \begin{cases} -1 & \text{if } \beta_k' < -1, \\ \beta_k' & \text{if } -1 < \beta_k' < 1, \\ 1 & \text{if } 1 < \beta_k'. \end{cases}$$

Note that, in (a-1), we use the absolute value of β_k' and that, in (a-2), we consider the sign of β_k'.

(b) Set

$$\phi_k^1 = ((L_k s_k)^T J_{k+1} s_k)^2 + \| L_k s_k \|^2 (s_k^T v_k),$$
$$\phi_k^2 = ((L_k s_k)^T J_{k+1} s_k)^2 - \| L_k s_k \|^2 (s_k^T v_k)$$

and

$$\beta_k^1 = \{-(L_k s_k)^T J_{k+1} s_k + \mathrm{sgn}((L_k s_k)^T J_{k+1} s_k)(\phi_k^1)^{1/2}\} / \| L_k s_k \|^2,$$
$$\beta_k^2 = \{-(L_k s_k)^T J_{k+1} s_k + \mathrm{sgn}((L_k s_k)^T J_{k+1} s_k)(\phi_k^2)^{1/2}\} / \| L_k s_k \|^2.$$

Then we have

(b-1)
$$\beta_k = \begin{cases} \min\{\max(|\beta_k^1|, |\beta_k^2|), 1\} & \text{if } \phi_k^1 \geqq 0 \text{ and } \phi_k^2 \geqq 0, \\ \min\{ |\beta_k^1|, 1\} & \text{if } \phi_k^1 \geqq 0 \text{ and } \phi_k^2 < 0, \\ \min\{ |\beta_k^2|, 1\} & \text{otherwise}; \end{cases}$$

or

(b-2) Set

$$\beta_k' = \begin{cases} \beta_k^1 & \text{if } (\phi_k^1 \geqq 0 \text{ and } \phi_k^2 \geqq 0 \text{ and } |\beta_k^1| \leqq |\beta_k^2|) \\ & \text{or if } (\phi_k^1 \geqq 0 \text{ and } \phi_k^2 < 0), \\ \beta_k^2 & \text{otherwise.} \end{cases}$$

For this β_k', we choose β_k by the same way as Strategy (a).

6. New Algorithm

Now we present a new factorized quasi-Newton algorithm.

(FACTNLS Algorithm)

Starting with a point $x_1 \in R^n$ and an m x n matrix L_1, the algorithm proceeds, for k = 1, 2, ..., as follows:

Step 1. Having x_k and L_k, find the search direction d_k by solving the linear system of equations (3.1).

Step 2. Choose a steplength α_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 a suitable updating formula for L_k.

7. 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 vector valued function (i.e. f(x)) evaluations.

The numerical calculations were carried out in double precision arithmetic on a NEC PC-9801VX personal computer, and the program was coded in FORTRAN 77. The iterative process is terminated

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

or

(2) if $| e_i^T J(x_{k+1})^T f(x_{k+1}) | \leqq \max(TOL2, \varepsilon) \| f(x_{k+1}) \| \| J(x_{k+1}) e_i \|$ for i=1,...,n and $\| x_{k+1} - x_k \|_\infty \leqq \max(TOL3, \varepsilon) \max(\| x_{k+1} \|_\infty, 1.0)$,

where e_i denotes the i-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 ε is machine epsilon. Further, the Jacobian matrix is evaluated by the forward difference approximation, and the bisection line search method with Armijo's rule

(7.1) $\quad F(x_k + \alpha_k d_k) \leqq F(x_k) + 0.1\alpha_k \nabla F(x_k)^T d_k$

is employed.

In the experiments, we set TOL1 = TOL2 = TOL3 = 10^{-4}, ITMAX = 500 and NFEMAX = 2000. In addition to (2.6), we used the following sizing parameter

(7.2) $\quad \beta_k = | f_{k+1}^T f_k | / \| f_k \|^2$.

Essentially, the DGW update is designed for use in a trust region framework, but we dare to use it in a line search framework from the point of view of discussing relative merits among several updates given in this paper. In addition, for the DGW and the Biggs updates, the modified Cholesky decomposition is employed to determine the search directions in case $J_k^T J_k + A_k$ in (2.1) is not positive definite. For all the methods, the initial matrices A_1 and L_1 were set to the zero matrices, respectively.

The names, the sizes and the starting points of the test problems [5], [6], together with the abbreviated problem names used in Tables 2-5, are listed in Table 1.

The computational results are summarized in Tables 2-5. Note that the numbers in Tables 3 and 5 include the number of vector valued function (i.e. f(x)) evaluations to evaluate J(x) by the forward difference approximation. In each table, we use the following symbols;

GN : the Gauss-Newton method,
Biggs : the Biggs update (2.5) and (2.6),
DGW : the DGW update (2.7) and (2.8),
BFGSF-0 : the FACTNLS algorithm with (3.3) and (3.5),
F-1 : the FACTNLS algorithm with (3.4) and (3.5),
F-2a : the FACTNLS algorithm with (3.4), (5.1) and (7.2),
F-2b : the FACTNLS algorithm with (3.4), (5.1) and (2.6),
F-3a : the FACTNLS algorithm with (3.4), (5.1) and (a-1),
F-3b : the FACTNLS algorithm with (3.4), (5.1) and (a-2),
F-4a : the FACTNLS algorithm with (3.4), (5.1) and (b-1),
F-4b : the FACTNLS algorithm with (3.4), (5.1) and (b-2),
DFP F-0 : the FACTNLS algorithm with (3.3) and (3.6),
F-1 : the FACTNLS algorithm with (3.4) and (3.6),
F-2a : the FACTNLS algorithm with (3.4), (5.2) and (7.2),
F-2b : the FACTNLS algorithm with (3.4), (5.2) and (2.6),
F-3a : the FACTNLS algorithm with (3.4), (5.2) and (a-1),
F-3b : the FACTNLS algorithm with (3.4), (5.2) and (a-2),
F-4a : the FACTNLS algorithm with (3.4), (5.2) and (b-1),
F-4b : the FACTNLS algorithm with (3.4), (5.2) and (b-2),

SZ F-0 : the FACTNLS algorithm with (3.3) and (4.19),
F-1 : the FACTNLS algorithm with (3.4) and (4.19),
F-2a : the FACTNLS algorithm with (3.4), (5.3) and (7.2),
F-2b : the FACTNLS algorithm with (3.4), (5.3) and (2.6),
F-3a : the FACTNLS algorithm with (3.4), (5.3) and (a-1),
F-3b : the FACTNLS algorithm with (3.4), (5.3) and (a-2),
F-4a : the FACTNLS algorithm with (3.4), (5.3) and (b-1),
F-4b : the FACTNLS algorithm with (3.4), (5.3) and (b-2),
G-N1 : if $\| f_k \| \leqq 10^{-1}$, then GN is used, otherwise SZF-1,
G-N2 : if $\| f_k \| \leqq 10^{-3}$, then GN is used, otherwise SZF-1,
G-N3 : if $\| f_k \| \leqq 10^{-5}$, then GN is used, otherwise SZF-1,
* : the method failed to converge in the specified number of function evaluations.

From these tables, we can see that the Gauss-Newton method performed very well for the zero or small residual problems, but did not necessarily perform well for the large residual problems. For all the problems, the structured quasi-Newton methods with the Biggs and the DGW updates performed well and were numerically stable. Roughly speaking,our numerical experiments show there is little difference between the Biggs and the DGW updates.

BFGSF-0 and DFPF-0 did not perform well for all the problems, and the latter was much worse than the former. Our numerical results show BFGSF-1 performed about as well as sized BFGS-like methods, even though BFGSF-1 does not employ a sizing technique. However, this tendency can not be observed between DFPF-1 and the sized DFP-like methods. The sized BFGS-like and the sized DFP-like methods performed well for all the problems. Sizing techniques take effect for the DFP-like methods better than for the BFGS-like methods. In addition, it is interesting that the BFGS-like methods with (3.4) perform well whether sizing techniques are employed or not.

The Songbai and Zhihong method performed better than our methods did. It is also interesting that the behavior of their method changes little whether sizing techniques are employed or not.

On the whole, sizing techniques similar to DGW's one which consider the sign of β_k', such as F-3b and F-4b, seem sensitive.

Acknowledgment

The authors would like to thank Professor Y. Oyanagi for sending us the PEAK problem for our numerical experiments.

References

[1]M.C.Bartholomew-Biggs, " The estimation of the Hessian matrix in nonlinear least squares problems with non-zero residuals", Math. Prog. 12 (1977) 67-80.

[2]A.Ben-Israel and T.N.E.Greville, Generalized Inverses: Theory and Applications (Robert E. Krieger Publishing Company, Huntington, 1980).

[3]J.E.Dennis,Jr., "Some computational techniques for the nonlinear least squares problem", in: G.D.Byrne and C.A.Hall, eds., Numerical Solution of Systems of Nonlinear Algebraic Equations (Academic Press, 1973), pp.157-183.

[4]J.E.Dennis,Jr. and R.B.Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations (Prentice-Hall, New Jersey, 1983).

[5]J.E.Dennis,Jr., D.M.Gay and R.E.Welsch, "An adaptive nonlinear least-squares algorithm", ACM Trans. on Mathematical Software 7, No.3 (1981) 348-368.

[6]T.Nakagawa and Y.Oyanagi,"Softwares for nonlinear least-squares methods", Jyohoshori 23, No.5, Information Processing Society of Japan (1982) pp.442-450, in Japanese.

[7]S.Songbai and Z.Zhihong,"A new secant method for nonlinear least squares problems", Technical Report NANOG-1988-03, Nanjing University (1988).

[8]T.Takahashi and H.Yabe, "Factorized quasi-Newton methods for nonlinear least squares problems", Kokyuroru 676, RIMS, Kyoto University (1988) pp.170-194.

[9]H.Yabe and T.Takahashi, "Structured quasi-Newton methods for nonlinear least squares problems", TRU Math. 24, No.2 (1988) pp.195-209.

Table 1. Test Problems

Abbrebiated Name	Name of Test Problem	m	n	Starting Point
WATSON6	Watson Problem with 6 variables	31	6	(0, 0, ..., 0)
WATSON9	Watson Problem with 9 variables	31	9	(0, 0, ..., 0)
WATSON12	Watson Problem with 12 variables	31	12	(0, 0, ..., 0)
WATSON20	Watson Problem with 20 variables	31	20	(0, 0, ..., 0)
ROSENBROCK	Rosenbrock Problem	2	2	(-1.2, 1.0)
HELIX	Helical Valley Problem	3	3	(-1, 0, 0)
POWELL	Powell's Singular Problem	4	4	(3, -1, 0, 1)
BEALE	Beale Problem	3	2	(0.1, 0.1)
FRDSTEIN1	Freudenstein and Roth Problem	2	2	(6, 6)
FRDSTEIN2	Freudenstein and Roth Problem	2	2	(15, -2)
BARD	Bard Problem	15	3	(1, 1, 1)
BOX	Box Problem	10	3	(0, 10, 20)
KOWALIK	Kowalik Problem	11	4	(0.25,0.39,0.415,0.39)
OSBORNE1	Osborne Problem	33	5	(0.5, 1.5, -1.0, 0.01, 0.02)
OSBORNE2	Osborne Problem	65	11	(1.3, 0.65, 0.65, 0.7, 0.6, 3.0, 5.0, 7.0, 2.0, 4.5, 5.5)
JENNRICH	Jennrich Problem	10	2	(0.3, 0.4)
PEAK	Peak Problem	51	5	(q, 2, 6, 3.5, 0.1) q = -2, -1, ..., 8

Table 2. Number of Iterations

	GN	Big	DGW	BFGS F-0	BFGS F-1	BFGS F-2a	BFGS F-2b	BFGS F-3a	BFGS F-3b	BFGS F-4a	BFGS F-4b
WATSON6	6	9	9	19	15	14	15	17	19	12	20
WATSON9	80	82	82	28	28	26	21	31	30	23	31
WATSON12	61	60	60	19	13	14	7	19	27	11	30
WATSON20	4	9	9	18	16	15	11	21	25	11	22
ROSENBROCK	11	20	17	29	22	13	16	16	13	15	12
HELIX	8	11	13	26	24	16	15	20	34	14	19
POWELL	9	14	14	20	14	14	14	14	14	14	14
BEALE	6	8	9	13	9	8	8	8	11	8	9
FRDSTEIN1	5	6	6	9	6	6	6	6	6	6	6
FRDSTEIN2	105*	6	6	9	7	7	7	11	27	10	31
BARD	5	12	10	20	9	8	8	8	8	8	7
BOX	4	6	5	10	5	5	5	5	5	5	5
KOWALIK	19	10	12	14	10	9	9	8	12	9	10
OSBORNE1	6	27	21	43	27	18	18	16	17	15	26
OSBORNE2	9	13	12	22	24	15	15	16	16	17	14
JENNRICH	138*	9	9	10	11	9	12	11	102*	10	16

	DFP F-0	DFP F-1	DFP F-2a	DFP F-2b	DFP F-3a	DFP F-3b	DFP F-4a	DFP F-4b
WATSON6	76	27	8	8	9	9	9	8
WATSON9	126	47	23	23	21	21	21	21
WATSON12	153*	25	6	6	8	9	8	7
WATSON20	79	25	6	6	8	9	8	7
ROSENBROCK	500*	54	19	19	18	17	19	21
HELIX	283	34	12	12	12	14	11	15
POWELL	20	14	14	14	14	14	14	14
BEALE	19	11	8	8	7	12	7	11
FRDSTEIN1	8	6	6	6	6	6	6	6
FRDSTEIN2	10	6	6	6	8	10	8	13
BARD	17	9	8	8	8	8	8	8
BOX	15	5	5	5	5	5	5	5
KOWALIK	62	10	9	9	8	10	8	8
OSBORNE1	332*	57	139	139	21	19	20	26
OSBORNE2	40	41	13	13	13	15	13	15
JENNRICH	500*	89	9	9	8	8	8	8

Table 2. (Continued)

	SZ F-0	SZ F-1	SZ F-2a	SZ F-2b	SZ F-3a	SZ F-3b	SZ F-4a	SZ F-4b	SZ G-N1	SZ G-N2	SZ G-N3
WATSON6	9	10	7	7	7	7	7	7	8	10	10
WATSON9	20	20	19	19	19	20	20	19	19	20	20
WATSON12	6	6	5	5	5	5	5	5	5	6	6
WATSON20	6	6	5	5	5	5	5	5	5	6	6
ROSENBROCK	24	14	14	14	14	14	14	14	13	14	14
HELIX	7	11	11	11	11	11	11	11	10	11	11
POWELL	14	10	10	10	10	10	10	10	9	10	10
BEALE	12	9	7	7	8	8	9	9	7	9	9
FRDSTEIN1	7	6	6	6	6	6	6	6	6	6	6
FRDSTEIN2	99*	27	35#	35#	44#	93#	26#	19#	27	27	27
BARD	23	7	5	5	5	6	5	5	6	7	7
BOX	7	5	5	5	5	5	5	5	4	5	5
KOWALIK	13	10	8	8	9	10	9	9	19	10	10
OSBORNE1	42	33	28	28	18	19	18	17	6	33	33
OSBORNE2	24	21	14	14	12	14	12	13	21	21	21
JENNRICH	13	9	7	7	7	36	7	15	9	9	9

Table 3. Number of Vector Valued Function Evaluations

	GN	Big	DGW	BFGS F-0	BFGS F-1	BFGS F-2a	BFGS F-2b	BFGS F-3a	BFGS F-3b	BFGS F-4a	BFGS F-4b
WATSON6	49	70	71	147	117	114	124	134	173	103	174
WATSON9	810	830	831	297	295	280	229	339	365	251	380
WATSON12	806	793	794	267	187	204	113	283	448	167	493
WATSON20	105	210	211	406	362	345	261	495	629	263	543
ROSENBROCK	62	73	87	96	78	48	88	79	59	77	57
HELIX	37	50	62	122	127	95	95	121	238	88	128
POWELL	50	75	75	105	75	75	75	75	75	75	75
BEALE	26	33	36	50	39	36	36	35	46	35	38
FRDSTEIN1	18	21	21	30	21	21	21	21	21	21	21
FRDSTEIN2	2004*	21	21	33	31	31	31	102	442	93	520
BARD	24	53	45	102	41	37	37	37	37	37	33
BOX	20	28	24	45	24	24	24	24	24	24	24
KOWALIK	103	59	70	94	61	56	56	51	74	56	61
OSBORNE1	44	172	148	274	181	128	128	118	127	107	219
OSBORNE2	125	171	159	286	310	200	200	216	222	232	194
JENNRICH	2001*	32	32	70	57	64	76	59	2013*	55	197

Table 3. (Continued)

	DFP F-0	DFP F-1	DFP F-2a	DFP F-2b	DFP F-3a	DFP F-3b	DFP F-4a	DFP F-4b
WATSON6	541	196	63	63	70	70	70	63
WATSON9	1279	480	240	240	220	220	220	220
WATSON12	2005*	342	91	91	117	130	117	104
WATSON20	1682	547	147	147	189	210	189	168
ROSENBROCK	1521*	180	68	68	73	69	80	82
HELIX	1142	143	60	60	59	66	56	74
POWELL	105	75	75	75	75	75	75	75
BEALE	65	42	33	33	30	45	30	41
FRDSTEIN1	27	21	21	21	21	21	21	21
FRDSTEIN2	33	21	21	21	39	49	38	63
BARD	77	41	37	37	37	37	37	37
BOX	64	24	24	24	24	24	24	24
KOWALIK	324	60	55	55	50	61	50	50
OSBORNE1	2003*	349	843	843	141	130	141	176
OSBORNE2	494	507	170	170	171	204	171	204
JENNRICH	1514*	272	32	32	29	29	29	29

	SZ F-0	SZ F-1	SZ F-2a	SZ F-2b	SZ F-3a	SZ F-3b	SZ F-4a	SZ F-4b	SZ G-N1	SZ G-N2	SZ G-N3
WATSON6	70	77	56	56	56	56	56	56	63	77	77
WATSON9	210	210	200	200	200	210	210	200	200	210	210
WATSON12	91	91	78	78	78	78	78	78	78	91	91
WATSON20	147	147	126	126	126	126	126	126	126	147	147
ROSENBROCK	126	61	61	61	61	61	61	61	57	61	61
HELIX	34	51	51	51	51	51	51	51	47	51	51
POWELL	77	55	55	55	55	55	55	55	50	55	55
BEALE	44	35	30	30	32	32	35	36	29	35	35
FRDSTEIN1	34	22	22	22	22	22	22	22	22	22	22
FRDSTEIN2	2020*	468	583g	583g	748g	1800g	384g	251g	468	468	468
BARD	107	32	24	24	24	28	24	24	28	32	32
BOX	35	24	24	24	24	24	24	24	20	24	24
KOWALIK	75	59	49	49	55	60	55	55	103	59	59
OSBORNE1	262	210	199	199	148	139	137	128	44	210	210
OSBORNE2	312	272	187	187	163	186	163	173	272	272	272
JENNRICH	44	32	30	30	26	616	26	109	32	32	32

g: the global minimum is obtained.

Table 4. Number of Iterations for PEAK Problem

	q=-2	q=-1	q=0	q=1	q=2	q=3	q=4	q=5	q=6
G-N	---	36#	9	6	4	4	6	9	---
Biggs	---	---	8	8	4	5	8	8	---
DGW	14#	9	9	7	4	5	7	12	31#
BFGS:F-0	21#	13	13#	9	6	6	10	12#	16#
F-1	14	21	13	8	4	5	9	16	21
F-2a	13	17	9	7	4	5	7	12	12
F-2b	13	17	9	7	4	5	7	13	12
F-3a	18	26	9	7	4	5	7	---	12
F-3b	13#	43	8	7	4	5	7	19	11
F-4a	14	19	10	7	4	5	7	---*	12
F-4b	14	13	9	7	5	5	7	113	13
S-Z:F-0	19	14	9	9	6	6	8	12	16
F-1	19	14	13	8	4	5	8	15	16
F-2a	14	9	9	6	4	5	7	12	11
F-2b	14	9	9	6	4	5	7	12	11
F-3a	16	9	9	7	4	5	7	9	13
F-3b	12	9	8	6	4	5	6	16	11
F-4a	14#	9	9	7	4	5	7	12	13
F-4b	16	9	9	6	5	5	7	14	14

\# : the negative Γ is obtained.
---: an another stationary point is obtained.

(Note)For the cases of q=7 and 8, an another stationary point is obtained by all the methods.

Table 5. Number of Vector Valued Function Evaluations for PEAK

	q=-2	q=-1	q=0	q=1	q=2	q=3	q=4	q=5	q=6
Gauss-Newton	---	475#	62	42	30	30	43	61	---
Biggs	---	---	55	60	30	36	54	56	---
DGW	103#	65	68	57	30	36	51	93	250#
BFGS:F-0	156#	97	94#	62	42	42	70	89#	124#
F-1	106	150	91	54	30	36	62	123	170
F-2a	101	134	65	48	30	36	50	104	102
F-2b	101	134	65	48	30	36	50	110	102
F-3a	151	299	67	48	30	36	50	---	98
F-3b	103#	574	59	48	30	36	50	190	99
F-4a	115	205	73	48	30	36	50	---*	97
F-4b	116	115	66	48	36	36	51	1756	112
S-Z:F-0	127	92	60	60	42	42	54	78	108
F-1	127	94	86	55	30	36	55	100	111
F-2a	109	63	61	43	30	36	49	87	80
F-2b	109	63	61	43	30	36	49	89	80
F-3a	127	62	63	49	30	36	49	62	100
F-3b	87	62	55	43	30	36	43	141	81
F-4a	108#	62	63	49	30	36	49	86	100
F-4b	131	62	61	43	36	36	49	104	108