

**Variations of Structured Broyden
Families for Nonlinear
Least Squares Problems**

Hiroshi Yabe

**CRPC-TR92220
June, 1992**

Center for Research on Parallel Computation
Rice University
P.O. Box 1892
Houston, TX 77251-1892

Variations of Structured Broyden Families for Nonlinear Least Squares Problems *

Hiroshi Yabe [†]

Abstract:

In this paper, we consider structured quasi-Newton methods for finding a local solution to nonlinear least squares problems. This paper is concerned with the line search globalization method. Recently, factorized versions of the structured quasi-Newton methods have been studied by Sheng and Zou, and Yabe and Takahashi in order to obtain a descent search direction for the objective function. In this paper, we first generalize the update of Sheng and Zou, and propose a new factorized family corresponding to the Broyden family (SZ-Broyden family). Second, we suggest a relationship between the structured quasi-Newton updates and the factorized versions. We use this relationship to show that the factorized Broyden family proposed by Yabe and Yamaki corresponds to the Engels and Martinez family, and we further obtain a new structured quasi-Newton update which corresponds to the SZ-Broyden family in the sense of this relation. Finally, we apply sizing techniques to these methods and present some numerical experiments.

Key words:

Nonlinear least squares problems, Structured quasi-Newton methods, Broyden family, Sheng and Zou method, Sizing techniques

1 Introduction

In this paper we consider methods for finding a local solution, x^* , say, to the nonlinear least squares problem

$$(1.1) \quad \text{minimize } f(x) = \frac{1}{2} \|r(x)\|^2 = \frac{1}{2} \sum_{j=1}^m (r_j(x))^2,$$

where $r_j : R^n \rightarrow R$, $j = 1, \dots, m$ ($m \geq n$) are twice continuously differentiable, $r(x) = (r_1(x), \dots, r_m(x))^T$ and $\| \cdot \|$ denotes the 2 norm. Such problems arise widely in data fitting and in the solution of well-determined and over-determined systems of equations. Among

*Parts of this work were done while the author was visiting Department of Mathematical Sciences and the Center for Research on Parallel Computation at Rice University, Houston, Texas.

[†]Faculty of Engineering, Science University of Tokyo, 1-3, Kagurazaka, Shinjuku-ku, Tokyo 162, Japan

many numerical methods, structured quasi-Newton methods seem very promising. These methods use the structure of the Hessian matrix of $f(x)$,

$$(1.2) \quad \nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^m r_j(x) \nabla^2 r_j(x),$$

where J is the Jacobian matrix of r , and approximate the second part of the Hessian by some matrix A instead of approximating the whole Hessian by some matrix B as usual quasi-Newton methods. The structured quasi-Newton methods were proposed in order to overcome the poor performance of the Gauss-Newton method for large residual problems. These methods were first proposed by Brown and Dennis[4], Broyden and Dennis (see Dennis[6]). On the other hand, quasi-Newton methods do not work as well as the Gauss-Newton method for zero or small residual problems, because quasi-Newton updates do not in general produce zero matrices. Consequently, some remedies have been considered. Bartholomew-Biggs[2], Dennis, Gay and Welsch[8] proposed sizing techniques, and Al-Baali and Fletcher[1] proposed the hybrid method which combined the structured quasi-Newton methods and the Gauss-Newton method. Recently, several studies on the structured quasi-Newton methods have been done, e.g. Dennis, Martinez and Tapia[9], Engels and Martinez[11] (see also Engels[10] and Martinez[14]), Fletcher and Xu[12], Xu[22]. Very recently, Huschens[13] has proposed a very effective method for both cases of zero and large residual problems.

For structured quasi-Newton methods, there are two types of globalization strategies. One is the line search descent method and the other is the trust region method. This paper is concerned with the former, which generates the sequence $\{x_k\}$ by

$$(1.3) \quad x_{k+1} = x_k + \alpha_k d_k,$$

where α_k is a step length and a search direction d_k is given by solving the linear system of equations

$$(1.4) \quad (J_k^T J_k + A_k) d = -J_k^T r_k,$$

where $r_k = r(x_k)$, $J_k = J(x_k)$, and the $n \times n$ matrix A_k is the approximation to the second part of the Hessian matrix. The matrix A_k is generated by some quasi-Newton updating formula, say, A -update. This system corresponds to the Newton equation

$$\nabla^2 f(x_k) d = -\nabla f(x_k),$$

or equivalently,

$$(1.5) \quad (J_k^T J_k + \sum_{j=1}^m r_j(x_k) \nabla^2 r_j(x_k)) d = -J_k^T r(x_k).$$

Since the coefficient matrix of (1.4) does not necessarily possess the hereditary positive definiteness property, Yabe and Takahashi[25] proposed computing the search direction d_k by solving the linear system of equations

$$(1.6) \quad (J_k + L_k)^T (J_k + L_k) d = -J_k^T r_k,$$

where the matrix L_k is an $m \times n$ correction matrix to the Jacobian matrix such that $(J_k + L_k)^T (J_k + L_k)$ approximates the Hessian and is generated by some updating formula,

say, L -update. Since the coefficient matrix is expressed by its factorized form, the search direction may be expected to be a descent direction for f . Following Dennis[7], we dealt with the secant condition

$$(1.7) \quad (J_{k+1} + L_{k+1})^T (J_{k+1} + L_{k+1}) s_k = z_k,$$

where

$$(1.8) \quad s_k = x_{k+1} - x_k, \quad z_k = (J_{k+1} - J_k)^T r_{k+1} + J_{k+1}^T J_{k+1} s_k.$$

We call this method the factorized quasi-Newton method. Yabe and Takahashi[25] proposed BFGS-like and DFP-like updates, and Yabe and Takahashi[26] proved the local and q-superlinear convergence of their methods. Further, Yabe and Yamaki[28] obtained a structured Broyden family for L_k that contained these updates.

On the other hand, Sheng and Zou[19] studied factorized versions of the structured quasi-Newton methods independently of us. They considered the approximation of $r(x)$ around x_k as follows:

$$r(x_k + d) \approx r(x_k) + (J_k + L_k)d,$$

and proposed obtaining a search direction d_k by solving the linear least squares problem

$$(1.9) \quad \text{minimize } \frac{1}{2} \|r_k + (J_k + L_k)d\|^2 \text{ with respect to } d.$$

In the case of $L_k = 0$, the above implies the Gauss-Newton model. The normal equation of (1.9) is represented by

$$(1.10) \quad (J_k + L_k)^T (J_k + L_k) d = -(J_k + L_k)^T r_k.$$

Since the vector $-L_k^T r_k$ exists in the righthand side, the above does not correspond to the Newton equation(1.5), so Sheng and Zou imposed the condition $L_{k+1}^T r_{k+1} = 0$ on the matrix L_{k+1} in addition to the secant condition (1.7). They obtained a BFGS-like update and showed the local and q-superlinear convergence of their method. In practical computations, they used the hybrid method with the Gauss-Newton method. The idea of Sheng and Zou seems very interesting to us, because their update includes a feature different from our factorized updates. Further, some numerical experiments given in Yabe and Takahashi[27] suggest the efficiency of their method.

Now we have two kinds of updates, an A -update and an L -update, each with merits and demerits. An A -update just needs an $n \times n$ symmetric square matrix and is calculated in $O(n^2)$ arithmetic cost, but the coefficient matrix in (1.4) is not necessarily positive definite for the line search strategy. On the other hand, an L -update may be expected to maintain the positive definiteness of the coefficient matrix in (1.6), but it needs an $m \times n$ rectangular matrix and is calculated in $O(mn)$ arithmetic cost. However both approaches should not compete each other, but should complement each other. By using a relationship between an A -update and an L -update, special features of an L -update can be reflected in an A -update. This is a motivation of this paper. The first purpose of this paper is to generalize the update of Sheng and Zou, and to propose a new update which corresponds to the Broyden family. We will call this new class of updates the SZ-Broyden family (L -update). The second purpose is to investigate the relationship between

an A -update and an L -update. The third purpose is to obtain a new A -update which corresponds to the SZ-Broyden family (L -update) by using this relationship. We will call this new class of updates the SZ-Broyden family (A -update). The final purpose is to examine the effectiveness of sizing techniques for Broyden-like families (A -update).

Sections 2 and 3 generalize the Sheng and Zou update, and construct the SZ-Broyden family (L -update). In Section 4, we investigate a relationship between an A -update and an L -update. By using this relation, we show that the structured Broyden family given by Yabe and Yamaki[28] corresponds to the structured secant update from the convex class proposed by Engels and Martinez[11] (see also Engels[10] and Martinez[14]). We also obtain a new A -update that corresponds to the SZ-Broyden family (L -update). Section 5 deals with sizing techniques, which were first proposed by Bartholomew-Biggs[2] and Dennis et al.[8]. Finally we show some numerical experiments of Broyden-like families for A -updates in Section 6, and examine the effectiveness of sizing techniques. Throughout the paper, for simplicity, we drop the subscript k and replace the subscript $k + 1$ by "+". Further $\| \cdot \|$ denotes the 2 norm.

2 Notations and Basic Properties

In this section, following Yabe[24], we generalize the update of Sheng and Zou, and give some basic properties of the update. In the next section we shall construct a new family which specifically corresponds to the Broyden family.

The conditions which Sheng and Zou imposed on the $m \times n$ matrix L_+ are the secant condition

$$(2.1) \quad (J_+ + L_+)^T (J_+ + L_+) s = z$$

and the orthogonality condition

$$(2.2) \quad L_+^T r_+ = 0,$$

which connects the Newton equation (1.5) and the normal equation (1.10). Letting $Y = J_+ + L_+$ and $h = Ys$, the conditions (2.1) and (2.2) can be written by the matrix equations of Y

$$(2.3) \quad CY = D, \quad Ys = h,$$

where

$$C = \begin{bmatrix} h^T \\ r_+^T \end{bmatrix}, \quad D = \begin{bmatrix} z^T \\ r_+^T J_+ \end{bmatrix}.$$

By using the results of Chapter 2, Section 1 in Ben-Israel and Greville[3], we have the following theorem.

Theorem 2.1 *The matrix equations (2.3) have a common solution if and only if each equation separately has a solution and*

$$(2.4) \quad Ch = Ds.$$

The matrix equation $CY = D$ is consistent if and only if $CC^{(1)}D = D$ and $CC^{(1)}C = C$ for some $C^{(1)}$, and the matrix equation $Ys = h$ is consistent if and only if $hs^{(1)}s = h$ and $ss^{(1)}s = s$ for some $s^{(1)}$. For these $C^{(1)}$ and $s^{(1)}$, the general solution of (2.3) is given by

$$(2.5) \quad Y = C^{(1)}D + (I - C^{(1)}C)hs^{(1)} + (I - C^-C)\Phi(I - ss^-),$$

where C^- is an arbitrary matrix such that $CC^-C = C$, s^- is an arbitrary vector such that $ss^-s = s$ and Φ is an arbitrary $m \times n$ matrix.

Note that the matrix $C^{(1)}D + (I - C^{(1)}C)hs^{(1)}$ is a particular solution of the inhomogeneous equations (2.3) and that $(I - C^-C)\Phi(I - ss^-)$ is a general solution of the homogeneous equations $CY = 0$ and $Ys = 0$. The preceding theorem suggests that we just consider the equations (2.3) for a vector h which satisfies $Ch = Ds$; i.e.

$$h^T h = s^T z \quad \text{and} \quad r_+^T h = r_+^T J_+ s.$$

We will use the following definitions:

$$Q = \frac{r_+ r_+^T}{\|r_+\|^2}, \quad P = I - Q = I - \frac{r_+ r_+^T}{\|r_+\|^2},$$

$$N = J_+ + PL, \quad B^\sharp = N^T N = (J_+ + PL)^T (J_+ + PL),$$

$$P^\sharp = N^T P N, \quad Q^\sharp = N^T Q N = \frac{J_+^T r_+ r_+^T J_+}{\|r_+\|^2} \quad \text{and} \quad z^\sharp = z - Q^\sharp s.$$

We will also assume that the following conditions hold:

(A1) r_+ is independent of h .

(A2) h satisfies $h^T h = s^T z > 0$ and $r_+^T h = r_+^T J_+ s$.

Now we present the following properties, which are useful in the construction of L -updates.

Theorem 2.2 Suppose that the assumptions (A1) and (A2) hold. Then

- (1) The matrix CC^T is nonsingular and $\det(CC^T) = \|r_+\|^2 \|Ph\|^2 > 0$.
- (2) $s^T z^\sharp > 0$.
- (3) $(I - C^\dagger C)h = 0$ and $(I - C^\dagger C)r_+ = 0$, where C^\dagger denotes the Moore-Penrose generalized inverse of C .
- (4) If $\text{rank } N = n$, the following statements are equivalent:
 - (a) P^\sharp is nonsingular (i.e. P^\sharp is symmetric positive definite).
 - (b) $r_+^T N(N^T N)^{-1} N^T r_+ \neq \|r_+\|^2$.
 - (c) r_+ cannot be spanned by the column vectors of N .

Proof.

(1) Since r_+ and h are linearly independent, the matrix CC^T is nonsingular and $Ph \neq 0$. Thus we have

$$\begin{aligned} \det(CC^T) &= \begin{vmatrix} h^T h & h^T r_+ \\ h^T r_+ & r_+^T r_+ \end{vmatrix} \\ &= \|r_+\|^2 \|Ph\|^2 > 0. \end{aligned}$$

(2) It follows from the definitions of Q^\sharp and z^\sharp that

$$s^T z^\sharp = s^T z - s^T Q^\sharp s = \|Ph\|^2 > 0.$$

(3) Since the matrix C^\dagger is represented by $C^T(CC^T)^{-1}$, we have

$$(I - C^\dagger C)[h, r_+] = (I - C^\dagger C)C^T = (I - C^T(CC^T)^{-1}C)C^T = 0,$$

which means that the matrix $I - C^\dagger C$ is the orthogonal projection matrix from R^n onto the orthogonal complement of the subspace spanned by h and r_+ .

(4)(i) (a) \iff (b) :

$$\begin{aligned} \det(P^\sharp) &= \det(N^T P N) \\ &= \det \left[N^T N - \left(\frac{N^T r_+}{\|r_+\|} \right) \left(\frac{N^T r_+}{\|r_+\|} \right)^T \right] \\ &= \det[N^T N] \det \left[I - (N^T N)^{-1} \left(\frac{N^T r_+}{\|r_+\|} \right) \left(\frac{N^T r_+}{\|r_+\|} \right)^T \right] \\ &= \det[N^T N] \left(1 - \frac{1}{\|r_+\|^2} r_+^T N (N^T N)^{-1} N^T r_+ \right). \end{aligned}$$

Since N is of column full rank, the matrix $N^T N$ is nonsingular. Thus (a) is equivalent to (b).

(ii) (a) \iff (c) :

(Proof of (a) \Rightarrow (c))

Assume that r_+ lies in the span of the column vectors of N . Since r_+ is formed by $r_+ = Nq$ for some vector q , we have

$$r_+^T N (N^T N)^{-1} N^T r_+ = q^T N^T N q = \|r_+\|^2.$$

By the equivalence of (a) and (b), the matrix P^\sharp is singular. Thus (a) implies (c).

(Proof of (c) \Rightarrow (a))

Assume that r_+ does not lie in the span of the column vectors of N . Since r_+ can be represented by $r_+ = w_1 + w_2$ for a vector w_1 in the range space of N and a nonzero vector w_2 in the null space of N^T , the Pythagorean theorem yields

$$\|w_1 + w_2\|^2 = \|w_1\|^2 + \|w_2\|^2.$$

Thus we have

$$\begin{aligned} \frac{r_+^T N (N^T N)^{-1} N^T r_+}{\|r_+\|^2} &= \frac{\|(N^T N)^{-1/2} N^T r_+\|^2}{\|r_+\|^2} = \frac{\|(N^T N)^{-1/2} N^T w_1\|^2}{\|w_1 + w_2\|^2} \\ &= \frac{\|w_1\|^2}{\|w_1 + w_2\|^2} = \frac{\|w_1\|^2}{\|w_1\|^2 + \|w_2\|^2} < 1, \end{aligned}$$

which means that P^\sharp is nonsingular. Thus (c) implies (a).

The proof is complete. ■

We will now compute a vector h satisfying the assumption (A2). First, we have a general form of h satisfying the condition $r_+^T h = r_+^T J_+ s$ as follows

$$\begin{aligned} h &= (r_+^T)^\dagger (r_+^T J_+ s) + (I - (r_+^T)^\dagger (r_+^T)) u' = \frac{r_+^T J_+ s}{\|r_+\|^2} r_+ + \left(I - \frac{r_+ r_+^T}{\|r_+\|^2} \right) u' \\ &= \frac{r_+ r_+^T}{\|r_+\|^2} N s + \left(I - \frac{r_+ r_+^T}{\|r_+\|^2} \right) u' = Q N s + P u', \end{aligned}$$

where u' is an arbitrary vector. Now we choose a vector u' satisfying the condition $h^T h = s^T z$. Setting $u = \tau u'$ and choosing a parameter τ such that

$$h^T h = s^T Q^\sharp s + \frac{1}{\tau^2} \|P u\|^2 = s^T z,$$

we have

$$h = Q N s + \frac{1}{\tau} P u$$

and

$$\frac{1}{\tau^2} \|P u\|^2 = s^T z - s^T Q^\sharp s = s^T z^\sharp.$$

Note that the result (2) in Theorem 2.2 guarantees the positivity of the righthand side of the expression above.

To end this section, we derive a useful particular solution from the general solution (2.5). Setting $C^{(1)}$ and C^- to the Moore-Penrose generalized inverses C^\dagger , and letting $\Phi = N$, the result (3) in Theorem 2.2 yields

$$Y = N + C^\dagger (D - C N) - (I - C^\dagger C) N s s^-.$$

Since

$$C^\dagger = \frac{1}{\|r_+\|^2 \|P h\|^2} [h, r_+] \begin{bmatrix} \|r_+\|^2 & -h^T r_+ \\ -h^T r_+ & h^T h \end{bmatrix}, \quad D - C N = \begin{bmatrix} z^T - h^T N \\ 0 \end{bmatrix},$$

we have a particular solution to (2.3)

$$(2.6) \quad Y = N + \left(\frac{P h}{\|P h\|^2} \right) (z - N^T h)^T - (I - C^\dagger C) N s s^-.$$

3 SZ-Broyden Family (L -Update)

In this section, we construct a BFGS-like update, which corresponds to the formula of Sheng and Zou, and a DFP-like update by using the particular solution given in the previous section. We will also present a new L -update which corresponds to the Broyden family.

Since

$$Ph = \frac{1}{\tau}Pu, \quad \|Ph\|^2 = \frac{1}{\tau^2}\|Pu\|^2 = s^T z^\sharp,$$

$$\frac{Ph}{\|Ph\|^2} = \frac{1}{\tau} \left(\frac{Pu}{s^T z^\sharp} \right) = \tau \left(\frac{Pu}{\|Pu\|^2} \right) \quad \text{and} \quad z - N^T h = z^\sharp - \frac{1}{\tau} N^T Pu,$$

it follows from (2.6) that

$$(3.1) \quad Y = N + \frac{1}{\tau} \left(\frac{Pu}{s^T z^\sharp} \right) \left(z^\sharp - \frac{1}{\tau} N^T Pu \right)^T - (I - C^\dagger C) N s s^-,$$

or, equivalently,

$$(3.2) \quad Y = N + \tau \left(\frac{Pu}{\|Pu\|^2} \right) \left(z^\sharp - \frac{1}{\tau} N^T Pu \right)^T - (I - C^\dagger C) N s s^-,$$

where

$$(3.3) \quad h = QNs + \frac{1}{\tau}Pu,$$

and u is an arbitrary vector such that $Pu \neq 0$ and τ is given by

$$(3.4) \quad \tau^2 = \frac{\|Pu\|^2}{s^T z^\sharp}.$$

We begin by constructing a BFGS-like update. In (3.1), set

$$u = Ns.$$

Then, we have

$$\begin{aligned} h &= QNs + \frac{1}{\tau}(I - Q)Ns \\ &= \left(1 - \frac{1}{\tau}\right)(r_+^T Ns) \frac{r_+}{\|r_+\|^2} + \frac{1}{\tau}Ns, \end{aligned}$$

and

$$\|Pu\|^2 = s^T P^\sharp s, \quad \tau^2 = \frac{s^T P^\sharp s}{s^T z^\sharp}.$$

Since the result (3) in Theorem 2.2 implies $(I - C^\dagger C)Ns = 0$, the solution (3.1) yields

$$Y = N + \frac{1}{\tau} \left(\frac{PNs}{s^T z^\sharp} \right) \left(z^\sharp - \frac{1}{\tau} N^T PNs \right)^T = N + \left(\frac{PNs}{s^T P^\sharp s} \right) (\tau z^\sharp - P^\sharp s)^T.$$

Setting $Y = J_+ + L_+$ and $N = J_+ + PL$, we have

$$L_+ = PL + \left(\frac{PNs}{s^T P^\sharp s} \right) (\tau z^\sharp - P^\sharp s)^T,$$

which is the update formula of Sheng and Zou.

Setting $B_+ = Y^T Y$ yields

$$\begin{aligned} B_+ &= B^\sharp + \left(\frac{P^\sharp s}{s^T P^\sharp s} \right) (\tau z^\sharp - P^\sharp s)^T + (\tau z^\sharp - P^\sharp s) \left(\frac{P^\sharp s}{s^T P^\sharp s} \right)^T \\ &\quad + \frac{1}{s^T P^\sharp s} (\tau z^\sharp - P^\sharp s)(\tau z^\sharp - P^\sharp s)^T \\ &= B^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} \\ &= \left(P^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} \right) + Q^\sharp. \end{aligned}$$

This corresponds to a BFGS update from intermediate matrices B^\sharp and P^\sharp to a new Hessian approximation B_+ , so we call this an SZ-BFGS update. We summarize as follows: (SZ-BFGS update)

$$(3.5) \quad L_+^{BFGS} = PL + \left(\frac{PNs}{s^T P^\sharp s} \right) (\tau z^\sharp - P^\sharp s)^T,$$

$$(3.6) \quad \tau^2 = \frac{s^T P^\sharp s}{s^T z^\sharp},$$

$$\begin{aligned} (3.7) \quad B_+^{BFGS} &= \left(P^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} \right) + Q^\sharp \\ &= B^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp}. \end{aligned}$$

Note that the intermediate matrices B^\sharp , P^\sharp and Q^\sharp contain the projection information for the orthogonality condition (2.2).

Next we consider a DFP-like update. We will now assume that the matrix P^\sharp is nonsingular. We should note that a necessary and sufficient condition for the matrix P^\sharp to be nonsingular is given in Theorem 2.2. In (3.2), setting

$$u = N(P^\sharp)^{-1} z^\sharp \quad \text{and} \quad s^- = \left(\frac{z^\sharp}{s^T z^\sharp} \right)^T,$$

we have

$$\|Pu\|^2 = (z^\sharp)^T (P^\sharp)^{-1} z^\sharp, \quad N^T P u = z^\sharp \quad \text{and} \quad \tau^2 = \frac{(z^\sharp)^T (P^\sharp)^{-1} z^\sharp}{s^T z^\sharp}.$$

The result (3) of Theorem 2.2 means

$$0 = (I - C^\dagger C)h = (I - C^\dagger C) \left(Ns - PN s + \frac{1}{\tau} PN(P^\sharp)^{-1} z^\sharp \right),$$

and we have

$$(I - C^\dagger C)Ns = -(I - C^\dagger C)PNb, \quad b = \frac{1}{\tau}(P^\sharp)^{-1}z^\sharp - s.$$

Noting that $C^\dagger CPNb = (b^T z^\sharp)(Pu/\|Pu\|^2)$, we have

$$\begin{aligned} Y &= N + (\tau - 1) \left(\frac{Pu}{\|Pu\|^2} \right) (z^\sharp)^T + PNb \left(\frac{z^\sharp}{s^T z^\sharp} \right)^T - (\tau - 1) \left(\frac{Pu}{\|Pu\|^2} \right) (z^\sharp)^T \\ &= N + PNb \left(\frac{z^\sharp}{s^T z^\sharp} \right)^T. \end{aligned}$$

So setting $Y = J_+ + L_+$ and $N = J_+ + PL$, we have an SZ-DFP update:
(SZ-DFP update)

$$(3.8) \quad L_+^{DFP} = PL + PN \left(\frac{1}{\tau}(P^\sharp)^{-1}z^\sharp - s \right) \left(\frac{z^\sharp}{s^T z^\sharp} \right)^T,$$

$$(3.9) \quad \tau^2 = \frac{(z^\sharp)^T (P^\sharp)^{-1} z^\sharp}{s^T z^\sharp},$$

$$\begin{aligned} (3.10) \quad B_+^{DFP} &= \left(P^\sharp - \frac{P^\sharp s (z^\sharp)^T + z^\sharp s^T P^\sharp}{s^T z^\sharp} + \left(1 + \frac{s^T P^\sharp s}{s^T z^\sharp} \right) \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} \right) + Q^\sharp \\ &= B^\sharp - \frac{P^\sharp s (z^\sharp)^T + z^\sharp s^T P^\sharp}{s^T z^\sharp} + \left(1 + \frac{s^T P^\sharp s}{s^T z^\sharp} \right) \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp}. \end{aligned}$$

Note that the expression (3.10) corresponds to a DFP update from intermediate matrices B^\sharp and P^\sharp to a new Hessian approximation B_+ , and that the intermediate matrices B^\sharp , P^\sharp and Q^\sharp contain the projection information for the orthogonality condition (2.2).

It is well known that the standard Broyden family for general nonlinear minimization problems can be expressed by the linear combination of the standard BFGS update and the standard DFP update. Yabe[23] studied the factorized versions of the standard Broyden family. Yabe constructed the factorized version of the standard Broyden family by using the convex combination of the factorized BFGS and the factorized DFP updates. In the remainder of this section, we construct a new L -update which corresponds to the Broyden family, by using the same technique as Yabe[23]. For a parameter τ , let

$$\begin{aligned} \hat{L}_+^{BFGS} &= PL + \left(\frac{PNs}{s^T P^\sharp s} \right) (\tau z^\sharp - P^\sharp s)^T, \\ \hat{L}_+^{DFP} &= PL + PN(\tau(P^\sharp)^{-1}z^\sharp - s) \left(\frac{z^\sharp}{s^T z^\sharp} \right)^T \end{aligned}$$

and

$$L_+ = (1 - \sqrt{\phi}) \hat{L}_+^{BFGS} + \sqrt{\phi} \hat{L}_+^{DFP},$$

where ϕ is a parameter such that $0 \leq \phi \leq 1$. Setting

$$B_+ = (J_+ + L_+)^T (J_+ + L_+)$$

yields

$$\begin{aligned}
B_+ &= (1 - \sqrt{\phi})^2 (J_+ + \hat{L}_+^{BFGS})^T (J_+ + \hat{L}_+^{BFGS}) + \phi (J_+ + \hat{L}_+^{DFP})^T (J_+ + \hat{L}_+^{DFP}) \\
&\quad + (1 - \sqrt{\phi}) \sqrt{\phi} \{ (J_+ + \hat{L}_+^{BFGS})^T (J_+ + \hat{L}_+^{DFP}) + (J_+ + \hat{L}_+^{DFP})^T (J_+ + \hat{L}_+^{BFGS}) \} \\
&= (1 - \phi) (J_+ + \hat{L}_+^{BFGS})^T (J_+ + \hat{L}_+^{BFGS}) + \phi (J_+ + \hat{L}_+^{DFP})^T (J_+ + \hat{L}_+^{DFP})
\end{aligned}$$

Choosing a parameter τ such that

$$\left[(1 - \phi) \frac{s^T z^\sharp}{s^T P^\sharp s} + \phi \frac{(z^\sharp)^T (P^\sharp)^{-1} z^\sharp}{s^T z^\sharp} \right] \tau^2 = 1,$$

the secant condition $B_+ s = z$ is satisfied and we have

$$\begin{aligned}
B_+ &= (1 - \phi) \left(B^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{\tau^2}{s^T P^\sharp s} z^\sharp (z^\sharp)^T \right) \\
&\quad + \phi \left(B^\sharp - \frac{P^\sharp s (z^\sharp)^T + z^\sharp s^T P^\sharp}{s^T z^\sharp} + \tau^2 \frac{(z^\sharp)^T (P^\sharp)^{-1} z^\sharp}{(s^T z^\sharp)^2} z^\sharp (z^\sharp)^T + \frac{s^T P^\sharp s}{(s^T z^\sharp)^2} z^\sharp (z^\sharp)^T \right) \\
&= B^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} + \phi (s^T P^\sharp s) v^\sharp (v^\sharp)^T,
\end{aligned}$$

where

$$v^\sharp = \frac{P^\sharp s}{s^T P^\sharp s} - \frac{z^\sharp}{s^T z^\sharp}.$$

The preceding corresponds to a Broyden family from intermediate matrices B^\sharp and P^\sharp to a new Hessian approximation B_+ , and the intermediate matrices B^\sharp , P^\sharp and Q^\sharp contain the projection information for the orthogonality condition (2.2). Now we obtain a new family, called an SZ-Broyden family, as follows:

(SZ-Broyden family (L -update))

$$\begin{aligned}
(3.11) \quad L_+ &= PL + (1 - \sqrt{\phi}) \left(\frac{PNs}{s^T P^\sharp s} \right) (\tau z^\sharp - P^\sharp s)^T \\
&\quad + \sqrt{\phi} PN (\tau (P^\sharp)^{-1} z^\sharp - s) \left(\frac{z^\sharp}{s^T z^\sharp} \right)^T,
\end{aligned}$$

where

$$(3.12) \quad 0 \leq \phi \leq 1, \quad \left[(1 - \phi) \frac{s^T z^\sharp}{s^T P^\sharp s} + \phi \frac{(z^\sharp)^T (P^\sharp)^{-1} z^\sharp}{s^T z^\sharp} \right] \tau^2 = 1,$$

$$\begin{aligned}
(3.13) \quad B_+ &= \left(P^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} + \phi (s^T P^\sharp s) v^\sharp (v^\sharp)^T \right) + Q^\sharp \\
&= B^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} + \phi (s^T P^\sharp s) v^\sharp (v^\sharp)^T \\
&= (1 - \phi) B_+^{BFGS} + \phi B_+^{DFP},
\end{aligned}$$

$$(3.14) \quad v^\sharp = \frac{P^\sharp s}{s^T P^\sharp s} - \frac{z^\sharp}{s^T z^\sharp},$$

and the matrices B_+^{BFGS} and B_+^{DFP} are given in (3.7) and (3.10), respectively.

Note that setting $\phi = 0$ and $\phi = 1$ in the SZ-Broyden family (3.11) yields the SZ-BFGS update (3.5) and the SZ-DFP update (3.8), respectively. Since the approach of Yabe and Takahashi does not contain the orthogonality condition (2.2), this new family is expected to possess a new feature which cannot be obtained in our approach.

4 Relation between A-Updates and L-Updates, and a new Broyden-like Family

The main subject of this section is the investigation of the relationship between A -updates and L -updates. Using this relationship, we show that the structured Broyden family given by Yabe and Yamaki[28] can be regarded as the factorized version of the structured secant update from the convex class proposed by Engels and Martinez[11]. On the other hand, in the previous section we proposed the SZ-Broyden family (L -update) based on the idea of Sheng and Zou. As mentioned in Sections 1 and 3, this family has a feature different from our factorized updates. An application of the relationship between A -updates and L -updates to the SZ-Broyden family (L -update) enables us to obtain a new A -update which has a feature different from the family of Engels and Martinez.

Consider the case where we do not impose the orthogonality condition $L_+^T r_+ = 0$ on the matrix L_+ for the SZ-BFGS update. In this case, we may regard $P = I$. We then have

$$N = J_+ + L, \quad Q = 0, \quad Q^\sharp = 0, \quad z^\sharp = z \quad \text{and} \quad P^\sharp = B^\sharp.$$

Thus the update (3.5) is reduced to the BFGS-like update given by Yabe and Takahashi[25]:

$$(4.1) \quad L_+ = L + \left(\frac{(J_+ + L)s}{s^T B^\sharp s} \right) (\tau z - B^\sharp s)^T,$$

$$(4.2) \quad \tau^2 = \frac{s^T B^\sharp s}{s^T z}, \quad B^\sharp = (J_+ + L)^T (J_+ + L),$$

$$(4.3) \quad B_+ = B^\sharp - \frac{B^\sharp s s^T B^\sharp}{s^T B^\sharp s} + \frac{z z^T}{s^T z}.$$

In L -updates, the matrix $(J_+ + L)^T (J_+ + L)$ is a new approximation to the Hessian matrix $\nabla^2 f(x_+)$, and in A -updates, the matrix $J_+^T J_+ + A_+$ is a new approximation to the Hessian. Furthermore, the matrices $(J_+ + L)^T (J_+ + L)$ and $J_+^T J_+ + A$ are intermediate matrices for L -updates and A -updates, respectively. Thus we can regard the matrices $(J_+ + L)^T (J_+ + L)$ and $(J_+ + L_+)^T (J_+ + L_+)$ as the matrices $J_+^T J_+ + A$ and $J_+^T J_+ + A_+$, respectively. So, setting

$$(4.4) \quad B^\sharp = J_+^T J_+ + A \quad \text{and} \quad B_+ = J_+^T J_+ + A_+$$

in (4.3), we have an A -update

$$(4.5) \quad A_+ = A - \frac{w w^T}{s^T w} + \frac{z z^T}{s^T z}, \quad w = (J_+^T J_+ + A)s,$$

which is the structured BFGS update given by Al-Baali and Fletcher[1]. Thus the expression (4.1) corresponds to the factorized form of the structured BFGS update of Al-Baali and Fletcher. This is a relationship between A -updates and L -updates. This relation was suggested by Yabe[24]. Independently of us, Takahashi[20] also pointed out the relation between (4.1) and (4.5).

Consider also the case where we do not impose the orthogonality condition $L_+^T r_+ = 0$ on the matrix L_+ for the SZ-DFP update. Then the update (3.8) is reduced to the DFP-like update given in Yabe and Takahashi[25]:

$$(4.6) \quad L_+ = L + (J_+ + L) \left(\frac{1}{\tau} (B^\sharp)^{-1} z - s \right) \left(\frac{z}{s^T z} \right)^T,$$

$$(4.7) \quad \tau^2 = \frac{z^T (B^\sharp)^{-1} z}{s^T z},$$

$$(4.8) \quad B_+ = B^\sharp - \frac{B^\sharp s z^T + z s^T B^\sharp}{s^T z} + \left(1 + \frac{s^T B^\sharp s}{s^T z} \right) \frac{z z^T}{s^T z}.$$

Substituting (4.4) into (4.8), we have an A -update

$$(4.9) \quad A_+ = A + \frac{(q - As)z^T + z(q - As)^T}{s^T z} - \frac{s^T(q - As)}{(s^T z)^2} z z^T,$$

$$q = (J_+ - J)^T r_+,$$

which is the revised form of the structured DFP update given by Dennis, Gay and Welsch(DGW)[8]. Thus the expression (4.6) corresponds to the factorized form of the A -update (4.9). Note that the vector $\nabla f(x_+) - \nabla f(x)$ was originally used instead of z in Dennis et al.[8].

Next, consider the case where we do not impose the orthogonality condition $L_+^T r_+ = 0$ on the matrix L_+ for the SZ-Broyden family. Then the family (3.11) reduces to the structured Broyden family given by Yabe and Yamaki[28]:

$$(4.10) \quad L_+ = L + (1 - \sqrt{\phi}) \left(\frac{(J_+ + L)s}{s^T B^\sharp s} \right) (\tau z - B^\sharp s)^T$$

$$+ \sqrt{\phi} (J_+ + L) (\tau (B^\sharp)^{-1} z - s) \left(\frac{z}{s^T z} \right)^T,$$

$$(4.11) \quad B_+ = B^\sharp - \frac{B^\sharp s s^T B^\sharp}{s^T B^\sharp s} + \frac{z z^T}{s^T z} + \phi (s^T B^\sharp s) v v^T,$$

where

$$(4.12) \quad \tau^2 = \frac{1}{(1 - \phi) \frac{s^T z}{s^T B^\sharp s} + \phi \frac{z^T (B^\sharp)^{-1} z}{s^T z}},$$

$$(4.13) \quad B^\sharp = (J_+ + L)^T (J_+ + L) \quad \text{and} \quad v = \frac{B^\sharp s}{s^T B^\sharp s} - \frac{z}{s^T z}.$$

In the same way as above, substituting (4.4) into (4.11) we obtain an A -update:

$$(4.14) \quad A_+ = A - \frac{w w^T}{s^T w} + \frac{z z^T}{s^T z} + \phi (s^T w) v v^T,$$

$$v = \frac{w}{s^T w} - \frac{z}{s^T z}, \quad w = (J_+^T J_+ + A)s, \quad 0 \leq \phi \leq 1,$$

which corresponds to the structured secant update from the convex class independently proposed by Engels[10] and Martinez[14] (see also Engels and Martinez[11]). Thus the expression (4.10) can be regarded as the factorized form of their family.

We have stated the relationship between A -updates and L -updates above. Now we are interested in what A -update corresponds to the SZ-Broyden family (L -update), so we apply the relation (4.4) to the SZ-Broyden family (3.13). Since $Q^\sharp = J_+^T Q J_+$ and $Q = r_+ r_+^T / \|r_+\|^2$, we have

$$P^\sharp s = (B^\sharp - Q^\sharp)s = As + J_+^T (I - Q) J_+ s$$

and

$$z^\sharp = z - Q^\sharp s = (J_+ - J)^T r_+ + J_+^T (I - Q) J_+ s.$$

Thus we obtain a new A -update:
(SZ-Broyden family (A -update))

$$(4.15) \quad A_+ = A - \frac{w^\sharp (w^\sharp)^T}{s^T w^\sharp} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp} + \phi (s^T w^\sharp) v^\sharp (v^\sharp)^T,$$

where

$$\begin{aligned} v^\sharp &= \frac{w^\sharp}{s^T w^\sharp} - \frac{z^\sharp}{s^T z^\sharp}, \\ w^\sharp &= As + J_+^T \left(I - \frac{r_+ r_+^T}{\|r_+\|^2} \right) J_+ s, \\ z^\sharp &= (J_+ - J)^T r_+ + J_+^T \left(I - \frac{r_+ r_+^T}{\|r_+\|^2} \right) J_+ s \end{aligned}$$

and

$$0 \leq \phi \leq 1.$$

For $\phi = 0$ and $\phi = 1$, we have an SZ-BFGS update (A -update) and an SZ-DFP update (A -update), respectively:

$$A_+ = A - \frac{w^\sharp (w^\sharp)^T}{s^T w^\sharp} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp}$$

and

$$A_+ = A - \frac{w^\sharp (z^\sharp)^T + z^\sharp (w^\sharp)^T}{s^T z^\sharp} + \left(1 + \frac{s^T w^\sharp}{s^T z^\sharp} \right) \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp}.$$

Note that the preceding updates contain the projection information of the orthogonality condition (2.2) for an L -update in the vectors w^\sharp and z^\sharp . Thus we may expect this new A -update to possess a feature different from the family of Engels and Martinez in practical computations.

5 Sizing Techniques

We know that, for large residual problems, the structured quasi-Newton methods perform well, but that for zero and small residual problems, the Gauss-Newton method performs better. Thus, in the latter case, it is desirable for the structured quasi-Newton methods to follow the Gauss-Newton method. For this purpose, Bartholomew-Biggs[2] and Dennis et al.[8] introduced sizing techniques, and Al-Baali et al.[1] and Fletcher et al.[12] considered the combination of the structured quasi-Newton methods and the Gauss-Newton method — hybrid methods. Though both strategies can be applied to all the methods given in the previous sections, we consider only sizing techniques in what follows.

For A -updates, Bartholomew-Biggs[2] proposed a sizing parameter (Biggs parameter)

$$(5.1) \quad \beta_k = \frac{r(x_{k+1})^T r(x_k)}{r(x_k)^T r(x_k)}$$

based on the idea such that if $r(x_{k+1}) = \beta_k r(x_k)$ for some β_k , $A_k = \sum_{i=1}^m r_i(x_k) \nabla^2 r_i(x_k)$ and each $r_i(x_k)$ is quadratic, then $\sum_{i=1}^m r_i(x_{k+1}) \nabla^2 r_i(x_{k+1}) = \beta_k A_k$. Dennis et al.[8] proposed a sizing parameter (DGW parameter)

$$(5.2) \quad \beta_k = \min \left(\left| \frac{s_k^T (J_{k+1} - J_k)^T r_{k+1}}{s_k^T A_k s_k} \right|, 1 \right)$$

based on the idea that the spectrum of the sized matrix $\beta_k A_k$ should overlap that of the second part of the Hessian matrix in the direction of s_k in the sense that

$$\left| \left[s_k^T \left(\sum_{i=1}^m r_i(x_{k+1}) \nabla^2 r_i(x_{k+1}) \right) s_k / s_k^T s_k \right] [s_k^T (\beta_k A_k) s_k / s_k^T s_k]^{-1} \right| \approx |s_k^T (J_{k+1} - J_k)^T r_{k+1} / s_k^T (\beta_k A_k) s_k| = 1.$$

Note that the factor $s_k^T (J_{k+1} - J_k)^T r_{k+1} / s_k^T A_k s_k$ corresponds to the factor given in Oren and Luenberger[18], and Oren[16].

Now we present an algorithm for structured quasi-Newton methods with sizing techniques.

(Algorithm 1 for A -updates)

Starting with a point $x_1 \in R^n$ and an $n \times n$ matrix A_1 (usually, $A_1 = 0$ and $\beta_1 = 1$), the algorithm proceeds, for $k = 1, 2, \dots$, as follows:

Step 1. Having x_k and A_k , find the search direction d_k by solving the linear system of equations

$$(5.3) \quad (J_k^T J_k + A_k) d = -J_k^T r_k.$$

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 A_{k+1} by using the following A -updates:

(Engels and Martinez family)

$$(5.4) \quad A_{k+1} = \beta_k A_k - \frac{w_k w_k^T}{s_k^T w_k} + \frac{z_k z_k^T}{s_k^T z_k} + \phi_k (s_k^T w_k) v_k v_k^T,$$

where

$$v_k = \frac{w_k}{s_k^T w_k} - \frac{z_k}{s_k^T z_k}, \quad w_k = (J_{k+1}^T J_{k+1} + \beta_k A_k) s_k,$$

or

(SZ-Broyden family (A -update))

$$(5.5) \quad A_{k+1} = \beta_k A_k - \frac{w_k^\sharp (w_k^\sharp)^T}{s_k^T w_k^\sharp} + \frac{z_k^\sharp (z_k^\sharp)^T}{s_k^T z_k^\sharp} + \phi_k (s_k^T w_k^\sharp) v_k^\sharp (v_k^\sharp)^T,$$

where

$$\begin{aligned} v_k^\sharp &= \frac{w_k^\sharp}{s_k^T w_k^\sharp} - \frac{z_k^\sharp}{s_k^T z_k^\sharp}, \\ w_k^\sharp &= \beta_k A_k s_k + J_{k+1}^T (I - (\|r_{k+1}\|^2)^\dagger r_{k+1} r_{k+1}^T) J_{k+1} s_k, \\ z_k^\sharp &= (J_{k+1} - J_k)^T r_{k+1} + J_{k+1}^T (I - (\|r_{k+1}\|^2)^\dagger r_{k+1} r_{k+1}^T) J_{k+1} s_k, \end{aligned}$$

and β_k is defined by the Biggs parameter (5.1) or the DGW parameter (5.2), ϕ_k is a parameter such that

$$0 \leq \phi_k \leq 1,$$

and $(\|r_{k+1}\|^2)^\dagger$ denotes the Moore-Penrose generalized inverse of $\|r_{k+1}\|^2$.

Though we use the A -updates for our numerical experiments in the next section, we will also present the algorithm for L -updates. For the factorized version of the structured quasi-Newton methods, we can no longer use the DGW sizing parameter, because it explicitly contains the matrix A . So Yabe and Takahashi[26] used a similar idea to Dennis et al., and they proposed a sizing parameter

$$(5.6) \quad \beta_k = \min \left(\frac{|-(L_k s_k)^T J_{k+1} s_k + \text{sgn}((L_k s_k)^T J_{k+1} s_k) \sqrt{\mu_k}|}{\|L_k s_k\|^2}, 1 \right),$$

where

$$\mu_k = ((L_k s_k)^T J_{k+1} s_k)^2 + \|L_k s_k\|^2 |s_k^T (J_{k+1} - J_k)^T r_{k+1}|$$

and the symbol $\text{sgn}(\zeta)$ denotes the sign of ζ . In fact, the preceding can be obtained based on the idea that the spectrum of the matrix $(\beta_k L_k)^T (\beta_k L_k) + (\beta_k L_k)^T J_{k+1} + J_{k+1}^T (\beta_k L_k)$ overlaps that of the second part of the Hessian matrix in the direction of s_k in the sense that

$$\left| \frac{s_k^T (J_{k+1} - J_k)^T r_{k+1}}{s_k^T [(\beta_k L_k)^T (\beta_k L_k) + (\beta_k L_k)^T J_{k+1} + J_{k+1}^T (\beta_k L_k)] s_k} \right| = 1.$$

We now present an algorithm of the factorized version of structured quasi-Newton methods with sizing techniques.

(Algorithm 2 for L -updates)

Starting with a point $x_1 \in R^n$ and an $m \times n$ matrix L_1 (usually, $L_1 = 0$ and $\beta_1 = 1$), the algorithm proceeds, for $k = 1, 2, \dots$, as follows:

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

$$(5.7) \quad (J_k + L_k)^T (J_k + L_k) d = -J_k^T r_k.$$

(or, following Sheng and Zou, find the search direction d_k by solving the normal equation of (1.9)

$$(5.8) \quad (J_k + L_k)^T (J_k + L_k) d = -(J_k + L_k)^T r_k.)$$

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 the following L -updates:

(the Broyden-like family proposed by Yabe and Yamaki)

$$(5.9) \quad L_{k+1} = \beta_k L_k + (1 - \sqrt{\phi_k}) \left(\frac{(J_{k+1} + \beta_k L_k) s_k}{s_k^T B_k^\# s_k} \right) (\tau_k z_k - B_k^\# s_k)^T \\ + \sqrt{\phi_k} (J_{k+1} + \beta_k L_k) (\tau_k (B_k^\#)^{-1} z_k - s_k) \left(\frac{z_k}{s_k^T z_k} \right)^T,$$

where

$$\tau_k = \left[(1 - \phi_k) \frac{s_k^T z_k}{s_k^T B_k^\# s_k} + \phi_k \frac{z_k^T (B_k^\#)^{-1} z_k}{s_k^T z_k} \right]^{-1/2},$$

$$B_k^\# = (J_{k+1} + \beta_k L_k)^T (J_{k+1} + \beta_k L_k), \quad v_k = \frac{B_k^\# s_k}{s_k^T B_k^\# s_k} - \frac{z_k}{s_k^T z_k},$$

or

(SZ-Broyden family (L -update))

$$(5.10) \quad L_{k+1} = \beta_k P_k L_k + (1 - \sqrt{\phi_k}) \left(\frac{P_k N_k s_k}{s_k^T P_k^\# s_k} \right) (\tau_k z_k^\# - P_k^\# s_k)^T \\ + \sqrt{\phi_k} P_k N_k (\tau_k (P_k^\#)^{-1} z_k^\# - s_k) \left(\frac{z_k^\#}{s_k^T z_k^\#} \right)^T,$$

where

$$\tau_k = \left[(1 - \phi_k) \frac{s_k^T z_k^\#}{s_k^T P_k^\# s_k} + \phi_k \frac{(z_k^\#)^T (P_k^\#)^{-1} z_k^\#}{s_k^T z_k^\#} \right]^{-1/2},$$

$$Q_k = (\|r_{k+1}\|^2)^\dagger r_{k+1} r_{k+1}^T, \quad P_k = I - Q_k, \quad N_k = J_{k+1} + \beta_k P_k L_k,$$

$$P_k^\sharp = N_k^T P_k N_k, \quad Q_k^\sharp = N_k^T Q_k N_k \quad \text{and} \quad z_k^\sharp = z_k - Q_k^\sharp s_k,$$

and β_k is defined by (5.1) or (5.6), ϕ_k is a parameter such that

$$0 \leq \phi_k \leq 1,$$

and $(\|r_{k+1}\|^2)^\dagger$ denotes the Moore-Penrose generalized inverse of $\|r_{k+1}\|^2$.

Since setting $A_k = 0$ and $L_k = 0$ in Step 1 yields the Gauss-Newton direction, it is reasonable that the initial matrices are set to be zero matrices in the both algorithms. The sizing techniques described above have a feature which enables us to apply Algorithms 1 and 2 to large residual problems and zero residual problems. Dennis et al. proposed their parameter in conjunction with the Oren-Luenburger self-scaling technique. Using this idea, we have another property of the sizing techniques. In fact, Oren[16] showed that the Oren-Luenburger self-scaling variable metric (SSVM) method has the favorable property that it is invariant under scaling of the objective function or of the variables. Thus we have a similar proposition.

Proposition 5.1 (1) *Let $\{A_k\}, \{x_k\}, \{\beta_k\}$ and $\{\hat{A}_k\}, \{\hat{x}_k\}, \{\hat{\beta}_k\}$ be the corresponding sequences generated by Algorithm 1 when it is applied to $r(x)$ and $ar(\hat{x}/b)$ respectively, where a and b are positive scalars. If $\hat{A}_1 = (a^2/b^2)A_1$, $\hat{x}_1 = bx_1$ and the sequences $\{\phi_k\}$ and $\{\alpha_k\}$ are identical for both cases, then $\hat{A}_k = (a^2/b^2)A_k$ and $\hat{x}_k = bx_k$ for all k .*

(2) *Let $\{L_k\}, \{x_k\}, \{\beta_k\}$ and $\{\hat{L}_k\}, \{\hat{x}_k\}, \{\hat{\beta}_k\}$ be the corresponding sequences generated by Algorithm 2 when it is applied to $r(x)$ and $ar(\hat{x}/b)$ respectively, where a and b are positive scalars. If $\hat{L}_1 = (a/b)L_1$, $\hat{x}_1 = bx_1$ and the sequences $\{\phi_k\}$ and $\{\alpha_k\}$ are identical for both cases, then $\hat{L}_k = (a/b)L_k$ and $\hat{x}_k = bx_k$ for all k .*

Proof.

(1) Noting

$$\hat{x} = bx, \quad \hat{r}(\hat{x}) = ar(x) = ar\left(\frac{1}{b}\hat{x}\right) \quad \text{and} \quad \hat{J}(\hat{x}) = \frac{a}{b}J(x),$$

it is easily shown that, for $k = 1$,

$$\hat{d}_1 = -(\hat{J}_1^T \hat{J}_1 + \hat{A}_1)^{-1} \hat{J}_1^T \hat{r}_1 = bd_1,$$

which implies

$$\hat{x}_2 = \hat{x}_1 + \alpha_1 \hat{d}_1 = b(x_1 + \alpha_1 d_1) = bx_2.$$

We assume as an induction hypotheses that the results hold for k . Then, for $k + 1$, we have

$$\hat{s}_k = \hat{x}_{k+1} - \hat{x}_k = bs_k,$$

$$\hat{z}_k = (\hat{J}_{k+1} - \hat{J}_k)^T \hat{r}_{k+1} + \hat{J}_{k+1}^T \hat{J}_{k+1} \hat{s}_k = \frac{a^2}{b} z_k,$$

$$\hat{\beta}_k = \frac{\hat{r}_{k+1}^T \hat{r}_k}{\hat{r}_k^T \hat{r}_k} = \beta_k$$

or

$$\hat{\beta}_k = \min \left(\left| \frac{\hat{s}_k^T (\hat{J}_{k+1} - \hat{J}_k)^T \hat{r}_{k+1}}{\hat{s}_k^T \hat{A}_k \hat{s}_k} \right|, 1 \right) = \min \left(\left| \frac{s_k^T (J_{k+1} - J_k)^T r_{k+1}}{s_k^T A_k s_k} \right|, 1 \right) = \beta_k,$$

$$\hat{w}_k = (\hat{J}_{k+1}^T \hat{J}_{k+1} + \hat{\beta}_k \hat{A}_k) \hat{s}_k = \frac{a^2}{b} w_k, \quad \text{and} \quad \hat{v}_k = \frac{\hat{w}_k}{\hat{s}_k^T \hat{w}_k} - \frac{\hat{z}_k}{\hat{s}_k^T \hat{z}_k} = \frac{1}{b} v_k.$$

Thus we have the matrix

$$\begin{aligned} \hat{A}_{k+1} &= \hat{\beta}_k \hat{A}_k - \frac{\hat{w}_k \hat{w}_k^T}{\hat{s}_k^T \hat{w}_k} + \frac{\hat{z}_k \hat{z}_k^T}{\hat{s}_k^T \hat{z}_k} + \phi_k (\hat{s}_k^T \hat{w}_k) \hat{v}_k \hat{v}_k^T \\ &= \frac{a^2}{b^2} A_{k+1} \end{aligned}$$

and the search direction

$$\hat{d}_{k+1} = -(\hat{J}_{k+1}^T \hat{J}_{k+1} + \hat{A}_{k+1})^{-1} \hat{J}_{k+1}^T \hat{r}_{k+1} = b d_{k+1},$$

which implies

$$\hat{x}_{k+2} = \hat{x}_{k+1} + \alpha_{k+1} \hat{d}_{k+1} = b(x_{k+1} + \alpha_{k+1} d_{k+1}) = b x_{k+2}.$$

Since the same argument can be applied to the case of the SZ-Broyden family (A -update) and (2), we omit the proof. ■

Since initial matrices are usually set to be zero matrices, i.e. $A_1 = 0$ and $L_1 = 0$, the assumptions in the above proposition, $\hat{A}_1 = (a^2/b^2)A_1$ and $\hat{L}_1 = (a/b)L_1$ are not restrictive in practical computation.

6 Computational Experiments

The purposes of our numerical experiments are to compare a new A -update (5.5) with the Engels and Martinez family (5.4), and to investigate how the computational performance depends on the choice of the parameters ϕ_k and β_k given in Algorithm 1 from the viewpoint of the number of iterations and the number of vector valued function (i.e. $r(x)$) evaluations. Note that there are different strategies among the structured quasi-Newton methods. Dennis et al.[8] combined the DGW update and the trust region globalization strategy, and Al-Baali et al.[1] proposed the hybrid method which combined the Gauss-Newton method and the structured BFGS update for the line search globalization strategy, and so forth. In this section, we just compare the performance of some updates for the line search globalization strategy.

The numerical calculations were carried out in double precision arithmetic on a SUN SPARC station 1+, and the program was coded in FORTRAN 77. The Jacobian matrix is evaluated by the forward difference approximation. In Algorithm 1, the initial matrix A_1 is set to zero matrix. The linear system of equations in Step 1 is solved by the modified

Cholesky method, i.e. when the coefficient matrix cannot be decomposed because of indefiniteness, a diagonal element of Cholesky factor was replaced by a small positive number. In Step 2, the bisection line search method with Armijo's rule

$$(6.1) \quad f(x_k + \alpha_k d_k) \leq f(x_k) + 0.1 \alpha_k \nabla f(x_k)^T d_k$$

is employed. Further, in Step 4, the iterative process is terminated

$$(T1) \text{ if } \|r(x_{k+1})\|_\infty \leq \max(\text{TOL1}, \varepsilon),$$

or

$$(T2) \text{ if } |e_i^T J(x_{k+1})^T r(x_{k+1})| \leq \max(\text{TOL2}, \varepsilon) \|r(x_{k+1})\| \|J(x_{k+1}) e_i\| \text{ for } i = 1, \dots, n \\ \text{and } \|x_{k+1} - x_k\|_\infty \leq \max(\text{TOL3}, \varepsilon) \max(\|x_{k+1}\|_\infty, 1.0), \text{ where } e_i \text{ denotes the } i\text{-th} \\ \text{column of the unit matrix,}$$

or

$$(T3) \text{ if the number of iterations exceeds the prescribed limit (ITMAX),}$$

or

$$(T4) \text{ if the number of function evaluations exceeds the prescribed limit (NFEMAX),}$$

where $\|\bullet\|_\infty$ denotes the maximum norm and ε is machine epsilon. The modified Cholesky method and the stopping criteria described above followed the code NOLLS1 in Tanabe and Ueda[21]. In the experiments, we set $\text{TOL1} = \text{TOL2} = \text{TOL3} = 10^{-4}$, $\text{ITMAX} = 500$ and $\text{NFEMAX} = 2000$. For the SZ-Broyden family (A -update) in Step 5, the Moor-Penrose generalized inverse $(\|r_{k+1}\|^2)^\dagger$ was numerically set to $(1/\|r_{k+1}\|^2)$ if $\|r_{k+1}\|^2 \geq 10^{-20}$, and 0 otherwise. Since the stopping criteria (T1) with $\text{TOL1} = 10^{-4}$ was used, $(\|r_{k+1}\|^2)^\dagger$ was not set to zero in our numerical experiments. In addition to the convex classes of the Broyden-like families mentioned in the previous sections, we used the Gauss-Newton method (GN) and the structured symmetric rank one (SR1) update for comparison. The structured SR1 update with sizing was first proposed by Bartholomew-Biggs[2], and is represented by

$$(6.2) \quad A_{k+1} = \beta_k A_k + \frac{(q_k - \beta_k A_k s_k)(q_k - \beta_k A_k s_k)^T}{s_k^T (q_k - \beta_k A_k s_k)^T},$$

where

$$q_k = (J_{k+1} - J_k)^T r_{k+1}.$$

Since the DGW sizing parameter (5.2) makes the denominator zero, we just applied the Biggs sizing parameter to the above.

The names, the sizes and the starting points of the test problems we used are listed in Table 1, together with the abbreviated problem names. These problems are given in Dennis et al.[8], and are in detail in Moré, Garbow and Hillstom[15]. In Table 1, (Z), (S) and (L) mean a zero residual problem, a small residual problem and a large residual problem, respectively. Tables 2, 3 and 4 are computational results for the Engels and Martinez family with no sizing, with DGW sizing and Biggs sizing parameters, respectively.

Tables 5, 6 and 7 are for the SZ-Broyden (A -update) with no sizing, DGW sizing and Biggs sizing parameters, respectively. The computational results for the Gauss-Newton method and the structured SR1 update are given in Table 8. In each table, the number of iterations and the number of function evaluations are written. The latter is written in a parenthesis in the tables, and contains the number to evaluate the Jacobian matrix by forward finite difference. Also, the asterisk * in each table means that the method failed to converge in the specified number of iterations or function evaluations. Finally we summarize these results in Tables 9 and 10. Tables 9 and 10 mean the total numbers of iterations and function evaluations for all the methods, respectively. In each table, the number in the parenthesis denotes the performance ratio of sizing techniques. For example, in the part of B(0.1) with DGW sizing in Table 9.1, the ratio 0.844 implies 304/360. The small ratio means that the sizing technique works very well. However we should note that this ratio depends on the choices of ITMAX and NFEMAX in the stopping criteria in the case where the symbol * is attached. In each table, "B(ξ)" means the results of the Broyden-like family with $\phi_k = \xi$. So, in the Engels and Martinez family, "B(0.0)" and "B(1.0)" corresponds to the results of the Al-Baali and Fletcher update and the revised DGW update, respectively. However, since Al-Baali and Fletcher proposed the hybrid method and Dennis et al. used the trust region strategy, we cannot make a direct comparison with their results.

The methods B(0.0) for WATSON12, B(0.8) for WATSON9, B(1.0) for ROSENBROCK, HELIX in Table 2; B(0.0) for WATSON12 in Table 3; B(0.5) for WATSON9, B(0.9) for WATSON20, B(1.0) for WATSON9, WATSON12, WATSON20, ROSENBROCK, HELIX, BEALE and OSBORNE1 in Table 5 failed in the sense that they did not converge to the solutions in the specified number of iterations or function evaluations. All the cases stopped near the minimum points but B(1.0) for ROSENBROCK. These poor performances were caused by the poor line search procedures. The method B(1.0) in Table 5 performed very poorly. In Table 8, The Gauss-Newton method for FRDSTEIN2 and JENNRICH failed because of the large residuals at the solutions, but they had a tendency to approach the minimum points.

The numerical results in Tables 2-8 suggest the efficiency of the sizing techniques. However, they did not always work well. In the Engels and Martinez family, the numerical results for POWELL, FRDSTEIN1 and BOX did not depend on sizing techniques. For some problems, sizing techniques spoiled the performance, e.g. B(0.4), B(0.6), B(0.7) and B(0.9) for WATSON12, B(0.4) and B(0.8) for ROSENBROCK, B(0.0), B(0.1), B(0.2), B(0.3) and B(0.4) for FRDSTEIN2, B(0.8) for KOWALIK, and B(0.1), B(0.2) and B(0.8) for OSBORNE1 in Table 3; B(0.0) and B(0.2) for WATSON9, B(0.4), B(0.6), B(0.7) and B(0.9) for WATSON12, B(0.2) for HELIX, B(0.1), B(0.2), B(0.3), B(0.4), B(0.7), B(0.8), B(0.9) and B(1.0) for OSBORNE1, and B(0.0), B(0.1) and B(0.2) for JENNRICH in Table 4; B(0.3) for WATSON12, B(0.2), B(0.3) and B(0.4) for FRDSTEIN2, B(0.9) for POWELL, and B(0.7) for FRDSTEIN2 in Table 6; B(0.3) for WATSON12 in Table 7; and SR1 for WATSON12 in Table 8.

Speaking of the choice of the parameter ϕ_k , the choice did not affect the performance for POWELL, FRDSTEIN1, FRDSTEIN2 and BOX in Tables 2; for POWELL, FRDSTEIN1, BARD and BOX in Tables 3 and 4; for POWELL, BEALE, BOX and JENNRICH in Table 6; and for POWELL, BEALE, FRDSTEIN1 and BOX in Table 7.

There seems no special choice of ϕ_k for which the results were best, and it depends on the situation.

From Tables 9 and 10, we summarize our numerical results as follows:

- (1) The structured quasi-Newton methods with sizing are more robust than the Gauss-Newton method.
- (2) The Engels and Martinez family matched with the DGW sizing parameter, and the SZ-Broyden family (A -update) matched with the Biggs sizing parameter. Further the SZ-Broyden family with Biggs sizing parameter worked better than the other families.
- (3) For both families with sizing, the cases of $\phi = 0.5$ were numerically stable.
- (4) As the parameter ϕ approached 1, the performance of sizing techniques increased. The DFP-like update without sizing was much inferior to other updates without sizing. On the other hand, the DFP-like update with sizing worked as well as the other sized updates.
- (5) The Bartholomew-Biggs update, i.e. the structured symmetric rank one update with Biggs parameter, worked well.

The result (2) suggests that an application of features of L -updates to A -updates is promising. In this paper, we suggested one relationship between A -updates and L -updates. This result encouraged us to study another relation and to propose a new A -update which corresponds to a new L -update. The result (3) is somewhat similar to the numerical results given in Oren[17], in which Oren applied his sizing parameter to the standard Broyden family for general minimization problems. The result (4) means that the DFP-like update needs sizing technique very much. However this does not mean that the other updates, e.g. BFGS-like update, need no sizing. There is a possibility of finding another kind of sizing parameter which is effective for the other updates. This result supports the research of Contreras and Tapia[5]. In their paper, they claimed that the standard DFP update needed to be sized for general minimization problems, and that the DFP update was much imposed when matched with the Oren-Luenberger sizing parameter. They proposed another kind of sizing parameter for the standard BFGS update. Their idea may be applied to the structured quasi-Newton methods. The result (5) encourages us to study a nonconvex class of the structured Broyden family, because the structured symmetric rank one update does not belong to the convex class but is a member of the Broyden-like family.

7 Concluding Remarks

This paper presents variations of structured Broyden families for nonlinear least squares problems. The main purposes of this paper are to investigate the relationship between A -updates and L -updates, and to obtain a new A -update which possess a new feature. First, we generalized the update of Sheng and Zou and proposed a new update which corresponds to the Broyden family, say, SZ-Broyden family (L -update). Second, we suggested the relation between A -updates and L -updates, and showed that the structured Broyden family proposed by Yabe and Yamaki can be regarded as the factorized version of the Engels and Martinez family. Third, we obtained the SZ-Broyden family (A -update) which corresponds to the SZ-Broyden family (L -update). Further, we applied the sizing techniques to these updates. Finally, we gave some numerical experiments. The numeri-

cal results show that the SZ-Broyden family (A -update) with the Biggs sizing parameter works well. These results also show that the DFP-like update needs sizing techniques very much and supports the research of Contreras and Tapia[5].

Further investigation of the relationship between A -updates and L -updates seems very promising to us. This may give us a new A -updates. Since L -updates enable us to obtain a descent search direction for the objective function, by investigating the relation we may expect to find conditions under which matrices $J_k^T J_k + A_k$ possess the hereditary positive definiteness property for A -updates. However, the relation mentioned in this paper is not exact yet, because the intermediate matrices of A -updates and L -updates do not in general correspond exactly. The results of Section 4 seem to give us a clue to understanding the relation.

A further topic of research is proof of convergence of the methods with the SZ-Broyden families (A -update and L -update). In this paper, we focused on A -updates, but the numerical implementation of the SZ-Broyden family (L -update) (5.10) also remains. This family is theoretically interesting, but it contains the inverse of the projection matrix P_k^\sharp in the calculation of the vector $c = (P_k^\sharp)^{-1} z_k^\sharp$. In practical computations, it is possible for the matrix P_k^\sharp to be singular. Though we gave the mathematically necessary and sufficient condition for P_k^\sharp to be positive definite in Theorem 2.2, we must devise an efficient calculation of the vector c for practical computations. Numerical results of the family (5.9) can be found in Yabe and Yamaki[28]. However when the numerical results between A -updates and L -updates are compared, the arithmetic costs should be considered. Since A -updates are calculated in $O(n^2)$ arithmetic cost and L -updates are calculated in $O(mn)$ arithmetic cost respectively, the numbers of iterations cannot be directly compared.

This paper mainly dealt with the convex classes of the Broyden-like families. As mentioned in the previous section, updates which are not contained in the convex classes are also promising. The structured SR1 update is especially interesting. In fact, setting

$$\phi_k = \frac{s_k^T z_k}{s_k^T (z_k - w_k)} \quad \text{and} \quad \phi_k = \frac{s_k^T z_k^\sharp}{s_k^T (z_k^\sharp - w_k^\sharp)}$$

in (5.4) and (5.5), respectively, we have

$$A_{k+1} = \beta_k A_k + \frac{(z_k - w_k)(z_k - w_k)^T}{s_k^T (z_k - w_k)}$$

and

$$A_{k+1} = \beta_k A_k + \frac{(z_k^\sharp - w_k^\sharp)(z_k^\sharp - w_k^\sharp)^T}{s_k^T (z_k^\sharp - w_k^\sharp)}.$$

Since $z_k - w_k = z_k^\sharp - w_k^\sharp = (J_{k+1} - J_k)^T r_{k+1} - \beta_k A_k s_k$, the above yields the structured SR1 update (6.2). This means that the structured SR1 update is a common member of the nonconvex classes of the Engels-Martinez family and the SZ-Broyden family (A -update). However, note that the projection information of the orthogonality condition (2.2) is no longer included in the structured SR1 update.

Acknowledgements:

Parts of this work were done while the author was visiting Department of Mathematical Sciences and the Center for Research on Parallel Computation (CRPC) at Rice University, Houston, Texas, USA. I appreciate CRPC for providing me a SUN SPARC station 1+ for the numerical experiments. I would like to thank Prof. John E. Dennis, Jr. and Prof. Richard A. Tapia, Dept. of Mathematical Sciences, Rice University, for providing valuable information and their encouragement for this research. I am also grateful to Dr. Robert Michael Lewis, Dept. of Mathematical Sciences, Rice University, for his valuable comments to the draft of this paper.

References

1. M.Al-Baali and R.Fletcher (1985) Variational methods for non-linear least squares. *Journal of the Operational Research Society* 36 , No.5, 405-421.
2. M.C.Bartholomew-Biggs (1977) The estimation of the Hessian matrix in nonlinear least squares problems with non-zero residuals. *Mathematical Programming* 12, 67-80.
3. A.Ben-Israel and T.N.E.Greville (1980) *Generalized Inverses - Theory and Applications* . Robert E. Krieger Publishing Company, Huntington.
4. K.M.Brown and J.E.Dennis,Jr. (1971) A new algorithm for nonlinear least-squares curve fitting, In *Mathematical Software*, J.Rice ed., Academic Press, New York, pp.391-396.
5. M.Contreras and R.A.Tapia (1991) *Sizing the BFGS and DFP updates: A numerical study*, Technical Report TR91-19, July, Dept. of Mathematical Sciences, Rice University, Houston, Texas, USA.
6. J.E.Dennis,Jr. (1973) Some computational techniques for the nonlinear least squares problem, In *Numerical Solution of Systems of Nonlinear Algebraic Equations*, G.D.Byrne and C.A.Hall, eds., Academic Press, New York, pp.157-183.
7. J.E.Dennis,Jr. (1976) A brief survey of convergence results for quasi-Newton methods. *SIAM-AMS Proceedings* 9, 185-199.
8. J.E.Dennis,Jr., D.M.Gay and R.E.Welsch (1981) An adaptive nonlinear least squares algorithm. *ACM Transactions on Mathematical Software* 7, No.3, 348-368.
9. J.E.Dennis,Jr., H.J.Martinez and R.A.Tapia (1989) Convergence theory for the structured BFGS secant method with an application to nonlinear least squares. *Journal of Optimization Theory and Applications* 61, No.2, 161-178.
10. J.R.Engels (1988) *Local convergence analysis for partially known quasi-Newton updates*, Technical Report 88-13, July, Dept. of Mathematics, Facultés Universitaires ND de la Paix, Belgium.
11. J.R.Engels and H.J.Martinez (1991) Local and superlinear convergence for partially known quasi-Newton methods. *SIAM J. on Optimization* 1, No.1, 42-56.
12. R.Fletcher and C.Xu (1987) Hybrid methods for nonlinear least squares. *IMA J. Numerical Analysis* 7, 371-389.
13. J.Huschens (1991) *On the use of product structure in secant methods for nonlinear least squares problems*, Manuscript, Universität Trier, FB IV-Mathematik, Postfach 3825, 5500 Trier, Federal Republic of Germany.
14. H.J.Martinez (1988) *Local and superlinear convergence of structured secant methods from the convex class* , PhD Thesis, April, Rice University, Houston, Texas, USA. (see also Technical Report TR88-01, January, Dept. of Mathematical Sciences, Rice University)
15. J.J.Moré, B.S.Garbow and K.E.Hillstom (1981) Testing unconstrained optimization software. *ACM Transactions on Mathematical Software* 7, No.1, 17-41.

16. S.S.Oren (1974) Self-scaling variable metric (SSVM) algorithms, Part 2: Implementation and Experiments. *Management Science* 20, No.5, 863-874.
17. S.S.Oren (1974) On the selection of parameters in self scaling variable metric algorithms. *Mathematical Programming* 7, 351-367.
18. S.S.Oren and D.G.Luenberger (1974) Self-scaling variable metric (SSVM) algorithms, Part 1: Criteria and sufficient conditions for scaling a class of algorithms. *Management Science* 20, No.5, 845-862.
19. S.Sheng and Z.Zou (1988) *A new secant method for nonlinear least squares problems*, Technical Report NANO-1988-03, March, Nanjing University, People's Republic of China.
20. T.Takahashi (1990) Private memorandum at Department of Mathematical Sciences, Rice University, Houston, Texas, USA.
21. K.Tanabe and S.Ueda (1981) NOLLS1, A FORTRAN subroutine for nonlinear least squares by a quasi-Newton method. *Computer Science Monographs* 17, The Institute of Statistical Mathematics, Tokyo, Japan.
22. C.X.Xu (1990) Hybrid method for nonlinear least-square problems without calculating derivatives. *Journal of Optimization Theory and Applications* 65, No.3, 555-574.
23. H.Yabe (1981) A family of variable-metric methods with factorized expressions. *TRU Mathematics* 17, No.1, 141-152.
24. H.Yabe (1990) Generalization of the Sheng Songbai and Zou Zhihong method for solving nonlinear least squares problems. *Studies in Liberal Arts and Sciences* 23, Science University of Tokyo, Tokyo, Japan, in *Japanese*, 119-130.
25. H.Yabe and T.Takahashi (1988) Structured quasi-Newton methods for nonlinear least squares problems. *TRU Mathematics* 24, No.2, 195-209.
26. H.Yabe and T.Takahashi (1991) Factorized quasi-Newton methods for nonlinear least squares problems. *Mathematical Programming* 51, No.1, 75-100.
27. H.Yabe and T.Takahashi (1991) Numerical comparison among structured quasi-Newton methods for nonlinear least squares problems. *Journal of the Operations Research Society of Japan* 34, No.3, 287-305.
28. H.Yabe and N.Yamaki (1991) Convergence of structured quasi-Newton methods with structured Broyden family, In *Nonlinear Optimization - Modeling and Algorithms*. Cooperative Research Report 29, March, The Institute of Statistical Mathematics, Tokyo, Japan, in *Japanese*, 120-136.

Table 1. Test Problems

Abbreviated Name	Name of Test Problem	m	n	Starting Point	Residual
WATSON6	Watson Problem	31	6	(0, 0, ..., 0)	(S)
WATSON9	Watson Problem	31	9	(0, 0, ..., 0)	(S)
WATSON12	Watson Problem	31	12	(0, 0, ..., 0)	(S)
WATSON20	Watson Problem	31	20	(0, 0, ..., 0)	(S)
ROSENBROCK	Rosenbrock Problem	2	2	(-1.2, 1.0)	(Z)
HELIX	Helical Valley Problem	3	3	(-1, 0, 0)	(Z)
POWELL	Powell's Singular Problem	4	4	(3, -1, 0, 1)	(Z)
BEALE	Beale Problem	3	2	(0.1, 0.1)	(Z)
FRDSTEIN1	Freudenstein and Roth Problem	2	2	(6, 6)	(Z)
FRDSTEIN2	Freudenstein and Roth Problem	2	2	(15, -2)	(L)
BARD	Bard Problem	15	3	(1, 1, 1)	(S)
BOX	Box Problem	10	3	(0, 10, 20)	(Z)
KOWALIK	Kowalik Problem	11	4	(0.25, 0.39, 0.415, 0.39)	(S)
OSBORNE1	Osborne Problem	33	5	(0.5, 1.5, -1.0, 0.01, 0.02)	(S)
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)	(S)
JENNRICH	Jennrich Problem	10	2	(0.3, 0.4)	(L)

Table 2. Engels and Martinez family without sizing

		BFGS									
		B	(0.0)	B	(0.1)	B	(0.2)	B	(0.3)	B	(0.4)
WATSON6	(S)	15(119)	17(128)	20(156)	16(125)	69(493)
WATSON9	(S)	74(757)	82(832)	68(698)	98(999)	94(958)
WATSON12	(S)	153(2009)*	65(860)	78(1036)	68(907)	19(271)
WATSON20	(S)	15(343)	17(380)	57(1228)	18(408)	20(453)
ROSENBROCK	(Z)	23(80)	27(93)	28(94)	31(102)	27(91)
HELIX	(Z)	22(99)	18(85)	22(97)	21(95)	20(87)
POWELL	(Z)	14(75)	14(75)	14(75)	14(75)	14(75)
BEALE	(Z)	12(48)	16(94)	13(50)	13(50)	10(41)
FRDSTEIN1	(Z)	6(21)	6(21)	6(21)	6(21)	6(21)
FRDSTEIN2	(L)	7(26)	7(24)	7(24)	7(24)	6(22)
BARD	(S)	9(41)	9(41)	9(41)	9(41)	10(45)
BOX	(Z)	5(24)	5(24)	5(24)	5(24)	5(24)
KOWALIK	(S)	12(75)	12(74)	12(70)	13(76)	13(75)
OSBORNE1	(S)	33(216)	26(172)	27(175)	32(206)	28(180)
OSBORNE2	(S)	22(286)	25(325)	23(301)	25(317)	22(278)
JENNRICH	(L)	9(50)	14(53)	15(70)	12(60)	17(75)

Table 2. (Continued)

		B	(0.5)	B	(0.6)	B	(0.7)	B	(0.8)	B	(0.9)	B	DFP (1.0)
WATSON6		18(148)	24(177)	18(134)	22(161)	20(147)	26(190)
WATSON9		96(976)	101(1025)	111(1121)	199(2007)*	97(982)	114(1152)
WATSON12		66(877)	48(642)	59(781)	60(802)	53(704)	62(821)
WATSON20		17(384)	34(740)	17(379)	22(492)	30(653)	49(1052)
ROSENBROCK		35(131)	46(170)	31(101)	24(79)	244(761)	500(1529)*
HELIX		27(126)	31(134)	25(106)	26(110)	38(163)	497(2002)*
POWELL		14(75)	14(75)	14(75)	14(75)	14(75)	14(75)
BEALE		10(60)	16(57)	17(60)	16(57)	27(93)	15(55)
FRDSTEIN1		6(21)	6(21)	6(21)	6(21)	6(21)	6(21)
FRDSTEIN2		6(22)	5(19)	6(22)	6(22)	6(22)	6(22)
BARD		11(49)	11(49)	11(49)	11(49)	12(53)	12(53)
BOX		5(24)	5(24)	5(24)	5(24)	5(24)	5(24)
KOWALIK		10(61)	11(65)	10(60)	9(55)	9(55)	9(55)
OSBORNE1		32(208)	37(236)	26(167)	24(156)	28(177)	26(191)
OSBORNE2		22(279)	26(328)	27(355)	26(326)	28(365)	39(482)
JENNRICH		20(66)	22(71)	24(77)	32(101)	41(128)	87(266)

Table 3. Engels and Martinez family with DGW sizing

		BFGS				
		B (0.0)	B (0.1)	B (0.2)	B (0.3)	B (0.4)
WATSON6	(S)	8(71)	8(65)	8(64)	10(77)	6(49)
WATSON9	(S)	73(749)	79(802)	81(821)	79(800)	81(820)
WATSON12	(S)	153(2011)*	65(860)	62(820)	61(806)	62(819)
WATSON20	(S)	10(240)	6(149)	6(148)	9(210)	5(126)
ROSENBROCK	(Z)	18(71)	22(111)	20(80)	21(77)	37(175)
HELIX	(Z)	15(73)	17(82)	14(77)	12(61)	13(59)
POWELL	(Z)	14(75)	14(75)	14(75)	14(75)	14(75)
BEALE	(Z)	11(45)	9(38)	9(37)	10(40)	12(46)
FRDSTEIN1	(Z)	6(21)	6(21)	6(21)	6(21)	6(21)
FRDSTEIN2	(L)	11(59)	8(32)	8(27)	8(29)	7(26)
BARD	(S)	8(37)	8(37)	8(37)	8(37)	8(37)
BOX	(Z)	5(24)	5(24)	5(24)	5(24)	5(24)
KOWALIK	(S)	10(65)	10(64)	11(65)	9(56)	10(61)
OSBORNE1	(S)	22(162)	27(198)	24(183)	26(203)	23(182)
OSBORNE2	(S)	15(203)	13(177)	13(182)	13(180)	13(171)
JENNRICH	(L)	8(46)	7(31)	9(36)	9(32)	13(63)

Table 3. (Continued)

		B (0.5)	B (0.6)	B (0.7)	B (0.8)	B (0.9)	DFP (1.0)
WATSON6		9(70)	9(70)	9(70)	9(70)	9(70)	8(63)
WATSON9		81(820)	82(830)	81(821)	81(820)	80(811)	79(800)
WATSON12		61(806)	61(806)	61(807)	60(793)	61(807)	60(793)
WATSON20		9(210)	8(189)	8(190)	7(168)	7(169)	7(168)
ROSENBROCK		16(85)	19(99)	18(77)	20(98)	15(102)	17(95)
HELIX		18(87)	14(73)	11(51)	13(58)	14(62)	15(66)
POWELL		14(75)	14(75)	14(75)	14(75)	14(75)	14(75)
BEALE		10(41)	9(37)	10(40)	10(40)	11(43)	9(36)
FRDSTEIN1		6(21)	6(21)	6(21)	6(21)	6(21)	6(21)
FRDSTEIN2		6(23)	6(22)	6(23)	6(23)	6(23)	7(26)
BARD		8(37)	8(37)	8(37)	8(37)	8(37)	8(37)
BOX		5(24)	5(24)	5(24)	5(24)	5(24)	5(24)
KOWALIK		10(60)	10(60)	10(62)	10(61)	8(50)	9(55)
OSBORNE1		26(210)	24(185)	21(156)	22(166)	21(161)	21(168)
OSBORNE2		13(170)	13(170)	13(170)	13(170)	13(170)	13(170)
JENNRICH		8(30)	7(26)	7(26)	8(29)	8(29)	9(32)

Table 4. Engels and Martinez family with Biggs sizing

		BFGS					
		B (0.0)	B (0.1)	B (0.2)	B (0.3)	B (0.4)	
WATSON6	(S)	10(82)	8(65)	8(64)	10(78)	8(63)	
WATSON9	(S)	76(775)	80(812)	82(831)	81(821)	81(820)	
WATSON12	(S)	74(980)	64(847)	62(820)	63(833)	62(819)	
WATSON20	(S)	8(194)	6(149)	5(127)	9(211)	7(168)	
ROSENBROCK	(Z)	25(84)	21(81)	21(90)	22(81)	21(79)	
HELIX	(Z)	14(72)	15(80)	28(135)	16(78)	13(70)	
POWELL	(Z)	14(75)	14(75)	14(75)	14(75)	14(75)	
BEALE	(Z)	8(34)	8(34)	11(45)	7(30)	7(31)	
FRDSTEIN1	(Z)	6(21)	6(21)	6(21)	6(21)	6(21)	
FRDSTEIN2	(L)	6(23)	7(24)	7(24)	6(21)	7(24)	
BARD	(S)	8(37)	8(37)	8(37)	8(37)	8(37)	
BOX	(Z)	5(24)	5(24)	5(24)	5(24)	5(24)	
KOWALIK	(S)	9(57)	11(68)	11(65)	9(55)	11(65)	
OSBORNE1	(S)	30(196)	32(211)	29(194)	35(228)	35(222)	
OSBORNE2	(S)	22(295)	18(243)	15(201)	14(183)	15(196)	
JENNRICH	(L)	10(69)	20(68)	21(72)	15(57)	18(60)	

Table 4. (Continued)

		B (0.5)	B (0.6)	B (0.7)	B (0.8)	B (0.9)	DFP (1.0)
WATSON6		9(70)	7(56)	10(77)	9(71)	10(78)	10(78)
WATSON9		80(810)	81(820)	81(821)	80(812)	81(821)	80(811)
WATSON12		61(806)	61(806)	61(807)	61(808)	61(807)	61(807)
WATSON20		8(189)	5(126)	8(190)	10(233)	9(211)	8(190)
ROSENBROCK		17(110)	21(81)	20(86)	18(79)	21(72)	21(72)
HELIX		13(67)	14(64)	13(61)	12(61)	12(59)	12(59)
POWELL		14(75)	14(75)	14(75)	14(75)	14(75)	14(75)
BEALE		8(34)	8(34)	10(41)	8(34)	9(35)	10(40)
FRDSTEIN1		6(21)	6(21)	6(21)	6(21)	6(21)	6(21)
FRDSTEIN2		6(22)	5(19)	6(22)	6(22)	6(22)	6(22)
BARD		8(37)	8(37)	8(37)	8(37)	8(37)	8(37)
BOX		5(24)	5(24)	5(24)	5(24)	5(24)	5(24)
KOWALIK		10(60)	9(55)	9(55)	9(55)	9(55)	9(55)
OSBORNE1		32(208)	28(185)	71(454)	85(538)	81(508)	28(207)
OSBORNE2		15(196)	15(196)	16(208)	16(208)	13(172)	15(196)
JENNRICH		8(48)	15(72)	11(38)	15(50)	18(59)	19(62)

Table 5. SZ-Broyden family (A-update) without sizing

		BFGS									
		B	(0.0)	B	(0.1)	B	(0.2)	B	(0.3)	B	(0.4)
WATSON6	(S)	18(139)	15(118)	16(125)	16(127)	18(137)
WATSON9	(S)	193(1947)	96(977)	91(927)	129(1309)	104(1054)
WATSON12	(S)	79(1047)	76(1008)	81(1074)	31(428)	63(836)
WATSON20	(S)	20(449)	17(386)	21(471)	24(535)	27(592)
ROSENBROCK	(Z)	27(95)	30(104)	109(337)	35(134)	39(146)
HELIX	(Z)	23(102)	38(167)	29(125)	39(163)	52(218)
POWELL	(Z)	16(85)	14(75)	16(85)	14(75)	16(85)
BEALE	(Z)	16(59)	21(73)	29(97)	29(134)	44(147)
FRDSTEIN1	(Z)	21(86)	7(24)	20(65)	34(129)	13(42)
FRDSTEIN2	(L)	7(31)	7(27)	7(26)	7(25)	7(25)
BARD	(S)	9(49)	16(71)	13(58)	13(58)	14(62)
BOX	(Z)	8(39)	12(63)	8(38)	11(48)	8(39)
KOWALIK	(S)	16(89)	16(89)	17(93)	18(98)	20(108)
OSBORNE1	(S)	31(205)	34(217)	39(248)	69(432)	33(207)
OSBORNE2	(S)	22(281)	31(399)	25(323)	29(373)	26(341)
JENNRICH	(L)	12(41)	13(44)	14(47)	16(53)	17(56)

Table 5. (Continued)

		B	(0.5)	B	(0.6)	B	(0.7)	B	(0.8)	B	(0.9)	B	DFP (1.0)
WATSON6		31(227)	22(165)	25(183)	34(256)	48(346)	102(722)
WATSON9		197(2001)*	114(1154)	135(1361)	136(1380)	154(1556)	199(2001)*
WATSON12		97(1295)	48(641)	81(1067)	68(921)	101(1335)	153(2003)*
WATSON20		46(1013)	51(1096)	55(1177)	72(1548)	94(2005)*	95(2017)*
ROSENBROCK		43(159)	35(112)	57(178)	183(582)	97(380)	500(1507)*
HELIX		61(249)	61(256)	90(367)	78(317)	166(672)	498(2002)*
POWELL		22(115)	21(110)	30(157)	38(198)	11(60)	14(75)
BEALE		45(151)	59(202)	99(330)	144(462)	229(717)	500(1518)*
FRDSTEIN1		18(57)	21(66)	25(78)	39(120)	13(42)	75(228)
FRDSTEIN2		7(24)	7(24)	7(24)	6(22)	6(22)	6(22)
BARD		13(58)	14(61)	15(77)	16(69)	22(92)	31(129)
BOX		9(42)	6(28)	7(32)	8(36)	9(40)	10(44)
KOWALIK		21(113)	25(133)	30(158)	38(198)	44(228)	164(828)
OSBORNE1		34(217)	45(282)	62(386)	57(360)	114(694)	332(2002)*
OSBORNE2		23(299)	26(338)	28(353)	31(394)	40(494)	51(626)
JENNRICH		19(62)	22(71)	25(80)	31(98)	42(131)	89(272)

Table 6. SZ-Broyden family (A-update) with DGW sizing

		BFGS					
		B (0.0)	B (0.1)	B (0.2)	B (0.3)	B (0.4)	
WATSON6	(S)	8(63)	11(86)	11(101)	10(88)	10(81)	
WATSON9	(S)	78(790)	84(852)	83(851)	85(870)	83(844)	
WATSON12	(S)	61(806)	60(795)	63(841)	61(816)	61(810)	
WATSON20	(S)	6(147)	10(233)	9(222)	10(242)	8(193)	
ROSENBROCK	(Z)	21(87)	19(79)	20(82)	23(81)	18(88)	
HELIX	(Z)	14(67)	13(60)	13(63)	15(78)	15(74)	
POWELL	(Z)	14(75)	14(75)	14(75)	14(75)	14(75)	
BEALE	(Z)	10(40)	11(42)	11(42)	12(46)	11(42)	
FRDSTEIN1	(Z)	6(22)	5(18)	6(21)	6(21)	5(18)	
FRDSTEIN2	(L)	6(28)	7(29)	8(32)	10(46)	11(54)	
BARD	(S)	10(44)	10(45)	12(54)	12(54)	12(54)	
BOX	(Z)	5(24)	5(24)	5(24)	6(28)	5(24)	
KOWALIK	(S)	12(70)	9(54)	9(53)	9(53)	10(59)	
OSBORNE1	(S)	25(190)	21(147)	22(165)	21(158)	25(202)	
OSBORNE2	(S)	14(185)	12(160)	13(173)	13(173)	13(172)	
JENNRICH	(L)	9(32)	9(32)	9(32)	9(32)	9(32)	

Table 6. (Continued)

		B (0.5)	B (0.6)	B (0.7)	B (0.8)	B (0.9)	DFP (1.0)
WATSON6		11(87)	10(79)	10(79)	10(79)	10(78)	10(78)
WATSON9		84(853)	83(842)	85(862)	77(788)	83(841)	85(862)
WATSON12		59(783)	61(808)	59(782)	12(171)	59(781)	61(808)
WATSON20		11(255)	9(212)	9(212)	12(275)	9(211)	11(254)
ROSENBROCK		19(86)	20(117)	13(72)	20(111)	20(78)	13(69)
HELIX		15(80)	14(65)	14(67)	15(68)	16(73)	14(68)
POWELL		14(75)	14(75)	14(75)	14(75)	14(75)	14(75)
BEALE		11(42)	11(42)	12(45)	12(45)	12(45)	12(45)
FRDSTEIN1		6(21)	6(21)	6(21)	6(21)	6(21)	6(21)
FRDSTEIN2		8(27)	8(27)	8(32)	6(22)	6(23)	6(23)
BARD		12(54)	13(57)	11(51)	11(48)	12(52)	15(65)
BOX		5(24)	5(24)	5(24)	5(24)	5(24)	5(24)
KOWALIK		10(59)	9(54)	10(59)	9(54)	11(64)	10(58)
OSBORNE1		22(175)	20(153)	22(163)	22(162)	21(152)	21(161)
OSBORNE2		13(171)	13(171)	12(159)	13(171)	13(171)	13(170)
JENNRICH		9(32)	9(32)	9(32)	9(32)	9(32)	9(32)

Table 7. SZ-Broyden family (A-update) with Biggs sizing

		BFGS				
		B (0.0)	B (0.1)	B (0.2)	B (0.3)	B (0.4)
WATSON6	(S)	8(64)	8(63)	8(63)	8(63)	8(63)
WATSON9	(S)	84(856)	79(800)	80(810)	81(820)	80(810)
WATSON12	(S)	66(872)	60(793)	60(793)	60(793)	60(793)
WATSON20	(S)	11(253)	7(168)	7(168)	7(168)	7(168)
ROSENBROCK	(Z)	22(81)	25(87)	19(68)	29(99)	27(112)
HELIX	(Z)	12(63)	15(69)	15(77)	15(68)	14(63)
POWELL	(Z)	13(70)	13(70)	13(70)	13(70)	13(70)
BEALE	(Z)	8(33)	8(33)	8(33)	8(33)	8(33)
FRDSTEIN1	(Z)	5(19)	6(21)	5(18)	5(18)	6(21)
FRDSTEIN2	(L)	7(31)	7(27)	7(26)	6(22)	6(22)
BARD	(S)	8(36)	8(36)	8(36)	8(36)	8(36)
BOX	(Z)	5(24)	5(24)	5(24)	5(24)	5(24)
KOWALIK	(S)	9(57)	9(54)	9(53)	10(58)	10(59)
OSBORNE1	(S)	30(199)	25(165)	31(203)	27(173)	22(141)
OSBORNE2	(S)	22(287)	15(202)	13(171)	13(171)	12(159)
JENNRICH	(L)	10(35)	11(38)	12(41)	12(41)	13(44)

Table 7. (Continued)

		B (0.5)	B (0.6)	B (0.7)	B (0.8)	B (0.9)	DFP (1.0)
WATSON6		8(63)	8(63)	8(63)	9(70)	9(70)	9(70)
WATSON9		80(810)	81(820)	81(820)	82(830)	81(820)	80(810)
WATSON12		60(793)	60(793)	60(793)	60(793)	60(793)	60(793)
WATSON20		7(168)	7(168)	7(168)	7(168)	7(168)	8(189)
ROSENBROCK		22(95)	35(115)	21(70)	36(117)	36(132)	21(72)
HELIX		13(58)	12(54)	12(54)	11(51)	20(101)	12(54)
POWELL		13(70)	14(75)	14(75)	14(75)	14(75)	14(75)
BEALE		8(33)	8(33)	8(33)	8(33)	8(33)	8(33)
FRDSTEIN1		6(21)	6(21)	6(21)	6(21)	6(21)	6(21)
FRDSTEIN2		7(24)	7(24)	7(24)	7(24)	6(22)	6(22)
BARD		9(40)	9(40)	10(44)	12(52)	14(60)	15(64)
BOX		5(24)	5(24)	5(24)	5(24)	5(24)	5(24)
KOWALIK		11(64)	12(69)	13(73)	15(83)	15(83)	17(93)
OSBORNE1		24(153)	26(166)	20(134)	25(160)	45(293)	28(177)
OSBORNE2		15(195)	15(195)	15(195)	15(195)	15(195)	16(207)
JENNRICH		13(44)	14(47)	15(50)	16(53)	17(56)	19(62)

Table 8. Gauss-Newton and structured SR1 methods

		GN		SR1 +Biggs	no	SR1 sizing	
WATSON6	(S)	6(49)	9(70)	10(77)
WATSON9	(S)	80(810)	82(830)	82(830)
WATSON12	(S)	61(806)	60(793)	59(780)
WATSON20	(S)	4(105)	9(210)	9(210)
ROSENBROCK	(Z)	11(62)	20(73)	20(73)
HELIX	(Z)	8(37)	11(50)	15(65)
POWELL	(Z)	9(50)	14(75)	14(75)
BEALE	(Z)	6(26)	8(33)	10(39)
FRDSTEIN1	(Z)	5(18)	6(21)	6(21)
FRDSTEIN2	(L)	106(2009)*	6(21)	6(21)
BARD	(S)	5(24)	12(53)	12(53)
BOX	(Z)	4(20)	6(28)	7(32)
KOWALIK	(S)	19(103)	10(59)	10(59)
OSBORNE1	(S)	6(44)	27(172)	27(175)
OSBORNE2	(S)	9(125)	13(171)	20(254)
JENNRICH	(L)	139(2011)*	9(32)	9(32)

Table 9.1 Total Number of Iterations (Engels and Martinez family)

		BFGS					
		B(0.0)	B(0.1)	B(0.2)	B(0.3)	B(0.4)	
No	Sizing	431*	360	404	388	380	
DGW	Sizing	387*	304	298	300	315	
		(0.898)	(0.844)	(0.738)	(0.773)	(0.829)	
Biggs	Sizing	325	323	333	320	318	
		(0.754)	(0.897)	(0.824)	(0.825)	(0.837)	

Table 9.1 (Continued)

							DFP
		B(0.5)	B(0.6)	B(0.7)	B(0.8)	B(0.9)	B(1.0)
No	Sizing	395	437	407	502*	658	1467*
DGW	Sizing	300	295	288	292	286	287
		(0.759)	(0.675)	(0.708)	(0.582)	(0.435)	(0.196)
Biggs	Sizing	300	302	349	362	363	312
		(0.759)	(0.691)	(0.857)	(0.721)	(0.552)	(0.213)

Table 9.2 Total Number of Iterations (SZ-Broyden family(A-update))

		BFGS				
		B(0.0)	B(0.1)	B(0.2)	B(0.3)	B(0.4)
No	Sizing	518	443	535	514	501
DGW	Sizing	299	300	308	316	310
		(0.577)	(0.677)	(0.576)	(0.615)	(0.619)
Biggs	Sizing	320	301	300	307	299
		(0.618)	(0.679)	(0.561)	(0.597)	(0.597)

Table 9.2 (Continued)

							DFP
		B(0.5)	B(0.6)	B(0.7)	B(0.8)	B(0.9)	B(1.0)
No	Sizing	686*	577	771	979	1190*	2819*
DGW	Sizing	309	305	299	253	306	305
		(0.450)	(0.529)	(0.388)	(0.258)	(0.257)	(0.108)
Biggs	Sizing	301	319	302	328	358	324
		(0.439)	(0.553)	(0.392)	(0.335)	(0.301)	(0.115)

Table 9.3 Total Number of Iterations (Gauss-Newton, SR1)

		GN	SR1
No	Sizing	478*	316
Biggs	Sizing	—	302
			(0.956)

Table 10.1 Total of Function Evaluations (Engels and Martinez family)

		BFGS				
		B(0.0)	B(0.1)	B(0.2)	B(0.3)	B(0.4)
No	Sizing	4269*	3281	4160	3530	3189
DGW	Sizing	3952*	2766	2697	2728	2754
		(0.926)	(0.843)	(0.648)	(0.773)	(0.864)
Biggs	Sizing	3018	2839	2825	2833	2774
		(0.707)	(0.865)	(0.679)	(0.803)	(0.870)

Table 10.1 (Continued)

		B(0.5)	B(0.6)	B(0.7)	B(0.8)	B(0.9)	DFP B(1.0)
No	Sizing	3507	3833	3532	4537*	4423	7990*
DGW	Sizing	2769	2724	2650	2653	2654	2629
		(0.790)	(0.711)	(0.750)	(0.585)	(0.600)	(0.329)
Biggs	Sizing	2777	2671	3017	3128	3056	2756
		(0.792)	(0.697)	(0.854)	(0.689)	(0.691)	(0.345)

Table 10.2 Total of Function Evaluations (SZ-Broyden family(A-update))

		BFGS B(0.0)	B(0.1)	B(0.2)	B(0.3)	B(0.4)
No	Sizing	4744	3842	4139	4121	4095
DGW	Sizing	2670	2731	2831	2861	2822
		(0.563)	(0.711)	(0.684)	(0.694)	(0.689)
Biggs	Sizing	2980	2650	2654	2657	2618
		(0.628)	(0.690)	(0.641)	(0.645)	(0.639)

Table 10.2 (Continued)

		B(0.5)	B(0.6)	B(0.7)	B(0.8)	B(0.9)	DFP B(1.0)
No	Sizing	6082*	4739	6008	6961	8814*	15996*
DGW	Sizing	2824	2779	2735	2146	2721	2813
		(0.464)	(0.586)	(0.455)	(0.308)	(0.309)	(0.176)
Biggs	Sizing	2655	2707	2641	2749	2946	2766
		(0.437)	(0.571)	(0.440)	(0.395)	(0.334)	(0.173)

Table 10.3 Total of Function Evaluations (Gauss-Newton, SR1)

		GN	SR1
No	Sizing	6299*	2796
Biggs	Sizing	—	2691
			(0.962)