数理解析研究所講究録
第 746 巻 1991 年 156-165

Variations of Structured Broyden Families for Nonlinear Least Squares Problems

東京理科大学・工学部　　矢部　博　(Hiroshi Yabe)

1 Introduction

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

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

where $r_j : R^n \to R$, $j = 1, ..., m\,(m \geq n)$ are twice continuously differentiable. Such problems arise widely in data fitting and in the solution of well-determined and over-determined systems of equations. Among many numerical methods, structured quasi-Newton methods seem very promising, which use the structure of the Hessian matrix of $f(x)$ such as

$$\text{(2)} \qquad \nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^{m} r_j(x)\nabla^2 r_j(x).$$

These methods were proposed by Broyden and Dennis and were developed by Bartholomew-Biggs[2], Dennis, Gay and Welsch[6]. Recently several studies have been suggested, e.g., Al-Baali and Fletcher[1], Dennis, Songbai and Vu[7], Dennis, Martinez and Tapia[8], Fletcher and Xu[9], Martinez[10], Xu[12].

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

$$\text{(3)} \qquad x_{k+1} = x_k + \alpha_k d_k,$$

where α_k is a step length and a search direction d_k is given by solving the Newton equation

$$\text{(4)} \qquad (J_k^T J_k + A_k)d_k = -\,J_k^T r_k,$$

$r(x) = (r_1(x), \ldots, r_m(x))^T, r_k = r(x_k), J_k = J(x_k)$ (Jacobian matrix of r) and a matrix A_k is the approximation to the second part of the Hessian matrix. Since the coefficient matrix of (4) does not necessarily possess the hereditary positive definiteness property, Yabe and Takahashi[15], [17] proposed to compute the search direction by solving the linear system of equations

$$\text{(5)} \qquad (L_k + J_k)^T(L_k + J_k)d_k = -J_k^T r_k,$$

where the matrix L_k is an $m \times n$ correction matrix to the Jacobian matrix such that $(L_k + J_k)^T(L_k + J_k)$ is the approximation of the Hessian and overcomes the difficulty of the Gauss-Newton method. Since the coefficient matrix is expressed by the factorized form, the search direction may be expected to be a descent direction for f. Following to [4], we dealt with the secant condition

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

where

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

We called this method the factorized quasi-Newton method and obtained two updating formulae for L_k, which corresponds to the BFGS and the DFP updates.

On the other hand, Sheng Songbai and Zou Zhiong[11] have studied factorized versions of the structured quasi-Newton methods independently of us. They proposed the approximation of $r(x)$ around x_k as follows:

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

and obtained a search direction by solving the linear least squares problem

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

where $\| \; \|$ denotes the 2 norm. In the case of $L_k = \mathbf{0}$, the above implies the Gauss-Newton model. The normal equation of (8) is represented by

(9) $$(J_k+L_k)^T(J_k+L_k)d = -(J_k+L_k)^T r_k.$$

Since the above dose not correspond to the Newton equation(5), Sheng Songbai et al. imposed the condition $L_k^T r_k = \mathbf{0}$ on a matrix L_k, in addition to the secant condition (6) and obtained an BFGS-like update.

Now we present an algorithm of factorized quasi-Newton methods.

(Algorithm of factorized quasi-Newton methods)
Starting with a point $x_1 \in R^n$ and an $m \times n$ matrix L_1 (usually $L_1 = \mathbf{0}$), the algorithm proceeds, for $k = 1, 2, \ldots$, as follows:

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

(10) $$(L_k+J_k)^T(L_k+J_k)\,d \;=\; -J_k^T r_k.$$

(or, following to Sheng Songbai and Zou Zhihong, find the search direction d_k by solving the normal equation of (8)

(11) $$(L_k+J_k)^T(L_k+J_k)\,d \;=\; -(L_k+J_k)^T r_k.\,)$$

Step 2. Choose a steplength α_k by a suitable line serach 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.

The idea of Sheng Songbai et al. seems very interesting to us and some numerical experiments given in [16] suggest the efficiency of their method. So, in this paper, we generalize the update of Sheng Songbai et al. and propose a new update which corresponds to the Broyden family. Further, in Section 5, we introduce the structured Broyden family given by Yabe and Yamaki [18] and obtain a family for A_k. This family for A_k corresponds to the structured secant update from the convex class proposed by Martinez [10]. Throughout the paper, for simplicity, we drop the subscript k and replace the subscript $k+1$ by "+". Further $\| \; \|$ denotes a 2 norm.

2 Notations and Basic Properties

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

(12) $$(L_+ + J_+)^T(L_+ + J_+)s = z$$

and the condition

(13) $$L_+^T r_+ = \mathbf{0},$$

which connects the Newton equation (5) and the normal equation (9). Letting $Y = L_+ + J_+$ and $h = Ys$, the conditions (12) and (13) can be written by the matrix equations of Y

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

where

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

By using Chapter 2 in [3], we have the following theorem.

Theorem 2.1 *The matrix equations* (14) *have a common solution if and only if each equation separately has a solution and*

(16) $$Ch = Ds,$$

where 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 $s^{(1)}$. In which case, for these $C^{(1)}$ and $s^{(1)}$, the general solution of (14) *is*

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

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 equation (14) and that $(I - C^-C)\Phi(I - ss^-)$ is a general solution of the homogeneous equations $CY = \mathbf{0}$ and $Ys = \mathbf{0}$. The above theorem suggests that we just consider the equation (14) for a vector h which satisfies $Ch = Ds$, i.e.,

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

In the below, we use the following notations

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

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

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

Suppose the assumptions:

(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 usefull in the construction of updating formulae for L. The proof of this is shown in [14].

Theorem 2.2 **(1)** *The matrix* CC^T *is nonsingular and* $det(CC^T) = ||r_+||^2||Ph||^2 > 0$.

(2) $s^T z^\sharp > 0$.

(3) $(I - C^\dagger C)h = \mathbf{0}$ *and* $(I - C^\dagger C)r_+ = \mathbf{0}$ *hold, where* $C^\dagger$ *denotes the Moore-Penrose generalized inverse of* C.

(4) *If* rank $N = n$, *the following statements are equivalent:*

- **(a)** $P^\sharp$ *is nonsingular.*
- **(b)** $r_+^T N(N^TN)^{-1}N^T r_+ \neq ||r_+||^2$.
- **(c)** r_+ *can not be spanned by the column vectors of* N.

3 Construction of Partucular Solutions

In the general solution (17), 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

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

Since

$$(18) \qquad C^\dagger = \frac{1}{||r_+||^2||Ph||^2}[h\,, r_+]\begin{bmatrix} ||r_+||^2 & -h^T r_+ \\ -h^T r_+ & h^T h \end{bmatrix}, \quad D - CN = \begin{bmatrix} z^T - h^T N \\ \mathbf{0} \end{bmatrix},$$

we have

$$(19) \qquad Y = N + \left(\frac{Ph}{||Ph||^2}\right)(z - N^T h)^T - (I - C^\dagger C)Nss^-.$$

In the below, we obtain 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

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

where u' is an arbitrary vector satisfying the condition $h^T h = s^T z$. Secondly, setting $u = \tau u'$ and choosing a parameter τ such that

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

we have

$$h = QNs + \frac{1}{\tau}Pu,$$

where τ satisfies

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

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

Since

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

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

finally, it follows from (19) that
[Particular solution 1]

$$Y = N + \frac{1}{\tau}\left(\frac{Pu}{s^T z^\natural}\right)\left(z^\natural - \frac{1}{\tau}N^T Pu\right)^T - (I - C^\dagger C)Nss^-, \tag{20}$$

and
[Particular solution 2]

$$Y = N + \tau\left(\frac{Pu}{||Pu||^2}\right)\left(z^\natural - \frac{1}{\tau}N^T Pu\right)^T - (I - C^\dagger C)Nss^-, \tag{21}$$

where

$$h = QNs + \frac{1}{\tau}Pu, \tag{22}$$

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

$$\tau^2 = \frac{||Pu||^2}{s^T z^\natural}. \tag{23}$$

From (22), we have the following corollary. The proof of this is shown in [14].

Corollary 3.1 **(1)** *The linear independence of r_+ and h are equivalent to the linear independence of r_+ and u.*

(2) $(I - C^\dagger C)u = \mathbf{0}$.

4 SZ-Broyden Family

In this section, we construct a BFGS-like update, which corresponds to the formula of Sheng Songbai and Zou Zhihong, and a DFP-like update by using the particular solutions given in the previous section. Further we propose a structured family, which corresponds to the Broyden family.

In (20), set

$$u = Ns.$$

Then Corollary 3.1 implies $(I - C^\dagger C)Ns = \mathbf{0}$ and we have

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

So the solution (20) yields

$$Y = N + \frac{1}{\tau}\left(\frac{PNs}{s^T z^\natural}\right)\left(z^\natural - \frac{1}{\tau}N^T PNs\right)^T = N + \left(\frac{PNs}{s^T P^\natural s}\right)(\tau z^\natural - P^\natural s)^T. \tag{24}$$

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

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

which is the updating formula of Sheng Songbai and Zou Zhihong.

Here, setting $B_+ = Y^T Y$ gives

$$B_+ = \left(P^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp}\right) + Q^\sharp.$$

The above corresponds to a BFGS update from $P^\sharp$ to B_+ except for $Q^\sharp$, so we call this a SZ-BFGS update. We summarize as follows;
(SZ-BFGS update)

$$L_+^{BFGS} = PL + \left(\frac{PNs}{s^T P^\sharp s}\right)\left(\tau z^\sharp - P^\sharp s\right)^T, \tag{25}$$

$$\tau^2 = \frac{s^T P^\sharp s}{s^T z^\sharp}, \tag{26}$$

$$\begin{aligned} B_+^{BFGS} &= \left(P^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp}\right) + Q^\sharp \\ &= B^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp (z^\sharp)^T}{s^T z^\sharp}. \end{aligned} \tag{27}$$

In the below, we assume that the matrix $P^\sharp$ is nonsingular. In (21), 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

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

The result (3) of Theorem 2.2 means

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

and

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

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

$$\begin{aligned} Y &= N + (\tau - 1)\left(\frac{Pu}{\|Pu\|^2}\right)(z^\sharp)^T + PNb\left(\frac{z^\sharp}{s^T z^\sharp}\right)^T - (\tau - 1)\left(\frac{Pu}{\|Pu\|^2}\right)(z^\sharp)^T \\ &= N + PNb\left(\frac{z^\sharp}{s^T z^\sharp}\right)^T. \end{aligned} \tag{28}$$

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

(SZ-DFP update)

$$L_+^{DFP} = PL + PN\left(\frac{1}{\tau}(P^\sharp)^{-1}z^\sharp - s\right)\left(\frac{z^\sharp}{s^T z^\sharp}\right)^T, \tag{29}$$

$$\tau^2 = \frac{(z^\sharp)^T(P^\sharp)^{-1}z^\sharp}{s^T z^\sharp}, \tag{30}$$

$$\begin{aligned} B_+^{DFP} &= \left(P^\sharp - \frac{P^\sharp s(z^\sharp)^T + z^\sharp s^T P^\sharp}{s^T z^\sharp} + \left(1 + \frac{s^T P^\sharp s}{s^T z^\sharp}\right)\frac{z^\sharp(z^\sharp)^T}{s^T z^\sharp}\right) + Q^\sharp \\ &= B^\sharp - \frac{P^\sharp s(z^\sharp)^T + z^\sharp s^T P^\sharp}{s^T z^\sharp} + \left(1 + \frac{s^T P^\sharp s}{s^T z^\sharp}\right)\frac{z^\sharp(z^\sharp)^T}{s^T z^\sharp}. \end{aligned} \tag{31}$$

Note that the expression (31) corresponds to a DFP update from $P^\sharp$ to B_+ except for $Q^\sharp$.

It is well known that the standard Broyden family for general nonlinear optimization can be expressed by the linear combination of the standard BFGS update and the standard DFP update[5]. Further, Yabe[13], Yamaki and Yabe[19] have 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 structured family which corresponds to the Broyden family, by using the same technique as Yabe[13]. For a parameter τ, let

$$\hat{L}_+^{BFGS} = PL + \left(\frac{PNs}{s^T P^\sharp s}\right)(\tau z^\sharp - P^\sharp s)^T, \tag{32}$$

$$\hat{L}_+^{DFP} = PL + PN(\tau(P^\sharp)^{-1}z^\sharp - s)\left(\frac{z^\sharp}{s^T z^\sharp}\right)^T \tag{33}$$

and

$$L_+ = (1 - \sqrt{\phi})\hat{L}_+^{BFGS} + \sqrt{\phi}\,\hat{L}_+^{DFP}, \tag{34}$$

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

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

gives

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

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

$$B_+ = \left(P^\sharp - \frac{P^\sharp s s^T P^\sharp}{s^T P^\sharp s} + \frac{z^\sharp(z^\sharp)^T}{s^T z^\sharp} + \phi(s^T P^\sharp s)v^\sharp(v^\sharp)^T\right) + Q^\sharp,$$

where

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

Now we obtain a new family, called a SZ-Broyden family, as follows:
(SZ-Broyden family)

$$(35)\quad L_+ = PL + (1-\sqrt{\phi})\left(\frac{PNs}{s^TP^\sharp s}\right)(\tau z^\sharp - P^\sharp s)^T + \sqrt{\phi}\,PN(\tau(P^\sharp)^{-1}z^\sharp - s)\left(\frac{z^\sharp}{s^Tz^\sharp}\right)^T,$$

where

$$(36)\qquad 0 \le \phi \le 1, \quad \left[(1-\phi)\frac{s^Tz^\sharp}{s^TP^\sharp s} + \phi\frac{(z^\sharp)^T(P^\sharp)^{-1}z^\sharp}{s^Tz^\sharp}\right]\tau^2 = 1,$$

$$\begin{aligned}(37)\qquad B_+ &= \left(P^\sharp - \frac{P^\sharp ss^TP^\sharp}{s^TP^\sharp s} + \frac{z^\sharp(z^\sharp)^T}{s^Tz^\sharp} + \phi(s^TP^\sharp s)v^\sharp(v^\sharp)^T\right) + Q^\sharp \\ &= (1-\phi)B_+^{BFGS} + \phi B_+^{DFP}, \\ (38)\qquad v^\sharp &= \frac{P^\sharp s}{s^TP^\sharp s} - \frac{z^\sharp}{s^Tz^\sharp},\end{aligned}$$

and the matrices B_+^{BFGS} and B_+^{DFP} are given in (27) and (31), respectively.

Note that the expression (37) corresponds to a Broyden family from $P^\sharp$ to B_+ except for $Q^\sharp$. Setting $\phi = 0$ and $\phi = 1$ yields the SZ-BFGS update (25) and the SZ-DFP update (29), respectively.

5 Other types of Structured Broyden Family

In the previous section, we propose the SZ-Broyden family based on the idea of Sheng Songbai and Zou Zhihong. The subjects of this section are to introduce the structured Broyden family given by Yabe and Yamaki[18] and to consider the relationship between our family and the structuted secant update from the convex class proposed by Martinez[10].

Consider the case where we do not impose the condition $L_+^Tr_+ = \mathbf{0}$ on the matrix L_+ for the SZ-BFGS update. Since $P = I$, we have

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

Thus the update (25) is reduced to the BFGS-like update given by Yabe and Takahashi[15]:

$$\begin{aligned}(39)\qquad L_+ &= L + \left(\frac{(L+J_+)s}{s^TB^\sharp s}\right)(\tau z - B^\sharp s)^T, \\ (40)\qquad \tau^2 &= \frac{s^TB^\sharp s}{s^Tz}, \quad B^\sharp = (L+J_+)^T(L+J_+), \\ (41)\qquad B_+ &= B^\sharp - \frac{B^\sharp ss^TB^\sharp}{s^TB^\sharp s} + \frac{zz^T}{s^Tz}.\end{aligned}$$

Here we can regard the matrices $(L+J_+)^T(L+J_+)$ and $(L_+ + J_+)^T(L_+ + J_+)$ as the matrices $J_+^TJ_+ + A$ and $J_+^TJ_+ + A_+$, respectively. So, setting

$$(42)\qquad B^\sharp = J_+^TJ_+ + A, \quad \text{and} \quad B_+ = J_+^TJ_+ + A_+$$

in (41), we have an updating formula for A

$$(43)\qquad A_+ = A - \frac{ww^T}{s^Tw} + \frac{zz^T}{s^Tz}, \quad w = (J_+^TJ_+ + A)s,$$

whish is the structured BFGS update given by Al-Baali and Fletcher[1]. Thus the expression (39) corresponds to the factorized form of the structured BFGS update of Al-Baali and Fletcher.

Consider the case where we do not impose the condition $L_+^T r_+ = \mathbf{0}$ on the matrix L_+ for the SZ-DFP update. Then the update (29) is reduced to the DFP-like update given by Yabe and Takahashi[15]:

$$
\begin{aligned}
(44)\qquad L_+ &= L + (L+J_+)\left(\frac{1}{\tau}(B^\natural)^{-1}z - s\right)\left(\frac{z}{s^T z}\right)^T, \\
(45)\qquad \tau^2 &= \frac{z^T (B^\natural)^{-1} z}{s^T z}, \\
(46)\qquad B_+ &= B^\natural - \frac{B^\natural s z^T + z s^T B^\natural}{s^T z} + \left(1 + \frac{s^T B^\natural s}{s^T z}\right)\frac{z z^T}{s^T z}.
\end{aligned}
$$

Substituting (42) for (46), we have an updating formula for A

$$
\begin{aligned}
(47)\qquad A_+ &= A + \frac{(q - As)z^T + z(q-As)^T}{s^T z} - \frac{s^T (q - As)}{(s^T z)^2} z z^T, \\
q &= (J_+ - J)^T r_+,
\end{aligned}
$$

which is the revised update of Dennis, Gay and Welsch[6]. Thus the expression (44) corresponds to the factorized form of the DGW update.

Further, consider the case where we do not impose the condition $L_+^T r_+ = \mathbf{0}$ on the matrix L_+ for the SZ-Broyden family. Then the family (35) is reduced to the structured Broyden family given by Yabe and Yamaki[18]:

$$
\begin{aligned}
(48)\qquad L_+ &= L + (1 - \sqrt{\phi})\left(\frac{(L+J_+)s}{s^T B^\natural s}\right)(\tau z - B^\natural s)^T \\
&\quad + \sqrt{\phi}(L+J_+)(\tau (B^\natural)^{-1} z - s)\left(\frac{z}{s^T z}\right)^T, \\
(49)\qquad B_+ &= B^\natural - \frac{B^\natural s s^T B^\natural}{s^T B^\natural s} + \frac{z z^T}{s^T z} + \phi (s^T B^\natural s) v v^T,
\end{aligned}
$$

where

$$
(50)\qquad \tau^2 = \frac{1}{(1-\phi)\dfrac{s^T z}{s^T B^\natural s} + \phi \dfrac{z^T (B^\natural)^{-1} z}{s^T z}}, \qquad v = \frac{B^\natural s}{s^T B^\natural s} - \frac{z}{s^T z}.
$$

As the same way of the above, substituting (42) for (49), we have an updating formula for A

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

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

which is the structured secant update from the convex class proposed by Martinez[10]. Thus the expression (48) corresponds to the factorized form of the structured secant update from the convex class.

Finally, we apply sizing techniques given by Bartholomew-Biggs[2], Dennis, Gay and Welsch[6] to the above and obtain the following sized family:

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

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

Here β is defined by the Bartholomew-Biggs' parameter

$$\beta = \frac{r_+^T r}{r^T r} \tag{53}$$

or by the DGW parameter

$$\beta = \min\left(\left| \frac{s^T q}{s^T A s} \right|, 1 \right). \tag{54}$$

References

[1] M.Al-Baali and R.Fletcher : Variational methods for non–linear least squares, *Journal of the Operational Research Society*, Vol.36 (1985), pp.405–421.

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

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

[4] J.E.Dennis,Jr. : A brief survey of convergence results for quasi–Newton methods, *SIAM-AMS Proceedings*, Vol.9 (1976), pp.185–199.

[5] J.E.Dennis,Jr. and R.B.Schnabel : *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice–Hall, New Jersey, 1983.

[6] J.E.Dennis,Jr., D.M.Gay and R.E.Welsch : An adaptive nonlinear least squares algorithm, *ACM Transactions on Mathematical Software*, Vol.7, No.3, (1981), pp.348–368.

[7] J.E.Dennis,Jr., S.Songbai and P.A.Vu : *A memoryless augmented Gauss–Newton method for nonlinear least–squares problems*, Technical Report 85–1, Department of Mathematical Sciences, Rice University (1985).

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

[9] R.Fletcher and C.Xu : Hybrid methods for nonlinear least squares, *IMA J. Numerical Analysis*, Vol.7 (1987), pp.371–389.

[10] H.J.Martinez R. : *Local and superlinear convergence of structured secant methods from the convex class*, PhD Thesis, Rice University, 1988.

[11] Sheng Songbai and Zou Zhihong : *A new secant method for nonlinear least squares problems*, Technical Report NANOG–1988–03, Nanjing University (1988).

[12] C.X.Xu : Hybrid method for nonlinear least–square problems without calculating derivatives, *Journal of Optimization Theory and Applications*, Vol.65, No.3 (1990), pp.555–574.

[13] H.Yabe : A family of variable–metric methods with factorized expressions, *TRU Mathematics*, Vol.17, No.1 (1981), pp.141–152.

[14] H.Yabe : Generalization of the Sheng Songbai and Zou Zhihong method for solving nonlinear least squares problems, *Studies in Liberal Arts and Sciences*, Vol.23, Science University of Tokyo, *in Japanese* (1990) to appear.

[15] H.Yabe and T.Takahashi : Structured quasi–Newton methods for nonlinear least squares problems, *TRU Mathematics*, Vol.24, No.2, (1988), pp.195–209.

[16] H.Yabe and T.Takahashi : Numerical comparison among structured quasi–Newton methods for nonlinear least squares problems, *Research Report (Kokyuroku)*, Vol.717, Research Institute for Mathematical Sciences, Kyoto University (1990), pp.144–158.

[17] H.Yabe and T.Takahashi : Factorized quasi–Newton methods for nonlinear least squares problems, *Mathematical Programming*, to appear.

[18] H.Yabe and N.Yamaki : Convergence of structured quasi–Newton methods with structured Broyden family, *Research Report*, The Institute of Statistical Mathematics, *in Japanese* (1991) to appear.

[19] N.Yamaki and H.Yabe : On factorized variable–metric methods, *TRU Mathematics*, Vol.17, No.2 (1981), pp.285–294.