A note on the equivalence of a class of factorized Broyden families for nonlinear least squares problems

Hideho Ogasawara

Department of Applied Mathematics, Faculty of Science, Science University of Tokyo, 1-3 Kagurazaka, Shinjuku-ku, Tokyo 162-0825, Japan

Received 25 September 1998; accepted 3 July 1999

Submitted by R.A. Brualdi

Abstract

This note is concerned with the recently proposed two factorized quasi-Newton updates for solving nonlinear least squares problems. One is a class of updates proposed by Yabe and Yamaki (SIAM Journal on Optimization 5 (1995) 770–791), which was designed as a factorized version of the structured Broyden family with a nonnegative parameter. The other is a class of updates of Xu, Ma and Kong (Journal of Computational Mathematics 14 (1996) 143–158) which was derived as a class of updates satisfying the secant condition in factorized form. It is shown that these classes of updates are essentially the same. © 1999 Elsevier Science Inc. All rights reserved.

AMS classification: 65K05; 49D37; 90C30

Keywords: Factorized Broyden family; Quasi-Newton update; Nonlinear least squares problem

1. Introduction

We consider the nonlinear least squares problem

\[
\text{minimize } f(x) = \frac{1}{2} \sum_{i=1}^{m} r_i(x)^2 = \frac{1}{2} \|r(x)\|^2,
\]

\[\text{subject to } x \in \mathbb{R}^n.\]
where \( r(x) = (r_1(x), \ldots, r_m(x))^T \), each \( r_i : \mathbb{R}^n \to \mathbb{R} \) is twice continuously differentiable for \( i = 1, \ldots, m \) \((m > n)\), \( \|r(x)\| \) denotes the Euclidean norm of \( r(x) \), and the superscript \( T \) denotes the transpose. We assume that there exists a local minimizer of the problem, denoted by \( x_+ \). For general unconstrained minimization problems, Newton’s method is probably the best-known numerical method. However, since computing the overall Hessian matrix at every iteration is often expensive, many quasi-Newton methods have been proposed which proceed approximating the Hessian matrix by using the first-order information only. The standard quasi-Newton methods including the BFGS and DFP methods have proved to be very efficient, and they can be applied to problem (1.1) as well, but they do not take the structure of the problem into account at all. The Hessian matrix of the function \( f(x) \) has a special structure of the form

\[
\nabla^2 f(x) = J(x)^T J(x) + G(x),
\]

with

\[
G(x) = \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x),
\]

where \( J(x) \) denotes the \( m \times n \) Jacobian matrix of \( r(x) \). We note that the first term of the Hessian matrix in (1.2) is computable exactly since the gradient of \( f(x) \) is given by

\[
\nabla f(x) = J(x)^T r(x)
\]

and hence the gradient-based methods, such as the quasi-Newton methods, require the Jacobian matrix \( J(x) \), while the second term \( G(x) \) is generally hard to compute due to the presence of the Hessian matrices of all or some of components \( r_i(x) \).

The Gauss–Newton method is one of the well-known algorithms for solving problem (1.1). This method neglects \( G_k \) and uses \( J_k^T J_k \) as an approximation of the Hessian matrix \( \nabla^2 f(x_k) \), where \( G_k = G(x_k), J_k = J(x_k) \) and \( x_k \) is the current iterate. The Gauss–Newton method performs well for small residual or nearly linear problems since \( G(x_k) \) is close to zero in this case. However, for large residual or strongly nonlinear problems, it is well-known that the method performs poorly.

For the nonlinear least squares problem (1.1), structured quasi-Newton methods have been found to be promising. These methods compute a step \( s_k \) by solving the following equation

\[
(I_k^T J_k + A_k)s = -J_k^T r_k
\]

at the \( k \)th iteration, and generate the next iterate \( x_{k+1} \) by

\[
x_{k+1} = x_k + s_k,
\]

where \( r_k = r(x_k) \) and the matrix \( A_k \) is an approximation to \( G_k \).

Several kinds of structured quasi-Newton updates were proposed and their local convergence theories were developed. Dennis, Gay and Welsch [2] proposed the
structured DFP update, and Dennis and Walker [4] showed local and \( q \)-superlinear convergence of this method. Dennis, Martínez and Tapia [3] derived the structure principle, and proved local and \( q \)-superlinear convergence of the structured BFGS update suggested by Al-Baali and Fletcher [1]. Engels and Martínez [5] extended these updates to the structured Broyden family and established local and \( q \)-superlinear convergence of the convex class of this family. Recently, Yab e and Yamaki [10] further extended the results to Broyden’s bounded \( \phi \)-class which includes the convex class.

Within the line search strategy, it is very important to obtain a descent search direction for the objective function. For general unstructured quasi-Newton updates, successful updates such as the BFGS update have hereditary positive definiteness property which ensures that the computed search direction is a descent direction. However, for structured quasi-Newton updates, the matrix \( A_k \) is not always updated in such a way that the coefficient matrix \( B_k = J_k^T J_k + A_k \) of (1.3) is positive definite, so the resulting direction \( d_k \) is not necessarily a descent direction for \( f(x) \).

As a remedy to overcome this deficiency, Yab e and Takahashi [7] proposed a factorized form of structured quasi-Newton methods. Their methods use \( B_k = (J_k + W_k)^T (J_k + W_k) \) in place of \( J_k^T J_k + A_k \), where the matrix \( W_k \) is an \( m \times n \) correction matrix to \( J_k \) such that \( (J_k + W_k)^T (J_k + W_k) \) is an approximation to \( \nabla^2 f(x_k) \), and is updated so that \( J_k^T W_k + W_k^T J_k + W_k^T W_k \) approximates the second-order term \( G_k \) of the Hessian matrix of \( f \). Yab e and Takahashi [7,8] derived the factorized BFGS and factorized DFP updates for \( W_k \), and proved local and \( q \)-superlinear convergence of their methods. Yab e and Yamaki [9] extended the results to the factorized Broyden family with a nonnegative parameter \( \phi \). Independently, Xu, Ma and Kong [6] proposed a class of factorized quasi-Newton updates involving the factorized BFGS and DFP updates proposed by Yab e and Takahashi as special cases of the class, and proved local and \( q \)-superlinear convergence of their methods.

In Section 2, we will show that the factorized quasi-Newton updates proposed by Xu et al. [6] are identical with a slightly generalized version of the factorized Broyden family proposed by Yab e and Yamaki [9]. Also we will show that, this family seemingly has four different forms but they reduce to two forms.

2. Factorized quasi-Newton updates

In the following, we suppress the subscript \( k \) and replace \( k + 1 \) by \( + \). In factorized quasi-Newton methods, \( W_+ \) is constructed to satisfy the secant condition

\[
(J_+ + W_+)^T (J_+ + W_+) s = y^2
\]

provided \( s^T y^2 \geq 0 \), where

\[
s = x_+ - x, \quad y^2 = J_+^T J_+ s + (J_+ - J)^T r_+.
\]

Set

\[
L^2 = J_+ + W, \quad B^2 = L^2^T L^2
\]
and assume that $L^z$ has full column rank. Then $B^z$ is symmetric positive definite. Also we assume that $s^Ty^z > 0$.

Xu, Ma and Kong [6] have derived a class of factorized quasi-Newton updates

$$ W^{XMK}_+ \equiv W + (a - d) \frac{L^z s y^z s^T}{s^T y^z} + b \frac{L^z B^{z^{-1}} y^z y^z s^T}{s^T y^z} - c \frac{L^z s s^T B^{z^{-1}} s}{s^T B^{z^{-1}} s} \quad (2.1) $$

where $a$, $b$, $c$ and $d$ are parameters satisfying

$$ a s^T B^z y^z + 2 b s^T y^z + b^2 y^z s^T B^{z^{-1}} y^z = s^T y^z, \quad (2.2) $$

$$ \frac{ad}{s^T y^z} = \frac{bc}{s^T B^z s}, \quad (2.3) $$

$$ c + d = 1. \quad (2.4) $$

Independently, Yabe and Yamaki [9] have already proposed a factorized Broyden family of updates as follows:

$$ W^{YY}_+ \equiv (1 - \sqrt{\phi}) \hat{W}^{BFGS}_+ + \sqrt{\phi} \hat{W}^{DFP}_+, \quad (2.5) $$

where

$$ \hat{W}^{BFGS}_+ \equiv W + \frac{L^z s}{s^T B^z s} (\sqrt{\phi} y^z - B^z s)^T, $$

$$ \hat{W}^{DFP}_+ \equiv W + L^z (\sqrt{\lambda} B^{z^{-1}} y^z - s) \frac{y^z s^T}{s^T y^z}. $$

$$ \phi \geq 0, \quad \lambda = \left( (1 - \sqrt{\phi}) s^T y^z \sqrt{\phi} y^z + \frac{\sqrt{\phi} y^z s^T B^{z^{-1}} y^z}{s^T y^z} \right)^{-1}. $$

We notice here that $\lambda = \lambda_\phi$, i.e., $\lambda$ is a function of $\phi$, and $\lambda > 0$ for $\phi \geq 0$ from our assumptions that $s^T y^z > 0$ and $B^z$ is positive definite using the Cauchy–Schwarz inequality.

From its derivation, it is not difficult to see that, $W^{YY}_+$ with $\pm \sqrt{\phi}$ and $\pm \sqrt{\lambda}$ replacing $\sqrt{\phi}$ and $\sqrt{\lambda}$, respectively, also gives a factorized Broyden family.

We present this family formally in the following, and denote it by $W^{YY(\pm \sqrt{\phi}, \pm \sqrt{\lambda})}_+$.

$$ W^{YY(\pm \sqrt{\phi}, \pm \sqrt{\lambda})}_+ \equiv (1 \pm \sqrt{\phi}) \hat{W}^{BFGS}_+ \pm \sqrt{\phi} \hat{W}^{DFP}_+, \quad (2.6) $$

where

$$ \hat{W}^{BFGS}_+ \equiv W + \frac{L^z s}{s^T B^z s} (\pm \sqrt{\lambda} y^z - B^z s)^T, $$

$$ \hat{W}^{DFP}_+ \equiv W + L^z \left( \pm \sqrt{\lambda} B^{z^{-1}} y^z - s \right) \frac{y^z s^T}{s^T y^z}, $$

$$ \phi \geq 0, \quad \lambda = \left( (1 - \sqrt{\phi}) s^T y^z \sqrt{\phi} y^z \pm \frac{\sqrt{\phi} y^z s^T B^{z^{-1}} y^z \pm \sqrt{\lambda}}{s^T y^z} \right)^{-1}. \quad (2.7) $$
We will call this update (2.6) the extended Yabe and Yamaki’s update. According to the combination of the signs, there are generally four forms of the update. Yabe and Yamaki’s original update $W_{+}^{YY}$ of (2.5) is therefore $W_{+}^{YY(+\sqrt{\lambda_0},+\sqrt{\lambda_1})}$ in our notation. We emphasize here that all results of local and $q$-superlinear convergence for $W_{+}^{YY}$ obtained by Yabe and Yamaki [9] carry over to the extended update (2.6).

In particular, setting $\phi = 0$ and $\phi = 1$ in (2.6) yields factorized BFGS and DFP updates, respectively. Specifically, we have

$$W_{+}^{\text{BFGS}+} \equiv W_{+}^{YY(0,+,\sqrt{\lambda_0})}$$

$$= W + \frac{L_s^2 y \cdot \left( \pm \sqrt{\frac{s^T B \cdot y \cdot y^T}{s^T y^2} - B^2} \right)}{s^T y^2}$$

$$W_{+}^{\text{DFP}+1} \equiv W_{+}^{YY(+1,+,\sqrt{\lambda_1})}$$

$$= W + L_s^2 \left( \pm \sqrt{\frac{s^T y \cdot y^T}{y^2 B^{-1} y^T y^T y^T - y^T}} \right) y^T$$

$$W_{+}^{\text{DFP}+2} \equiv W_{+}^{YY(-1,+,\sqrt{\lambda_1})}$$

$$= W + 2L_s^2 \left( \pm \sqrt{\frac{s^T y \cdot y^T}{y^2 B^{-1} y^T y^T y^T - y^T}} \right) y^T$$

where $\lambda_0$ and $\lambda_1$ are, respectively, the values of $\lambda$ with $\phi = 0$ and $\phi = 1$, i.e., the quantities in the square roots. It is interesting to note that, the BFGS update has two factorized correction forms (2.8) of rank-one, whereas the DFP update has two factorized correction forms (2.9) of rank-two besides two factorized correction forms (2.9) of rank-one.

Now we will show that the two updates $W_{+}^{XMK}$ and $W_{+}^{YY(\pm\sqrt{\lambda},\pm\sqrt{\lambda})}$ are the same.

**Theorem 1.** Xu et al.’s update (2.1) is the same as the extended Yabe and Yamaki’s update (2.6).

**Proof.** Update (2.6) can be written as

$$W_{+}^{YY(\pm\sqrt{\lambda},\pm\sqrt{\lambda})} = W + \left( \pm \sqrt{\lambda} \cdot \pm \sqrt{\phi} \right) \frac{s^T y \cdot \left( \pm \sqrt{\frac{s^T B \cdot y \cdot y^T}{s^T y^2} - B^2} \right)}{s^T y^2}$$

$$+ \left( \pm \sqrt{\phi} \right) \cdot \left( \pm \sqrt{\lambda} \right) \frac{L_s^2 y \cdot y^T}{s^T y^2}$$

$$- (1 \mp \sqrt{\phi}) \cdot \frac{L_s^2 s \cdot y \cdot y^T}{s^T y^2}.$$
Comparing (2.1) with (2.11) and taking condition (2.4) into account, we can take $a, b, c$ and $d$ as
\[ a = \pm \sqrt{\lambda} \left( 1 \mp \phi \right) \frac{s^T y^z}{s^T B^z s}, \quad b = \left( \pm \sqrt{\phi} \right) \left( \pm \sqrt{\lambda} \right), \]
\[ c = 1 \mp \sqrt{\phi}, \quad d = \pm \sqrt{\phi}. \]  
(2.12)
By direct substitution, it is easily verified that these values of $a, b, c$ and $d$ satisfy the remaining parameter conditions (2.2)–(2.3). At this point, we see that update (2.6) is a member of update (2.1). Actually, update (2.6) is an explicit form of update (2.1) in terms of a nonnegative parameter $\phi$. We next show this by solving (2.2)–(2.4).
Suppose first that $d = 0$. Then, from (2.4) and (2.3), we have $c = 1$, $b = 0$. Thus, from (2.2), we know that
\[ a = \pm \sqrt{\lambda} \frac{s^T y^z}{s^T B^z s}. \]
Similarly, if $c = 0$, then we have $d = 1$, $a = 0$, and
\[ b = \pm \sqrt{\lambda} \frac{s^T y^z}{y^{2z} B^{-1} y^z}. \]
Suppose that $c \neq 0$ and $d \neq 0$. Bearing (2.12) in mind, we set the quantity of (2.3) to
\[ \tau^2 = \frac{c d}{s^T B^z s}. \]
Then we can write as
\[ a = \tau c \frac{s^T y^z}{s^T B^z s}, \quad b = \tau d. \]
(2.13)
Substituting (2.13) into (2.2), we have
\[ \tau^2 \left[ (\phi^2 + 2cd) \frac{(s^T y^z)^2}{s^T B^z s} + d^2 (y^{2z} B^{-1} y^z) \right] = s^T y^z, \]
so that, by using (2.4),
\[ \tau^2 \left[ (1 - d^2) \frac{s^T y^z}{s^T B^z s} + d^2 \frac{s^T B^{-1} y^z}{s^T y^z} \right] = 1. \]
(2.14)
Now let $\phi = d^2$. Then obviously $\phi \geq 0$, and from (2.14), we get $\tau^2 = \lambda$, where $\lambda$ is defined in (2.7). Therefore, with (2.13) and (2.4), we finally obtain (2.12), whose expressions include the case $c = 0$ or $d = 0$. $\square$

Clearly, if $L_+ = J_+ + W_+$ is a factor of $B_+$ such that $B_+ = L_+^T L_+$, then, for any orthogonal matrix $Q$, $QL_+$ is also another factor of $B_+$, i.e., $Q$ is an obvious
free factor. Therefore, we may consider that \( QL_+ = Q(J_+ + W_+) \) is the ‘same’ as 
\( L_+ = J_+ + W_+ \), in a sense. Let \( W_1^+ \) and \( W_2^+ \) be two given updates of \( W \). We say that the update \( W_1^+ \) is essentially the same as the update \( W_2^+ \) if there exists an orthogonal matrix \( Q \) such that \( J_+ + W_1^+ = Q(J_+ + W_2^+) \).

Let 
\[
Q = I - \frac{2L^2s^T L^ST}{s^TBs}.
\]

Then \( Q \) is an orthogonal matrix, and we can verify that 
\[
Q(J_+ + W_+^{BFGS+}) = J_+ + W_+^{BFGS-}.
\]

Therefore, the update \( W_+^{BFGS-} \) is essentially the same as the update \( W_+^{BFGS+} \). Similarly, letting 
\[
Q_1 = I - \frac{2L^2B^2-1 y^Ty^TB^{-1}L^TS}{y^TB^{-1}y^T},
\]
\[
Q_2 = I - \frac{2(s^Ty^2)}{y^TB^{-1}y^T} \left( \frac{2L^2s}{s^TB^2s} - \frac{L^2B^2-1 y^T}{s^TB^2s} \right) \left( \frac{2L^2s}{s^TB^2s} - \frac{L^2B^2-1 y^T}{s^TB^2s} \right)^T,
\]
we have 
\[
Q_1(J_+ + W_+^{DFP1+}) = J_+ + W_+^{DFP1-}, \quad Q_2(J_+ + W_+^{DFP2+}) = J_+ + W_+^{DFP2-}.
\]

Hence, the updates \( W_+^{DFP1-} \) and \( W_+^{DFP2-} \) are, respectively, essentially the same as the updates \( W_+^{DFP1+} \) and \( W_+^{DFP2+} \). More generally, we obtain the following.

**Theorem 2.** The update \( W_+^{YY(\pm \sqrt{\phi}, -\sqrt{\phi})} \) is essentially the same as the update \( W_+^{YY(\pm \sqrt{\phi}, +\sqrt{\phi})} \).

**Proof.** We first recall that Yabe and Yamaki [9] derive a general factorized update (see expression (14) in [9])
\[
W_+[u] \equiv W - \frac{L^2su^TL^S}{u^TL^2s} + \frac{1}{\sqrt{s^Ty}} \frac{u}{\|u\|} y^T,
\]
where \( u \) is an arbitrary vector such that \( u^TL^2s \neq 0 \). Since \( u \) is arbitrary, we can replace \( u \) by \(-u\) in (2.15) to get a factorized update as follows:
\[
W_+[-u] = W - \frac{L^2su^TL^S}{u^TL^2s} - \frac{1}{\sqrt{s^Ty}} \frac{u}{\|u\|} y^T.
\]

Let \( \phi \geq 0 \) and consider
\[
u^\pm \sqrt{\phi} = (1 \mp \sqrt{\phi}) \frac{L^2s}{s^TB^2s} \pm \sqrt{\phi} \frac{L^2B^2-1 y^T}{s^TB^2s}.
\]

Then we have
\[(u^{\pm\sqrt{\sigma}})^T L^z S = 1, \quad \|u^{\pm\sqrt{\sigma}}\| = \frac{1}{\sqrt{s^T y^s \sqrt{s}}}. \quad (2.18)\]

We note that Yabe and Yamaki’s original update (2.5) is obtained by substituting the vector \(u = u^{+\sqrt{\sigma}}\) of (2.17) into (2.15), i.e., \(W^{YY}_+ = W_+u^{+\sqrt{\sigma}}\) (see (28) in [9]).

Similarly, we can easily check that the updates \(W^{YY}_+(\pm\sqrt{\sigma}, +\sqrt{\sigma})\) and \(W^{YY}_+(\pm\sqrt{\sigma}, -\sqrt{\sigma})\) are, respectively, obtained by substituting \(u = u^{+\sqrt{\sigma}}\) into (2.15) and (2.16), i.e.,

\[
W^{YY}_+(\pm\sqrt{\sigma}, +\sqrt{\sigma}) = W_+u^{+\sqrt{\sigma}}, \quad W^{YY}_+(\pm\sqrt{\sigma}, -\sqrt{\sigma}) = W_+[-u^{\pm\sqrt{\sigma}}]. \quad (2.19)
\]

Now, for any nonzero \(u\), define

\[Q[u] \equiv I - \frac{2uu^T}{\|u\|^2}\]

Obviously, \(Q[u]\) is an orthogonal matrix and \(Q[u]u = -u\) holds. It follows that

\[
Q[u] \left( L^z - \frac{L^z su^T L^z}{u^T L^z S} + \frac{1}{\sqrt{s^T y^s \|u\|^2}} \right) u \|u\|^2 \right) y^T
\]

\[
= L^z - \frac{L^z su^T L^z}{u^T L^z S} + \frac{Q[u]u}{\sqrt{s^T y^s \|u\|^2}} y^T
\]

\[
= L^z - \frac{L^z su^T L^z}{u^T L^z S} - \frac{2u}{\|u\|^2} \left( u^T L^z - u^T L^z S \frac{u^T L^z}{u^T L^z S} \right) - \frac{1}{\sqrt{s^T y^s \|u\|^2}} u y^T
\]

Thus, for any \(u\) such that \(u^T L^z S \neq 0\), we have

\[Q[u](J_+ + W_+[u^{\pm\sqrt{\sigma}}]) = J_+ + W_+[-u^{\pm\sqrt{\sigma}}]. \quad (2.20)\]

In particular, by (2.18), we can take \(u = u^{\pm\sqrt{\sigma}}\) in (2.20) to get

\[Q[u^{\pm\sqrt{\sigma}}](J_+ + W_+[u^{\pm\sqrt{\sigma}}]) = J_+ + W_+[-u^{\pm\sqrt{\sigma}}].\]

This is precisely the relation that

\[Q[u^{\pm\sqrt{\sigma}}](J_+ + W_+^{YY}(\pm\sqrt{\sigma}, +\sqrt{\sigma})) = J_+ + W_+^{YY}(\pm\sqrt{\sigma}, -\sqrt{\sigma})\]

by virtue of (2.19). The proof is complete. \(\square\)

3. Concluding remarks

In this paper, we have shown that the class of factorized quasi-Newton updates derived by Xu, Ma and Kong [6] is just the same as a slightly generalized version of the factorized Broyden family proposed by Yabe and Yamaki [9]. This version
seemingly has four forms. However, we have shown that the update with $-\sqrt{\tau}$ is essentially the same as the update with $\sqrt{\tau}$, and hence the forms of the family reduce to two types. At present, we do not know whether the update with $-\sqrt{\tau}$ is essentially the same as the update with $\sqrt{\tau}$, namely, whether we can write
\[
J_+ + W_+^{YY(-\sqrt{\tau} + \sqrt{\tau})} = Q \left( J_+ + W_+^{YY(\sqrt{\tau} + \sqrt{\tau})} \right)
\]
for some $Q$ orthogonal. If it were so, the factorized Broyden family (2.6) would reduce to the only one update $W_+^{YY(\sqrt{\tau} + \sqrt{\tau})}$ in essence, which is the update $W_+^{YY}$ of (2.5) first proposed by Yabe and Yamaki [9].

Acknowledgements

I would like to thank the referee for the valuable comments and suggestions which enabled me to improve the exposition of this paper.

References