Journal of Computational Mathematics, Vol.24, No.5, 2006, 665–674.

A GENERALIZED QUASI-NEWTON EQUATION AND COMPUTATIONAL EXPERIENCE *1)

Lei-hong Zhang Ping-qi Pan
(*Department of Mathematics, Southeast University, Nanjing 210096, China*)

Abstract

The quasi-Newton equation has played a central role in the quasi-Newton methods for solving systems of nonlinear equations and/or unconstrained optimization problems. Instead, Pan suggested a new equation, and showed that it is of the second order while the traditional of the first order, in certain approximation sense [12]. In this paper, we make a generalization of the two equations to include them as special cases. The generalized equation is analyzed, and new updates are derived from it. A DFP-like new update outperformed the traditional DFP update in computational experiments on a set of standard test problems.

Mathematics subject classification: 65H10, 65K05, 90C26
Key words: System of nonlinear equations, Unconstrained optimization, Quasi-Newton equation; Second-order Quasi-Newton equation, Update formula.

1. Introduction

For solving the system of nonlinear equations $F(x) = 0$, where $F : D \subset R^n \to R^n$, or the unconstrained optimization problem $\min f(x)$ (with $F(x) = \nabla f(x)$), the iteration form

$$x_{k+1} = x_k - \alpha_k B_k^{-1} F(x_k), \quad k = 0, 1, \cdots \tag{1.1}$$

is widely used, where $\{B_k\}$ satisfies the quasi-Newton equation (also known as the quasi-Newton condition):

$$B_{k+1} s_k = y_k, \tag{1.2}$$

where

$$s_k = x_{k+1} - x_k, \ y_k = F(x_{k+1}) - F(x_k).$$

A large number of formulae satisfying (1.2) have been proposed, among which the most famous two are BFGS (independently by Broyden(1969, 1970), Fletcher (1970), Goldfarb (1970), Shanno (1970)) and DFP (independently by Davidon (1959), Fletcher and Powell (1963)). Besides, there are also some modifications that do not satisfy (1.2) but generate sequences $\{x_k\}$ linearly or superlinearly converging to x^*, a zero point of $F(x)$. Powell, for instance, proposed two formulae ([14], [15])

$$B_{k+1} = B_k + \theta_k \frac{(y_k - B_k s_k) s_k^T}{s_k^T s_k}, \tag{1.3}$$

* Received March 30, 2005; final revised December 14, 2005.
1) Project 10371017 supported by National Natural Science Foundation of China.

$$B_{k+1} = B_k + \theta_k \frac{s_k(y_k - B_k s_k)^T + (y_k - B_k s_k)s_k^T}{s_k^T s_k} - \theta_k^2 \frac{(y_k - B_k s_k)^T s_k s_k s_k^T}{(s_k^T s_k)^2}, \tag{1.4}$$

where $\theta_k \in R$. Moré and Trangenstein ([10]) then proved that if $F(x) = Gx - b$, where G is a nonsingular and symmetric matrix and $\alpha_k = 1$ ($\forall k$), the generated sequence $\{x_k\}$ using (1.3) or (1.4) can be globally and superlinearly convergent. All those not only imply that the classical quasi-Newton equation (1.2) may be unnecessary for the *iteration scheme* (1.1), but also encourage us to establish some other well-performed quasi-Newton-type formulae. In fact, some recently reported works (e.g. [1], [8], [9], [16], [17], [19], [20]) also contributed to the so called "Modified Quasi-Newton Methods". This is one of the motivations of this paper.

The direct elicitation, however, is the work of Pan ([12]). Introducing a function approximating $F(x)$, he derived equation

$$B_{k+1} s_k = 2y_k - B_k s_k, \tag{1.5}$$

and showed that the preceding is of a second order while (1.2) of a first order, in certain approximation sense. Note that both (1.3) and (1.4) do not satisfy (1.5).

This paper is intended to make a generalization of the two equations to include them as special cases. It is organized as follows. In section 2, we first generalize (1.5) by introducing an extra matrix parameter T_k to Pan's approximation function. Then, in section 3, we analyze the generalized equation. In section 4, we derive associated updates. Finally, in section 5, we report our computational experience with a DFP-like new update on a set of standard test problems, demonstrating its superiority to the traditional DFP update.

2. The Generalized Quasi-Newton Equation

Let us drop subscript and consider two points $\widehat{x}$, $\widetilde{x}$ ($\widetilde{x} \neq \widehat{x}$) in R^n. Assume we know the values of F at them and the Jacobian $F'(\widehat{x})$ of F at $\widehat{x}$, and denote them respectively by

$$\widehat{F} = F(\widehat{x}), \quad \widetilde{F} = F(\widetilde{x}), \quad B = F'(\widehat{x}).$$

Introduce the notation

$$s = \widetilde{x} - \widehat{x}, \quad y = \widetilde{F} - \widehat{F}. \tag{2.1}$$

We will derive an approximation $\widetilde{B}$ to the Jacobian $F'(\widetilde{x})$ of F at $\widetilde{x}$. As was in [12], we define a *one-reduction matrix* of s as follows.

Definition 2.1. *Let $s \in R^n$. $A \in R^{n\times n}$ is called a one-reduction matrix of s if*

$$s^T A s = 1. \tag{2.2}$$

Refer to [12] for some examples and properties of the one-reduction matrix.

Based on Taylor's theorem, the original approximate quadratic function in [12] is

$$Q(x) = \widehat{F} + B(x - \widehat{x}) + \frac{1}{2}(x - \widehat{x})^T A (x - \widehat{x}) y - \frac{1}{2}(x - \widehat{x})^T \overline{A} (x - \widehat{x}) B s.$$

Taking into account of the last two terms, we see that the requirement of (2.2), however, potentially limits the degree of approximation. For a better approximation, consider the following quadratic mapping

$$Q(x) = \widehat{F} + B(x - \widehat{x}) + \frac{1}{2}(x - \widehat{x})^T A (x - \widehat{x}) T y - \frac{1}{2}(x - \widehat{x})^T \overline{A} (x - \widehat{x}) T B s, \tag{2.3}$$

where A and $\overline{A}$ are any two one-reduction matrices of s and $T = T(s, y) \in R^{n\times n}$. After some manipulation, it follows that

$$\begin{aligned} Q(x) \quad = \quad & [\widetilde{F} + (\frac{T}{2} - I)(y - Bs)] + [B + \frac{1}{2}Tys^T(A + A^T) - \frac{1}{2}TBss^T(\overline{A} + \overline{A}^T)](x - \widetilde{x}) \\ & + \frac{1}{2}(x - \widetilde{x})^T A(x - \widetilde{x})Ty - \frac{1}{2}(x - \widetilde{x})^T \overline{A}(x - \widetilde{x})TBs. \end{aligned} \tag{2.4}$$

From (2.4), we get

$$Q(\widehat{x}) = \widehat{F}, \quad Q^{'}(\widehat{x}) = B, \quad Q(\widetilde{x}) = \widetilde{F} + (\frac{T}{2} - I)(y - Bs).$$

Therefore, $Q(x)$ may be regarded as an approximation to $F(x)$, and hence, $Q^{'}(\widetilde{x})$ as an approximation to $F^{'}(\widetilde{x})$, i.e.,

$$\widetilde{B} = Q^{'}(\widetilde{x}) = B + \frac{1}{2}Tys^T(A + A^T) - \frac{1}{2}TBss^T(\overline{A} + \overline{A}^T). \tag{2.5}$$

Since both A and $\overline{A}$ are one-reduction matrices of s, postmultiplying by s the two sides of (2.5) leads to

$$\widetilde{B}s = Bs + T(y - Bs), \tag{2.6}$$

which is referred to as *generalized quasi-Newton equation*. Two specific cases of it are given as follows:

1. Setting $T = I$ in (2.6) leads to the classical quasi-Newton equation $\widetilde{B}s = y$. For the associated approximation function, it holds that

$$Q(\widehat{x}) = \widehat{F}, \quad Q^{'}(\widehat{x}) = B, \tag{2.7}$$

but $Q(\widetilde{x}) = \widetilde{F} + \frac{1}{2}(Bs - y)$ is not equal to $\widetilde{F}$, in general.

2. Setting $T = 2I$ in (2.6) yields Pan's second order quasi-Newton equation $\widetilde{B}s = 2y - Bs$ [12] (see (1.5)). In such case, the associated function not only satisfies (2.7) but also an extra condition $Q(\widetilde{x}) = \widetilde{F}$.

So, the function $Q(x)$ associated with Pan's quasi-Newton equation should be a better approximation to $F(x)$ near $\widehat{x}$ or $\widetilde{x}$ than that associated with the classical one.

3. Analysis of the Generalized Quasi-Newton Equation

We analyze the generalized quasi-Newton equation (2.6) by examining to which extent $Q(x)$ approximates $F(x)$ near $\tilde{x}$, under certain hypotheses on $T(s, y)$. We will use $\|\cdot\|$ to denote the l_2 norm of a vector.

Theorem 3.1. *Assume that $F : D \subset R^n \to R^n$ is Fréchet-differentiable at $\widetilde{x} \in int(D)$ and $Q(x)$ is defined by (2.3). If $\lim_{s\to 0} \|T(s, y)\| < \infty$, then for any non-zero and sufficiently small $t \in R$, it holds that*

$$Q(\widetilde{x} + ts) - F(\widetilde{x} + ts) = E(t, s), \tag{3.1}$$

where

$$\lim_{s\to 0}\frac{E(t,s)}{\|s\|}=0,\quad \lim_{t\to 0}E(t,s)=\overline{E}(s),\quad and\quad \lim_{s\to 0}\frac{\overline{E}(s)}{\|s\|}=0; \tag{3.2}$$

i.e.,

$$E(t,s)=o(\|s\|),\ \overline{E}(s)=o(\|s\|).$$

Proof. Using the second-order version of Taylor's theorem yields

$$F(\widetilde{x}+ts)=\widetilde{F}+tF^{'}(\widetilde{x})s+\frac{1}{2}F^{''}(\widetilde{x})(ts)(ts)-R_1(ts), \tag{3.3}$$

where $F^{'}(x)$ and $F^{''}(x)$ denote the first and second Fréchet-derivatives at $x\in int(D)$ respectively, and

$$\lim_{ts\to 0}\frac{R_1(ts)}{(t\|s\|)^2}=0.$$

Taking $t=-1$ in (3.3) leads to

$$\widehat{F}=\widetilde{F}-F^{'}(\widetilde{x})s+\frac{1}{2}F^{''}(\widetilde{x})ss-R_2(s), \tag{3.4}$$

where

$$\lim_{s\to 0}\frac{R_2(s)}{\|s\|^2}=0.$$

The second-order Taylor's expansion also leads to

$$\widetilde{F}=\widehat{F}+Bs+\frac{1}{2}F^{''}(\widehat{x})ss+R_3(s), \tag{3.5}$$

where

$$\lim_{s\to 0}\frac{R_3(s)}{\|s\|^2}=0.$$

By virtue of (2.4), (3.3), (3.4), (3.5), we have

$$\begin{aligned}
E(t,s) &= Q(\widetilde{x}+ts)-F(\widetilde{x}+ts)\\
&= Bs-y+t(B-F^{'}(\widetilde{x}))s+(\frac{T}{2}+tT+T\frac{t^2}{2})(y-Bs)-\frac{t^2}{2}F^{''}(\widetilde{x})ss+R_1(ts)\\
&= -\frac{1}{2}F^{''}(\widehat{x})ss-R_3(s)+t(-\frac{1}{2}F^{''}(\widetilde{x})ss+R_2(s)-\frac{1}{2}F^{''}(\widehat{x})ss-R_3(s))\\
&\quad +(\frac{T}{2}+tT+\frac{t^2}{2}T)(\frac{1}{2}F^{''}(\widehat{x})ss+R_3(s))-\frac{t^2}{2}F^{''}(\widetilde{x})ss+R_1(ts)\\
&= \frac{t+t^2}{2}(\frac{T}{2}F^{''}(\widehat{x})ss-F^{''}(\widetilde{x})ss)+\frac{1}{2}(\frac{T}{2}+\frac{t}{2}T-I-tI)F^{''}(\widehat{x})ss\\
&\quad +R_1(ts)+tR_2(s)+(\frac{T}{2}+tT+\frac{t^2}{2}T-I-tI)R_3(s).
\end{aligned} \tag{3.6}$$

Therefore,

$$
\begin{aligned}
\tfrac{\|E(t,s)\|}{\|s\|} \quad &\leq \left|\frac{t+t^2}{2}\right| \cdot \frac{\|\frac{T}{2}F''(\widehat{x})ss - F''(\widetilde{x})ss\|}{\|s\|} \\
&+\frac{1}{2}\frac{\|(\frac{T}{2} - I + \frac{t}{2}T - tI)F''(\widehat{x})ss\|}{\|s\|} + \frac{\|R_1(ts)\|}{\|s\|} + \frac{\|tR_2(s)\|}{\|s\|} \\
&+\frac{\|(\frac{T}{2} + tT + \frac{t^2}{2}T - I - tI)R_3(s)\|}{\|s\|}.
\end{aligned}
$$

Under the assumption $\lim_{s\to 0} \|T(s,y)\| < \infty$, we have

$$\lim_{s\to 0} \frac{E(t,s)}{\|s\|} = 0$$

and

$$\lim_{t\to 0} E(t,s) = \frac{1}{2}(\frac{T}{2} - I)F''(\widehat{x})ss + (\frac{T}{2} - I)R_3(s) = \overline{E}(s).$$

Apparently, $\lim_{s\to 0} \frac{\overline{E}(s)}{\|s\|} = 0$, and the proof is completed.

Note that the associated $T(s,y)$ equals I for the classical quasi-Newton equation and the $T(s,y)$ equals $2I$ for Pan's equation (see the last paragraphs of Section 2). By Theorem 3.1, their associated $Q(x)$ is at least a first order approximation to $F(x)$ at $\tilde{x}$. Furthermore, we come to a second order approximation by strengthening the condition on $T(s,y)$:

Theorem 3.2. *All assumptions are the same as those in Theorem 3.1 except for* $\lim_{s\to 0} \|T(s,y)\| < \infty$ *which is replaced by* $\lim_{s\to 0} \frac{T(s,y)}{2} = I$. *Then for any non-zero and sufficiently small* $t \in R$, *it holds that*

$$Q(\widetilde{x} + ts) - F(\widetilde{x} + ts) = R(t,s),$$

where

$$\lim_{s\to 0} \frac{R(t,s)}{\|s\|^2} = 0, \quad \lim_{t\to 0} R(t,s) = \overline{R}(s), \quad \textit{and} \quad \lim_{s\to 0} \frac{\overline{R}(s)}{\|s\|^2} = 0;$$

i.e.,

$$R(t,s) = o(\|s\|^2), \ \overline{R}(s) = o(\|s\|^2).$$

Proof. From (3.6), we have

$$
\begin{aligned}
R(t,s) \quad = \quad & \frac{t+t^2}{2}(\frac{T}{2}F''(\widehat{x})ss - F''(\widetilde{x})ss) + \frac{1}{2}(\frac{T}{2} + \frac{t}{2}T - I - tI)F''(\widehat{x})ss \\
& +R_1(ts) + tR_2(s) + (\frac{T}{2} + tT + \frac{t^2}{2}T - I - tI)R_3(s).
\end{aligned}
$$

The assumption $\lim_{s\to 0} \frac{T(s,y)}{2} = I$ ensures

$$\lim_{s\to 0} \frac{R(t,s)}{\|s\|^2} = 0,$$

and

$$\lim_{t\to 0} R(t,s) = \frac{1}{2}(\frac{T}{2} - I)F''(\widehat{x})ss + (\frac{T}{2} - I)R_3(s) \triangleq \overline{R}(s).$$

Therefore, $\lim_{s\to 0} \frac{\overline{R}(s)}{\|s\|^2} = 0$, and we complete the proof.

It is seen that for $Q(x)$ associated with Pan's equation $\widetilde{B}s = 2y - Bs$, Theorem 3.2 holds, since condition $\lim_{s\to 0} \frac{T(s,y)}{2} = I$ is fulfilled while this is not so for the classical quasi-Newton equation $\widetilde{B}s = y$. This coincides with [12], where the former equation is described as *second order* while the latter as *first order*.

4. Associated Updates

A vast number of updating formulae could be derived from (2.5), satisfying (2.6). With specific T's, we only present four families of updates as follows.

Family 1: $T = I$. (Classical)

- Setting in (2.5)

$$\begin{aligned} A &= \varphi[\frac{y^Ts + s^TBs}{(y^Ts)^3}yy^T - \frac{B}{y^Ts}] + (1-\varphi)\frac{yy^T}{(y^Ts)^2} \triangleq \varphi A^{DFP} + (1-\varphi)A^{BFGS}, \\ \bar{A} &= \varphi\frac{yy^T}{(y^Ts)^2} + (1-\varphi)\frac{B}{s^TBs} \triangleq \varphi\bar{A}^{DFP} + (1-\varphi)\bar{A}^{BFGS}, \end{aligned}$$

  gives the updates of Broyden Family in which $\varphi = 1$ corresponds to the DFP update, and $\varphi = 0$ the BFGS.

- Setting $A = \overline{A} = \frac{ss^T}{(s^Ts)^2}$ yields

$$\widetilde{B} = B + \frac{(y - Bs)s^T}{s^Ts}. \tag{4.1}$$

  It is known that the iteration scheme (1.1) using this update locally and superlinearly converges to x^*(see [5]).

- Setting $A = \overline{A} = \frac{(y-Bs)(y-Bs)^T}{((y-Bs)^Ts)^2}$, where $(y - Bs)^Ts \neq 0$, leads to

$$\widetilde{B} = B + \frac{(y - Bs)(y - Bs)^T}{(y - Bs)^Ts}, \tag{4.2}$$

  which is the symmetric rank-one update, firstly published by Davidon (see [4]).

Family 2: $T = \theta I$. $(\theta \in R)$

- Setting in (2.5)

$$\begin{aligned} A &= \varphi[\frac{y^Ts + s^TBs}{(y^Ts)^3}yy^T - \frac{B}{y^Ts}] + (1-\varphi)\frac{yy^T}{(y^Ts)^2} \triangleq \varphi A^{DFP} + (1-\varphi)A^{BFGS}, \\ \bar{A} &= \varphi\frac{yy^T}{(y^Ts)^2} + (1-\varphi)\frac{B}{s^TBs} \triangleq \varphi\bar{A}^{DFP} + (1-\varphi)\bar{A}^{BFGS}, \end{aligned}$$

  leads to a class of updates corresponding to the Broyden family. In particular, such updates with $\theta = 2$ may be termed as second-order Broyden Family, where $\varphi = 1$ and $\theta = 2$ together correspond to the second-order DFP update while $\varphi = 0$ and $\theta = 2$, to the second-order BFGS (see [12]).

- Setting $A = \overline{A} = \frac{ss^T}{(s^Ts)^2}$ yields

$$\widetilde{B} = B + \theta\frac{(y - Bs)s^T}{s^Ts}, \tag{4.3}$$

which is known as the modification of Broyden family, proposed firstly by Powell [14].

- Setting $A = \overline{A} = \frac{(y-Bs)(y-Bs)^T}{((y-Bs)^T s)^2}$, where $(y-Bs)^T s \neq 0$, we come to

$$\widetilde{B} = B + \theta \frac{(y-Bs)(y-Bs)^T}{(y-Bs)^T s},$$

which can be viewed as a kind of the modification of Davidon's symmetric rank-one update.

Family 3: $T = \frac{(\rho y - Bs)c^T}{(y-Bs)^T c}$, where $\rho \in R,\ c \in R^n,\ (y-Bs)^T c \neq 0$.

- This setting leads to

$$\tilde{B}s = \rho y,$$

which corresponds to the equation satisfied by Huang Family ([7]).

Family 4: $T = [\theta I + (\theta - \theta^2)\frac{cs^T}{c^T s}]$. $(c \in R^n,\ c^T s \neq 0)$

- Setting $c = s$ gives

$$\widetilde{B} = B + \theta \frac{s(y-Bs)^T + (y-Bs)s^T}{s^T s} - \theta^2 \frac{(y-Bs)^T s s s^T}{(s^T s)^2}. \tag{4.4}$$

If $\theta \in (0,2)$, then the preceding may be viewed as a modification of the PSB, appearing originally in [15]. The local and superlinear convergence of the scheme (1.1) using update (4.4) together with $\alpha_k = 1$ has been shown by Moré and Trangenstein [10]. Furthermore, when $F(x)$ is defined as $F(x) = Gx - b$, where G is a symmetric and nonsingular matrix, the global and superlinear convergence can also be ensured. Besides, if $\theta = 2$, it is easy to verify its linear convergence.

- Setting $c = y$ implies the following new update,

$$\widetilde{B} = B + \theta \frac{y(y-Bs)^T + (y-Bs)y^T}{y^T s} - \theta^2 \frac{(y-Bs)^T s y y^T}{(y^T s)^2}, \tag{4.5}$$

which is referred to as DFP-like update because it becomes the DFP update when $\theta = 1$.

It can be shown under appropriate assumptions that the preceding update and some conventional ones (like DFP and BFGS) share properties such as the hereditary positive definiteness when $\theta_k \in [0,2]$, the locally linear convergence when $\theta_k \in (0,1]$ and superlinear convergence when $\theta_k \in (1 - \sqrt{\frac{1}{3}}, 1]$. We will handle this topic separately in [18].

5. Computational Results

We report in this section computational results obtained from our computational experiments, showing that the DFP-like new update outperforms the DFP in terms of both the convergence rate and stability.

We used the following algorithm model:

Algorithm 1. *Let θ be given. Determine an initial point x_0 and set $B_0 := I,\ k := 0,\ g_k = \nabla f(x_k),\ s_k := 0,\ y_k := 0$, and $\varepsilon = 10E-09$.*

step 1. If $\|g_k\| < \varepsilon$, stop; otherwise, apply Cholesky decomposition to B_k to obtain $B_k = H_k H_k^T$, and solve two triangular systems $H_k z = g_k, H_k^T d_k = z$ for the direction d_k.

step 2. Determine α_k by Armijo-Goldstein rule ($\rho = 0.4$).
step 3. Set $x_{k+1} := x_k - \alpha_k d_k,\ s_k := -\alpha_k d_k,\ y_k := g_{k+1} - g_k$.
step 4. Update B_k by the DFP-like formula, i.e.,

$$B_{k+1} := B_k + \theta \frac{y_k(y_k - B_k s_k)^T + (y_k - B_k s_k) y_k^T}{y_k^T s_k} - \theta^2 \frac{(y_k - B_k s_k)^T s_k y_k y_k^T}{(y_k^T s_k)^2}. \tag{5.1}$$

step 5. Set $k := k + 1$ and go to step 1.

Our computational tests involve four codes based on the preceding algorithm:
1. DFP: Algorithm 1 with $\theta = 1$.
2. R-DPF-L: Algorithm 1 with θ taken as a random value in $(1 - \sqrt{\frac{1}{3}}, 1]$.
3. DFP-L1: Algorithm 1 with $\theta = 0.65$.
4. DFP-L2: Algorithm 1 with $\theta = 0.85$.

Our test set of problems involves the following 21 standard test problems, all of which are of the form

$$\min_{x \in R^n} f(x), \tag{5.2}$$

whose solution x^* satisfies the first order necessary condition $\bigtriangledown f(x^*) \triangleq F(x^*) = 0$:
Tf.1 Rosenbrock function, $x_0 = (-1.2, 1)^T$;
Tf.2 Freudenstein and Roth function, $x_0 = (0.5, -2)^T$;
Tf.3 Powell badly scaled function, $x_0 = (0, 1)^T$;
Tf.4 Jennrich and Sampson function (n=2, m=2), $x_0 = (0.3, 0.4)^T$;
Tf.5 Brown badly scaled function, $x_0 = (1, 1)^T$;
Tf.6 Box three-dimensional function (n=m=3), $x_0 = (0, 10, 20)^T$;
Tf.7 Variably dimensioned function (n=2, m=4), $x_0 = (0.5, 0)^T$;
Tf.8 Broyden tridiagonal function (n=m=2), $x_0 = (-1, -1)^T$;
Tf.9 Wood function (n=4, m=6), $x_0 = (-3, -1, -3, -1)^T$;
Tf.10 Penalty function (n=2, m=3), $x_0 = (1, 2)^T$;
Tf.11 Brown almost-linear function (n=m=2), $x_0 = (0.5, 0.5)^T$;
Tf.12 Discrete boundary value function (n=m=2), $x_0 = (2, 5)^T$;
Tf.13 Linear function-rank-1 (n=m=2), $x_0 = (1, 1)^T$;
Tf.14 Beale function, $x_0 = (1, 1)^T$;
Tf.15 Trigonometric function, $x_0 = (0.5, 0.5)^T$;
Tf.16 Penalty function II (n=2, m=4), $x_0 = (0.5, 0.5)^T$;
Tf.17 Brown and Dennis function (n=4, m=4), $x_0 = (25, 5, -5, -1)^T$;
Tf.18 Biggs EXP6 function, $x_0 = (1, 2, 1, 1, 1, 1)^T$;
Tf.19 Gaussian function (n=3, m=15), $x_0 = (0.3, 1.3, 0)^T$;
Tf.20 Watson function (n=2, m=31), $x_0 = (0, 0)^T$;
Tf.21 Extended Rosenbrock function (n=4, m=4), $x_0 = (-1.2, 1, -1.2, 1)^T$;

All tests were conducted on a P2 personal computer with machine precision $10E - 16$. The three new codes were tested, and compared with the traditional DFP code. Iteration counts required to solve each problem by them are listed in Table 1.

Note: *Each of the test problems was solved 20 times with R-DFP-L. In the table, listed is only the smallest number of iterations required by it. In addition, "25;206" implies that two local minimums were found; the same to "31;72;64;91".*

This table tells how differently the DFP-like performs with different values of θ. Overall, DFP-L2 ($\theta = 0.85$) outperformed DFP in terms of both convergence rate and stability. Also, results obtained with "R-DFP-L" suggest that it would be advantageous to use some kind of self-scaling technique with respect to θ.

Table 1. Iteration counts

Problem	DFP	R-DFP-L	DFP-L1	DFP-L2
$Tf.1$	2263	131	211	345
$Tf.2$	28	25;206	26	27
$Tf.3$	601	216	273	389
$Tf.4$	57	39	46	54
$Tf.5$	60	55	65	60
$Tf.6$	13	31;72;64;91	288	67
$Tf.7$	8	8	14	11
$Tf.8$	21	15	17	20
$Tf.9$	-	7720	-	-
$Tf.10$	6	6	11	7
$Tf.11$	20	13	20	20
$Tf.12$	16	13	15	16
$Tf.13$	20	19	19	20
$Tf.14$	46	24	34	41
$Tf.15$	35	11	14	13
$Tf.16$	18	15	18	18
$Tf.17$	614	188	3728	736
$Tf.18$	-	-	-	-
$Tf.19$	31	31	31	31
$Tf.20$	21	21	21	21
$Tf.21$	-	550	2900	7834
Total	3277	735	4578	1507

6. Concluding Remarks

We made a generalization of the quasi-Newton equation, analyzed it, and derived some updates from it. While most of the updates are analogues to some existing ones, there might still be some updates unknown—the Broyden-lie ones for example—that are better than those derived in this paper. In particular, the DFP-like update (4.5) appears to be very promising, and deserves further investigation.

References

[1] C. Brezinski, A classification of quasi-Newton methods, *Numerical Algorithms*, **33** (2003), 123–135.

[2] C. G. Broyden, A class of methods for solving nonlinear simultaneous equations, *Math. Comp.*, **19** (1965), 577–593.

[3] C. G. Broyden, J. E. Dennis and J. J. Moré, On the local and superlinear convergence of quasi-Newton methods, *Math. Comp.*, **12** (1973), 223–246.

[4] W. C. Davidon, Variable metric method for minimization, *Rep. ANL-5990 Rev, Argonne National Laboratories, Argonne, III,* 1959.

[5] J. E. Dennis, JR. and J. J. Moré, Quasi-Newton methods, motivation and theory, *SIAM Review*, **19**:1 (1977), 46–89.

[6] J. E. Dennis, On the convergence of Broyden's method for nonlinear systems of equations, *Math. Comp.,* **25** (1971), 559–567.

[7] H. Y. Huang, Unified approach to quadratically convergent algorithms for function minimization, *Journal of Optimization Theory and Applications*, **5** (1970), 405–423.

[8] A. P. Liao, Modifying the BFGS method, *Operations Research Letters*, **20** (1997), 171–177.

[9] D. H. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, *Journal of Computational and Applied Mathematics*, **129** (2001), 15–35.

[10] J. J. Moré and J. A. Trangenstein, On the global convergence of Broyden's method, *Math, Comp.*, **30** (1976), 523–540.

[11] J. J. Moré, B. S. Garbow and K. E. Hillstrom, Testing unconstrained optimization software, *ACM Transactions on Mathematical Software*, **7** (1981), 17–41.

[12] P. Q. Pan, On advance of second order Quasi-Newton methods, *Jounal of Computational Mathematics*, **2**:1 (1984), 24–32.

[13] P. Q. Pan and Y. T. Zhao, Is it possible to be superior to the BFGS method? (in Chinese), *In Operations Research and Decision Sciences, Chendu University of Technology and Science Press,* (1992), 623–631.

[14] M. J. D. Powell, A FORTRAN subroutine for solving systems of nonlinear algebraic equations, *Numerical Methods for Nonlinear Algebraic Equations, Gordon and Breach, London(1970)*, 115–161.

[15] M. J. D. Powell, A new algorithm for unconstrained optimization, *Nonlinear Programming, J. B. Rosen, O. L. Mangasarian and K. Ritter. eds., Academic Press, New York*, 1970.

[16] Z. X. Wei, G. H. Yu, G. L. Yuan, and Z. G. Lian, The superlinear convergence of a modified BGFS-type method for uncontrained optimization, *Computational Optimization and Application*, **29** (2004), 315–332.

[17] C. X. Xu and J. Z. Zhang, A Survey of quasi-Newton equations and quasi-Newton methods for optimization, *Annals of Operations Research*, **103** (2001), 213–234.

[18] L. H. Zhang and P. Q. Pan, On the local and superlinear convergence of a new DFP-like formula, *(Submitted).*

[19] J. Z. Zhang, N. Y. Deng, and L. H. Chen, New quasi-Newton equation and related methods for unconstrained optimization, *Journal of Optimization Theory and Applications*, **102** (1999), 147–167.

[20] J. Z. Zhang and C. X. Xu, Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations, *Journal of Computational and Applied Mathematics*, **137** (2001), 269–278.