A NOTE ON PAN’S SECOND-ORDER QUASI-NEWTON UPDATES *

Lei-Hong Zhang† Ping-Qi Pan‡
Department of Mathematics, Southeast University, Nanjing, 210096, P.R.China.

Abstract
This note, attempts to further Pan’s second-order quasi-Newton methods([1]). To complement the numerical implementation, the linear convergence of a rank-one second-order update and the least change property are presented.

Keywords: Second-order quasi-Newton update, linear convergence, least change property.
AMS classification: 65H10, 65K05, 90C26

1 Introduction
In 1959, Davidon first introduced the quasi-Newton method([2]) for solving the unconstrained minimization

$$\min_{x \in \mathbb{R}^n} f(x). \quad (1.1)$$

Since then, there has been an ever-expanding interest in quasi-Newton methods both in literature(e.g., [3], [4], [5], [6], [7]) and in the implementation(e.g.,[8], [9], [10]). The most general form of this method is

$$x_{k+1} = x_k - B_k^{-1}\nabla f(x_k), \quad k = 0, 1, \cdots \quad (1.2)$$

where $B_k$ satisfies the well-known quasi-Newton equation,

$$B_{k+1}y_k = y_k. \quad (1.3)$$

---

*This work was supported by the National Science Foundation of China (No.10371017).
†Department of Mathematics, Southeast University,Email:longzlh@163.com
‡Department of Mathematics, Southeast University,Email:panpq@seu.edu.cn
with the notation
\[ s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k). \] (1.4)

Since the advent of (1.3), a vast number of quasi-Newton updates which satisfy (1.3) have been proposed. The most efficient and practical one so far, is BFGS (independently by Broyden(1969, 1970), Fletcher (1970), Goldfard (1970), Shanno (1970)). However, on the other hand, the effort to construct some updates not satisfying (1.3) never stops(e.g., [11], [12]). In 1984, Pan([1]) suggested a new quasi-Newton equation,
\[ B_{k+1}s_k = 2y_k - B_k s_k, \] (1.5)
and he also showed that in an approximate sense, this new equation is of second order, in contrast to the first order of the classical one(1.3). This equation, hence, is called the second-order quasi-Newton equation and its relevant updates the second-order quasi-Newton formulae. Although some computational tests implemented in [1] illustrate its attraction for further investigation, some theoretical analysis is still deficient even the convergent analysis. This note, then, is intended to be a step to develop the second-order quasi-Newton equation/updates. The local linear convergence of a rank-one updating formula is described first, and then follows the least change property, ie., the property of the "closest" to the preceding update.

2 Linear Convergence

For easy of expression, we shall use \( F(x) \) to denote \( \nabla f(x) \); also, we shall omit suffixes and shall denote all quantities corresponding to \( k + 1 \) and \( k \) by using a tilde superfix, no superfix, so that \( F(x_{k+1}) \) is written \( \tilde{F} \), and \( F(x_k) \) is written \( F \), and the same to others.

As a special second-order quasi-Newton update, Pan ([1]) gave a rank-one formula,
\[ \tilde{B} = B + \frac{2(y - Bs)s^T}{s^Ts}, \] (2.1)
We will show the linear convergence of the iteration (1.2) with the update (2.1). So, first we assume \( F(x) \) has the following hypotheses.

(A). The mapping \( F(x) \) is continuously differentiable in an open convex set \( D \).
(B). There is an \( x^* \) in \( D \) such that \( F(x^*) = 0 \) and \( F'(x^*) \) is nonsingular.
F′(x) satisfies a Lipschitz condition at x∗, that is, there is a constant λ such that
\[ \| F′(x) - F′(x^*) \| \leq \lambda \| x - x^* \| \quad x \in D. \] (2.2)

The proof of the linear convergence follows from Lemma 1 and Lemma 2 below.

**Lemma 1** (Broyden, Dennis and Moré, [13], 1973) Let \( F : \mathbb{R}^n \rightarrow \mathbb{R}^n \) satisfy the hypotheses (A), (B), (C) of \( F(x) \), and let \( U \) be an update function for \( F(x) \) such that for all \( (x, B) \in \text{dom } U \) and \( B \in U(x, B) \)

\[ \| B - F′(x^*) \| \leq [1 + \alpha_1 \sigma(x, \bar{x})] \| B - F′(x^*) \| + \alpha_2 \sigma(x, \bar{x}) \] (2.3)

for some constants \( \alpha_1 \) and \( \alpha_2 \) where \( \bar{x} = x - B^{-1}F(x) \) and

\[ \sigma(x, \bar{x}) = \max\{ \| \bar{x} - x^* \|, \| x - x^* \| \} \]

Then there are positive constants \( \varepsilon \) and \( \delta \) such that if \( x_0 \in D \) and \( B_0 \in D_M \) satisfy \( \| x_0 - x^* \| \leq \varepsilon \) and \( \| B_0 - F′(x^*) \| \leq \delta \), then iteration (1.2) is well defined and converges linearly to \( x^* \).

This lemma still holds if we take the Frobenius norm.

**Lemma 2** Let \( F : \mathbb{R}^n \rightarrow \mathbb{R}^n \) satisfy the hypotheses (A), (B) of \( F(x) \). Then for any \( x^* \in D \),

\[ \| F(v) - F(u) - F′(x^*)(v - u) \| \leq \gamma \max\{ \| v - x^* \|, \| u - x^* \| \} \| v - u \| \]

for all \( v \) and \( u \) in \( D \).

This conclusion follows from a standard result (Ortega & Rheinboldt, 1970:70).

**Theorem 1** Let \( F : \mathbb{R}^n \rightarrow \mathbb{R}^n \) satisfy the assumptions (A), (B), (C) of \( F(x) \), consider the iteration scheme (2.4)

\[
\begin{align*}
x_{k+1} &= x_k - B_k^{-1}F(x_k) \\
B_{k+1} &= B_k + 2\frac{(y_k - B_k s_k)s_k^T}{s_k^T s_k}, \quad k = 0, 1, \ldots
\end{align*}
\] (2.4)

with \( y_k = F(x_{k+1}) - F(x_k) \), and \( s_k = x_{k+1} - x_k \). Then this method is well defined and locally convergent at \( x^* \).
Proof. According to (2.4), it follows

\[ B_{k+1} - F'(x^*) = B_k - F'(x^*) + \frac{2(y_k - B_k s_k) s_k^T}{s_k^T s_k} \]

\[ = [B_k - F'(x^*)][I - \frac{2s_k s_k^T}{s_k^T s_k}] + \frac{2(y_k - F'(x^*)s_k) s_k^T}{s_k^T s_k}. \]

Applying \( \| I - \frac{2s_k s_k^T}{s_k^T s_k} \|_2 = 1 \) and lemma 2 gives

\[ \| B_{k+1} - F'(x^*) \|_F \leqslant \| [B_k - F'(x^*)] \|_F + \| \frac{2(y_k - F'(x^*)s_k)}{s_k} \|_F \]

\[ \leqslant \| [B_k - F'(x^*)] \|_F + 2\gamma \sigma(x_k, x_{k+1}), \]

which completes the proof with the aid of lemma 1. \( \Box \)

In fact, this result can be extended to a larger group of updates in which a parameter \( \theta \in (0, 2] \) replaces ”2” in (2.4), ie.,

\[ \tilde{B} = B + \frac{\theta(y - Bs)s^T}{s^T s}. \] (2.5)

This formula originally suggested by Powell [14] as the modification of Broyden’s rank-one update([7]). The proof is the same when noting \( \| I - \theta \frac{s s^T}{s^T s} \| = 1 \) with \( \theta \in (0, 2] \). Correspondingly, the second-order equation shall then become

\[ \tilde{B}s = Bs + \theta(y - Bs). \] (2.6)

3 Least Change Property

Now, we are in a position to describe another property related also to the update (2.1).

Theorem 2 Given \( B \in R^{n \times n}, y \in R^n \) and some nonzero \( s \in R^n \), define \( \tilde{B} \) by (2.1), then \( \tilde{B} \) is the unique solution to the problem,

\[ \min_{\tilde{B}} \{ \| \tilde{B} - B \|_F : \tilde{B}s = 2y - Bs \}. \] (3.1)

Proof. Let \( \overline{B} \) be an arbitrary such that \( \overline{B}s = 2y - Bs \), which implies

\[ y - Bs = \frac{1}{2} (\overline{B} - B)s. \] (3.2)
It follows from (3.2) that
\[
\| \tilde{B} - B \|_F = \| 2 \frac{(y - Bs)s^T}{s^Ts} \|_F = \| \frac{(Bs - Bs)s^T}{s^Ts} \|_F \leq \| B - B \|_F.
\]
The conclusion that \( \tilde{B} \) is the unique solution results from the fact that the mapping \( f : R^{n \times n} \rightarrow R \) defined by \( f(\bar{B}) = \| \bar{B} - B \|_F \) is strictly convex in the convex set \( \{ \bar{B} : \bar{B}s = 2y - Bs \} \). □

The counterpart of theorem 2 concerns a rank-two update, another second-order quasi-Newton formula,
\[
\tilde{B} = B + 2c(y - Bs)^T + (y - Bs)c^T + 2 \frac{(y - Bs)^T sc c^T}{(c^Ts)^2}, \quad (0 \neq c \in R^n) \tag{3.3}
\]
Analogous to theorem 2, the least change property is also true to this update.

**Theorem 3** Let \( B \in R^{n \times n} \) be symmetric, and \( c, s, y \in R^n \) satisfy \( c^Ts > 0 \); assume also that \( M \in R^{n \times n} \) is some nonsingular symmetric matrix such that \( Mc = M^{-1}s \), then the matrix \( \tilde{B} \) defined by (3.3) is the unique solution to the problem,
\[
\min_{\bar{B}} \{ \| \bar{B} - B \|_{M,F}: \bar{B}s = 2y - Bs, \bar{B}^T = \bar{B} \}, \tag{3.4}
\]
where \( \| B \|_{M,F} = \| MBM \|_F \).

**Proof.** Assume \( \bar{B} \) be a symmetric matrix such that \( \bar{B}s = 2y - Bs \), also implying (3.2). Let \( Mc = M^{-1}s = z \), \( E = M(\tilde{B} - B)M \), \( \tilde{E} = M(\tilde{B} - B)M \). Multiplying by \( M \) on the two sides of (3.3), and taking account of (3.2) leads to,
\[
\tilde{E} = M(\tilde{B} - B)M = Ezz^T + z z^T E z^T z - \frac{z z^T E z z^T}{(z^T z)^2}.
\]
Obviously, \( \| \tilde{E}z \|_2 = \| Ez \|_2 \); and moreover, if \( v \perp z \), from (3.5), it follows
\[
\| \tilde{E}v \|_2 \leq \| Ev \|_2,
\]
which support the conclusion that \( \tilde{B} \) defined by (3.3) is the solution of (3.4). The uniqueness of this solution results from the fact that the mapping \( f : R^{n \times n} \rightarrow R \) defined by \( f(\bar{B}) = \| \bar{B} - B \|_{M,F} \) is strictly convex in the convex set \( \{ \bar{B} : \bar{B}s = 2y - Bs, \bar{B}^T = \bar{B} \} \).

Thus we complete the proof. □

In particular, setting \( c = s \) in (3.3) yields the second-order PSB update, and taking \( c = y \), then, gives the second-order DFP (see [1]).

More generally, theorem 2 and theorem 3 still hold if "2" in (2.1) and (3.3) is replaced by a real parameter \( \theta \neq 0 \), and we just summarize them below.
Theorem 4 Given $B \in \mathbb{R}_{\times n}$, $y \in \mathbb{R}^n$ and some nonzero $s \in \mathbb{R}^n$, define $\tilde{B}$ by (2.5), then $\tilde{B}$ is the unique solution to the problem,

$$\min_{\overline{B}} \{ \| \overline{B} - B \|_F : \overline{B}s = Bs + \theta(y - Bs) \}.$$

Theorem 5 Let $B \in \mathbb{R}^{n\times n}$ be symmetric, and $c, s, y \in \mathbb{R}^n$ satisfy $c^T s > 0$; assume also that $M \in \mathbb{R}^{n\times n}$ is some nonsingular symmetric matrix such that $Mc = M^{-1}s$, then the matrix

$$\tilde{B} = B + \theta \frac{c(y - Bs)^T + (y - Bs)c^T}{c^T s} - \theta \frac{(y - Bs)^T sc c^T}{(c^T s)^2},$$

is the unique solution to the problem,

$$\min_{\overline{B}} \{ \| \overline{B} - B \|_{M,F} : \overline{B}s = Bs + \theta(y - Bs), \overline{B}^T = \overline{B} \},$$

where $\| B \|_{M,F} = \| MBM \|_F$.

4 Concluding Remarks

For a credible algorithm, the convergence analysis is necessary. As the first step to develop the second-order quasi-Newton methods, the proposed rank-one update (2.1) possesses the property of linear convergence. The proposition described in section 3 shows the evidence that the proposed updates (2.1) and (3.3) are the best candidates for $\tilde{B}$ as the rank-one and rank-two updates, respectively. In practice, however, some other rank-two updates are still attractive (refer to [1]), and the linear and further, the superlinear convergence remain to be open.

References


[8] FLETCHER, R. and POWELL, M.J.D. On the modification of $LDLT^T$ factorization, 
Math, Comp., v. 28, 1974, pp.1067-1087.

[9] BRODLIE, K.W. An assessment of two approaches to variable metric methods, Math, 
Prog., v. 12, 1977, pp.344-355.


convergence of a modified BGFS-type method for unconstrained optimization. Computational 

[12] JIANZHONG ZHANG, CHENGXIAN XU, Properties and numerical performance of quasi-
Newton methods with modified quasi-Newton equations. Journal of Computational and 
Applied Mathematics., 137, 2001, 269-278.

[13] G.G. BROYDEN. J. E. DENNIS ADN J. J. MORÉ. On the local and superlinear conver-


1999.