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])

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

* 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

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