Vol. 2 No. 1 JOURNAL OF COMPUTATIONAL MATHEMATICS January 1984

ON ADVANCE OF SECOND ORDER QUASI-NEWTON METHODS*

PAN PING-QI (潘平奇)

(*Nanjing Technological College of Forest Products, Nanjing, China*)

1. Introduction

The key to the quasi-Newton methods for solving nonlinear equations in several variables or optimization problems lies in how the approximation to the Jacobian or Hessian matrix (or its inverse) is determined.

We denote by R^n and $R^{n\times n}$ the real n-dimensional linear space of all column vectors and that of all real square matrices of order n, respectively. The mapping with domain D in R^n and range in R^n is denoted by $F: D\subset R^n \to R^n$, which is the left of the system of equations or the gradient of the objective function.

Let $\hat{x}$ and $\tilde{x}$ be two distinct points in D and denote

$$\hat{F}=F(\hat{x}), \quad \tilde{F}=F(\tilde{x}), \quad B=F'(\hat{x}), \tag{1.1}$$

where $F'(x)$ is the Jacobian matrix of F at x, and

$$\delta=\tilde{x}-\hat{x}, \quad \gamma=\tilde{F}-\hat{F}. \tag{1.2}$$

It is our purpose to approximate the Jacobian matrix of F at $\tilde{x}$.

Evidently, the linear mapping

$$L(x)=\tilde{F}+\tilde{B}(x-\tilde{x}), \tag{1.3}$$

where $\tilde{B}\in R^{n\times n}$ and

$$\tilde{B}\delta=\gamma \tag{1.4}$$

satisfies

$$L(\hat{x})=\hat{F}, \; L(\tilde{x})=\tilde{F}. \tag{1.5}$$

So, if $\|\delta\|$ is small enough, $L(x)$ can be regarded as a reasonable approximation to $F(x)$ at $\tilde{x}$ and $\tilde{B}$ as an approximate Jacobian matrix of F at the same point. As a matter of fact, we have

Theorem 1. *Suppose that $F:D\subset R^n\to R^n$ is Fréchet-differentiable at $\tilde{x}\in \operatorname{int}(D)$ and that $L(x)$ is determined by* (1.3) *and* (1.4). *Then, for any $t\in R^1$, which is non-zero and sufficiently small,*

$$L(\tilde{x}+t\delta)-F(\tilde{x}+t\delta)=tE(t,\delta), \tag{1.6}$$

where $\lim\limits_{\delta\to 0} E(t,\delta)/\|\delta\|=0$, $\lim\limits_{t\to 0} E(t,\delta)=\bar{E}(\delta)$ *and* $\lim\limits_{\delta\to 0}\bar{E}(\delta)/\|\delta\|=0$.

Proof.

$$F(\tilde{x}+t\delta)=\tilde{F}+F'(\tilde{x})(t\delta)-E_1(t\delta), \tag{1.7}$$

where
$$\lim_{t\delta\to 0} E_1(t\delta)/\|t\delta\|=0.$$

$$\hat{F}=\tilde{F}-F'(\tilde{x})\delta-\bar{E}(\delta), \tag{1.8}$$

* Received January 5, 1983.

where
$$\lim_{\delta\to 0} \bar{E}(\delta)/\|\delta\| = 0.$$

By (1.3), (1.7), (1.4) and (1.8), we have

$$L(\tilde{x}+t\delta)-F(\tilde{x}+t\delta)=t\tilde{B}\delta-tF'(\tilde{x})\delta+E_1(t\delta)$$
$$=t(\tilde{F}-\hat{F}-F'(\tilde{x})\delta)+E_1(t\delta)=t[\bar{E}(\delta)+E_1(t\delta)/t]. \tag{1.9}$$

If we let
$$E(t,\ \delta)=\bar{E}(\delta)+E_1(t\delta)/t, \tag{1.10}$$

then
$$\lim_{\delta\to 0} E(t,\ \delta)/\|\delta\| = \lim_{\delta\to 0} \bar{E}(\delta)/\|\delta\| + \lim_{\delta\to 0} E_1(t\delta)/(t\|\delta\|)=0$$

and
$$\lim_{t\to 0} E(t,\ \delta)=\bar{E}(\delta)+\lim_{t\to 0} E_1(t\delta)/t=\bar{E}(\delta).$$

The theorem is then proved.

In view of the above, we regard (1.3) as the first order approximation to F at $\tilde{x}$.

Let $\tilde{B}$ be nonsingular and $\tilde{H}$ be its inverse. Then (1.4) is equivalent to
$$\delta=\tilde{H}\gamma. \tag{1.11}$$

Equation (1.4) or (1.11) has been central to the development of quasi-Newton methods, and therefore it is often termed the quasi-Newton equation. But in view of Theorem 1, we call it first order quasi-Newton equation, and all of the known updates, which can be derived from it and used in quasi-Newton methods, are called, correspondingly, the first order.

Since there is no B or H in (1.4) or (1.11), $\tilde{B}$ or $\tilde{H}$ relates only to γ and δ rather than to B or H. Thus $\tilde{B}=\gamma q^T/q^T\delta$, where $q\in R^n$ and $q^T\delta\neq 0$, satisfies (1.4) and is independent of B. In fact, B or H in the first order updates can be, theoretically, replaced by any other positive definite matrices (for optimization problems).

In this paper, many new updates are derived by means of new approximations to F and shown to be of second order in a certain sense. They have not only higher precision than the old ones but also the same simplicity; moreover, B or H there cannot be replaced by other matrices, and therefore the approximate Jacobian (or Hessian) matrix or its inverse at the initial point must be used as the initial matrix, or theoretically so to say the least. In addition, it is much more significant that the so-called second order quasi-Newton equation derived later must become the new starting point of quasi-Newton methods.

2. One-Reduction Matrices of δ

Let us introduce the new concept needed in the ensuing sections:

Definition. *Let $\delta\in R^n$, $A\in R^{n\times n}$ is a one-reduction matrix of δ if*
$$\delta^T A\delta=1. \tag{2.1}$$

Two examples

(1) $$U(P,\ S)=\frac{1+\delta^T S\delta}{\delta^T P\delta}P-S,\quad \forall P,\ S\in R^{n\times n} \text{ and } \delta^T P\delta\neq 0. \tag{2.2}$$

(2) $$V(P,\ q)=\frac{1-\delta^T q}{\delta^T P\delta}P+\frac{qq^T}{\delta^T q},\quad \forall P\in R^{n\times n} \text{ and } \delta^T P\delta\neq 0,\ q\in R^n \text{ and } \delta^T q\neq 0. \tag{2.3}$$