数理解析研究所講究録
990 巻 1997 年 62-71

Quadratic and Superlinear Convergence of the Huschens Method for Nonlinear Least Squares Problems

非線形最小2乗問題に対するHuschens法の2次および超1次収束性

東京理科大学・理学部　小笠原 英穂　(Hideho OGASAWARA)
東京理科大学・工学部　矢部 博　(Hiroshi YABE)

1 Introduction

In this paper, we consider iterative methods for solving the nonlinear least squares problem

$$\text{(1.1)} \qquad \underset{x\in R^n}{\text{minimize}}\ f(x)=\frac{1}{2}\|r(x)\|^2,$$

where $r(x)=(r_1(x),\ldots,r_m(x))^T$, and each $r_i: R^n \to R$ is twice continuously differentiable for $i=1,\ldots,m$ $(m\geq n)$, and $\|r(x)\|$ denotes the l_2 norm of $r(x)$. We assume that there exists a local minimum of the problem, denoted by x_*. Several kinds of Newton-based methods for solving problem (1.1) have been proposed, and most of them deal with the structure of the Hessian matrix of the function $f(x)$

$$\text{(1.2)} \qquad \nabla^2 f(x)=C(x)+G(x),$$

with

$$C(x)=J(x)^TJ(x) \quad \text{and} \quad G(x)=\sum_{i=1}^m r_i(x)\nabla^2 r_i(x),$$

where $J(x)$ denotes the $m\times n$ Jacobian matrix of $r(x)$.

Among these methods, structured quasi-Newton methods are considered as promising methods. These methods compute a step s_k by solving the Newton equation

$$\text{(1.3)} \qquad (C(x_k)+A_k)s=-\nabla f(x_k)$$

at k-th iteration, and generate a sequence $\{x_k\}$, which approximates the solution x_*, such that

$$\text{(1.4)} \qquad x_{k+1}=x_k+s_k,$$

where ∇f denotes the gradient vector of f. Here the matrix A_k is a k-th approximation to the matrix $G(x_k)$, and several kinds of structured quasi-Newton updates were proposed and their local convergence properties were discussed. For example, Dennis, Gay and

Welsch [6] proposed the structured DFP update, and Dennis and Walker [5] showed local and superlinear convergence of this method. Dennis, Martinez and Tapia [7] derived the structure principle, and proved local and superlinear convergence of the structured BFGS update, which was proposed by Al-Baali and Fletcher [1]. Engels and Martinez [8] later unified these methods, and proposed the structured Broyden family and showed local and superlinear convergence of the convex class of this family. Very recently, Yabe and Yamaki [15] have proved local and superlinear convergence of a wider class of the family in the way different from [8]. Futhermore, the factorized forms of structured quasi-Newton methods were studied by Yabe and Takahashi [13], and Yabe and Yamaki [14] with the aim of constructing a descent search direction for $f(x)$.

These methods overcome the weakness so that the Gauss-Newton method may perform poorly for large residual problems. On the other hand, the Gauss-Newton method possesses the quadratic convergence property in the case of zero residual problems, i.e., $r(x_*) = 0$ and then $G(x_*) = 0$, while structured quasi-Newton methods do not perform as well as the Gauss-Newton method does. This is caused by the fact that a matrix A_k generated by quasi-Newton updates does not approach the zero matrix. As practical remedies for this difficulty, there are two typical strategies, a sizing technique and a hybrid method. The former was independently proposed by Bartholomew-Biggs [2] and Dennis, Gay and Welsch [6], and this strategy multiplies A_k by a sizing factor before updating. The latter was proposed by Al-Baali and Fletcher [1] and Fletcher and Xu [9], and this combines the structured quasi-Newton method and the Gauss-Newton method. However, there is no theoretical convergence property for these two strategies.

In order to overcome this difficulty, Huschens [10] proposed a new structured quasi-Newton method that converges quadratically to x_* for zero residual problems and converges superlinearly to x_* for nonzero residual problems. Specifically, he incorporated a self-sizing strategy into structured quasi-Newton methods without losing a fast rate of convergence for large residual problems. By considering that the matrix $G(x)$ is represented by

$$G(x) = \|r(x)\| \sum_{i=1}^{m} \frac{r_i(x)}{\|r(x)\|} \nabla^2 r_i(x),$$

Huschens proposed to approximate the matrix $G(x_k)$ by $\|r(x_k)\| A_k$, where the matrix $A_k \in R^{n\times n}$ is the k-th approximation to the part $\sum_{i=1}^{m}(r_i(x)/\|r(x)\|)\nabla^2 r_i(x)$ so that

$$\nabla^2 f(x_k) \approx C(x_k) + \|r(x_k)\| A_k.$$

In this case, the step s_k can be computed by solving

$$(C(x_k) + \|r(x_k)\| A_k)s = -\nabla f(x_k). \tag{1.5}$$

Based on this idea, he gave the following condition that a new approximation to the Hessian matrix should satisfy

$$(C(x_{k+1}) + \|r(x_{k+1})\| A_{k+1})s_k = z_k, \tag{1.6}$$

where

$$s_k = x_{k+1} - x_k, \tag{1.7}$$

$$z_k = C(x_{k+1})s_k + \|r(x_{k+1})\|y_k^\sharp$$

and

$$y_k^\sharp = (J(x_{k+1}) - J(x_k))^T \frac{r(x_{k+1})}{\|r(x_k)\|}.$$

Thus the matrix A_k is updated such that the new matrix A_{k+1} satisfies the secant condition

$$A_{k+1}s_k = y_k^\sharp. \tag{1.8}$$

Following the above condition, Huschens proposed a family with one parameter ϕ_k that corresponds to the Broyden family, say Huschens-Broyden family, and showed convergence properties of the convex class, i.e., $0 \le \phi_k \le 1$, of this family. Specifically, he showed local and quadratic convergence for zero residual problems, and local and superlinear convergence for nonzero residual problems.

In this paper, we will extend the results of Huschens to a wider class of the Huschens-Broyden family and give local convergence properties in a way different from the proofs by him. We emphasize that for zero residual problems, Huschens used a restricted value of the parameter ϕ_k in the convex class, while we will deal with the case where ϕ_k can vary fully in the wider class. This paper is organized as follows. In Section 2, we briefly review the Huschens method. In Section 3, we present some useful lemmas to show convergence properties in the following sections. In Sections 4 and 5, we show quadratic and superlinear convergence properties of the method for zero and for nonzero residual problems, respectively.

Throughout this paper, $\|\cdot\|$ denotes the l_2 norm for vectors or matrices, and $\|\cdot\|_F$ and $\|\cdot\|_{F,M}$ denote the Frobenius norm and the weighted Frobenius norm for some nonsingular matrix M, which are defined by

$$\|Q\|_F = \sqrt{\mathrm{Trace}(QQ^T)} \quad \text{and} \quad \|Q\|_{F,M} = \|M^{-1}QM^{-1}\|_F,$$

respectively.

2 Huschens method

In this section, we review the Huschens method. In what follows, we omit subscript k and simply denote "$k+1$" by "+" if not necessary.

Let ϕ be a scalar parameter and

$$B = J(x)^T J(x) + \|r(x)\|A \quad \text{and} \quad B^\sharp = J(x_+)^T J(x_+) + \|r(x_+)\|A. \tag{2.1}$$

Based on the structure principle given by Dennis et al. [7], Huschens [10] derived the structured Broyden family:

$$B_+ = B^\sharp - \frac{B^\sharp ss^T B^\sharp}{s^T B^\sharp s} + \frac{zz^T}{s^T z} + \phi(s^T B^\sharp s)vv^T, \tag{2.2}$$

where

$$v = \frac{z}{s^T z} - \frac{B^\sharp s}{s^T B^\sharp s}. \tag{2.3}$$

From this, an A-update can be obtained as follows:

$$A_+ = A + \frac{1}{\|r(x_+)\|}\Big[- \frac{B^\sharp s s^T B^\sharp}{s^T B^\sharp s} + \frac{zz^T}{s^T z} + \phi(s^T B^\sharp s) v v^T \Big]. \tag{2.4}$$

We call the families (2.2) and (2.4) *Huschens-Broyden families* of B and A, respectively. Note that these families contain important members. For example, we call the cases of $\phi = 0$ and $\phi = 1$ *structured BFGS* and *structured DFP updates* in the sense of Huschens, respectively. By using the preceding families, Huschens proposed the following clever method.

Totally Structured Secant Method (TSSM):

Given $x \in R^n$, $A \in R^{n\times n}$ symmetric, solve

$$Bs = -\nabla f(x) \tag{2.5}$$

and set

$$x_+ = x + s, \tag{2.6}$$

and calculate $B^\sharp, A_+, y^\sharp, z$ and B_+ by (2.1), (2.4),

$$y^\sharp = (J(x_+) - J(x))^T \frac{r(x_+)}{\|r(x)\|}, \tag{2.7}$$

$$z = J(x_+)^T J(x_+) s + \|r(x_+)\| y^\sharp \tag{2.8}$$

and

$$B_+ = J(x_+)^T J(x_+) + \|r(x_+)\| A_+, \tag{2.9}$$

respectively.

His essential ideas are that the matrices A in B and $B^\sharp$ are multiplied by the sizing factors $\|r(x)\|$ and $\|r(x_+)\|$, respectively, and that $\|r(x)\|$ is used as the denominator of $y^\sharp$ instead of $\|r(x_+)\|$. These ideas lead to the self-sizing property of the method, and consequently enable us to possess the quadratic convergence property for zero residual problems. He dealt with the least change formulation of secant updates

$$B_+ = B^\sharp + \frac{(z - B^\sharp s)w^T + w(z - B^\sharp s)^T}{w^T s} - \frac{(z - B^\sharp s)^T s}{(w^T s)^2} w w^T, \tag{2.10}$$

where

$$w = z + \tau \sqrt{\frac{z^T s}{s^T B^\sharp s}}\, B^\sharp s, \quad \tau \in [0, 1].$$

For zero residual problems, he showed local and quadratic convergence of the method with a fixed τ in (2.10), which corresponds to the subclass in the convex class of (2.2). On the other hand, for nonzero residual problems, he proved local and superlinear convergence of the method with any ϕ in the convex class of the family (2.2). Then by using the fact that there exists a bijective mapping between $\tau \in [0, 1]$ in (2.10) and $\phi \in [0, 1]$ in (2.2), which was proved by Schnabel [11], he showed local and superlinear convergence of the method with any $\tau \in [0, 1]$ in the family (2.10).

In the following sections, we will deal with a wider class of the family than that of Huschens. Specifically, we only impose a boundedness condition on the parameter ϕ. We will directly deal with the family (2.2) for both cases, and show quadratic and superlinear convergence properties for zero and nonzero residual problems, respectively. We will use the way of proof similar to that by Stachurski [12] or by Yabe and Yamaki [15].

We note that the Huschens-Broyden family (2.2) can be rewritten by

$$B_+ = B_+^{DFP} + (\phi - 1)\Delta B, \tag{2.11}$$

where

$$B_+^{DFP} = B^\sharp + \frac{(z - B^\sharp s)z^T + z(z - B^\sharp s)^T}{s^T z} - \frac{s^T(z - B^\sharp s)}{(s^T z)^2} zz^T, \tag{2.12}$$

$$\Delta B = (s^T B^\sharp s)\left(\frac{z}{s^T z} - \frac{B^\sharp s}{s^T B^\sharp s}\right)\left(\frac{z}{s^T z} - \frac{B^\sharp s}{s^T B^\sharp s}\right)^T. \tag{2.13}$$

Notice that the matrix ΔB is the difference of the structured DFP and structured BFGS updates.

We define here the difference matrix

$$\Delta(a, b, X) = (a^T X a)\left(\frac{b}{a^T b} - \frac{Xa}{a^T X a}\right)\left(\frac{b}{a^T b} - \frac{Xa}{a^T X a}\right)^T, \tag{2.14}$$

so that

$$\Delta B = \Delta(s, z, B^\sharp). \tag{2.15}$$

If we note that from the definitions of z and $B^\sharp$

$$z - B^\sharp s = \|r(x_+)\|(y^\sharp - As), \tag{2.16}$$

then we also have

$$A_+ = A_+^{DFP} + \frac{\phi - 1}{\|r(x_+)\|}\Delta B, \tag{2.17}$$

where

$$A_+^{DFP} = A + \frac{(y^\sharp - As)z^T + z(y^\sharp - As)^T}{s^T z} - \frac{s^T(y^\sharp - As)}{(s^T z)^2} zz^T, \tag{2.18}$$

and ΔB is defined in (2.13).

The structured DFP updates B_+^{DFP} and A_+^{DFP} are connected with the relation

$$B_+^{DFP} = J(x_+)^T J(x_+) + \|r(x_+)\| A_+^{DFP}. \tag{2.19}$$

These forms (2.11) and (2.17) are useful ones for our analysis of convergence properties in the following sections.

3 Preliminaries

In this section, we give assumptions and useful lemmas to show local convergence properties. Let D be an open convex subset of R^n, which contains a local minimizer x_*. We assume the following standard conditions.

(A1) There exists a positive constant ξ such that

$$\begin{aligned} \|\nabla^2 f(x) - \nabla^2 f(x_*)\| &\leq \xi\|x - x_*\|, \\ \|C(x) - C(x')\| &\leq \xi\|x - x'\|, \\ \|J(x) - J(x')\| &\leq \xi\|x - x'\|, \\ \|r(x) - r(x')\| &\leq \xi\|x - x'\| \end{aligned}$$

for any x and x' in D.

(A2) $\nabla^2 f$ is symmetric positive definite at x_*, i.e., there exist positive constants ν_1 and ν_2 such that

$$\nu_1\|u\|^2 \leq u^T \nabla^2 f(x_*) u \leq \nu_2\|u\|^2 \qquad \text{for all } u \in R^n.$$

We mention here a few technical assumptions. We include for easy reference the conditions of the Lipschitz-continuity of $C(x)$ and $r(x)$ in (A1), though these are implied by the condition of $J(x)$. Furthermore, when we consider the sequence of iterates $\{x_k\}$, we always assume for convenience that finite convergence does not occur, i.e., $\nabla f(x_k) \neq 0$, which implies $x_k \neq x_*$ for all k.

It follows easily from assumption (A1) that, for $x \in D$,

$$\|\nabla f(x) - \nabla f(x_*) - \nabla^2 f(x_*)(x - x_*)\| \leq \frac{\xi}{2}\|x - x_*\|^2, \tag{3.1}$$

$$\|r(x) - r(x_*) - J(x_*)(x - x_*)\| \leq \frac{\xi}{2}\|x - x_*\|^2, \tag{3.2}$$

(see Lemma 4.1.12 in [4]) and

$$\|J(x)\| \leq \xi\|x - x_*\| + \|J(x_*)\|. \tag{3.3}$$

We define throughout the paper

$$D_\varepsilon = \{x \mid \|x - x_*\| < \varepsilon\}, \tag{3.4}$$

$$\sigma(x, x_+) = \max(\|x - x_*\|, \|x_+ - x_*\|), \tag{3.5}$$

$$M = \nabla^2 f(x_*)^{\frac{1}{2}}, \tag{3.6}$$

and set

$$\widehat{A} = M^{-1} A M^{-1}, \quad \widehat{B}^\sharp = M^{-1} B^\sharp M^{-1}, \quad \widehat{B}_+ = M^{-1} B_+ M^{-1}, \tag{3.7}$$

$$\widehat{y}^\sharp = M^{-1} y^\sharp, \quad \widehat{z} = M^{-1} z \quad \text{and} \quad \widehat{s} = M s. \tag{3.8}$$

Note that assumptions (A1) and (A2) ensure that x_* is an isolated local mimimizer, i.e., there exists some positive constant ε_* such that

$$\|r(x_*)\| < \|r(x)\| \quad \text{for all } x \in D_{\varepsilon_*} \subset D,\ x \neq x_*, \tag{3.9}$$

because D is an open subset of R^n. In what follows, we assume the choice of $\varepsilon \leq \varepsilon_*$ for ε_* defined above.

We note further that by the equivalence of norms, for any $n \times n$ matrix Q, there exists a positive constant η such that

$$\frac{1}{\eta}\|Q\|_{F,M} \leq \|Q\| \leq \eta\|Q\|_{F,M}. \tag{3.10}$$

Now we give the following two lemmas. These lemmas are elementary but useful in the analysis presented in Sections 4 and 5.

Lemma 1 *If two vectors a and b satisfy the following inequality*

$$\|b-a\| \leq \kappa\|a\| \tag{3.11}$$

for a nonnegative constant $\kappa < 1$, then it holds that

$$(1-\kappa)\|a\| \leq \|b\| \leq (1+\kappa)\|a\|, \tag{3.12}$$

$$(1-\kappa)\|a\|^2 \leq a^T b \leq (1+\kappa)\|a\|^2, \tag{3.13}$$

$$\frac{1}{1+\kappa}\|b\|^2 \leq a^T b \leq \frac{1}{1-\kappa}\|b\|^2. \tag{3.14}$$

Lemma 2 *If*

$$a^T X a > 0 \quad \textit{and} \quad 0 < \frac{\|a\|\|b\|}{a^T b} \leq \beta, \tag{3.15}$$

then

$$\|\Delta(a,b,X)\|_F \leq 4\beta^2\frac{\|b-Xa\|^2}{a^T X a}, \tag{3.16}$$

where $\Delta(a,b,X)$ is the difference matrix defined by (2.14).

4 Quadratic convergence for zero residual problems

To show local and quadratic convergence of TSSM, we begin by giving the following three lemmas.

Lemma 3 *Suppose that assumptions* (A1) *and* (A2) *hold in a zero residual case, i.e., $\|r(x_*)\| = 0$. Let B, s, x_+, z be given by (2.1), (2.5), (2.6), (2.8), and let $\widehat{z}$ and $\widehat{s}$ be defined by (3.8) . Assume that $\|A\| \leq \tau$ for some positive constant τ. Then there exist positive constants ε, K, $\zeta^\sharp$, $\widehat{\zeta}$ and β such that if $0 < \|x - x_*\| < \varepsilon$, then it holds that*

$$\|x_+ - x_*\| \leq K\|x - x_*\|^2, \tag{4.1}$$

$$\|y^\sharp\| \le \zeta^\sharp \sigma(x, x_+)\|s\|, \tag{4.2}$$

$$\|\hat{z} - \hat{s}\| \le \hat{\zeta}\sigma(x, x_+)\|\hat{s}\|, \tag{4.3}$$

$$\frac{1}{\beta}\|\hat{s}\| \le \|\hat{z}\| \le \beta\|\hat{s}\|, \tag{4.4}$$

$$\frac{1}{\beta}\max(\|\hat{s}\|^2, \|\hat{z}\|^2) \le \hat{s}^T\hat{z} \le \beta\min(\|\hat{s}\|^2, \|\hat{z}\|^2) \tag{4.5}$$

and

$$\frac{1}{\beta}\|\hat{s}\|\|\hat{z}\| \le \hat{s}^T\hat{z}. \tag{4.6}$$

The following lemma states the bounded deterioration property of the matrix A_+^{DFP} in (2.18).

Lemma 4 *Suppose that the assumptions of Lemma 3 hold and that $s \neq 0$. Then A_+^{DFP} given by (2.18) is well defined and*

$$\|A_+^{DFP}\|_{F,M} \le \|A\|_{F,M} + \omega_1\sigma(x, x_+),$$

where

$$\omega_1 = \|M^{-1}\|^2\beta\{(2+\beta)\zeta^\sharp + 4\tau(\beta+1)\hat{\zeta}\}.$$

The following lemma gives an estimate of the part $\Delta B/\|r(x_+)\|$ in (2.17).

Lemma 5 *Suppose that the assumptions of Lemma 3 hold and that $s \neq 0$, $\|r(x_+)\| > 0$. Then for ε sufficiently small,*

$$\frac{\|\Delta B\|_{F,M}}{\|r(x_+)\|} \le \omega_2\sigma(x, x_+),$$

where ω_2 is some positive constant.

Now we present the local and quadratic convergence theorem of TSSM.

Theorem 1 *Suppose that the standard assumptions* (A1) *and* (A2) *are satisfied in a zero residual case. Assume that there exists a positive constant ϕ' such that $|\phi_k| \le \phi'$. Let the sequence $\{x_k\}$ be generated by TSSM. Then there exist positive constants ε and δ such that for $x_0 \in R^n$ and symmetric $A_0 \in R^{n\times n}$ satisfying*

$$\|x_0 - x_*\| < \varepsilon \quad \textit{and} \quad \|A_0\|_{F,M} < \delta,$$

the sequence $\{x_k\}$ is well defined and converges quadratically to x_*.

5 Superlinear convergence for nonzero residual problems

This section is devoted to the study of the superlinear convergence property of TSSM for nonzero residual nonlinear least squares problems. For this purpose, we give the following two lemmas.

Lemma 6 *Suppose that assumptions* (A1) *and* (A2) *hold in a nonzero residual case, i.e.,* $\|r(x_*)\| > 0$. *Let* B, s, x_+, z *be given by (2.1), (2.5), (2.6), (2.8), and let* $\widehat{z}$ *and* $\widehat{s}$ *be defined by (3.8). Then there exist positive constants* ε, $\zeta^\sharp$, $\widehat{\zeta}$ *and* β *such that if* $\|x - x_*\| < \varepsilon$ *and* $\|x_+ - x_*\| < \varepsilon$, *then it holds that*

$$\left\| y^\sharp - \sum_{i=1}^{m} \frac{r_i(x_*)}{\|r(x_*)\|} \nabla^2 r_i(x_*) s \right\| \le \zeta^\sharp \sigma(x, x_+) \|s\|, \tag{5.1}$$

$$\|\widehat{z} - \widehat{s}\| \le \widehat{\zeta} \sigma(x, x_+) \|\widehat{s}\|, \tag{5.2}$$

$$\frac{1}{\beta} \|\widehat{s}\| \le \|\widehat{z}\| \le \beta \|\widehat{s}\|, \tag{5.3}$$

$$\frac{1}{\beta} \max(\|\widehat{s}\|^2, \|\widehat{z}\|^2) \le \widehat{s}^T \widehat{z} \le \beta \min(\|\widehat{s}\|^2, \|\widehat{z}\|^2) \tag{5.4}$$

and

$$\frac{1}{\beta} \|\widehat{s}\| \|\widehat{z}\| \le \widehat{s}^T \widehat{z}. \tag{5.5}$$

The following lemma implies the bounded deterioration property of the matrix B_+^{DFP} in (2.12).

Lemma 7 *Assume that, for some positive constants* $\delta^\sharp$ *and* $\tau^\sharp$,

$$0 < \|B^\sharp - \nabla^2 f(x_*)\|_{F,M} \le \delta^\sharp \qquad \text{and} \qquad \|\widehat{B}^\sharp\| \le \tau^\sharp.$$

Suppose that the assumptions of Lemma 6 hold. Then

$$\|B_+^{DFP} - \nabla^2 f(x_*)\|_{F,M} \le \|B^\sharp - \nabla^2 f(x_*)\|_{F,M} - \frac{\|\widehat{z} - \widehat{B}^\sharp \widehat{z}\|^2}{2\delta^\sharp \|\widehat{z}\|^2} + \omega \sigma(x, x_+),$$

where

$$\omega = \widehat{\zeta} \beta \Big(1 + 4\tau^\sharp(\beta + 1)\Big).$$

Finally we give local and superlinear convergence theorem of TSSM.

Theorem 2 *Suppose that the standard assumptions* (A1) *and* (A2) *are satisfied in a nonzero residual case. Assume that there exists a positive constant* ϕ' *such that* $|\phi_k| \le \phi'$. *Let the sequence* $\{x_k\}$ *be generated by TSSM. Then there exist positive constants* ε *and* δ *such that if*

$$\|x_0 - x_*\| < \varepsilon \qquad \text{and} \qquad \|B_0 - \nabla^2 f(x_*)\|_{F,M} < \delta,$$

the sequence $\{x_k\}$ *is well defined and converges superlinerly to* x_*.

References

[1] M. Al-Baali and R. Fletcher, "Variational methods for non-linear least squares," Journal of the Operational Research Society, vol.36, pp.405–421, 1985.

[2] M.C. Bartholomew-Biggs, "The estimation of the Hessian matrix in nonlinear least squares problems with non-zero residuals," Mathematical Programming, vol.12, pp.67–80, 1977.

[3] J.E. Dennis, Jr. and J.J. Moré, "A characterization of superlinear convergence and its application to quasi-Newton methods," Mathematics of Computation, vol.28, pp.549–560, 1974.

[4] J.E. Dennis, Jr. and R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, New Jersey, 1983.

[5] J.E. Dennis, Jr. and H.F. Walker, "Convergence theorems for least-change secant update methods," SIAM Journal on Numerical Analysis, vol.18, pp.949–987, 1981.

[6] J.E. Dennis, Jr., D.M. Gay and R.E. Welsch, "An adaptive nonlinear least-squares algorithm," ACM Transactions on Mathematical Software, vol.7, pp.348–368, 1981.

[7] J.E. Dennis, Jr., H.J. Martinez and R.A. Tapia, "Convergence theory for the structured BFGS secant method with an application to nonlinear least squares," Journal of Optimization Theory and Applications, vol.61, pp.161–178, 1989.

[8] J.R. Engels and H.J. Martinez, "Local and superlinear convergence for partially known quasi-Newton methods," SIAM Journal on Optimization, vol.1, pp.42–56, 1991.

[9] R. Fletcher and C. Xu, "Hybrid methods for nonlinear least squares," IMA Journal of Numerical Analysis, vol.7, pp.371–389, 1987.

[10] J. Huschens, "On the use of product structure in secant methods for nonlinear least squares problems," SIAM Journal on Optimization, vol.4, pp.108–129, 1994.

[11] R.B. Schnabel, "Analyzing and improving quasi-Newton methods for unconstrained optimization," Ph.D. thesis, Cornell University, Ithaca, NY, 1977.

[12] A. Stachurski, "Superlinear convergence of Broyden's bounded θ-class of methods," Mathematical Programming, vol.20, pp.196–212, 1981.

[13] H. Yabe and T. Takahashi, "Factorized quasi-Newton methods for nonlinear least squares problems," Mathematical Pogramming, vol.51, pp.75–100, 1991.

[14] H. Yabe and N. Yamaki, "Convergence of a factorized Broyden-like family for nonlinear least squares problems," SIAM Journal on Optimization, vol.5, pp.770–791, 1995.

[15] H. Yabe and N. Yamaki, "Local and superlinear convergence of structured quasi-Newton methods for nonlinear optimization," to appear in Journal of the Operations Research Society of Japan.