Journal of Computational Mathematics, Vol.11, No.3, 1993, 236–249.

ON THE CONVERGENCE OF THE FACTORIZATION UPDATE ALGORITHM*

Bai Zhong-zhi   Wang De-ren
(Shanghai University of Science and Technology, Shanghai, China)

Abstract

In this paper, we make a Kantorovich-type analysis for the sparse Johnson and Austria's algorithm given in [2], which is called factorization update algorithm. When the mapping is linear, it is shown that a modification of that algorithm leads to global and Q-superlinear convergence. Finally, we point out the modification is also of local and Q-superlinear convergence for nonlinear systems of equations and give its corresponding Kantorovich-type convergence result.

§1. Introduction

For a large sparse nonlinear system of equations

$$F(x) = 0, \qquad F : D \subset R^n \longrightarrow R^n, \tag{1.1}$$

Johnson and Austria gave a kind of direct secant updates employing matrix factorizations to get its solution and proved its local and Q-superlinear convergence property(see [1]). By changing the updates in matrix forms, the authors[3] set up its Kantorovich-type analysis. Since Johnson and Austria's algorithm does not maintain the sparse structures of the triangular factors, the authors proposed its modified version, obtained a factorization update algorithm which has the sparse transitivity property of triangular factors H and U, and proved that this algorithm is of local and Q-superlinear convergence (see [2]).

In this paper, in order to complete the convergence theory of the factorization update algorithm, we make a Kantorovich-type analysis paralleling the one given in [3]. When $F(x)$ is linear, the Kantorovich-type convergence theorem naturally leads to global and Q-superlinear convergence for the modified factorization update algorithm. Finally, we show that the modified algorithm is convergent locally and Q-superlinearly for solving the large sparse nonlinear system of equations (1.1).

In the remainder of this section, we restate the factorization update algorithm for reference.

First, we introduce the following notations given in [2]:

* Received October 8, 1991.

Set $P_n \subset R^n \times R^n, P_n = \{(i,j)|i \neq j, 1 \leq i,j \leq n\}$; for certain $P \subset P_n$, subspace $\mathcal{L} \subset R^n \times R^n, \mathcal{L} = \{L \in L(R^n)|L$ is a unit lower triangular matrix, $L_{ij} = 0$, for $(i,j) \in P$ and $i > j, 1 \leq i,j \leq n\}$; subspace $\mathcal{U} \subset R^n \times R^n$, $\mathcal{U} = \{U \in L(R^n)|U$ is an upper triangular matrix, $U_{ij} = 0$ for $(i,j) \in P$ and $i \leq j, 1 \leq i,j \leq n\}$; matrices $\Lambda_1 = (\lambda_{ij}^{(1)}), \Lambda_2 = (\lambda_{ij}^{(2)}) \in L(R^n)$,

$$\lambda_{ij}^{(1)} = \begin{cases} 1, & \text{if } U_{ij} \neq 0, \\ 0, & \text{otherwise,} \end{cases} \qquad \lambda_{ij}^{(2)} = \begin{cases} 1, & \text{if } H_{ij} \neq 0, \text{ and } i > j, \\ 0, & \text{otherwise} \end{cases}$$

and n-dimensional vectors

$$s(i)^T = (s(i)_1, s(i)_2, \cdots, s(i)_n),$$

$$y(i)^T = (y(i)_1, y(i)_2, \cdots, y(i)_n),$$

with

$$s(i)_j = \begin{cases} s_j, & \text{if } \lambda_{ij}^{(1)} = 1, \\ 0, & \text{otherwise,} \end{cases} \qquad y(i)_j = \begin{cases} y_j, & \text{if } \lambda_{ij}^{(2)} = 1, \\ 0, & \text{otherwise} \end{cases}$$

where s_j, y_j are the j-th components of vectors s, y, respectively.

Now, the factorization update algorithm is of the form

$$\begin{cases} x^{(k+1)} = x^{(k)} + s^{(k)}, & s^{(k)} = x^{(k+1)} - x^{(k)}, \\ U_k s^{(k)} = -H_k F(x^{(k)}), \end{cases} \tag{1.2}$$

where, given an initial value $x^{(0)}$ and initial approximation $H_0^{-1}U_0$ to $F'(x^{(0)})$, $H_0 \in \mathcal{L}, U_0 \in \mathcal{U}$, the matrices $\{H_k\}, \{U_k\}$ are generated by

$$\begin{cases} H_{k+1} = H_k - \sum_{i=1}^{n} \langle v_k(i), v_k(i) \rangle^+ e_i^T (H_k y_k - U_k s^{(k)}) e_i y_k(i)^T, \\ U_{k+1} = U_k + \sum_{i=1}^{n} \langle v_k(i), v_k(i) \rangle^+ e_i^T (H_k y_k - U_k s^{(k)}) e_i s^{(k)}(i)^T \end{cases} \tag{1.3}$$

and

$$\begin{cases} y_k = F(x^{(k+1)}) - F(x^{(k)}), \\ v_k(i)^T = (y_k(i)_1, \cdots, y_k(i)_{i-1}, s^{(k)}(i)_i, \cdots, s^{(k)}(i)_n), \end{cases} \tag{1.4}$$

where α^+ denotes $1/\alpha$ for $\alpha \neq 0$ or 0 for $\alpha = 0$, respectively, $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product, and e_i is the i-th unit vector in R^n.

In the following discussion, we always assume that $F : D \subset R^n \to R^n$ is continuously differentiable and F' satisfies the Lipschitz condition with the Lipschitz constant γ on an open convex set $D_0 \subset D$, and there exist $H(x) \in \mathcal{L}, U(x) \in \mathcal{U}$ such that $H(x)F'(x) = U(x)$ holds for all $x \in D_0$. In addition, we denote $\|\cdot\|$ as $\|\cdot\|_2$, and $\|\cdot\|_F$ as the Frobenius norm.

§2. Preliminaries

To simplify notations for a given iteration, no subscript will denote the current iterate, and the subscript "+" will indicate the next iterate.

The following results are indispensable for our discussion.

Lemma 2.1[2]. *Let triangular matrices $H \in \mathcal{L}$, $U \in \mathcal{U}$, and H_+, U_+ be defined by* (1.3). *Then for any triangular matrices $\hat{H} \in \mathcal{L}, \hat{U} \in \mathcal{U}$,*

$$\|H_+ - \hat{H} + U_+ - \hat{U}\|_F \leq \|H - \hat{H} + U - \hat{U}\|_F + \|V(\hat{H}y - \hat{U}s)\|, \tag{2.1}$$

where $V = \operatorname{diag}\left(\frac{1}{\|v(1)\|}, \cdots, \frac{1}{\|v(n)\|}\right), v(i) \neq 0,$ *for* $1 \leq i \leq n$.

Lemma 2.2. *There exist positive constants δ, ρ_1, ρ_2, such that when $\|x_+ - x\| < \delta$, the inequality*

$$\frac{1}{\rho_1} \leq \frac{\|s\|}{\|v(i)\|} \leq \rho_2, \qquad 1 \leq i \leq n, \tag{2.2}$$

holds, where $s = x_+ - x, v(i) \neq 0, 1 \leq i \leq n$.

Proof. The vector y can be expressed as

$$y = Y + F'(x)s, \qquad Y = F(x_+) - F(x) - F'(x)s.$$

With the Lipschitz condition and the mean value theorem, it is easy to estimate that

$$\|Y\| \leq \frac{1}{2}\gamma\|s\|^2.$$

By introducing sparse projection operators

$$\bar{S}_i = \operatorname{diag}(\lambda_{i1}^{(1)}, \lambda_{i2}^{(1)}, \cdots, \lambda_{in}^{(1)}),$$

$$\bar{Y}_i = \operatorname{diag}(\lambda_{i1}^{(2)}, \lambda_{i2}^{(2)}, \cdots, \lambda_{in}^{(2)}),$$

such that $\bar{S}_i s = s(i), \bar{Y}_i y = y(i)$, and matrix

$$G_i = [F'(x)^T e_1, \cdots, F'(x)^T e_{i-1}, e_i, \cdots, e_n]^T,$$

we can express the vector $v(i)$ as

$$v(i) = y(i) + s(i) = \bar{Y}_i Y + \bar{Y}_i G_i s. \tag{2.3}$$

Thus, we have the estimate

$$\|v(i)\| \leq \|\bar{Y}_i Y\| + \|\bar{Y}_i G_i\|\|s\| \leq \left(\frac{1}{2}\gamma\delta + \|\bar{Y}_i G_i\|\right)\|s\|, \tag{2.4}$$

where we made use of the inequality $\|\bar{Y}_i Y\| \leq \|Y\|$. Taking

$$\rho_1 = \max_{1 \leq i \leq n} \left(\frac{1}{2}\gamma\delta + \|\bar{Y}_i G_i\|\right),$$

it follows from (2.4) that

$$\|v(i)\| \leq \rho_1\|s\|.$$

This is the left inequality of (2.2).

On the other hand, from (2.3), we have

$$\bar{Y}_i G_i s = v(i) - \bar{Y}_i Y. \tag{2.5}$$

Since the generalized inverse of matrix $\bar{Y}_i G_i$ always exists, from (2.5) we have

$$\begin{aligned} \|s\| &\le \|(\bar{Y}_i G_i)^+\|\|v(i)\| + \|(\bar{Y}_i G_i)^+\|\|\bar{Y}_i Y\| \\ &\le \|(\bar{Y}_i G_i)^+\|\|v(i)\| + \frac{1}{2}\gamma\delta\|(\bar{Y}_i G_i)^+\|\|s\|. \end{aligned} \tag{2.6}$$

Let δ be so small that

$$1 - \frac{1}{2}\gamma\delta \max_{1\le i\le n} \|(\bar{Y}_i G_i)^+\| > 0.$$

Then from (2.6) we obtain

$$\|s\| \le \rho_2\|v(i)\|, \tag{2.7}$$

where

$$\rho_2 = \Big(\max_{1\le i\le n} \|(\bar{Y}_i G_i)^+\|\Big) \Big/ \Big(1 - \frac{1}{2}\gamma\delta \max_{1\le i\le n} \|(\bar{Y}_i G_i)^+\|\Big).$$

(2.7) is the right inequality of (2.2).

Lemma 2.3. *Suppose the condition of Lemma 2.1 is satisfied. Then there holds*

$$\|H_+ - H(x) + U_+ - U(x)\|_F \le \|H - H(x) + U - U(x)\|_F + \frac{1}{2}\gamma\rho_2\|H(x)\|_F\|s\|,$$

where ρ_2 is the same as in Lemma 2.2.

Proof. By using inequalities (2.1) and (2.2), this result is obvious.

Lemma 2.4[3]**.** *There exist positive constants c, d, such that*

a) $\|H(x) - H(y)\|_F \le c\|x - y\|$,

b) $\|U(x) - U(y)\|_F \le d\|x - y\|$,

c) $\|H(x) - H(y) + U(x) - U(y)\|_F \le \bar{\gamma}\|x - y\|$,

for all $x, y \in D_0$, where $\bar{\gamma} = c + d$.

The following result is the local and Q-superlinear convergence result for the factorization update algorithm. This result was proved in [2].

Lemma 2.5[2]**.** *Let $x^* \in D$ satisfy $F(x^*) = 0$. Suppose that there exist $H^* \in \mathcal{L}$, $U^* \in \mathcal{U}$ such that $F'(x^*) = (H^*)^{-1}U^*$, and for the initial approximations $x^{(0)} \in D_0, H_0 \in \mathcal{L}, U_0 \in \mathcal{U}$, there exist positive constants $\varepsilon, \delta_0, \eta$, such that*

$$\|x^{(0)} - x^*\| \le \varepsilon, \qquad \|H_0 - H^* + U_0 - U^*\|_F \le \delta_0,$$

$$\max\{\|H^*\|_F, \|(H^*)^{-1}\|_F, \|U^*\|_F, \|(U^*)^{-1}\|_F\} = \eta.$$

Then the sequence $\{x^{(k)}\}$ generated by the algorithm locally Q-superlinearly converges to x^* provided ε, δ_0 are suitably chosen, and $\|H_k\|_F, \|H_k^{-1}\|_F, \|U_k\|_F, \|U_k^{-1}\|_F$ are uniformly bounded with respect to k.

§3. Kantorovich-Type Analysis

The estimation in the following theorem shows the relationship of the error of $H_k^{-1}U_k$ as an approximation to $F'(x^{(k)})$ with the error of the initial approximations. Therefore, it is still important for our Kantorovich-type analysis.

Theorem 3.1. *Let the sequences $\{x^{(k)}\},\{H_k\}$ and $\{U_k\}$ be generated by the factorization update algorithm. For the initial approximations $x^{(0)}\in D_0, H_0\in\mathcal{L}, U_0\in\mathcal{U}$, if there exist positive constants æ, δ_0 such that*

$$\max\{\|H(x)^{-1}\|_F,\|H(x)\|_F,\|U(x)\|_F\}\le \text{æ}\,(\forall\, x\in D_0),$$

$$\|H(x^{(0)})-H_0+U(x^{(0)})-U_0\|_F\le\delta_0,$$

then as well as $\{x^{(j)}\}_{j=0}^k\subset D_0$, and æ $(\delta_0+\alpha_1\sum_{j=1}^k\|x^{(j)}-x^{(j-1)}\|)<1$, the inequality

$$\|H_k^{-1}U_k-F'(x^k)\|_F\le\sqrt{2}\frac{\text{æ}\,(1+\text{æ}^2)}{1-\text{æ}\left(\delta_0+\alpha_1\sum\limits_{j=1}^k\|x^{(j)}-x^{(j-1)}\|\right)}\left(\delta_0+\alpha_1\sum_{j=1}^k\|x^{(j)}-x^{(j-1)}\|\right)$$

holds, where $\alpha_1=\frac{1}{2}\gamma\rho_2\,\text{æ}+\bar\gamma$, and ρ_2 is the same as defined in Lemma 2.2.

Proof. By Lemma 2.3 we have

$$\|H_k-H(x^{(k)})+U_k-U(x^{(k)})\|_F\le\|H_{k-1}-H(x^{(k)})+U_{k-1}-U(x^{(k)})\|_F$$
$$+\frac{1}{2}\gamma\rho_2\,\text{æ}\,\|x^{(k)}-x^{(k-1)}\|,$$

and by Lemma 2.4, we also have

$$\|H_{k-1}-H(x^{(k)})+U_{k-1}-U(x^{(k)})\|_F\le\|H_{k-1}-H(x^{(k-1)})+U_{k-1}-U(x^{(k-1)})\|_F$$
$$+\bar\gamma\|x^{(k)}-x^{(k-1)}\|.$$

Thus, it holds that

$$\|H_k-H(x^{(k)})+U_k-U(x^{(k)})\|_F\le\|H_{k-1}-H(x^{(k-1)})+U_{k-1}-U(x^{(k-1)})\|_F$$
$$+\alpha_1\|x^{(k)}-x^{(k-1)}\|.$$

From this inequality, by regressing it successively, we get

$$\|H_k-H(x^{(k)})+U_k-U(x^{(k)})\|_F\le\delta_0+\alpha_1\sum_{j=1}^k\|x^{(j)}-x^{(j-1)}\|.$$

On the other hand, since

$$\|H_k-H(x^{(k)})\|_F\le\|H_k-H(x^{(k)})+U_k-U(x^{(k)})\|_F$$
$$\le\delta_0+\alpha_1\sum_{j=1}^k\|x^{(j)}-x^{(j-1)}\|,$$

by the Banach lemma, we know that H_k is nonsingular, and

$$\|H_k^{-1}\|_F \le \frac{æ}{1 - æ\,(\delta_0 + \alpha_1 \sum_{j=1}^{k} \|x^{(j)} - x^{(j-1)}\|)}.$$

Finally, we obtain

$$\begin{aligned}\|H_k^{-1}U_k - F'(x^{(k)})\|_F &\le \|H_k^{-1}\|_F \max\{1, \|H(x^{(k)})^{-1}\|_F \|U(x^{(k)})\|_F\} \\ &\quad \times (\|U_k - U(x^{(k)})\|_F + \|H_k - H(x^{(k)})\|_F) \\ &\le \sqrt{2}\|H_k^{-1}\|_F[1 + \|H(x^{(k)})^{-1}\|_F\|U(x^{(k)})\|_F]\|H_k - H(x^{(k)}) + U_k - U(x^{(k)})\|_F \\ &\le \sqrt{2}\frac{æ\,(1 + æ^2)}{1 - æ\,(\delta_0 + \alpha_1 \sum_{j=1}^{k} \|x^{(j)} - x^{(j-1)}\|)}\Big(\delta_0 + \alpha_1 \sum_{j=1}^{k} \|x^{(j)} - x^{(j-1)}\|\Big).\end{aligned}$$

Now, we give the Kantorovich-type theorem, which asserts the existence of a zero of F. The proof closely follows the techniques in [3].

Theorem 3.2. *Suppose that for given initial values $x^{(0)} \in D_0, H_0 \in \mathcal{L}, U_0 \in \mathcal{U}$, there exist positive constants $a_i (1 \le i \le 4)$, such that*

$$\max\{\|H(x)^{-1}\|_F, \|H(x)\|_F, \|U(x)\|_F\} \le a_1, \quad \|U_0^{-1}H_0\|_F \le a_2,$$

$$\|U_0^{-1}H_0F(x^{(0)})\| \le a_3, \quad \|H(x^{(0)}) - H_0 + U(x^{(0)}) - U_0\|_F \le a_4.$$

a) *Define*

$$h' = \frac{\gamma a_2 a_3 (1 - a_1 a_4)^2}{[1 - a_1 a_4 - \sqrt{2}a_1 a_2 a_4 (1 + a_1^2)]^2},$$

$$r'_\pm = \frac{1 - a_1 a_4[1 + \sqrt{2}a_2(1 + a_1^2)]}{\gamma a_2 (1 - a_1 a_4)}(1 \pm \sqrt{1 - 2h'}).$$

If $0 < a_1 a_4[1 + \sqrt{2}a_2(1 + a_1^2)] < 1, h' \le 1/2$ and $\bar{N}(x^{(0)}, r'_-) \subset D_0$, then F has a root $x^* \in \bar{N}(x^{(0)}, r'_-)$. If $h' < 1/2$, then x^* is unique in $N(x^{(0)}, r'_+) \cap D_0$. The sequence $\{\bar{x}^{(k)}\}$ generated by

$$\bar{x}^{(k+1)} = \bar{x}^{(k)} - U_0^{-1}H_0F(\bar{x}^{(k)}), \qquad k = 0, 1, 2, \cdots$$

converges to x^* from any $\bar{x}^{(0)} \in D_0 \cap N(x^{(0)}, r'_+)$.

b) *Define*

$$\gamma_0 = \max\Big\{\gamma, \frac{1}{2}\gamma\rho_2 a_1 + \bar{\gamma}\Big\}, \qquad p = 1 + 3\sqrt{2}a_2(1 + a_1^2),$$

$$R_0 = \frac{1 - a_1 a_4 p}{\gamma_0 a_1 p}, \qquad \delta = \frac{\sqrt{2}a_1(1 + a_1^2)}{1 - a_1(a_4 + \gamma_0 R_0)}a_4,$$

$$l = \max\Big\{\gamma_0, \frac{\sqrt{2}a_1(1 + a_1^2)}{1 - a_1(a_4 + \gamma_0 R_0)}\gamma_0\Big\}, \quad h = \frac{a_2 a_3 l}{(1 - 3a_2\delta)^2},$$

$$r = \frac{1 - 3a_2\delta}{3a_2 l}(1 - \sqrt{1 - 6h}).$$

If $3a_2\delta < 1, h \le 1/6$ and $\bar{N}(x^{(0)}, r) \subset D_0$, then the sequence $\{x^{(k)}\}$ generated by the algorithm is well defined in $\bar{N}(x^{(0)}, r)$ and converges to a solution $x^* \in \bar{N}(x^{(0)}, r)$ of $F(x) = 0$.

Proof. a) Define $G : D_0 \subset R^n \longrightarrow R^n$ by $G(x) = x - U_0^{-1} H_0 F(x)$; define $g : R^1 \longrightarrow R^1$, $g \in C[0, t'], t' \in [r'_-, r'_+]$ by $g(t) = t + a_2 f(t)$, where $f : R^1 \longrightarrow R^1$ is given by

$$f(t) = \frac{1}{2}\gamma t^2 + \Big(\frac{\sqrt{2}a_1(1 + a_1^2)a_4}{1 - a_1 a_4} - \frac{1}{a_2}\Big)t + \frac{a_3}{a_2}.$$

Note that f is convex and quadratic and that the hypotheses guarantee two positive real roots, $r'_{\pm}$.

By direct evaluation, we can get $\|G(x^{(0)}) - x^{(0)}\| \le g(0) - 0 \le a_3$. Because

$$\|H_0 - H(x^{(0)})\|_F \le \|H_0 - H(x^{(0)}) + U_0 - U(x^{(0)})\|_F \le a_4,$$

H_0^{-1} exists and $\|H_0^{-1}\|_F \le a_1/(1 - a_1 a_4)$, as

$$\|H_0^{-1}U_0 - F'(x^{(0)})\|_F \le \sqrt{2}\|H_0^{-1}\|_F[1 + \|H(x^{(0)})^{-1}\|_F\|U(x^{(0)})\|_F]\|H_0$$
$$- H(x^{(0)}) + U_0 - U(x^{(0)})\|_F \le \frac{\sqrt{2}a_1(1 + a_1^2)a_4}{1 - a_1 a_4}.$$

So

$$\|G'(x)\|_F = \|I - U_0^{-1}H_0 F'(x)\|_F \le a_2(\|H_0^{-1}U_0 - F'(x^{(0)})\|_F + \|F'(x^{(0)}) - F'(x)\|_F)$$
$$\le a_2\Big(\frac{\sqrt{2}a_1(1 + a_1^2)a_4}{1 - a_1 a_4} + \gamma\|x - x^{(0)}\|\Big)(t \in [\|x - x^{(0)}\|, t'])$$
$$\le a_2\Big(\frac{\sqrt{2}a_1(1 + a_1^2)a_4}{1 - a_1 a_4} + \gamma t\Big) = g'(t).$$

With the definition of f, we know that r'_- is the unique solution of f in $[0, t')$, and $g(t') - t' = a_2 f(t') \le 0$. From Kantorovich's majorizing principle it follows that x^* is the only fixed point of G in $N(x^{(0)}, r'_+) \cap D_0$, and the iterate sequence $\{\bar{x}^{(k)}\}$ converges to x^* from any $\bar{x}^{(0)} \in D_0 \cap N(x^{(0)}, r'_+)$.

b) Consider the scalar sequence $\{t_k\}$ defined by

$$\begin{cases} t_0 = 0, \qquad t_{k+1} = t_k + a_2 f(t_k), \qquad k = 0, 1, 2, \cdots, \\ f(t) = \dfrac{3}{2}lt^2 - \dfrac{1 - 3a_2\delta}{a_2}t + \dfrac{a_3}{a_2}. \end{cases}$$

The function f is a convex and quadratic with two real positive roots, the smaller of which is r. Since

$$
\begin{aligned}
t_{k+1} - t_k &= a_2\Big[\frac{t_k - t_{k-1}}{a_2} + \Big(3lt_{k-1} - \frac{1-3a_2\delta}{a_2}\Big)(t_k - t_{k-1}) + \frac{3}{2}l(t_k - t_{k-1})^2\Big] \\
&= 3a_2\Big[\frac{1}{2}l(t_k - t_{k-1})^2 + (lt_{k-1} + \delta)(t_k - t_{k-1})\Big] \\
&= 3a_2\Big[\frac{1}{2}l(t_k + t_{k-1}) + \delta\Big](t_k - t_{k-1}).
\end{aligned}
$$

Now, by induction, we know that t_k is strictly increasing. Also, as

$$
\begin{aligned}
r - t_{k+1} &= r - t_k + a_2[f(r) - f(t_k)] = \Big[1 + a_2\Big(3l\bar{\xi} - \frac{1-3a_2\delta}{a_2}\Big)\Big](r - t_k) \\
&= 3a_2(l\bar{\xi} + \delta)(r - t_k), \qquad t_k < \bar{\xi} < r
\end{aligned}
$$

it is obvious that, for all k, $t_k \leq r$ and $\lim\limits_{k\to\infty} t_k = r$. Now, we want to prove the following facts by induction:

$$
\begin{cases}
\|x^{(k+1)} - x^{(k)}\| \leq t_{k+1} - t_k, \\
x^{(k+1)} \in \bar{N}(x^{(0)}, r),
\end{cases}
\qquad \text{for all} \quad k.
$$

For $k = 0$ the facts are obviously true. Suppose that for all $k \leq m - 1$, the facts are correct. By Theorem 3.1 we have

$$
\begin{aligned}
\|H_i^{-1}U_i - F'(x^{(i)})\|_F &\leq \sqrt{2}\frac{a_1(1+a_1^2)}{1 - a_1[a_4 + \gamma_0\sum\limits_{j=1}^{i}(t_j - t_{j-1})]}\Big[a_4 + \gamma_0\sum_{j=1}^{i}(t_j - t_{j-1})\Big] \\
&\leq \sqrt{2}\frac{a_1(1+a_1^2)}{1 - a_1[a_4 + \gamma_0 R_0]}[a_4 + \gamma_0 R_0] = \delta + lR_0 = \frac{1}{3a_2}.
\end{aligned}
$$

From this, we get

$$
\begin{aligned}
\|H_i^{-1}U_i - H_0^{-1}U_0\|_F &\leq \|H_i^{-1}U_i - F'(x^{(i)})\|_F + \|F'(x^{(i)}) - F'(x^{(0)})\|_F \\
&+ \|F'(x^{(0)}) - H_0^{-1}U_0\|_F \leq \frac{1}{3a_2} + \gamma\|x^{(i)} - x^{(0)}\| + \frac{\sqrt{2}a_1(1+a_1^2)a_4}{1 - a_1a_4} \\
&\leq \frac{1}{3a_2} + \gamma_0 R_0 + \delta \leq \frac{2}{3a_2}.
\end{aligned}
$$

Using the Banach lemma again, we get that $H_i^{-1}U_i$ is nonsingular and $\|U_i^{-1}H_i\|_F \leq 3a_2$.

For $k = m$, because

$$\begin{aligned}
\|x^{(m+1)} - x^{(m)}\| &= \|U_m^{-1} H_m F(x^{(m)})\| \le 3a_2 \|F(x^{(m)})\| \\
&\le 3a_2[\|F(x^{(m)}) - F(x^{(m-1)}) - F'(x^{(m-1)})(x^{(m)} - x^{(m-1)})\| \\
&+ \|F'(x^{(m-1)}) - H_{m-1}^{-1} U_{m-1}\|_F \|x^{(m)} - x^{(m-1)}\| \\
&\le 3a_2[\frac{1}{2}\gamma\|x^{(m)} - x^{(m-1)}\| + \delta + l t_{m-1}]\|x^{(m)} - x^{(m-1)}\| \\
&\le 3a_2[\frac{1}{2}l(t_m - t_{m-1}) + l t_{m-1} + \delta](t_m - t_{m-1}) \\
&= t_{m+1} - t_m,
\end{aligned}$$

$$\|x^{(m+1)} - x^{(0)}\| \le \sum_{j=1}^{m+1} \|x^{(j)} - x^{(j-1)}\| \le \sum_{j=1}^{m+1} (t_j - t_{j-1}) = t_{m+1} < r.$$

These complete the induction.

These show that there exists $x^* \in \bar{N}(x^{(0)}, r)$ such that $\lim_{k\to\infty} x^{(k)} = x^*$, and

$$\begin{aligned}
\|F(x^{(k)})\| &\le [\|H_k^{-1}U_k - H_0^{-1}U_0\|_F + \|H_0^{-1}U_0\|_F]\|x^{(k+1)} - x^{(k)}\| \\
&\le (\frac{2}{3a_2} + a_2)\|x^{(k+1)} - x^{(k)}\| \longrightarrow 0, \quad k \to \infty.
\end{aligned}$$

Thus $F(x^*) = 0$.

In the conditions and the process of proving the theorem, we should note that the inequalities

$$a_1(a_4 + \gamma_0 R_0) < 1, \quad R_0 > 0, \quad r \le R_0 \text{ and } r \le r'_+ \left(h \le \frac{1}{6}\right)$$

can be derived by $3a_2\delta < 1$ and $lR_0 + \delta = 1/(3a_2)$.

§4. Convergence Analysis for the Linear System of Equations

If $F: R^n \longrightarrow R^n$ is affine and defined by

$$F(x) = Ax - b, \qquad A \in L(R^n) \text{ nonsingular and } b \in R^n, \tag{4.1}$$

with $F'(x) = A, H(x) \equiv H^*, U(x) \equiv U^*$, and the Lipschitz constants $\gamma, \bar{\gamma}$ are zeros, then the following result is direct from Theorem 3.2 and Lemma 2.5.

Theorem 4.1. *Let $F: R^n \longrightarrow R^n$ be given by (4.1) and consider the factorization update algorithm. For given initial approximations $x^{(0)} \in D_0, H_0 \in \mathcal{L}, U_0 \in \mathcal{U}$, there exist positive constants $a_i (1 \le i \le 4)$, such that*

$$\max\{\|(H^*)^{-1}\|_F, \|H^*\|_F, \|U^*\|_F\} \le a_1, \quad \|U_0^{-1}H_0\|_F \le a_2,$$

$$\|U_0^{-1}H_0F(x^{(0)})\| \le a_3, \quad \|H^* - H_0 + U^* - U_0\|_F \le a_4.$$

If $0 < a_1a_4[1 + 3\sqrt{2}a_2(1 + a_1^2)] < 1$, *then the factorization update algorithm converges Q-superlinearly to* $x^* = A^{-1}b$ *from any* $x^{(0)} \in R^n$.

Proof. If $s^{(k)} = 0$, then $F(x^{(k)}) = 0$ and $x^{(k)} = x^*$.

If $s^{(k)} \neq 0$, by using Theorem 3.1 we have

$$\begin{aligned}\|H_{k+1}^{-1}U_{k+1} - A\|_F &\leq \sqrt{2}\|H_{k+1}^{-1}\|_F[1 + \|(H^*)^{-1}\|_F\|U^*\|_F]\|H_{k+1} - H^* + U_{k+1} - U^*\|_F \\ &\leq \sqrt{2}\frac{a_1a_4(1 + a_1^2)}{1 - a_1a_4},\end{aligned}$$

as

$$\begin{aligned}\|H_{k+1}^{-1}U_{k+1} - H_0^{-1}U_0\|_F &\leq \|H_{k+1}^{-1}U_{k+1} - A\|_F + \|A - H_0^{-1}U_0\|_F \\ &\leq 2\sqrt{2}\frac{a_1a_4(1 + a_1^2)}{1 - a_1a_4}.\end{aligned}$$

We know that $H_{k+1}^{-1}U_{k+1}$ is nonsingular and

$$\|U_{k+1}^{-1}H_{k+1}\|_F \leq \frac{a_2(1 - a_1a_4)}{1 - a_1a_4[1 + 2\sqrt{2}a_2(1 + a_1^2)]}.$$

By using (1.2), it follows that for $k = 0, 1, 2, \cdots$,

$$\begin{aligned}\|s^{(k+1)}\| = \|x^{(k+2)} - x^{(k+1)}\| &\leq \|U_{k+1}^{-1}H_{k+1}\|_F\|F(x^{(k+1)}) - F(x^{(k)}) - H_k^{-1}U_ks^{(k)}\|_F \\ &\leq \|U_{k+1}^{-1}H_{k+1}\|_F\|A - H_k^{-1}U_k\|_F\|s^{(k)}\| \\ &\leq \frac{\sqrt{2}a_1(1 + a_1^2)a_2a_4}{1 - a_1a_4[1 + 2\sqrt{2}a_2(1 + a_1^2)]}\|s^{(k)}\|.\end{aligned}$$

since

$$0 < \frac{\sqrt{2}a_1(1 + a_1^2)a_2a_4}{1 - a_1a_4[1 + 2\sqrt{2}a_2(1 + a_1^2)]} < 1.$$

The above fact implies that $\sum_{k=0}^{\infty} \|s^{(k)}\| < \infty$, proving the local and linear convergence. Superlinear convergence follows just as in the proof of Theorem 3.2 in [2].

This theorem shows that unless the convergence of the algorithm is global for initial approximation $x^{(0)}$ and local for approximations H_0 and U_0, the matrices U_k generated by (1.3) may be singular. To avoid singularity in U_{k+1}, set

$$\begin{cases} H_{k+1} = H_k - \sum_{i=1}^{n} \theta_i^k \langle v_k(i), v_k(i)\rangle^+ e_i^T(H_ky_k - U_ks^{(k)})e_iy_k(i)^T, \\ U_{k+1} = U_k + \sum_{i=1}^{n} \theta_i^k \langle v_k(i), v_k(i)\rangle^+ e_i^T(H_ky_k - U_ks^{(k)})e_is^{(k)}(i)^T \end{cases} \tag{4.2}$$

where $\theta_i^k (1 \leq i \leq n)$ are chosen so that U_{k+1} is nonsingular. To be more precise, we form the following theorem:

Theorem 4.2. *For* $\{H_k\}$ *and* $\{U_k\}$ *defined by* (4.2) *and* $\sigma \in (0, 1)$, *there exist* $\theta_i^k (1 \leq i \leq n)$ *such that*

a) $H_{k+1} \in \mathcal{L}$ *is nonsingular,*

b) $U_{k+1} \in \mathcal{U}, |detU_{k+1}| \geq \sigma|\det U_k|$. (4.3)

Proof. The proof can be immediately obtained by defining

$$\beta_i^{(k)} = \frac{< v_k(i), v_k(i) >+ e_i^T(H_k y_k - U_k s^{(k)}) e_i^T s^{(k)}}{e_i^T U_k e_i},$$

and choosing

$$\theta_i^k = \begin{cases} 1, & \text{if } |1+\beta_i^{(k)}| \geq \sigma^{\frac{1}{n}}, \\ -\dfrac{1-\sigma^{\frac{1}{n}}}{\beta_i^{(k)}}, & \text{if } |1+\beta_i^{(k)}| < \sigma^{\frac{1}{n}}, \end{cases} \qquad 1 \leq i \leq n.$$

It is not too difficult to show that these choices of θ_i^k $(1 \leq i \leq n)$ provide numbers closest to unity and belonging to $(0,1)$, so that (4.3) is satisfied. In the rest of the paper, we will only assume that θ_i^k $(1 \leq i \leq n)$ are chosen to satisfy

$$U_{k+1} \in \mathcal{U} \quad \text{nonsingular}, \quad 0 < \hat{\theta} < \sup_k \max_i \theta_i^k \leq 1. \tag{4.4}$$

Theorem 4.3. *Suppose that* $\{H_k\} \subset \mathcal{L}, \{U_k\} \subset \mathcal{U}$ *are generated by* (4.2) *and* (4.4). *Then for any* $\hat{H} \in \mathcal{L}$, $\hat{U} \in \mathcal{U}$, *the following inequality holds:*

$$\begin{aligned} \|H_{k+1} - \hat{H} + U_{k+1} - \hat{U}\|_F^2 &\leq \|H_k - \hat{H} + U_k - \hat{U}\|_F^2 \\ &- \hat{\theta}\|V_k[(H_k y_k - U_k s^{(k)}) - (\hat{H} y_k - \hat{U} s^{(k)})]\|^2 + \|V_k(\hat{H} y_k - \hat{U} s^{(k)})\|^2. \end{aligned} \tag{4.5}$$

Proof. The proof, similar to those in [2] and [3], can be done by direct and technical estimations, so we omit it.

Theorem 4.4. *Let* $F : R^n \longrightarrow R^n$ *be given by* (4.1) *and consider the modified factorization algorithm as defined by* (1.2), (1.4), (4.2) *and* (4.4). *Suppose that there exist* $H^* \in \mathcal{L}, U^* \in \mathcal{U}$, *such that* $H^*A = U^*$. *Then the algorithm is globally and* Q*-superlinearly convergent to* $x^* = A^{-1}b$.

Proof. By Theorem 4.3, we easily obtain

$$\begin{aligned} \|H_{k+1} - H^* + U_{k+1} - U^*\|_F &\leq \|H_k - H^* + U_k - U^*\|_F \leq \cdots \\ &\leq \|H_0 - H^* + U_0 - U^*\|_F, \end{aligned}$$

$$\|H_k\|_F \leq \|H_k - H^*\|_F + \|H^*\|_F \leq \|H_0 - H^* + U_0 - U^*\|_F + \|H^*\|_F,$$

and

$$\|H_{k+1} - H^* + U_{k+1} - U^*\|_F^2 \leq \|H_k - H^* + U_k - U^*\|_F^2 - \hat{\theta}\frac{\|H_k y_k - U_k s^{(k)}\|^2}{\rho_1^2\|s^{(k)}\|^2}.$$

So

$$\hat{\theta}\frac{\|H_k y_k - U_k s^{(k)}\|^2}{\rho_1^2\|s^{(k)}\|^2} \leq \|H_k - H^* + U_k - U^*\|_F^2 - \|H_{k+1} - H^* + U_{k+1} - U^*\|_F^2.$$

Taking summation in this inequality for k from 0 to an integer m, we get

$$\hat{\theta}\sum_{k=0}^{m}\frac{\|H_k y_k - U_k s^{(k)}\|^2}{\rho_1^2\|s^{(k)}\|^2} \le \|H_0 - H^* + U_0 - U^*\|_F^2 - \|H_{m+1} - H^* + U_{m+1} - U^*\|_F^2$$

$$\le 2\|H_0 - H^* + U_0 - U^*\|_F^2.$$

Then

$$\lim_{k\to\infty}\frac{\|H_k y_k - U_k s^{(k)}\|^2}{\|s^{(k)}\|^2} = 0,$$

as

$$\lim_{k\to\infty}\frac{\|y_k - H_k^{-1}U_k s^{(k)}\|}{\|s^{(k)}\|} \le \lim_{k\to\infty}\|H_k^{-1}\|_F\frac{\|H_k y_k - U_k s^{(k)}\|}{\|s^{(k)}\|} = 0,$$

that is

$$\varepsilon_k = \frac{\|(A - H_k^{-1}U_k)s^{(k)}\|}{\|s^{(k)}\|} \longrightarrow 0, \quad k \longrightarrow \infty.$$

Since

$$(H_k^{-1}U_k - A)s^{(k)} = -F(x^{(k+1)}) = -A(x^{(k+1)} - x^*),$$

hence

$$\|x^{(k+1)} - x^*\| \le \|A^{-1}\|_F\|A(x^{(k+1)} - x^*)\| = \|A^{-1}\|_F\|(H_k^{-1}U_k - A)s^{(k)}\|$$

$$\le \|A^{-1}\|_F\varepsilon_k\|s^{(k)}\| \le \|A^{-1}\|_F\varepsilon_k(\|x^{(k+1)} - x^*\| + \|x^{(k)} - x^*\|).$$

Let k be sufficiently large such that $\|A^{-1}\|_F\varepsilon_k < 1$. Then

$$\|x^{(k+1)} - x^*\| \le \frac{\varepsilon_k\|A^{-1}\|_F}{1 - \varepsilon_k\|A^{-1}\|_F}\|x^{(k)} - x^*\|.$$

As

$$\lim_{k\to\infty}\frac{\varepsilon_k\|A^{-1}\|_F}{1 - \varepsilon_k\|A^{-1}\|_F} = 0,$$

we complete the proof of the theorem.

Furthermore, by taking δ in Lemma 2.5 so small that $\theta_i^k = 1$ $(1 \le i \le n)$ satisfies either (4.3) or (4.4), we also see that the modification of the factorization update algorithm defined by (1.2), (1.4), (4.2) and (4.4) is locally and Q-superlinearly convergent. However, even if $\theta_i^k = 1$ $(1 \le i \le n)$ is not chosen, we still have local and Q-superlinear convergence.

Theorem 4.5. *Let $F : R^n \to R^n$ satisfy the assumptions of Lemma 2.5. Then the modification method is locally and Q-superlinearly convergent.*

Proof. By substituting H^*, U^* for $\hat{H}, \hat{U}$ in (4.5) respectively, this result follows from a modification of the proof of Lemma 2.5 given in [2], so we omit it.

§5. Numerical Tests

In this section, three algorithms, simple Newton method(s-Newton), sparse LU method(s-LU) and the factorization update algorithm (FUA), are compared by use of two types of numerical examples. The initial triangular factors H_0 and U_0 are the triangular factors of $F'(x^{(0)})$. No pivoting is used in any of the methods. The important comparisons are in IT, NFEV and NFAC, where IT = number of iterations, NFEV = number of function evaluations, NFAC = number of matrix factorizations, and PROB = the problem's serial number.

Problems 1–11.

$$f_i(x) = (k_1 - k_2 x_i^2)x_i + 1 - k_3 \sum_{\substack{j=i-r_1 \\ j \neq i}}^{i+r_2} (x_j + x_j^2),$$

$$x_i = 0, \qquad \text{for} \quad i < 1 \quad \text{or} \quad i > n.$$

We choose the initial approximations $x_i^{(0)} = -1, 1 \leq i \leq n$, for Problems 1–11. We have done computational experiments by the s-Newton, the s-LU of Dennis and Marwil and the FUA on IBM-PC/XT. The computation terminates if iterations achieve $\|F\| \leq 2\times 10^{-6}$ for all algorithms. The numerical results are given in the follwing table.

							s-Newton	s-LU	FUA
							NFAC=1	NFAC=1	NFAC=1
PROB	n	r_1	r_2	k_1	k_2	k_3	IT=NFEV	IT=NFEV	IT=NFEV
1	5	3	2	2	1	0.5	16	17	14
2	5	2	2	2	1	0.5	17	22	14
3	5	2	2	1	1	0.5	12	26	11
4	5	3	3	1	1	0.5	10	49	10
5	8	2	2	2	1	0.5	18	19	15
6	8	3	2	2	1	0.5	17	19	13
7	12	2	4	1	1	1	23	39	18
8	12	2	2	1	1	0.5	12	47	11
9	15	2	4	1	1	1	26	33	24
10	15	2	2	1	1	0.5	12	44	12
11	20	2	2	2	1	1	15	23	15

Problems 12–19.

$$f_i(x) = (3 - k_1 x_i)x_i - x_{i-1} - 2x_{i+1} + 1, \qquad x_i = 0, \text{ for } i < 1 \text{ or } i > n.$$

We choose the initial approximations $x_i^{(0)} = -1, 1 \leq i \leq n$, for these problems. The computation terminates if iterations achieve $\|x^{(k+1)} - x^{(k)}\| < \varepsilon$. The numerical results are given in the following table.

				s-Newton	s-LU	FUA
				NFAC=1	NFAC=1	NFAC=1
PROB	n	k_1	ε	IT=NFEV	IT=NFEV	IT=NFEV
12	5	0.5	2E-10	18	31	13
13	5	0.5	2E-12	21	41	15
14	5	1.0	2E-10	20	33	15
15	5	1.0	2E-12	24	37	18
16	10	0.5	2E-10	17	35	13
17	10	0.5	2E-9	16	33	12
18	20	0.5	2E-10	17	35	14
19	20	0.5	2E-9	16	33	12

Obviously, from the above two tables we see that the FUA is superior to the other ones.

References

[1] G.W.Johnson and N.H.Austria, A quasi-Newton method employing direct secant updates of matrix factorizations, *SIAM J. Numer. Anal.*, **20** (1983), 315–325.

[2] Wang De-ren and Bai Zhong-zhi, A class of factorization update algorithm for solving systems of sparse nonlinear equations, 1989, to appear.

[3] Wang De-ren and Bai Zhong-zhi, Convergence analysis for a class of matrix factorization update Quasi-Newton methods, *Comm. on Appl. Math.*, **5** (1991), 50–60.

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

[5] E.Marwil, Convergence results for Schubert's method for solving spares nonlinear equations, *SIAM J. Numer. Anal.*, **16** (1979), 558–604.

[6] B.Lam, on the convergence of Quasi-Newton method for sparse nonlinear systems, *Math. Comp.*, **32** (1978), 447–451.