HIROSHIMA MATH. J.
16 (1986), 553–571

Componentwise error estimates for approximate solutions to systems of equations with the aid of Dahlquist constants

Hisayoshi SHINTANI

(Received September 20, 1985)

1. Introduction

For f: $D \subset R^n \to R^n$ consider the equation

$$x = f(x), \quad x \in D, \tag{1.1}$$

where D is a path-connected set. Let the l.u.b. Lipschitz constant of f over D be defined by

$$L(f) = \sup_{x, y \in D, x, \neq y} \|f(x) - f(y)\| / \|x - y\|, \tag{1.2}$$

where $\|\cdot\|$ is a given norm in R^n. Then it is known [1] that if $L(f) < 1$ and there exists an $x^{(0)} \in D$ such that

$$S = \{h;\ \|h - x^{(0)}\| \leqq L(f)\|x^{(1)} - x^{(0)}\|/(1 - L(f))\} \subset D, \tag{1.3}$$

then (1.1) has exactly one solution x^* in D and

$$\|x^{(1)} - x^*\| \leqq L(f)\,\|x^{(1)} - x^{(0)}\|/(1 - L(f)), \tag{1.4}$$

where $x^{(1)} = f(x^{(0)})$.

When $L(f)$ is finite, the Dahlquist constant of f over D (see [3]) is defined by

$$d(f) = \lim_{h \to +0} (L(I + hf) - 1)/h. \tag{1.5}$$

Söderlind [2] has shown that if x^* is a solution of (1.1), $x^{(0)}$, $x^{(1)} \in D$ and $d(f) < 1$, then

$$\|x^* - x^{(1)}\| \leqq L(f)\,\|x^{(1)} - x^{(0)}\|/(1 - d(f)), \tag{1.6}$$

where $x^{(1)} = f(x^{(0)})$. Since $d(f) \leqq L(f)$, the estimate (1.6) gives a smaller error bound than (1.4) when $d(f) < L(f)$ and especially when $d(f) < 0$.

For $x = (x_1, x_2, \dots, x_n)^T$ and $y = (y_1, y_2, \dots, y_n)^T \in C^n$, let $v(x) = (|x_1|, |x_2|, \dots, |x_n|)^T$ and write $x \geqq y$ if x_i and y_i are real and $x_i \geqq y_i$ $(i = 1, 2, \dots, n)$. Denote by $\sigma(A)$ the spectral radius of an $n \times n$ matrix A. For real $n \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ write $A \geqq B$ if $a_{ij} \geqq b_{ij}$ $(i, j = 1, 2, \dots, n)$. Then Urabe [5] has shown that if there exists a matrix $K \geqq 0$ such that

$$(1.7)\qquad v(f(x)-f(y)) \leqq Kv(x-y) \qquad \text{for} \quad x,\ y\in D,$$

$\sigma(K)<1$ and there exists an $x^{(0)}\in D$ such that

$$(1.8)\qquad S = \{h\colon v(h-x^{(1)}) \leqq (I-K)^{-1}Kv(x^{(0)}-x^{(1)})\} \subset D,$$

then (1.1) has exactly one solution x^* in D and

$$(1.9)\qquad v(x^{(1)}-x^*) \leqq (I-K)^{-1}Kv(x^{(0)}-x^{(1)}),$$

where $x^{(1)}=f(x^{(0)})$ and I is the identity matrix.

When the condition (1.7) is satisfied by $K\geqq 0$, there exists a real matrix M ($M\leqq K$) with nonnegative offdiagonal entries such that for any $\varepsilon>0$ a number $\delta>0$ can be chosen so that if $0<h<\delta$ then

$$(1.10)\qquad v[(I+hf)(x)-(I+hf)(y)] - v(x-y) \leqq h(M+\varepsilon I)v(x-y) \quad \text{for } x,\ y\in D.$$

This matrix M plays the role of $d(f)$ in componentwise error estimates just as K does that of $L(f)$ in (1.9). The first object of this paper is to show that, if there exists a matrix $K\geqq 0$ satisfying (1.7) and $\sigma(K)>1$, then

$$(1.11)\qquad v(x^{(1)}-x^*) \leqq (I-M)^{-1}Kv(x^{(1)}-x^{(0)}).$$

The estimate (1.11) gives a smaller error bound than (1.9) when $M\neq K$ and especially when M has negative diagonal entries.

A generalized Newton method for solving the equation

$$(1.12)\qquad g(x)=0, \quad x\in D$$

transforms (1.12) into (1.1) with $f(x)=x-H(x)g(x)$ and performs the iteration $x^{(k+1)}=f(x^{(k)})$ ($k=0, 1,\ldots$), where $H(x)$ is a matrix approximating $J(x)^{-1}$ and $J(x)$ is the Jacobian matrix of $g(x)$. Urabe [5] and Yamamoto [6, 7] have shown the existence and uniqueness of the solution of (1.12) by this method and obtained componentwise error estimates. The second object of this paper is to obtain error bounds of approximate solutions of (1.12) in terms of matrices corresponding to Dahlquist constants.

Although the results in this paper are stated and proved only for real functions, they are valid also for complex functions f and g. Finally the results are illustrated by numerical examples.

2. Preliminaries

Denote by $L(R^n)$ or $L(C^n)$ the set of all real or complex $n\times n$ matrices respectively and by $N(R^n)$ the set of all matrices $A\in L(R^n)$ with nonnegative offdiagonal entries. Denote by $\|x\|$ the l_p-norm ($1\leqq p\leqq\infty$) of $x\in C^n$ and by $\|\cdot\|_*$

the dual norm of $\|\cdot\|$. For $A=(a_{ij})\in L(C^n)$ let $\|A\|=\sup_{x\neq 0}\|Ax\|/\|x\|$ and

(2.1) $$v(A) = (|a_{ij}|), \quad \mu(A) = (c_{ij}),$$

where

(2.2) $$c_{ij} = \begin{cases} |a_{ij}| & (i\neq j) \\ \operatorname{Re} a_{ii} & (i=j). \end{cases}$$

Let

(2.3) $$e = (1, 1,\dots, 1)^T,$$

and for A, $B\in L(C^n)$ let

(2.4) $$c(A) = (\|a_1\|_*, \|a_2\|_*,\dots, \|a_n\|_*)^T,$$

(2.5) $$r(B) = (\|b_1\|, \|b_2\|,\dots, \|b_n\|),$$

where

$$A^T = (a_1, a_2,\dots, a_n), \quad B = (b_1, b_2,\dots, b_n).$$

Then we have $c(I)=e$ and $r(I)=e^T$. Notations $c_p(A)$ and $r_p(A)$ are also used to specify the l_p-norm.

Denote by $d(A)$ or $d_p(A)$ the Dahlquist constant of $A\in L(C^n)$ with respect to the l_p-norm. Then the following results are known (see [4]):

(2.6) $$|d(A)| \leqq \|A\|,$$

(2.7) $$d_1(A) = \max_j (\operatorname{Re} a_{jj} + \textstyle\sum_{i\neq j}|a_{ij}|),$$

(2.8) $$d_\infty(A) = \max_i (\operatorname{Re} a_{ii} + \textstyle\sum_{j\neq i}|a_{ij}|),$$

(2.9) $$d_2(A) = \text{maximal eigenvalue of } (A+A^*)/2,$$

where * stands for conjugation and transposition.

LEMMA 1. *For $A\in L(C^n)$ and $b\in C^n$*

(2.10) $$v(Ab) \leqq c(A)\|b\|.$$

PROOF. Let $A^T=(a_1, a_2,\dots, a_n)$. Then by Hölder's inequality we have

$$|a_i^T b| \leqq \|a_i\|_*\|b\| \quad (i=1, 2,\dots, n).$$

From this (2.10) follows.

COROLLARY. *For A and $B\in L(C^n)$*

(2.11) $$v(AB) \leqq c(A)r(B).$$

Let B: $C^n \times C^n \to C^n$ be a bilinear operator such that

(2.12) $$Bxy = [B_1x, B_2x, \dots, B_nx]y, \quad x, y \in C^n, \quad B_i \in L(C^n) \quad (i=1, 2, \dots, n).$$

Then

(2.13) $$v(Bxy) \leqq \|x\| \, \|y\| b(B),$$

where

(2.14) $$b(B) = c(\hat{B}), \quad \hat{B} = [c(B_1), c(B_2), \dots, c(B_n)],$$

because by Lemma 1

$$\begin{aligned} v(Bxy) &\leqq [\|x\| c(B_1), \|x\| c(B_2), \dots, \|x\| c(B_n)]v(y) \\ &\leqq \|x\| \hat{B} v(y) \leqq \|x\| \, \|y\| c(\hat{B}). \end{aligned}$$

We write $B \geqq 0$ if $B_i \geqq 0$ $(i=1, 2, \dots, n)$.

LEMMA 2. *For $A=(a_{ij}) \in L(R^n)$ let*

$$D = \mathrm{diag}(d_1, d_2, \dots, d_n), \quad S = \mu(A) + D, \quad Q = (I+D)^{-1}S,$$

where

$$d_i = \begin{cases} 0 & (a_{ii} \geqq 0) \\ -a_{ii} & (a_{ii} < 0) \end{cases} \quad (i=1, 2, \dots, n).$$

If $\sigma(Q)<1$, then $(I-\mu(A))^{-1}$ and $(I-A)^{-1}$ exist and

(2.15) $$v((I-A)^{-1}) \leqq (I-\mu(A))^{-1}.$$

PROOF. Let

$$P = (p_{ij}), \quad N = -D + P - A, \quad R = (I+D)^{-1}(P-N),$$

where

$$p_{ij} = \begin{cases} a_{ij} & (a_{ij} \geqq 0) \\ 0 & (a_{ij} < 0) \end{cases} \quad (i, j=1, 2, \dots, n).$$

Then $P \geqq 0$, $N \geqq 0$ and $S=N+P$. As $Q \geqq 0$ and $\sigma(Q)<1$, $(I-Q)^{-1}$ exists and $(I-Q)^{-1} \geqq 0$ (see [5]). We have $I-\mu(A)=(I+D)(I-Q)$ and $(I+D^{-1} \geqq 0$, so that $(I-\mu(A))^{-1}$ exists and is nonnegative. Since

$$\begin{aligned} \nu(I+R+R^2+\cdots) &\leqq I + \nu(R) + \nu(R)^2 + \cdots \\ &\leqq I + Q + Q^2 + \cdots \leqq (I-Q)^{-1}, \end{aligned}$$

$(I-R)^{-1}$ exists and $\nu((I-R)^{-1}) \leqq (I-Q)^{-1}$. Hence $(I-A)^{-1}$ exists and (2.15) follows, because $I-A=(I+D)(I-R)$.

COROLLARY 1. *If $\sigma(\nu(A))<1$, then $(I-\mu(A))^{-1}$ exists and*

$$0 \leqq (I-\mu(A))^{-1} \leqq (I-\nu(A))^{-1}. \tag{2.16}$$

PROOF. Since

$$0 \leqq (I+D)^{-1}S \leqq S \leqq S + D \leqq \nu(A)$$

and $\sigma(\nu(A))<1$, it follows that $\sigma(Q)<1$ and by Lemma 2 $(I-\mu(A))^{-1}$ exists and is nonnegative. Hence (2.16) is obtained because $I-\nu(A) \leqq I-\mu(A)$ and $(I-\nu(A))^{-1} \geqq 0$.

COROLLARY 2. *For $K \in L(R^n)$ and $M \in N(R^n)$ if $M \leqq K$, $K \geqq 0$ and $\sigma(K)<1$, then $(I-M)^{-1}$ exists and*

$$0 \leqq (I-M)^{-1} \leqq (I-K)^{-1}. \tag{2.17}$$

PROOF. Let D be defined as in the proof of Lemma 2 and put $S=M+D$. Then $S \geqq 0$ and $0 \leqq (I+D)^{-1}S=Q \leqq S \leqq K$. Since $\sigma(K)<1$, it follows that $\sigma(Q)<1$ and by Lemma 2 $(I-M)^{-1}$ exists and $(I-M)^{-1} \geqq 0$. We have (2.17) because $I-K \leqq I-M$ and $(I-K)^{-1} \geqq 0$.

LEMMA 3. *For $A \in L(C^n)$ if $d(A)<1$, then $(I-A)^{-1}$ exists and*

$$\|(I-A)^{-1}\| \leqq 1/(1-d(A)). \tag{2.18}$$

PROOF. From the definition of $d(A)$ for any $\varepsilon>0$ there exists a $\delta>0$ such that if $0<h<\delta$ then

$$\|(I+hA)x\| - \|x\| \leqq h(d(A)+\varepsilon)\|x\| \qquad \text{for} \quad x \in C^n.$$

Put $\delta_1=\delta/(1+\delta)$ and $k=h/(1-h)$ for $0<h<\delta_1$. Then for $0<h<\delta_1$

$$\begin{aligned} -h\|(A-I)x\| &= \|x\| - h\|(A-I)x\| - \|x\| \\ &\leqq \|x+h(A-I)x\| - \|x\| \\ &\leqq (1-h)[\|(I+kA)x\| - \|x\|] - h\|x\| \\ &\leqq h(d(A)+\varepsilon-1)\|x\|, \end{aligned}$$

so that

$$\|(I-A)x\| \geqq (1-d(A)-\varepsilon)\|x\| \quad \text{for any} \quad \varepsilon > 0.$$

This yeidls

$$\|(1-A)x\| \geqq (1-d(A))\|x\|.$$

Since $d(A)<1$, $(I-A)x=0$ implies $x=0$. Hence $(I-A)^{-1}$ exists and

$$\|y\| \geqq (1-d(A))\|(1-A)^{-1}y\| \qquad \text{for all} \quad y \in C^n,$$

from which (2.18) follows.

LEMMA 4. *For* $A \in L(C^n)$

$$v((I-A)x) \geqq (I-\mu(A))v(x) \qquad \text{for} \quad x \in C^n. \tag{2.19}$$

PROOF. Let $A=(a_{ij})$ and $b_i=\operatorname{Re} a_{ii}$ $(i=1, 2,\dots, n)$. Then

$$\begin{aligned} |(1-a_{ii})x_i - \textstyle\sum_{j\neq i}a_{ij}x_j| &\geqq |1-a_{ii}|\,|x_i| - |\textstyle\sum_{j\neq i}a_{ij}x_j| \\ &\geqq (1-b_i)|x_i| - \textstyle\sum_{j\neq i}|a_{ij}|\,|x_j| \quad (i=1, 2,\dots, n). \end{aligned}$$

From this (2.19) follows.

3. Error estimates

3.1. Nonlinear equations

Consider the equation (1.1) and introduce the following conditions.

CONDITION L: There exists a $K \in L(R^n)$ such that $K \geqq 0$ and

$$v(f(x)-f(y)) \leqq Kv(x-y) \qquad \text{for} \quad x, y \in D. \tag{3.1}$$

CONDITION D: There exists an $M \in N(R^n)$ such that for any $\varepsilon>0$ a number $\delta>0$ can be chosen so that

$$v[(I+hf)(x)-(I+hf)(y)] - v(x-y) \leqq h(M+\varepsilon I)v(x-y) \tag{3.2}$$
$$\text{for} \quad x, y \in D, \quad 0<h<\delta.$$

CONDITION I: For $M \in N(R^n)$ satisfying (3.2), $(I-M)^{-1}$ exists and $(I-M)^{-1} \geqq 0$.

Urabe [5] proved the following

THEOREM 1. *Suppose that Condition L is satisfied,* $\sigma(K)<1$ *and that there exists an* $x^{(0)} \in D$ *such that*

$$S = \{h\colon v(h-x^{(1)}) \leqq (I-K)^{-1}Kv(x^{(1)}-x^{(0)})\} \subset D,$$

where $x^{(1)}=f(x^{(0)})$. *Then a sequence* $\{x^{(k)}\}$ *in* S *is defined by*

$$x^{(k+1)} = f(x^{(k)}) \quad (k=0, 1,\dots), \tag{3.3}$$

it converges to a limit x^*, *which is the unique solution of* (1.1) *in* D, *and*

$$v(x^*-x^{(k)}) \leqq (I-K)^{-1}K^k v(x^{(1)}-x^{(0)}) \quad (k=1, 2,\dots). \tag{3.4}$$

COROLLARY. *Under the assumptions of the theorem*

$$v(x^*-x^{(1)}) \leqq (I-K)^{-1}Kv(x^{(1)}-x^{(0)}). \tag{3.5}$$

LEMMA 5. *If Condition L is satisfied, then Condition D is satisfied with* $M \leqq K$.

PROOF. *For* $h>0$ *we have by* (3.1)

$$\begin{aligned} &v[(I+hf)(x)-(I+hf)(y)] - v(x-y) \\ &\quad \leqq v(x-y) + hv(f(x)-f(y)) - v(x-y) \leqq hKv(x-y) \qquad \text{for} \quad x, y \in D. \end{aligned}$$

This proves the lemma.

LEMMA 6. *Suppose that Conditions D and I are satisfied. Then* $(I-f)^{-1}$ *exists on* $R=(I-f)(D)$ *and*

$$v[(I-f)^{-1}(u)-(I-f)^{-1}(v)] \leqq (I-M)^{-1}v(u-v) \qquad \textit{for} \quad u, v \in R. \tag{3.6}$$

PROOF. Put $\delta_1=\delta/(1+\delta)$ and $k=h/(1-h)$ for $0<h<\delta_1$. Then for $0<h<\delta_1$

$$\begin{aligned} &-hv[(f-I)(x)-(f-I)(y)] = v(x-y) - hv[(f-I)(x)-(f-I)(y)] - v(x-y) \\ &\quad \leqq v[(I+h(f-I))(x)-(I+h(f-I))(y)] - v(x-y) \\ &\quad \leqq (1-h)\,[v\{(I+kf)(x)-(I+kf)(y)\}-v(x-y)] - hv(x-y) \\ &\quad \leqq h(M+\varepsilon I-I)v(x-y). \end{aligned}$$

Hence

$$v[(I-f)(x)-(I-f)(y)] \geqq (I-M-\varepsilon I)v(x-y) \quad \text{for any} \quad \varepsilon > 0,$$

which yields

$$v[(I-f)(x)-(I-f)(y)] \geqq (I-M)v(x-y).$$

Since $(I-M)^{-1} \geqq 0$, we have

$$v(x-y) \leqq (I-M)^{-1}v[(I-f)(x)-(I-f)(y)] \qquad \text{for} \quad x, y \in D.$$

This implies that $I-f$: $D \to R$ is one-to-one. Hence $(I-f)^{-1}$ exists and (3.6) follows.

THEOREM 2. *Suppose that Conditions L and I are satisfied and that* (1.1) *has a solution* $x^* \in D$. *Then* x^* *is the unique solution of* (1.1) *in D. If* $x^{(0)} \in D$ *and* $x^{(1)} = f(x^{(0)}) \in D$, *then*

$$v(x^{(1)} - x^*) \leqq (I-M)^{-1} K v(x^{(1)} - x^{(0)}). \tag{3.7}$$

PROOF. Condition D is satisfied by Lemma 5 and the uniqueness of the solution to (1.1) follows from Lemma 6. Since

$$(I-f)(x^{(1)}) = f(x^{(0)}) - f(x^{(1)}), \quad (I-f)(x^*) = 0,$$

we have

$$x^{(1)} = (I-f)^{-1}[f(x^{(0)}) - f(x^{(1)})], \quad x^* = (I-f)^{-1}(0)$$

and (3.7) follows from (3.6).

COROLLARY 1. *Under the assumptions of Theorem* 1 *the estimate* (3.7) *holds.*

PROOF. Condition D is satisfied by Lemma 5. By Corollary 2 to Lemma 2 Condition I is satisfied. By Theorem 1 (3.1) has a solution $x^* \in D$. Hence by Theorem 2 (3.7) is valid. This completes the proof.

Under Condition L let $\hat{K} \in L(R^n)$ satisfy $K \geqq \hat{K} \geqq 0$ and

$$v(f(x) - f(x^{(1)})) \leqq \hat{K} v(x - x^{(1)}) \qquad \text{for} \quad x \in D \tag{3.8}$$

and let $\hat{M} \in N(R^n)$ satisfy $\hat{M} \leqq \hat{K}$ and the following condition: for any $\varepsilon > 0$ there exists $\delta > 0$ such that if $0 < h < \delta$ then

$$v[(I+hf)(x) - (I+hf)(x^{(1)})] - v(x - x^{(1)}) \leqq h(\hat{M} + \varepsilon I) v(x - x^{(1)}) \qquad \text{for} \quad x \in D. \tag{3.9}$$

Then we have

COROLLARY 2. *Under the assumptions of Theorem* 1

$$v(x^{(1)} - x^*) \leqq d_2 \leqq d_1 \leqq d, \tag{3.10}$$

where

$$\begin{aligned} d &= (I-K)^{-1} K v(x^{(1)} - x^{(0)}), \quad d_1 = (I-\hat{K})^{-1} \hat{K} v(x^{(1)} - x^{(0)}), \\ d_2 &= (I-\hat{M})^{-1} \hat{K} v(x^{(1)} - x^{(0)}). \end{aligned} \tag{3.11}$$

PROOF. Since $0 \leqq \hat{K} \leqq K$ and $\sigma(K) < 1$, we have $\sigma(\hat{K}) < 1$ and by Corollary 2 to Lemma 2

$$0 \leqq (I-\hat{M})^{-1} \leqq (I-\hat{K})^{-1} \leqq (I-K)^{-1}.$$

From (3.9) it follows that

$$v[(I-f)(x^*)-(I-f)(x^{(1)})] \geqq (I-\hat{M})v(x^*-x^{(1)}).$$

Hence

$$v(x^*-x^{(1)}) \leqq (I-M)^{-1}v(f(x^{(1)})-f(x^{(0)})) \leqq d_2$$

and (3.10) follows.

Let $\tilde{K} \geqq 0$ and $\tilde{M}$ ($\tilde{M} \leqq \tilde{K}$) satisfy (3.8) and (3.9) respectively with $x^{(1)}$ replaced by $x^{(0)}$. Then in the same manner we can show the following

THEOREM 3. *Suppose that Condition L is satisfied,* $\sigma(K)<1$ *and* $T=\{h: v(h-x^{(0)}) \leqq (I-K)^{-1}v(x^{(1)}-x^{(0)})\} \subset D$. *Then* (1.1) *has a unique solution* x^* *in* D *and*

$$v(x^{(0)}-x^*) \leqq (I-\tilde{M})^{-1}v(x^{(1)}-x^{(0)}) \leqq (I-\tilde{K})^{-1}v(x^{(1)}-x^{(0)}).$$

3.2. Generalized Newton method

For $g(x)$: $D \subset R^n \to R^n$ consider the equation

$$g(x) = 0, \quad x \in D \tag{3.12}$$

and a generalized Newton method which transforms (3.12) into (1.1) with

$$f(x) = x - Hg(x) \tag{3.13}$$

and performs the iteration (3.3) with $x^{(0)} \in D$, where H is a matrix approximating $J(x^{(0)})^{-1}$ and $J(x)$ is the Jacobian matrix of $g(x)$. We introduce the following

CONDITION G: 1°. $g(x)$ is continuously differentiable on a convex set D.
2°. There exists a symmetric bilinear operator $B \geqq 0$ such that

$$v[H(J(x)-J(x^{(0)}))] \leqq Bv(x-x^{(0)}) \qquad \text{for} \quad x \in D. \tag{3.14}$$

Let

$$F = I - HJ(x^{(0)}), \quad K = v(F), \quad x^{(1)} = f(x^{(0)}), \quad a = v(x^{(1)}-x^{(0)}), \tag{3.15}$$
$$w = v(F(x^{(1)}-x^{(0)})),$$

$$L = K + Ba, \quad b = b(B), \quad c = w + Baa/2, \quad t = (1-\|L\|)^2 - 2\|b\|\,\|c\|. \tag{3.16}$$

Then we have

THEOREM 4. *Suppose that Condition G is satisfied and that*

$$\|L\| < 1, \quad t \geqq 0, \quad S = \{h \colon v(h - x^{(1)}) \leqq \beta\} \subset D, \tag{3.17}$$

where

$$\beta = (I-L)^{-1}(c + \alpha^2 b/2), \quad \alpha = 2\|c\|/(1 - \|L\| + \sqrt{t}). \tag{3.18}$$

Then (3.12) *has a solution* x^* *in* S.

PROOF. We have

$$\begin{aligned} Hg(x) - Hg(y) &= HJ(x^{(0)})(x-y) \\ &\quad + \int_0^1 H[J(y+t(x-y)) - J(x^{(0)})](x-y)dt. \end{aligned} \tag{3.19}$$

Hence

$$v(f(x)-f(y)) \leqq Lv(x-y) + Bv(y-x^{(1)})v(x-y) + Bv(x-y)v(x-y)/2. \tag{3.20}$$

Since α satisfies the equation $\alpha = \|L\|\alpha + \|b\|\alpha^2/2 + \|c\|$, it follows that

$$\|\beta\| \leqq (\|c\| + \alpha^2\|b\|/2)/(1 - \|L\|) \leqq \alpha \tag{3.21}$$

$$\beta = c + \alpha^2 b/2 + L\beta. \tag{3.22}$$

Define $\{u^{(k)}\}$ by

$$u^{(k+1)} = Lu^{(k)} + Bu^{(k)}u^{(k)}/2 + c \quad (k=1, 2,\ldots) \tag{3.23}$$

with $u^{(1)}=0$. Then for $k=1, 2,\ldots$

$$0 \leqq u^{(k)} \leqq u^{(k+1)} \leqq \beta. \tag{3.24}$$

In fact (3.24) holds for $k=1$ because $0=u^{(1)} \leqq c = u^{(2)} \leqq \beta$. Assume that (3.24) is valid for $k=1, 2,\ldots, m-1$. Then

$$\begin{aligned} u^{(m+1)} - u^{(m)} &= L(u^{(m)} - u^{(m-1)}) + Bu^{(m)}(u^{(m)} - u^{(m-1)})/2 \\ &\quad + B(u^{(m)} - u^{(m-1)})u^{(m-1)}/2 \geqq 0 \end{aligned}$$

and by (3.21) and (3.22)

$$u^{(m+1)} \leqq Lu^{(m)} + \|u^{(m)}\|^2 b/2 + c \leqq L\beta + \alpha^2 b/2 + c = \beta.$$

Hence (3.24) holds for $k=m$. This completes the induction.

From (3.24) it follows that there exists a u^* such that $u^{(k)} \to u^*$ $(k \to \infty)$ and $u^* \leqq \beta$. This u^* satisfies the equation

$$u = Lu + Buu/2 + c. \tag{3.25}$$

We shall show by induction that the sequence $\{x^{(k)}\}$ is defined by (3.3) and for $k=1, 2, \ldots$

(3.26) $$v(x^{(k+1)}-x^{(k)}) \leqq u^{(k+1)} - u^{(k)}, \quad x^{(k+1)} \in S.$$

Since

$$v(x^{(2)}-x^{(1)}) \leqq c \leqq u^{(2)} - u^{(1)}, \quad x^{(2)} \in S,$$

(3.26) holds for $k=1$. Assume that (3.26) is valid for $k=1, 2, \ldots, m-1$. Then

$$v(x^{(m)}-x^{(1)}) \leqq u^{(m)} - u^{(1)} \leqq u^{(m)}.$$

By (3.20)

$$\begin{aligned} v(x^{(m+1)}-x^{(m)}) &\leqq L(u^{(m)}-u^{(m-1)}) + B(u^{(m)}-u^{(m-1)})(u^{(m)}-u^{(m-1)})/2 \\ &\quad + Bu^{(m-1)}(u^{(m)}-u^{(m-1)}) \\ &\leqq u^{(m+1)} - u^{(m)}. \end{aligned}$$

Hence

(3.27) $$v(x^{(m+1)}-x^{(1)}) \leqq u^{(m+1)} - u^{(1)} \leqq u^{(m+1)} \leqq \beta$$

and (3.26) holds for $k=m$. This completes the induction.

From (3.26) it follows that $\{x^{(k)}\}$ is a Cauchy sequence. Hence there exists an x^* such that $x^{(k)} \to x^*$ $(k \to \infty)$ and $v(x^{(1)}-x^*) \leqq \beta$ by (3.27). From (3.3) and (3.13) we have $Hg(x^*)=0$ by continuity of $g(x)$. Since $\|F\| \leqq \|K\| \leqq \|L\| < 1$, $I-F=HJ(x^{(0)})$ is invertible, so that H^{-1} exists and we have $g(x^*)=0$.

COROLLARY 1. *Under the assumptions of the theorem x^* is the unique solution of* (3.12) *in* $S_1=\{h\colon v(h-x^{(1)}) \leqq u^*\}$, *where u^* is the unique solution of* (3.25) *obtained by the iteration* (3.23) *with* $u^{(1)}=0$.

PROOF. Let $\tilde{x}$ be any solution of (3.12) in S_1. We shall show by induction that for $k=1, 2, \ldots.$

(3.28) $$v(\tilde{x}-x^{(k)}) \leqq u^* - u^{(k)}.$$

For $k=1$ (3.28) is valid because $v(\tilde{x}-x^{(1)}) \leqq u^* \leqq u^*-u^{(1)}$. Put $v(\tilde{x}-x^{(j)})=w^{(j)}$ $(j=1, 2, \ldots)$ and assume that (3.28) holds for $k=1, 2, \ldots, m$. Then by (3.20)

$$\begin{aligned} w^{(m+1)} &\leqq Lw^{(m)} + Bw^{(m)}w^{(m)}/2 + Bv(x^{(m)}-x^{(1)})w^{(m)} \\ &\leqq L(u^*-u^{(m)}) + B(u^*-u^{(m)})(u^*-u^{(m)})/2 + Bu^{(m)}(u^*-u^{(m)}) \\ &\leqq u^* - u^{(m+1)}. \end{aligned}$$

Hence (3.28) is valid also for $k=m+1$ and this completes the induction.

Letting $k\to\infty$ in (3.28), we have $\nu(\tilde{x}-x^*)=0$, which implies $\tilde{x}=x^*$ and the proof is completed.

COROLLARY 2. *Under the assumptions of the theorem let* $\{\beta^{(k)}\}$ *be defined by*

$$\beta^{(k+1)} = c + L\beta^{(k)} + B\beta^{(k)}\beta^{(k)}/2 \quad (k=0, 1, 2,\dots) \tag{3.29}$$

with $\beta^{(0)}=\beta$. *Then*

$$\beta^{(k)} \geqq \beta^{(k+1)} \geqq u^* \quad (k=0, 1, 2,\dots), \tag{3.30}$$

so that there exists a β^* *such that* $\beta^{(k)}\to\beta^*$ $(k\to\infty)$ *and*

$$\nu(x^{(1)}-x^*) \leqq \beta^{(k)} \quad (k=0, 1, 2,\dots). \tag{3.31}$$

PROOF. It can be shown by induction that

$$\beta^{(k)} \geqq \beta^{(k+1)}, \quad \beta^{(k+1)} \geqq u^{(k+1)} \quad (k=0, 1, 2,\dots).$$

Hence there exists a β^* such that $\beta^{(k)}\to\beta^*$ $(k\to\infty)$ and $u^*\leqq\beta^*$. From (3.27) it follows that $\nu(x^{(1)}-x^*)\leqq u^*$ and (3.31) holds.

REMARK 1. If $t>0$ in (3.17), then u^* is the unique solution of (3.25) satisfying $0\leqq u^*\leqq\beta$. In fact let $\tilde{u}$ be such a solution and put $v=u^*-\tilde{u}$. Then v satisfies the equation

$$v = (I-L)^{-1}[B(u^*+\tilde{u})/2]v = Gv.$$

Since

$$\|G\| \leqq \|b\|\alpha/(1-\|L\|) \leqq (1-\|L\|-\sqrt{t})/(1-\|L\|) < 1,$$

it follows that $v=0$, so that $u^*=\tilde{u}$. Thus if $t>0$, then we have $u^*=\beta^*$ in Corollary 2.

REMARK 2. Let $\{\delta^{(k)}\}$ be defined by

$$\delta^{(k+1)} = (I-L)^{-1}(c+B\delta^{(k)}\delta^{(k)}/2) \quad (k=0, 1, 2,\dots) \tag{3.32}$$

with $\delta^{(0)}=\beta$. Then it can be shown that

$$\nu(x^{(1)}-x^*) \leqq u^* \leqq \delta^{(k)} \leqq \beta^{(k)} \quad (k=0, 1, 2,\dots).$$

Let

$$M = \mu(F), \quad L_1 = M + Ba, \quad t_1 = (1-d(L_1))^2 - 2\|b\|\,\|c\|. \tag{3.33}$$

Then we have

THEOREM 5. *Under the assumptions of Theorem* 4 *with* $t>0$ *in* (3.17) *let*

$\{\gamma^{(k)}\}$ be defined by

(3.34) $$\gamma^{(k+1)} = (I-L_1)^{-1}(c+B\gamma^{(k)}\gamma^{(k)}/2) \quad (k=0, 1, 2,\dots)$$

with $\gamma^{(0)}=\gamma$, where

$$\gamma = (I-L_1)^{-1}(c+\alpha_1^2 b/2), \quad \alpha_1 = 2\|c\|/(1-d(L_1)+\sqrt{t_1}).$$

Then

$$\gamma^{(k)} \geqq \gamma^{(k+1)}, \quad \beta^{(k)} \geqq \gamma^{(k)} \quad (k=0, 1, 2,\dots),$$

so that there exists a γ^* such that $\gamma^{(k)}\to\gamma^*$ $(k\to\infty)$ and

(3.35) $$\nu(x^{(1)}-x^*) \leqq \gamma^{(k)} \leqq \beta^{(k)} \quad (k=0, 1, 2,\dots).$$

PROOF. From

$$x^* - x^{(1)} = x^* - x^{(0)} - H[g(x^*)-g(x^{(0)})]$$

it follows that

$$(I-F)(x^*-x^{(1)}) = F(x^{(1)}-x^{(0)}) - \int_0^1 H[J(x^{(0)}+t(x^*-x^{(0)}))-J(x^{(0)})](x^*-x^{(0)})dt.$$

Put $r=\nu(x^*-x^{(1)})$. Then by Lemma 4 we have

$$(I-M)r \leqq c + Bar + Baa/2.$$

By (2.6)

$$d(L_1) \leqq \|L_1\| \leqq \|L\| < 1,$$

so that $(I-L_1)^{-1}$ exists and by Corollary 2 to Lemma 2

$$0 \leqq (I-L_1)^{-1} \leqq (I-L)^{-1}.$$

Hence $t_1 \geqq t$ and $\alpha_1 \leqq \alpha$, so that $\gamma \leqq \beta$.

Define $\{v^{(k)}\}$ and $\{w^{(k)}\}$ by

$$v^{(k+1)} = (I-L_1)^{-1}(c+Bv^{(k)}v^{(k)}/2), \quad w^{(k+1)} = (I-L)^{-1}(c+Bw^{(k)}w^{(k)}/2) \quad (k=1, 2,\dots)$$

with $v^{(1)}=w^{(1)}=0$. Then it can be shown by induction that

$$v^{(k)} \leqq v^{(k+1)} \leqq \gamma, \quad w^{(k)} \leqq w^{(k+1)} \leqq \beta, \quad v^{(k+1)} \leqq w^{(k+1)} \quad (k=1, 2,\dots).$$

Hence there exist v^* and w^* such that

$$v^{(k)} \longrightarrow v^*, \quad w^{(k)} \longrightarrow w^* \ (k\to\infty), \quad v^* \leqq \gamma, \quad w^* \leqq \beta, \quad v^* \leqq w^*.$$

Since w^* is a solution of (3.25) satisfying $0 \leqq w^* \leqq \beta$, by Remark 1 we have $w^* = u^*$, so that $r \leqq v^* \leqq u^* = \beta^*$.

Let $\{\delta^{(k)}\}$ be defined by (3.32) with $\delta^{(0)} = \beta$. Then it can be shown by induction that

$$v^{(k)} \leqq \gamma^{(k+1)} \leqq \gamma^{(k)} \leqq \delta^{(k)}, \quad \delta^{(k+1)} \leqq \delta^{(k)} \leqq \beta^{(k)} \quad (k=1, 2,\dots).$$

Hence there exist γ^* and δ^* such that

$$\gamma^{(k)} \longrightarrow \gamma^*, \quad \delta^{(k)} \longrightarrow \delta^* \ (k\to\infty), \quad v^* \leqq \gamma^* \leqq \delta^*, \quad \delta^* \leqq \beta^*.$$

By Remark 1 we have $\delta^* = \beta^*$, so that $v^* \leqq \gamma^* \leqq \beta^*$. This completes the proof.

Let

$$s = (1-\|K\|)^2 - 2\|b\|\,\|a\|, \quad s_1 = (1-d(M))^2 - 2\|b\|\,\|a\|. \tag{3.36}$$

Then the following theorem can be proved in the same manner.

THEOREM 6. *Suppose that Condition G is satisfied and that*

$$\|K\| < 1, \quad s > 0, \quad T = \{h\colon v(h-x^{(0)}) \leqq \delta\} \subset D, \tag{3.37}$$

where

$$\delta = (I-K)^{-1}(a+\gamma^2 b/2), \quad \gamma = 2\|a\|/(1-\|K\|+\sqrt{s}). \tag{3.38}$$

Then (3.12) *has a solution* x^* *in* T.

Let $\{\varepsilon^{(k)}\}$ *be defined by*

$$\varepsilon^{(k+1)} = (I-M)^{-1}(a+B\varepsilon^{(k)}\varepsilon^{(k)}/2) \quad (k=0, 1,\dots) \tag{3.39}$$

with $\varepsilon^{(0)} = \varepsilon$, *where*

$$\varepsilon = (I-M)^{-1}(a+\gamma_1^2 b/2), \quad \gamma_1 = 2\|a\|/(1-d(M)+\sqrt{s_1}).$$

Then

$$v(x^{(0)}-x^*) \leqq \varepsilon^{(k+1)} \leqq \varepsilon^{(k)} \quad (k=0, 1,\dots). \tag{3.40}$$

3.3. Systems of linear algebraic equations

THEOREM 7. *For a nonsingular* $A \in L(C^n)$ *let* T *be a matrix approximating* A^{-1} *and put*

$$R = I - AT, \quad E = A^{-1} - T, \quad a = d(R). \tag{3.41}$$

If $a < 1$, *then*

$$(3.42)\qquad v(E) \leqq c(T)r(R)/(1-a),$$

$$(3.43)\qquad v(E) \leqq v(TR) + c(TR)r(R)/(1-a),$$

$$(3.44)\qquad v(E) \leqq v(T(I+R)R) + c(TR^2)r(R)/(1-a).$$

PROOF. By Lemma 3 $\|(I-R)^{-1}\| \leqq 1/1(-a)$. Since

$$E = T(1-R)^{-1}R = TR + TR(I-R)^{-1}R$$
$$= T(I+R)R + TR^2(I-R)^{-1}R,$$

by Corollary to Lemma 1 estimates (3.42), (3.43) and (3.44) are obtained.

Besides (3.24), (3.43) and (3.44) we have

$$(3.45)\qquad v(E) \leqq c(TR)e^T/(1-a),$$

$$(3.46)\qquad v(E) \leqq v(TR) + c(T)r(R^2)/(1-a),$$

$$(3.47)\qquad v(E) \leqq v(T(I+R)R) + c(TR)r(R^2)/(1-a)$$

and so on.

THEOREM 8. *For $b \in C^n$ and a nonsingular $A \in L(C^n)$, let x^* be the solution of the equation $Ax=b$, $\tilde{x}$ be an approximate solution and T be a matrix approximating A^{-1} and put*

$$(3.48)\qquad d = x^* - \tilde{x},\quad r = b - A\tilde{x},\quad R = I - AT,\quad a = d(R).$$

If $a<1$, then

$$(3.49)\qquad v(d) \leqq c(T)\,\|r\|/(1-a),$$

$$(3.50)\qquad v(d) \leqq v(Tr) + c(TR)\,\|r\|/(1-a),$$

$$(3.51)\qquad v(d) \leqq v(T(I+R)r) + c(TR^2)\,\|r\|/(1-a).$$

PROOF. By Lemma 3 $\|(I-R)^{-1}\| \leqq 1/(1-a)$. Since

$$d = A^{-1}r = T(I-R)^{-1}r = Tr + TR(I-R)^{-1}r$$
$$= T(1+R)r + TR^2(I-R)^{-1}r,$$

by Lemma 1 estimates (3.49), (3.50) and (3.51) are obtained.

Besides (3.50) and (3.51) we have

$$(3.52)\qquad v(d) \leqq v(Tr) + c(T)\,\|Rr\|/(1-a),$$

$$(3.53)\qquad v(d) \leqq v(T(I+R)r) + c(TR)\,\|Rr\|/(1-a)$$

and so on.

4. Numerical examples

Example 1. Consider the equation

$$(x_1, x_2)^T = f(x_1, x_2) = (-2x_1^2+x_2+3,\ -x_1-2x_2^2+4)^T/6 \tag{4.1}$$

on the set

$$D = \{(x_1, x_2)\colon 0.4 \leqq x_1,\ x_2 \leqq 0.6\}.$$

Then

$$K = \frac{1}{30}\begin{pmatrix} 12 & 5 \\ 5 & 12 \end{pmatrix}, \quad M = \frac{1}{30}\begin{pmatrix} -8 & 5 \\ 5 & -8 \end{pmatrix}.$$

Choosing $x^{(0)}=(0.46, 0.54)^T$, we have

$$\begin{aligned} & x^{(1)} = (1.5584, 1.4784)^T/3, \\ & u = Kv(x^{(1)}-x^{(0)}) = (0.28488, 0.25912)^T/9, \\ & (I-K)^{-1}u \leqq \|u\|_\infty e/(1-\|K\|_\infty) \leqq 0.0730462e. \end{aligned} \tag{4.2}$$

Since

$$S = \{(x_1, x_2)\colon 0.4464 \leqq x_1 \leqq 0.5926,\ 0.4197 \leqq x_2 \leqq 0.5659\} \subset D,$$

by Theorem 1 (4.1) has a unique solution x^* in D.

We have for (3.5)

$$(I-K)^{-1}u \leqq u + \|Ku\|_\infty e/(1-\|K\|_\infty) \leqq (0.0719458, 0.0690831)^T \tag{4.3}$$

and for (3.7)

$$(I-M)^{-1}u \leqq (I+M)u + \|M^2u\|_\infty e/(1-d_\infty(M)) \leqq (0.0285301, 0.0269081)^T. \tag{4.4}$$

The estimate (4.4) gives a smaller bound of $v(x^{(1)}-x^*)$ than (4.2) and (4.3).

Example 2. Consider the equation

$$g(x_1, x_2) = (x_1^3-3x_1x_2^2-1,\ 3x_1^2x_2-x_2^3)^T = 0 \tag{4.5}$$

on the set

$$D = \{(x_1, x_2)\colon 0.9 \leqq x_1 \leqq 1.2,\ -0.1 \leqq x_2 \leqq 0.1\}.$$

Choose

$$x^{(0)} = (0.96, 0.04)^T, \quad H = 0.4\,I.$$

Then

$$B = \left(\begin{array}{cc|cc} 2.592 & 0.168 & 0.168 & 2.592 \\ 0.168 & 2.592 & 2.592 & 0.168 \end{array}\right), \quad F = \begin{pmatrix} -0.104 & 0.09216 \\ -0.09216 & -0.104 \end{pmatrix},$$

$$x^{(1)} = (1.0079488, -0.0042112)^T, \quad a = (0.0479488, 0.0442112)^T,$$
$$c = (0.0149301, 0.0060311)^T.$$

We use l_∞-norm to have

(4.6) $b = 2.76e, \quad \|b\| = 5.52, \quad \|c\| = 0.0149301, \quad \|L\| = 0.450522,$

$\alpha = 0.0324658, \quad d = c + \alpha^2 b/2 \leqq (0.0163847, 0.0074856)^T.$

(4.7) $\beta = (I - L)^{-1} d \leqq d + \|d\| c(L)/(1 - \|L\|) \leqq (0.0298187, 0.0209196)^T = \beta_0,$

$d(L_1) = 0.242522, \quad \alpha_1 = 0.0213751,$

$d_1 = c + \alpha_1^2 b/2 \leqq (0.0161912, 0.0072921)^T,$

(4.8) $\gamma = (I - L_1)^{-1} d_1 \leqq d_1 + \|d_1\| c(L_1)/(1 - d(L_1))$

$\leqq (0.0213751, 0.0124760)^T = \gamma_0.$

Since

$$S = \{(x_1, x_2) \colon 0.9781 \leqq x_1 \leqq 1.0378, \ -0.02514 \leqq x_2 \leqq 0.01671\} \subset D,$$

by Theorem 4 (4.5) has a solution x^* in S. γ_0 gives a smaller bound of $v(x^{(1)} - x^*)$ than β_0. By one iteration we have improved bounds

$$c + L\beta_0 + B\beta_0\beta_0/2 \leqq (0.0282768, 0.0190958)^T = \beta_1,$$
$$c + L_1\gamma_0 + B\gamma_0\gamma_0/2 \leqq (0.0190412, 0.0117111)^T = \gamma_1.$$

Example 3. Consider the system of linear algebric equations

$$Ax = b \tag{4.9}$$

where

$$A = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}, \quad b = (4, 3)^T.$$

Let

$$\tilde{x} = (1.05, 0.95)^T, \quad T = \begin{pmatrix} 1.06 & -1.01 \\ -2.01 & 3.15 \end{pmatrix}$$

be approximations to the solution x^* of (4.9) and to A^{-1} respectively and put

$$d = x^* - \tilde{x}, \quad E = A^{-1} - T.$$

Then

$$R = I - AT = -\begin{pmatrix} 0.17 & 0.12 \\ 0.11 & 0.13 \end{pmatrix}, \quad r = b - A\tilde{x} = (-0.1, -0.05)^T.$$

We have

$$\begin{aligned} v(E) &\leqq v(T(I+R)R) + c_\infty(TR^2)r_\infty(R)/(1 - \|R\|_\infty) \\ &\leqq \begin{pmatrix} 0.0623031 & 0.0152995 \\ 0.0243947 & 0.1535894 \end{pmatrix}, \end{aligned} \tag{4.10}$$

$$\begin{aligned} v(E) &\leqq v(T(I+R)R) + c_\infty(TR^2)r_\infty(R)/(1 - d_\infty(R)) \\ &\leqq \begin{pmatrix} 0.0609799 & 0.0142876 \\ 0.0214930 & 0.1511705 \end{pmatrix}, \end{aligned} \tag{4.11}$$

$$\begin{aligned} v(d) &\leqq v(T(I+R)r) + c_\infty(TR)\|Rr\|_\infty/(1 - \|R\|_\infty) \\ &\leqq (0.0511334, 0.0579246)^T, \end{aligned} \tag{4.12}$$

$$\begin{aligned} v(d) &\leqq v(T(I+R)r) + c_\infty(TR)\|Rr\|_\infty/(1 - d_\infty(R)) \\ &\leqq (0.0504456, 0.0562983)^T. \end{aligned} \tag{4.13}$$

Estimates (4.11) and (4.13) provide slightly smaller bounds than those which employ $\|R\|_\infty$.

Combining (4.11) and (4.13) with the estimate

$$v(E) \leqq v(T(I+R)R) + c_1(TR^3)e^T/(1 - d_1(R)) \leqq \begin{pmatrix} 0.0605504 & 0.0146054 \\ 0.0202280 & 0.1515440 \end{pmatrix}, \tag{4.14}$$

we have

$$v(E) \leqq \begin{pmatrix} 0.0605504 & 0.0142876 \\ 0.0202280 & 0.1511705 \end{pmatrix}. \tag{4.15}$$

Error bounds (4.15) and (4.13) are to be compared with

$$v(E) = \begin{pmatrix} 0.06 & 0.01 \\ 0.01 & 0.15 \end{pmatrix}, \quad v(d) = (0.05, 0.05)^T.$$

References

[1] L. Collatz, *Functional analysis and numerical mathematics*, Academic Press, New York-San Francisco-London, 1966.

[2] G. Söderlind, *An error bound for fixed point iterations*, BIT, **24** (1984), 391–393.

[3] G. Söderlind, *On nonlinear difference and differential equations*, BIT, **24** (1984), 667–680.

[4] T. Ström, *On logarithmic norms*, SIAM J. Numer. Anal., **12** (1975), 741–753.

[5] M. Urabe, *Component-wise error analysis of iterative methods practiced on a floating point system*, Mem. Fac. Sci. Kyushu Unv. Ser. A, **27** (1973), 23–64.

[6] T. Yamamoto, *Componentwise error estimates for approximate solutions of nonlinear equations*, J. Information Process., **2** (1979), 121–126.

[7] T. Yamamoto, *Error bounds for approximate solutions of systems of equations*, Japan J. Appl. Math., **1** (1984), 157–171.

Department of Mathematics,
Faculty of School Education,
Hiroshima University