SOME ERROR ESTIMATES BY THE ALTERNATIVE METHOD*

By
L. CESARI (University of Michigan)
AND
T. T. BOWMAN (University of Florida, Gainesville)

We consider here, from the viewpoint of numerical analysis, the simple process for estimating the error between an approximate solution $x_0$ and the exact solution $x$ of a given boundary value problem in ordinary or partial differential equations, which arises from the alternative method (cf., e.g., [3]), and therefore from considerations of functional analysis. We shall limit ourselves to self-adjoint problems in the terminology of [3], where $x_0$ is a Galerkin approximation. In particular, for the problem of periodic solutions of periodic ordinary differential equations $x_0$ may be a harmonic balance approximation.

We know that, in the frame of the alternative method, it may be possible to answer the two relevant questions: (a) Is there an exact solution $x$ in a neighborhood $W$ of an approximate solution $x_0$? (b) Can we estimate the error $|x - x_0|$ between the exact solution $x$ and the given approximate solution $x_0$? In the present numerical analysis context, the process, still based on functional analysis, appears to have wider applications.

An outline of the computer program is given, together with actual numerical examples fully computed. In one example which had been previously computed in [5] we obtain results for larger nonlinearities.

1. The auxiliary and the bifurcation equations. Let $S$ be a Banach space over the reals of elements $x$ and norm $||x||$, let $E : D(E) \to S$, $N : D(N) \to S$ be operators in $S$, $E$ linear not necessarily bounded, $N$ not necessarily linear, with domains $D(E), D(N) \subset S$, $D(E) \cap D(N) \neq \emptyset$. We are concerned with solving the equation

$$Ex = Nx, \quad x \in D(E) \cap D(N).$$

As in [3] we assume that there is a projection operator $P : S \to S$ (i.e., linear, bounded, idempotent, or $P^2 = P$), with range $S_0 = R(P) \subset D(E)$, ker $E \subset S_0$, $S_1 = \ker P$, and a linear, bounded operator $H : S_1 \to S_1$, a partial inverse of $E$, with range $R(H) \subset D(E) \cap S_1$, satisfying the following relations:

\begin{align*}
(h_1) \quad H(I - P)Ex &= (I - P)x \text{ for all } x \in D(E), \\
(h_2) \quad PEx &= EPx \text{ for all } x \in D(E), \\
(h_3) \quad EH(I - P)Nx &= (I - P)Nx \text{ for all } x \in D(E) \cap D(N),
\end{align*}

where $I$ is the identity operator in $S$.

Applying $H(I - P)$ and $P$ to Eq. (1), we obtain

$$H(I - P)Ex = H(I - P)Nx, \quad PEx = PNx.$$
By (h1) the first of these equations becomes $(I - P)x = H(I - P)Nx$. Thus, any solution $x \in D(E) \cap D(N)$ of (1) is also a solution of the system

\begin{align*}
x &= Px + H(I - P)Nx, \\ P(Ex - Nx) &= 0.
\end{align*}

These equations are called the auxiliary and the bifurcation equations respectively. If $F$ and $T$ denote the operators $F = H(I - P)N$ and $T = P + F$, then the auxiliary equation becomes $x = Tx$. In other words, $x$ is a solution of the auxiliary equation if and only if $x$ is a fixed point of $T$. Let us prove that every solution $x \in D(E) \cap D(N)$ of the system (2), (3) is a solution of (1). Indeed, by applying $E$ to Eq. (2) and by force of (h2), (h3) we have $Ex = EPx + EH(I - P)Nx$, or

\[ Ex - Nx = P(Ex - Nx). \]

By (3) we have now $Ex - Nx = 0$. Thus, Eq. (1) and system (2), (3) are equivalent under assumptions (h123). Note that we have also proved that every solution of (2) is a solution of (4). We shall write (2), (3) in the form

\begin{align*}
x &= x^* + H(I - P)Nx, \\ P(Ex - Nx) &= 0,
\end{align*}

where $x^* = Px \in S_0$.

2. Remarks on the auxiliary equation. Let $x_0$ be an element of $S_0$ that we shall consider as an approximate solution of Eq. (1); thus $\Delta x_0 = Ex_0 - Nx_0$ is the corresponding error in the same equation. Then $\delta x_0 = H(I - P)\Delta x_0$ is an element of $S$, that we call the integrated error. We have

\[ (HI - P)Ex_0 = H(I - P)Nx_0 + \delta x_0, \]

and by (h1) also $x_0 = Px_0 + Fx_0 + \delta x_0$. Since $x_0 \in S_0$, then $Px_0 = x_0$, and we have $\delta x_0 = -Fx_0$.

**Theorem 1.** Let $x_0 \in S_0$, let $W$ be a neighborhood of $x_0$ in $S$ with $W \subset D(N)$, and let $N$ be Lipschitzian in $W$ of constant $L$, i.e. $||Nx - Ny|| \leq L ||x - y||$ for $x, y \in W$. Let $||H(I - P)|| = k$ and $||\delta x_0|| = \rho$, and let us assume that $kL < 1$. Let $c, d$ be real numbers, $0 < c < d$, such that $(1 - kL)^{-1}(c + \rho) \leq d$, and let us assume that $W$ contains the ball of center $x_0$ and radius $d$ in $S$. Let $W_0 = \{x^* \in S_0 | ||x^* - x_0|| \leq c\}$, and for $x^* \in W_0$ let $V_0 = \{x \in S | Px = x^*, ||x - x_0|| \leq d\}$, so that $W_0 \subset W$, and $V_0 \subset W$ for any $x^* \in W_0$. Then $T$ maps $V^*$ into itself, $T$ is a contraction on $V^*$, and hence $T$ has a unique fixed point $x = Tx = 3x^* \in V^*$ satisfying

\[ ||x - x^*|| \leq kLd + \rho \leq d - c. \]

Moreover, the map $3 : W_0 \to S$ mapping each $x^* \in W_0$ into the fixed point $x = Tx = 3x^*$ of $T$ in $V$, $P3x^* = x^*$, is continuous.

**Proof.** First, it will be shown that $T$ maps $V^*$ into itself. Indeed, if $x \in V^*$, then $PTx = P(x^* + Fx) = x^* + PH(I - P)Nx = x^*$. Also,

\[ ||Tx - x_0|| = ||x^* + Fx - (x_0 + Fx_0 + \delta x_0)|| \]

\[ \leq ||x^* - x_0|| + ||Fx - Fx_0|| + ||\delta x_0|| \]

\[ \leq c + kLd + \rho \leq d, \]
and thus $T$ maps $V^*$ into itself. If $y_i = Tx_i$, $y_2 = Tx_2$ for $x_1, x_2 \in V^*$, we also have
\[ ||y_1 - y_2|| = ||H(I - P)(N x_1 - N x_2)|| \leq k L ||x_1 - x_2||, \]
where $k L < 1$ and $T$ is a contraction restricted to $V^*$. By Banach's fixed-point theorem, there is a unique fixed point $x = Tx$ for $T$ in $V^*$. For this fixed point we have
\[ ||x - x^*|| = ||(x + F x) - x^*|| = ||F x - F x_0 - \delta x_0|| \leq ||F x - F x_0|| + ||\delta x_0|| \leq k L d + \rho \leq d - c. \]

We shall define the map $3$ from $W_0$ into $S$ by sending each $x^* \in W_0$ into the unique fixed point $x = Tx \in P^{-1} x^*$ in $V^*$. The map $3$ was shown to be continuous in [4, p. 39], and the proof of Theorem 1 is complete.

Then, the bifurcation equation (3) takes the form
\[ P(E - N)3x^* = 0, \quad x^* \in W_0 \subset S, \]
an equation in the space $S_0$.

3. Remarks on the bifurcation equation. First we note that $x = Tx = 3x^*$ is a solution of the auxiliary equation, with $P 3x^* = x^*$. Then, by (4) we have
\[ (E - N)3x^* = P(E - N)3x^* = EP3x^* - PN3x^* = Ex^* - PN3x^*, \quad (6) \]
\[ P(E - N)x^* = EPx^* - PNx^* = Ex^* - PNx^*. \quad (7) \]

Thus, by subtracting (7) from (6) we have
\[ P(E - N)3x^* - P(E - N)x^* = -PN(3x^* - x^*), \quad (8) \]
\[ ||P(E - N)(3x^* - x^*)|| \leq ||P|| L ||3x^* - x^*||. \]

We assume that, of the operator $E$, we know certain real eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$ and corresponding eigenelments $\phi_1, \phi_2, \cdots, \phi_m$, so that $E \phi_i + \lambda_i \phi_i = 0$, $i = 1, \cdots, m$. Let $S_0$ be the $m$-dimensional space spanned by $\phi_1, \cdots, \phi_m$; then $x_0$, an element of $S_0$, has a unique representation
\[ x_0 = a_1 \phi_1 + \cdots + a_m \phi_m. \]

If $\phi = (\phi_1, \cdots, \phi_m)$, $a = (a_1, \cdots, a_m)$, we may simply write $x_0 = a \phi$. We shall assume here that there is some constant $\sigma > 0$ such that, for all $y \in S_0$, or $y = b \phi$, we have $|b| \leq \sigma ||b \phi|| = \sigma ||y||$, where $|| \cdot ||$ is the Euclidean norm in $R^m$ and $|| ||$, as stated, is the norm in $S$. For instance, for $m = 1$ we may take $\sigma = ||\phi_1||^{-1}$. If it happens that $S_0$ can be thought of as contained in a Hilbert space $Z$ with inner product $(,)$ so that $||y|| = (y, y)^{1/2}$ for every $y \in S_0 \subset Z$ (and $|| ||$ is the norm in $S_0$ as well as in $Z$), then we can assume $\phi_1, \cdots, \phi_m$ to be orthonormal and then $\sigma = 1$.

If $x^* = b_1 \phi_1 + \cdots + b_m \phi_m = b \phi$, then $P(E - N)3x^*$ and $P(E - N)x^*$ are elements of $S_0$, and hence
\[ P(E - N)3x^* = U_1(b) \phi_1 + \cdots + U_m(b) \phi_m = U(b) \phi, \]
\[ P(E - N)x^* = u_1(b) \phi_1 + \cdots + u_m(b) \phi_m = u(b) \phi, \]
where $U(b), u(b) \in R^m$ are continuous functions of $b$. Then by (8) we have
\[ |U(b) - u(b)| \leq \sigma ||(U(b) - u(b)) \phi|| \leq \sigma ||P|| L ||3x^* - x^*||. \]
We note here that any element \( x^* \in S_0 \), \( x^* = b_1\phi_1 + \cdots + b_m\phi_m = b\phi \) such that \( P(Ex^* - Nx^*) = 0 \), that is, \( u(b) = 0 \), or \( u_1(b) = 0, \ldots, u_m(b) = 0 \), is said to be an \( m \)th Galerkin approximation for some possible solution \( x \in S \) of the equation \( Ex = Nx \). However, in no way below do we need to solve exactly the equation \( u(b) = 0 \). Thus, we shall denote by \( a \) some approximate solution of the equation \( u(b) = 0 \), and \( x_0 \) will be the approximate solution \( x_a = a\phi \). (An error in solving \( u(b) = 0 \) will just give a larger integrated error \( \delta x_a \).)

A number of remarks have been already made in [1, 2] for the numerical discussion of the equations \( u(b) = 0 \), \( U(b) = 0 \) in terms of topological degree of the corresponding maps \( u \) or \( U \). Here we restrict to the case \( m = 1 \). We note that for \( m = 1 \) and the problems of periodic solutions of periodic differential equations, the first Galerkin approximation is also called a harmonic balance solution.

Again, for \( m = 1 \), and having computer analysis applications in mind, we suppose that we can find a real number \( a \) and positive numbers \( c_1, c_2 > 0 \) such that \( u(a - c_1) \) and \( u(a + c_2) \) have opposite signs, and we take
\[
Q_1 = |u(a - c_1)|, \quad Q_2 = |u(a + c_2)|. \tag{9}
\]

**Theorem 2.** Given the assumptions above with \( d_i = (c_i + \rho)(1 - kL)^{-1}, i = 1, 2, \) if \( u(a - c_1), u(a + c_2) \) have opposite signs, and
\[
Q_i > ||\phi_i||^{-1} ||P|| L(d_i - c_i), \quad i = 1, 2, \tag{10}
\]
then there exists a number \( b \) such that \( a - c_1 < b < a + c_2 \), and \( b \) satisfies the bifurcation equation \( U(b) = 0 \); that is, there is an exact solution \( x \) to the equation \( Ex = Nx \), with \( x \in W, Px = x^* = b\phi \in W_0 \), and
\[
|b - a| \leq \text{max} (c_1, c_2), \quad ||x - x_0|| \leq \text{max} (d_1, d_2).
\]

**Proof.** We need only to show that \( U(b) \) has opposite signs at the points \( a - c_1 \) and \( a + c_2 \). Assume, for instance, that
\[
-Q_1 = u(a - c_1) < 0 < u(a + c_2) = Q_2.
\]
Then, for \( b = a + c_2 \) and \( x = Tx = 3x_b \), we also have
\[
U(b) = u(b) + (U(b) - u(b)) \geq u(b) - |U(b) - u(b)|
\]
\[
\geq Q_2 - ||\phi||^{-1} ||P|| L ||x - x_b||.
\]
By using (5) with \( ||x - x_b|| = ||x - x^*|| \leq kLd_2 + \rho = d_2 - c_2 \), we have
\[
U(b) \geq Q_2 - ||\phi||^{-1} ||P|| L(d_2 - c_2) > 0.
\]
Similarly, we have \( U(a - c_1) < 0 \). Then, \( U(a - c_1) \) and \( U(a + c_2) \) have opposite signs, and since \( U \) is continuous, there is a root \( b \) of \( U \) with \( a - c_1 \leq b \leq a + c_2 \), or \( |b - a| \leq \text{max} (c_1, c_2) \). On the other hand, \( x = Tx = 3x_b \in W \), hence \( ||x - x_0|| \leq \text{max} (d_1, d_2) \).

**Corollary.** Suppose \( kL < 1, x_0 = a\phi, \rho = ||\delta x_0||, c_1, c_2 > 0, u(a - c_1), u(a + c_1) \) of opposite signs, \( Q_1 = |u(a - c_1)|, \quad Q_2 = |u(a + c_2)| \), and
\[
||\phi||^{-1} ||P|| L(1 - kL)^{-1}(\rho + c_kL) < Q_i, \quad i = 1, 2.
\]
Then there is a point \( x_b = x^* = b\phi, a - c_1 \leq b \leq a + c_2 \), such that \( U(b) = 0 \), and \( x = 3x_b \) is an exact solution of (1), with
\[ |b - a| \leq \max (c_1, c_2), \]
\[ ||x - x_0|| \leq \max (d_1, d_2), \]
\[ d_i = (c_i + \rho)(1 - kL)^{-1}, \quad i = 1, 2. \]

Here \( L \) is a number such that \( ||N x_1 - N x_2|| \leq L ||x_1 - x_2|| \) in the neighborhood \( W \) of \( x_0 \) determined by \( ||x - x_0|| \leq d = \max (d_1, d_2) \).

Some examples.

We consider here an equation
\[ x'' + g(t, x) = 0, \quad -\infty < t < +\infty, \quad (11) \]
with \( g : \mathbb{R}^2 \to \mathbb{R} \) satisfying
\[ g(t + T, x) = g(t, x), \quad g(-t, -x) = g(t, x) \]
\[ g(4^{-1}T - t, x) = g(4^{-1}T + t, x). \]

and we seek solutions \( x(t), -\infty < t < +\infty \), satisfying
\[ x(t + T) = x(t), \quad x(-t) = -x(t), \quad (13) \]
\[ x(4^{-1}T - t) = x(4^{-1}T + t). \]

We take for \( S \) the Banach space of all real-valued bounded measurable functions \( x(t), -\infty < t < +\infty \), satisfying (13) with norm \( ||x|| = \text{Ess sup} |x(t)| \). Then, \( E \) is the differential operator \( E x = x'' \) with periodicity conditions of period \( T \) and further relations (13), and \( D(E) \) is, therefore, the set of all \( x \in S \) which are absolutely continuous together with \( x' \) in any finite interval, and with \( x'' \) bounded. Finally, let \( N \) be the Nemitski operator \( (Nx)(t) = -g(t, x(t)) \). We assume \( g \) to be locally Lipschitzian in \( x \), uniformly with respect to \( t \). We think of \( S \) as contained in the space \( X = L^2[0, T] \). Every element \( x \in S \) has a Fourier series
\[ x \sim \sum_{k=0}^{\infty} c_{2k+1} \sin (2k + 1)\omega t, \quad \omega = 2\pi T^{-1}, \]
\[ c_{\infty} = 2/T \int_0^T x(s) \sin n\omega s \, ds. \]

We take \( m = 1, S_0 \) the one-dimensional space spanned by \( \sin \omega t \), so that \( P x = c_1 \sin \omega t \), and \( S_1 = (I - P)S \) is the space of all elements \( x \in S \) with \( c_1 = 0 \), or
\[ x \sim \sum_{k=1}^{\infty} c_{2k+1} \sin (2k + 1)\omega t, \]
and we define \( H : S_1 \to S_1 \) by taking
\[ H x = -\sum_{k=1}^{\infty} (2k + 1)^{-2}\omega^{-2} c_{2k+1} \sin (2k + 1)\omega t. \]

Relations \((h_1), (h_2), (h_3)\) hold, and we know that \( R(H) \subset D(E), R(P) = S_0 \subset D(E) \).

Thus, for \( x \in S_1 = (I - P)S \), we have \( H(I - P)x = Hx \), and from [5] we have for \( y(t) = Hx \) the following representation:
so that by integration by parts we also have

\[ Hx(t) = t \int_0^t x(\beta) \, d\beta - \int_0^t \alpha x(\alpha) \, d\alpha - t \int_0^{T/4} x(\beta) \, d\beta. \]  

(14)

Also, from [5] we have the integral representation

\[ y(t) = Hx(t) = \int_0^{T/4} K(t, \beta) x(\beta) \, d\beta \]

\[ K(t, \beta) = -\beta + 4\pi^{-1} \omega^{-1} \sin \omega t \sin \omega \beta, \quad 0 \leq \beta \leq t \leq T/4, \]

\[ K(t, \beta) = -t + 4\pi^{-1} \omega^{-1} \sin \omega t \sin \omega \beta, \quad 0 \leq t \leq \beta \leq T/4. \]

By the use of this representation it was proved in [5] that

\[ ||H(I - P)|| = k_c = \pi^{-1/2} \omega^{1/2}, \quad k_0 = 0.14454. \]

Note that here we have \( ||P|| = 4/\pi \), and

\[ \lambda = \omega^2, \quad \phi(t) = 1/2 \sin \omega t, \quad \sigma^{-1} = ||\phi|| = \pi^{-1/2} \omega^{-1/2}. \]

For \( T = 2\pi \) we have of course \( \omega = 1, \lambda = 1. \)

I. For the problem of the \( 2\pi \)-periodic solutions of the equation,

\[ x'' + \sin x + C \sin t = 0, \quad (15) \]

where \( C \) is a constant, we have \( T = 2\pi, N x(t) = -\sin x(t) - C \sin t, L = 1, \) and thus, for \( m = 1, \phi = \pi^{-1/2} \sin t, x_b = b\phi = b\pi^{-1/2} \sin t, \)

\[ u(b) = -b - \int_0^{2\pi} N x_b(t) \pi^{-1/2} \sin t \, dt \]

\[ = -b + C \pi^{1/2} + \int_0^{2\pi} \sin (b \pi^{-1/2} \sin t) (\pi^{-1/2} \sin t) \, dt. \]  

(16)

For the integrated error \( \delta x_a = H(I - P)N x_a \) we have used relation (14) which now becomes

\[ \delta x_a(t) = -Hz(t), \]  

(17)

where \( z(t) = -\sin (a \pi^{-1/2} \sin t) - r \pi^{-1/2} \sin t, \) with \( r = \int_0^{2\pi} -\sin (a \pi^{-1/2} \sin t) \pi^{-1/2} \sin t \, dt, \) and we have to determine \( \rho = ||\delta x_a(t)|| = \max_t |\delta x_a(t)|. \)

Because of the symmetry properties of the problem, the maximum of \( |\delta x_a(t)| \) occurs in \([0, \pi/2]\). Finally, the constants \( c_1, c_2 \) must be determined so as to satisfy Theorem 2. Here \( ||\phi||^{-1} ||P|| L = \pi^{1/2}(4/\pi)(1) = 4\pi^{-1/2} \), and \( k = ||H(I - P)|| = 0.14454, \) as mentioned. Thus, \( c_1, c_2 \) have to be determined so that \( u(a - c_1) \) and \( u(a + c_2) \) have opposite signs, and

\[ 4\pi^{-1/2}(1 - 0.14454)^{-1}(\rho + 0.14454c_1) \leq Q_1 = |u(a - c_1)|, \]  

\[ 4\pi^{-1/2}(1 - 0.14454)^{-1}(\rho + 0.14454c_2) \leq Q_2 = |u(a + c_2)|. \]  

(18)

For various values of the constant \( C \) (see Table I below), we have performed the compu-
SOME ERROR ESTIMATES 127

tation by using an IBM 370. Column \((a_0)\) gives the value \(a_0 = a\pi^{-1/2}\), where \(a\) is an approximate root of \(u(b)\) (with an error \(\leq 0.0005\)). Thus, \(a_0 \sin t\) is the harmonic balance solution of (15). If \(b_0 \sin t\) denotes the first term in the Fourier series of the exact solution \(x(t)\) of (15), then \(b_0 = b\pi^{-1/2}, \lfloor b_0 - a_0 \rfloor \leq \pi^{-1/2} \max (c_1, c_2) = \eta, \) and this estimate \(\eta\) of \(|b_0 - a_0|\) is given in Table I. Finally, the error estimate \(\epsilon\) between \(x(t)\) and \(x_0(t), \) or

\[
||x - x_0|| = \text{Sup}_t |x(t) - x_0(t)| \leq (1 - 0.14454)^{-1}(\rho + \eta) = \epsilon
\]
is given in the last column of Table I.

Table I

<table>
<thead>
<tr>
<th>(C)</th>
<th>(a_0)</th>
<th>(\eta)</th>
<th>(\epsilon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.5</td>
<td>1.648</td>
<td>.067</td>
<td>0.100</td>
</tr>
<tr>
<td>.6</td>
<td>1.761</td>
<td>.068</td>
<td>0.104</td>
</tr>
<tr>
<td>.7</td>
<td>1.864</td>
<td>.068</td>
<td>0.107</td>
</tr>
<tr>
<td>.8</td>
<td>1.958</td>
<td>.068</td>
<td>0.111</td>
</tr>
<tr>
<td>.9</td>
<td>2.047</td>
<td>.068</td>
<td>0.115</td>
</tr>
<tr>
<td>1.0</td>
<td>2.130</td>
<td>.068</td>
<td>0.119</td>
</tr>
<tr>
<td>1.1</td>
<td>2.209</td>
<td>.068</td>
<td>0.122</td>
</tr>
<tr>
<td>1.2</td>
<td>2.285</td>
<td>.068</td>
<td>0.126</td>
</tr>
<tr>
<td>1.3</td>
<td>2.358</td>
<td>.068</td>
<td>0.129</td>
</tr>
<tr>
<td>1.4</td>
<td>2.428</td>
<td>.068</td>
<td>0.133</td>
</tr>
<tr>
<td>1.5</td>
<td>2.496</td>
<td>.068</td>
<td>0.136</td>
</tr>
</tbody>
</table>

Details of the program. The approximate root \(a\) of \(u(b)\) was determined by subdividing the interval \([0, 10]\) into 20 equal intervals, and computing \(u(b)\) at the points of subdivision by using (16) and Simpson's rule with division of \([0, 2\pi]\) into 40 equal parts. The interval at which \(u\) changes sign was then further subdivided, and the process repeated until divisions into intervals of length 0.0005 were reached, thus providing \(a\) with an error \(\leq 0.0005\). The values of \(\delta x_a(t)\) were determined by using (17), division of \([0, \pi/2]\) into 400 equal parts, computation of \(|\delta x_a(t)|\) at the points of subdivision \(t_i\), and noting the maximum of the values \(|\delta x_a(t_i)|\). Finally, the values \(c_1, c_2\) have been obtained by using (18), and by probing the values \(c_1, c_2 = 0.04n, n = 1, 2, \cdots, 50.\)

II. For the problem of the 2\(\pi\)-periodic solutions of the equation

\[
x'' + Ax^3 + B \sin t = 0,
\]
\(A, B\) constants, \(T = 2\pi, \) we have \(E x = x'', N x(t) = -Ax^3(t) - B \sin t. \) Here, \(x^3 - y^3 = (x - y)(x^2 + xy + y^2), \) and thus, for \(|x|, |y| \leq l, \) we have \(|x^3 - y^3| \leq 3l^2 |x - y| \). Hence, \(|Nx(t) - Ny(t)| \leq 3Al^2 |x(t) - y(t)|, \) provided \(|x(t)|, |y(t)| \leq l. \) Thus, we shall take \(L = 3Al^2\) with \(l \geq ||x_0(t)|| + d, d = \text{max} (d_1, d_2). \)

Again, as for example I, we take \(m = 1, \phi = \pi^{-1/2} \sin t, x_b = b\phi = b\pi^{-1/2} \sin t, \) and then

\[
u(b) = -b - \int_0^{2\pi} N x_b(t)\pi^{-1/2} \sin t \, dt
\]
\[
= -b - \int_0^{2\pi} [-A(b\pi^{-1/2} \sin t)^3 - B \sin t(\pi^{-1/2} \sin t) \, dt
\]
\[
= -b + B\pi^{1/2} + (3/4)Ab^3\pi^{-1}.
\]
For the integrated error $\delta x_a = -H(I - P)N x_a$, we have here

$$\delta x_a(t) = -Hz(t)$$

(21)

where $z(t) = -Aa^3 \pi^{-3/2} \sin^3 t + 3Aa^3 4^{-1} \pi^{-1/2} \sin t$.

Again we have to determine $\rho = \|\delta x_a(t)\| = \max_t |\delta x_a(t)|$, and this maximum occurs in $[0, \pi/2]$. Finally, the constants $c_1$, $c_2$ are to be determined so as to satisfy Theorem 2. Here $\|\phi\|^{-1} \|P\| L = \pi^{1/2}(4/\pi)L = 4\pi^{-1/2}L$, $k = \|H(I - P)\| = 0.14454$, and thus $c_1$, $c_2$ have to be determined in such a way that $u(a - c_1)$, $u(a + c_2)$ have opposite signs, and

$$4\pi^{-1/2}(1 - 0.14454)^{-1}(\rho + 0.14454c_1L) \leq Q_1 = |u(a - c_1)|,$$

$$4\pi^{-1/2}(1 - 0.14454)^{-1}(\rho + 0.14454c_1L) \leq Q_2 = |u(a + c_2)|.$$

(22)

For various values of $A$ and $B$ (see Table II below) we have performed the computations, and the corresponding values of $a_0$, $\eta \geq |b_0 - a_0|$, $\epsilon \geq \max |x(t) - x_a(t)|$ are displayed.

<table>
<thead>
<tr>
<th>$A$</th>
<th>$B$</th>
<th>$a_0$</th>
<th>$\eta$</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.2</td>
<td>0.2012</td>
<td>0.0100</td>
<td>0.01</td>
</tr>
<tr>
<td>0.2</td>
<td>0.4</td>
<td>0.4104</td>
<td>0.0095</td>
<td>0.01</td>
</tr>
<tr>
<td>0.2</td>
<td>0.6</td>
<td>0.6392</td>
<td>0.0082</td>
<td>0.01</td>
</tr>
<tr>
<td>0.2</td>
<td>0.8</td>
<td>0.9149</td>
<td>0.0143</td>
<td>0.02</td>
</tr>
<tr>
<td>0.4</td>
<td>0.2</td>
<td>0.2025</td>
<td>0.0099</td>
<td>0.01</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
<td>0.4227</td>
<td>0.0089</td>
<td>0.01</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6</td>
<td>0.7052</td>
<td>0.0143</td>
<td>0.02</td>
</tr>
<tr>
<td>0.6</td>
<td>0.2</td>
<td>0.2038</td>
<td>0.0098</td>
<td>0.01</td>
</tr>
<tr>
<td>0.6</td>
<td>0.4</td>
<td>0.4377</td>
<td>0.0081</td>
<td>0.01</td>
</tr>
<tr>
<td>0.8</td>
<td>0.2</td>
<td>0.2052</td>
<td>0.0097</td>
<td>0.01</td>
</tr>
<tr>
<td>0.8</td>
<td>0.4</td>
<td>0.4574</td>
<td>0.0072</td>
<td>0.01</td>
</tr>
<tr>
<td>1.0</td>
<td>0.2</td>
<td>0.2067</td>
<td>0.0095</td>
<td>0.01</td>
</tr>
<tr>
<td>1.0</td>
<td>0.4</td>
<td>0.4862</td>
<td>0.0234</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Details of the program. First note that, for $x_0 = x_a = a\pi^{-1/2} \sin t$, we have $\|x_0\| = a\pi^{-1/2}$, and thus, for $d = \max (d_1, d_2)$, $c = \max (c_1, c_2)$, we take $L = 3A|a| \pi^{-1/2} + d)^2$. For $l = |a| \pi^{-1/2} + d$, we have $kL = 0.14454 |3A|^2 < 1$ if and only if $l < 1.51860A^{-1/2}$, and thus $|a| \pi^{-1/2} + d < 1.51860A^{-1/2}$. Furthermore, to apply Theorem 1, we must have $(1 - kL)^{-1}(\rho + c) \leq d$, or $c \leq [-\rho + d(1 - kL)]$. Thus, $d$ must be large enough so that $-\rho + d(1 - kL) > 0$. We have set up the program in such a way that the successive values of $d = (0.1)n$, $n = 1, 2, \ldots$, are probed against these inequalities, and when these inequalities are satisfied, then relations (22) are checked, with $c_1 = c_2 = c$, $d_1 = d_2 = d$. The values for which the smallest error $\epsilon$ occurs are then collected. The remaining remarks as under I hold also in the present case.

For this joint paper the numerical computations have been carried out by Elinor Bowman.

References