

## STABILITY AND MULTIPLICITY OF SOLUTIONS TO DISCRETIZATIONS OF NONLINEAR ORDINARY DIFFERENTIAL EQUATIONS\*

WOLF-JÜRGEN BEYN† AND EUSEBIUS DOEDEL‡

**Abstract.** A large class of consistent and unconditionally stable discretizations of nonlinear boundary value problems is defined. The number of solutions to the discretizations is compared to the number of solutions of the continuous problem. We state conditions under which these numbers must agree for all sufficiently small mesh sizes. Various examples, including bifurcation problems, illustrate our theoretical results.

**Key words.** ordinary differential equations, bifurcation theory, discretization, collocation, stability theory, extraneous solutions

**1. Introduction.** We consider a large class of discretizations for the nonlinear boundary value problem

$$(1.1) \quad Nu \equiv u^{(n)} + f(x, u, u^{(1)}, \dots, u^{(n-1)}) = 0, \quad 0 \leq x \leq 1,$$

with boundary conditions

$$(1.1a) \quad B^l u \equiv \sum_{k=0}^{n-1} b_{k,l}^0 u^{(k)}(0) + \sum_{k=0}^{n-1} b_{k,l}^1 u^{(k)}(1) = 0, \quad 1 \leq l \leq n.$$

The discrete approximations all satisfy the conditions for consistency and stability, even on nonuniform meshes without the assumption of a bounded mesh ratio. These approximations are defined in § 2, where we also summarize their convergence properties.

Then, in § 3, we consider the relation between the number of solutions to the discrete problem and the number of solutions to the continuous problem. In general these need not be the same, not even asymptotically, as the mesh size goes to zero. The results of [2] indicate that these numbers of solutions can be guaranteed to agree only under rather restrictive assumptions on the differential equation. Essentially, the result in [2] is that extraneous solutions must disappear, as the mesh size approaches zero, when the lower order part of the differential operator in (1.1) is a sublinear function of its main arguments. We recover this result in § 3 with simple proofs adapted to the special nature of the approximations considered in this paper.

A variety of examples are given in § 4. In addition to problems that do not satisfy the sublinearity condition and that exhibit extraneous solutions, no matter how small the mesh size, we also give examples of discretized bifurcation problems that have extraneous solution branches. One of these examples shows that even if a bifurcation problem does satisfy the restrictive assumption of sublinearity, it still may have extraneous solution branches for large mesh sizes. Our theoretical results then explain why these branches must disappear (or straighten out) as the mesh size decreases. A final example illustrates the effect that discretization may have on the bifurcation diagram associated with a Hopf bifurcation problem.

**2. Definitions and basic convergence properties.** The discretizations studied can be defined as follows. Introduce a mesh  $\{0 = x_0 < x_1 < \dots < x_J = 1\}$ , with  $h_j \equiv x_j - x_{j-1}$ ,

\* Received by the editors April 21, 1980, and in revised form October 30, 1980. This work was supported in part by the U.S. Army under grant DAAG29-7B-G-0126 and by FCAC (Québec) under grant EQ-1438.

† Fakultät für Mathematik, Universität Konstanz, Konstanz, West Germany.

‡ Computer Science Department, Concordia University, Montréal, Québec, Canada.

$h \equiv (h_1, h_2, \dots, h_J)$  and  $|h| \equiv \max h_j$ . To each mesh point  $x_j$  ( $0 \leq j \leq J-n$ ) associate a polynomial  $p_j \in P_{n+m-1}$ . Here  $P_d$  is the space of all polynomials of degree less than or equal to  $d$ . Define  $p_h \equiv \{p_j\}_{j=0}^{J-n}$ . Let  $P_h^{n,m} \equiv P_h^{n,m}[\mu_1, \mu_2, \dots, \mu_n]$  denote the linear space of all  $p_h$  satisfying the boundary conditions (1.1a), in the sense that

$$(2.1) \quad \sum_{k=0}^{n-1} b_{k,l}^0 p_0^{(k)}(0) + \sum_{k=0}^{n-1} b_{k,l}^1 p_{J-n}^{(k)}(1) = 0, \quad 1 \leq l \leq n,$$

as well as *matching conditions* of the form

$$(2.2) \quad p_j(x_{j+\mu_k}) = p_{j+1}(x_{j+\mu_k}), \quad 1 \leq k \leq n, \quad 0 \leq j \leq J-n-1,$$

where  $1 \leq \mu_1 \leq \mu_2 \leq \dots \leq \mu_n \leq n$ . If not all integers  $\mu_k$  are distinct, then we define this to imply that derivatives also match, in the obvious manner. The discrete method now consists of finding  $p_h \in P_h^{n,m}$  satisfying the collocation equations

$$(2.3) \quad Np_i(z_{j,i}) = 0, \quad 1 \leq i \leq m, \quad 0 \leq j \leq J-n,$$

where for each  $j$  the  $z_{j,i}$  are distinct points in  $[x_j, x_{j+\mu_n}]$ . The  $z_{j,i}$  are assumed to be *locally semiuniform*, i.e.,  $\min_{i_1 \neq i_2} |z_{j,i_1} - z_{j,i_2}| \geq C|x_{j+\mu_n} - x_j|$ , for some constant  $C$  which is independent of  $j$  and  $h$ . This assumption is not strictly necessary (e.g., [6]), but it simplifies the argument somewhat. Note that it does not impose any restrictions on the mesh, if the  $z_{j,i}$  are chosen systematically with respect to  $[x_j, x_{j+\mu_n}]$ .

*Examples.* The orthogonal collocation methods correspond to taking  $\mu_i = 1$ ,  $1 \leq i \leq n$ , and Gauss  $z_{j,i}$  [1], [4]. Spline collocation methods can also be viewed as having  $\mu_i = 1$ ,  $1 \leq i \leq n$ , but with  $m = 2$ ,  $z_{j,1} = x_j$ ,  $z_{j,2} = x_{j+1}$  [17]. For a survey of such projection methods see [16]. The generalized finite difference methods of [5], [6], [13], [14], [15], [18] are obtained for general  $m$  and  $\mu_k = k$ ,  $1 \leq k \leq n$ .

The essential characteristic of all methods included in the present framework is the fact that all derivatives up to order  $n-1$  of each pair of consecutive polynomial components  $p_j$  and  $p_{j+1}$  match somewhere in  $[x_{j+1}, x_{j+\mu_n}]$ . But derivatives of order greater than  $n-1$  are not required to match. If  $\mu_k = 1$  ( $1 \leq k \leq n$ ), this is clear, since this is the case where all derivatives up to order  $n-1$  of each  $p_j, p_{j+1}$  match identically at  $x_{j+1}$ . If not all  $\mu_k$  are equal, this matching follows from repeated application of Rolle's theorem.

Introduce the following norm: If  $w_h \in P_h^{n,m}$ , then

$$\|w_h\|_p = \max_{0 \leq k \leq p} \max_{0 \leq j \leq J-n} \max_{x \in [x_j, x_{j+\mu_n}]} |w_j^{(k)}(x)|.$$

Also, for  $w \in C^p[0, 1]$ , we use the notation

$$\|w_h - w\|_p \equiv \max_{0 \leq k \leq p} \max_{0 \leq j \leq J-n} \max_{x \in [x_j, x_{j+\mu_n}]} |w_j^{(k)}(x) - w^{(k)}(x)|.$$

The proof of the following lemma is very similar to that of [6, Lemma 2.1] and will be omitted.

**LEMMA 2.1.** *Let  $\{h^\nu\}_{\nu=1}^\infty$  be a sequence of meshes with  $|h^\nu| \rightarrow 0$  as  $\nu \rightarrow \infty$ . For each  $\nu$  let  $w_{h^\nu} \in P_{h^\nu}^{n,m}$ , with  $\|w_{h^\nu}\|_n \leq C$ . Then there is a subsequence  $\{w_{h^{\nu'}}\}_{\nu'=1}^\infty$  and a function  $w \in C^{n-1}[0, 1]$  such that  $\|w_{h^{\nu'}} - w\|_{n-1} \rightarrow 0$  as  $\nu' \rightarrow \infty$ .*

*Remark.* For simplicity we do not notationally distinguish between sequence and subsequence.

Rewrite the collocation equations (2.3) in operator form as

$$N_h p_h = 0,$$

where  $N_h$  maps  $P_h^{n,m}$  into  $R_h^{n,m} \equiv \mathbb{R}^{m(J-n+1)}$ . For  $r_h \in R_h^{n,m}$ ,  $r_h \equiv \{r_{j,i}: 0 \leq j \leq J-n, 1 \leq i \leq m\}$ , let  $\|r_h\|_\infty \equiv \max_{i,j} |r_{j,i}|$ . (This is the usual max norm on  $\mathbb{R}^{m(J-n+1)}$ .) Let  $L_h[w_h]$  be the Fréchet derivative of  $N_h$  at  $w_h \in P_h^{n,m}$ . Thus  $L_h[w_h]$  maps  $P_h^{n,m}$  linearly into  $R_h^{n,m}$  as follows:  $L_h[w_h]q_h = r_h$ , where

$$r_{j,i} = q_j^{(n)}(z_{j,i}) + \sum_{k=0}^{n-1} f_{y_k}(z_{j,i}, w_j(z_{j,i}), \dots, w_j^{(n-1)}(z_{j,i}))q_j^{(k)}(z_{j,i}),$$

$$1 \leq i \leq m, \quad 0 \leq j \leq J-n.$$

Here the arguments of  $f$  are indicated as  $f \equiv f(x, y_0, y_1, \dots, y_{n-1})$ . The induced operator norm is given by

$$\|L_h[w_h]\| \equiv \max_{\|q_h\|_n=1} \|L_h[w_h]q_h\|_\infty.$$

Let  $u_h \equiv \{u_j(x)\}_{j=0}^{J-n} \in P_h^{n,m}$  interpolate a solution  $u$  of (1.1), (1.1a) at the mesh points and at certain additional points  $t_{j,i} \in [x_j, x_{j+\mu_n}]$  as follows:

$$u_j(x_{j+\mu_k}) = u(x_{j+\mu_k}), \quad 0 \leq k \leq n, \quad 0 \leq j \leq J-n,$$

$$u_j(t_{j,i}) = u(t_{j,i}), \quad 1 \leq i \leq m-1, \quad 0 \leq j \leq J-n,$$

where repeated points, if any, denote Hermite interpolation. Also define

$$B_\varepsilon^n(u_h) \equiv \{w_h \in P_h^{n,m}: \|w_h - u_h\|_n \leq \varepsilon\}.$$

We say that  $f(x, y_0, y_1, \dots, y_{n-1})$  has Lipschitz continuous derivatives with respect to  $y_0, y_1, \dots, y_{n-1}$  in a  $\rho$ -neighborhood of  $v \in C^{n-1}[0, 1]$  if

$$\max_{0 \leq k \leq n-1} |f_{y_k}(x, \alpha_0, \dots, \alpha_{n-1}) - f_{y_k}(x, \beta_0, \dots, \beta_{n-1})| \leq K_L \max_{0 \leq l \leq n-1} |\alpha_l - \beta_l|,$$

for all  $\{\alpha_l, \beta_l\}$  with  $\max_l |\alpha_l - v^{(l)}(x)| \leq \rho$ ,  $\max_l |\beta_l - v^{(l)}(x)| \leq \rho$ , and for all  $x \in [0, 1]$ . Here  $K_L \equiv K_L[v]$  and  $\rho \equiv \rho(v)$  are positive constants that do not depend on  $x$ .

LEMMA 2.2. Let  $f$  have Lipschitz continuous partial derivatives with respect to  $y_0, y_1, \dots, y_{n-1}$  in a  $\rho$ -neighborhood of  $u$ . Here  $u \in C^{n+m}[0, 1]$  is an exact solution of (1.1), (1.1a). Then for all sufficiently small  $|h|$  we have

$$\|L_h[v_h] - L_h[w_h]\| \leq nK_L \|v_h - w_h\|_{n-1},$$

whenever

$$v_h, w_h \in B_{\rho/2}^{n-1}(u_h).$$

A solution  $u \in C^n[0, 1]$  of (1.1), (1.1a) is called an isolated solution if the linearized problem

$$L[u]v \equiv v^{(n)} + \sum_{k=0}^{n-1} f_{y_k}(x, u, u^{(1)}, \dots, u^{(n-1)})v^{(k)} = 0,$$

subject to the boundary conditions  $B^l v = 0$ ,  $1 \leq l \leq n$ , admits only  $v(x) \equiv 0$  as solution. Below it will also be assumed that each  $f_{y_k}(x, y_0, \dots, y_{n-1})$  is continuous in  $x$ , in a  $\rho$ -neighborhood of  $u$ . By this is meant continuity in  $x$  for  $|y_l - u^{(l)}(x)| \leq \rho$ ,  $0 \leq l \leq n-1$ ,  $x \in [0, 1]$  with  $\rho \equiv \rho[u]$  independent of  $x$ . We can now state the following stability result:

**THEOREM 2.3.** *Let  $f$  have Lipschitz continuous derivatives with respect to  $y_0, y_1, \dots, y_{n-1}$  as in Lemma 2.2 and let each  $f_{y_k}$  be continuous in  $x$ , all in a  $\rho$ -neighborhood of an isolated solution  $u \in C^{n+m}[0, 1]$  of (1.1), (1.1a). Then there are positive constants  $\delta, \varepsilon$  and  $K$  such that*

$$\|v_h - w_h\|_n \leq K \|N_h v_h - N_h w_h\|_\infty,$$

for all  $v_h, w_h \in B_\varepsilon^n(u_h)$  and for all meshes  $h$  with  $|h| \in (0, \delta]$ .

For a proof of the statement above, one can first appeal to the general theory in [9]. This simplifies the problem to that of finding a corresponding stability result for the linearized problem. The linear case has been dealt with in [6], using a technique from [11]. Actually the discretizations in this work are slightly more general than those in [6], but this would be hardly noticeable in the proof.

A more direct approach is equally well possible and outlined below.

*Sketch of proof.* Suppose that the conclusion of the theorem does not hold. Then there exists a sequence of meshes  $\{h^\nu\}_{\nu=1}^\infty$  and corresponding  $\varepsilon^\nu > 0, K^\nu > 0$ , and  $v_{h^\nu}, w_{h^\nu} \in B_{\varepsilon^\nu}^n(u_{h^\nu})$ , with  $|h^\nu| \rightarrow 0, \varepsilon^\nu \rightarrow 0$  and  $K^\nu \rightarrow \infty$  as  $\nu \rightarrow \infty$ , such that

$$\|v_{h^\nu} - w_{h^\nu}\|_n > K^\nu \|N_{h^\nu} v_{h^\nu} - N_{h^\nu} w_{h^\nu}\|_\infty.$$

By the generalized mean value theorem

$$N_{h^\nu} v_{h^\nu} - N_{h^\nu} w_{h^\nu} = L_{h^\nu}[v_{h^\nu}, w_{h^\nu}](v_{h^\nu} - w_{h^\nu}),$$

where

$$L_{h^\nu}[v_{h^\nu}, w_{h^\nu}] \equiv \int_0^1 L_{h^\nu}[tv_{h^\nu} + (1-t)w_{h^\nu}] dt.$$

Thus, if we let  $e_{h^\nu} \equiv (\|v_{h^\nu} - w_{h^\nu}\|_n)^{-1}(v_{h^\nu} - w_{h^\nu})$ , then  $\|e_{h^\nu}\|_n = 1$  and  $\|L_{h^\nu}[v_{h^\nu}, w_{h^\nu}]e_{h^\nu}\|_\infty < (K^\nu)^{-1}$ . We may assume that  $\varepsilon^\nu \leq \rho/2$  for all  $\nu$ , so that  $v_{h^\nu}, w_{h^\nu} \in B_{\rho/2}^{n-1}(u_{h^\nu})$ , and also that  $|h^\nu|$  is sufficiently small for all  $\nu$ , so that the estimate of Lemma 2.2 is valid. Therefore

$$\begin{aligned} \|L_{h^\nu}[u_{h^\nu}]e_{h^\nu}\|_\infty &\leq \|L_{h^\nu}[u_{h^\nu}]e_{h^\nu} - L_{h^\nu}[v_{h^\nu}, w_{h^\nu}]e_{h^\nu}\|_\infty + \|L_{h^\nu}[v_{h^\nu}, w_{h^\nu}]e_{h^\nu}\|_\infty \\ &< \int_0^1 \|L_{h^\nu}[tu_{h^\nu} + (1-t)u_{h^\nu}] - L_{h^\nu}[tv_{h^\nu} + (1-t)w_{h^\nu}]\| dt + (K^\nu)^{-1} \\ &\leq \int_0^1 nK_L \|t(u_{h^\nu} - v_{h^\nu}) + (1-t)(u_{h^\nu} - w_{h^\nu})\|_{n-1} dt + (K^\nu)^{-1} \\ &\leq nK_L \varepsilon^\nu + (K^\nu)^{-1} \rightarrow 0 \quad \text{as } \nu \rightarrow \infty. \end{aligned}$$

Since  $\|e_{h^\nu}\|_n = 1$  for all  $\nu$ , it follows from Lemma 2.1 that there is a function  $e(x) \in C^{n-1}[0, 1]$  and a subsequence  $\{h^\nu\}_{\nu=1}^\infty$  such that  $\|e_{h^\nu} - e\|_{n-1} \rightarrow 0$  as  $\nu \rightarrow \infty$ . Using the techniques of [6] it is then not difficult to show, that in fact  $e \in C^n[0, 1]$ ,  $e \neq 0$  and  $L[u]e = 0$ . Also  $e(x)$  satisfies the homogeneous boundary conditions (1.1a). This contradicts the assumption that  $u$  is an isolated solution. Thus the statement of the theorem must be true.  $\square$

Given the result of Theorem 2.3, it is easy to establish convergence. For this purpose define the *truncation error*  $\tau_h \in R_h^{n,m}$  by

$$\tau_h \equiv N_h u_h.$$

Thus  $\tau_h$  has components  $\tau_{h,i} = Nu_i(z_{h,i})$ . In general, if  $u \in C^{n+m}[0, 1]$  then  $\tau_h$  satisfies an

estimate of the form

$$\|\tau_h\|_\infty \leq M_1[u]|h|^m,$$

but the order can normally be improved by suitable choice of the collocation points [4], [5], [13], [15]. For example the order is at least  $m + 1$  if for each  $j$  the  $z_{j,i}$  are chosen to coincide with the roots of

$$\left(\frac{d^n}{dx^n}\right) \left[ \prod_{k=0}^n (x - x_{j+\mu_k}) \prod_{k=1}^{m-1} (x - t_{j,k}) \right].$$

Below we also need a bound on the error in interpolating an exact solution  $u$  of (1.1), (1.1a) by  $u_h \in P_h^{n,m}$  in the manner indicated before. For this we have

$$\|u_h - u\|_n \leq M_2[u]|h|^m,$$

provided again that  $u \in C^{n+m}[0, 1]$ .

**THEOREM 2.4.** *Assume that the conditions of Theorem 2.3 are satisfied. Then there is a positive constant  $\delta$  such that  $N_h p_h = 0$  has a unique solution*

$$p_h \in P_h^{n,m} \text{ in } B_\epsilon^n(u_h) \text{ and such that}$$

$$\|p_h - u\|_n \leq (KM_1[u] + M_2[u])|h|^m, \text{ whenever } |h| \in (0, \delta).$$

*Proof.* The existence proof is identical to the corresponding part of the proof of [9, Thm. 3.6]. The error estimate is obtained as follows:

$$\begin{aligned} \|p_h - u\|_n &\leq \|p_h - u_h\|_n + \|u_h - u\|_n \\ &\leq K \|N_h p_h - N_h u_h\|_n + M_2[u]|h|^m \\ &\leq K \|N_h u_h\|_n + M_2[u]|h|^m \\ &\leq (KM_1[u] + M_2[u])|h|^m. \end{aligned}$$

**3. Multiplicity of solutions.** Theorem 2.4 in the preceding section guarantees that every isolated solution of the continuous problem is approximated by a solution of the discrete problem. Such a discrete solution is then itself necessarily isolated for all sufficiently small  $|h|$ , because stability implies isolation. The proof of this fact is essentially the same as the proof of [9, Thm. 2.5], where it is given for the continuous problem. Not much can be said if the continuous problem has a solution that is not isolated. For example, in such a case it is possible that the discretization has no solutions that converge to the given nonisolated solution of the continuous problem. Most of these difficulties can be avoided however by considering solution branches instead of single solutions, and following these with a continuation procedure.

Below we state a global convergence theorem. It does not exclude the possibility of existence of extraneous solutions to the discrete problem, not even for small  $|h|$ . However, under some mild continuity assumptions, the theorem shows that if such extraneous solutions exist, they must tend to infinity in the appropriate norm as  $|h| \rightarrow 0$ . The proof of this fact bears resemblance to that of Thm. 2.3, although it deals directly with the nonlinear equation.

**THEOREM 3.1.** *Assume that the problem (1.1), (1.1a), has exactly  $N$  solutions  $u^{[i]}(x) \in C^{n+m}[0, 1]$ ,  $N < \infty$ , each of which is isolated. Let  $f$  have Lipschitz continuous derivatives  $f_{y_k}$ ,  $0 \leq k \leq n - 1$ , and let each  $f_{y_k}$  be continuous in  $x$ ; all in a  $\rho[u^{[i]}]$ -neighborhood of each  $u^{[i]}$ . In addition assume that  $f$  is continuous in all its variables in a  $\rho$ -neighborhood of every  $v \in C^{n-1}[0, 1]$ ;  $\rho[v] > 0$ . Then for each  $R >$*

$\max_{1 \leq l \leq N} \|u^{[l]}\|_n$ , there exists a  $\delta_R > 0$  such that the discrete problem  $N_h p_h = 0$  has exactly  $N$  solutions  $p_h^{[l]} \in P_h^{n,m}$  in  $B_R^n(0)$ , for all  $h$  with  $|h| \in (0, \delta_R]$ .

*Proof.* The assumptions above include those of Theorem 2.4. Hence each  $u^{[l]}$  is approximated by a solution  $p_h^{[l]}$  of  $N_h p_h = 0$ . For all sufficiently small  $|h|$  these  $p_h^{[l]}$  must be distinct, due to the isolated nature of each of the  $u^{[l]}$ . Also, since  $R > \max_l \|u^{[l]}\|_n$ , the  $p_h^{[l]}$  must eventually (for small  $|h|$ ) lie in  $B_R^n(0)$ . Assume now that we can find a sequence of meshes  $\{h^\nu\}_{\nu=1}^\infty$  with  $|h^\nu| \rightarrow 0$  as  $\nu \rightarrow \infty$ , and for each  $\nu$  solutions  $p_{h^\nu} \neq p_{h^\nu}^{[l]}$ ,  $1 \leq l \leq N$ , of  $N_{h^\nu} p_{h^\nu} = 0$ , with  $\|p_{h^\nu}\|_n \leq R$ . Then by Lemma 2.1 there is a subsequence  $\{p_{h^\nu}\}_{\nu=1}^\infty$  and a function  $p \in C^{n-1}[0, 1]$ , such that  $\|p_{h^\nu} - p\|_{n-1} \rightarrow 0$  as  $\nu \rightarrow \infty$ .

Let  $s \in [0, 1]$ , and for each  $\nu$  let  $j \equiv j_\nu$  be such that  $s \in [x_j, x_{j+\mu_n}]$ . For each fixed  $j$  let the polynomials  $\psi_i \equiv \psi_{j_\nu, i}$  denote the Lagrange interpolating coefficients for the points  $z_i \equiv z_{j_\nu, i}$ ,  $i = 1, 2, \dots, m$ . Since the collocation points are assumed to be locally semi-uniform throughout this work, we have

$$\max_{i,j} \max_{[x_j, x_{j+\mu_n}]} |\psi_{i,j}(x)| \leq K_\psi,$$

for some constant  $K_\psi$  that does not depend on  $h$ . Write  $w^* \equiv (w, w^{(1)}, \dots, w^{(n-1)})$ . Consider now the following estimate:

$$\begin{aligned} & |p_j^{(n)}(s) + f(s, p^*(s))| \\ &= \left| - \sum_{i=1}^m \psi_i(s) f(z_i, p_j^*(z_i)) + f(s, p^*(s)) \right| \\ &= \left| \sum_{i=1}^m \psi_i(s) \{f(s, p^*(s)) - f(z_i, p_j^*(z_i))\} \right| \\ &\leq mK_\psi \max_i \{|f(s, p^*(s)) - f(z_i, p^*(s))| \\ &\quad + |f(z_i, p^*(s)) - f(z_i, p_j^*(s))| + |f(z_i, p_j^*(s)) - f(z_i, p_j^*(z_i))|\}. \end{aligned}$$

The added continuity assumptions can now be used to show that the right-hand side of this final inequality tends to zero uniformly in  $s$  as  $\nu \rightarrow \infty$ . From integration, letting  $\nu \rightarrow \infty$  and differentiation it follows that  $p \in C^n[0, 1]$ ,  $Np = 0$  and  $B^l p = 0$ . Again the details proceed much like those given in the proof of [6, Thm. 2.2]. Thus  $p_{h^\nu}$  tends to another solution  $p$  of the continuous problem. This new solution  $p$  must be distinct from the other  $N$  solutions, for otherwise the stability result of Thm. 2.3 is violated for one of the  $u^{[l]}$ . Hence we have derived a contradiction.  $\square$

Under rather restrictive assumptions on the form of the differential operator it is sometimes possible to derive a priori bounds of the form  $\|p_h\|_n \leq K$  for all solutions of  $N_h p_h = 0$  and for all meshes  $h$  with  $|h|$  sufficiently small. It then follows from Theorem 3.1 that extraneous solutions must disappear as  $|h| \rightarrow 0$ . This is the case, for example, in the following class of problems:

**THEOREM 3.2.** Assume that the conditions of Theorem 3.1 hold. Also assume that for all  $h$  with  $|h|$  sufficiently small (say  $|h| \in (0, \delta]$ ,  $\delta > 0$ ) we have

- (i)  $N_h = L_h + G_h$ ,  $L_h$  linear, invertible,  $\|L_h^{-1}\| \leq K_s$ , and
- (ii) For each  $\epsilon > 0$  there exists a positive number  $K(\epsilon)$  such that

$$(3.1) \quad \|G_h p_h\|_\infty \leq K(\epsilon) + \epsilon \|p_h\|_n \text{ for all solutions } p_h \text{ of } N_h p_h = 0.$$

Here  $K_s$ ,  $K(\epsilon)$  and  $\epsilon$  are independent of  $h$ . Then there is a  $\delta_1 > 0$  such that the discrete problem  $N_h p_h = 0$  has exactly  $N$  solutions for all meshes  $h$  with  $|h| \in (0, \delta_1]$ .

*Proof.* We need only establish that all solutions of  $N_h p_h = 0$  are uniformly bounded. But this is trivial since

$$\|p_h\|_n \leq \|L_h^{-1} G_h p_h\|_n \leq K_s \|G_h p_h\|_\infty \leq K_s [K(\epsilon) + \epsilon \|p_h\|_n].$$

Thus if we choose  $\epsilon < 1/K_s$ , then

$$\|p_h\|_n < \frac{K_s K(\epsilon)}{1 - K_s \epsilon}.$$

**COROLLARY 3.3.** *In the problem (1.1), (1.1a) let  $Lu \equiv u^{(n)}$ , and assume that the linear homogeneous problem  $Lu = 0$ , subject to (1.1a), admits the trivial solution only. Also assume that*

$$|f(x, y_0, y_1, \dots, y_{n-1})| \leq C_1 + C_2 \left\{ \max_{0 \leq k \leq n-1} |y_k| \right\}^\alpha, \quad x \in [0, 1], \quad y_k \in \mathbb{R},$$

with  $0 \leq \alpha < 1$ . Further, let the conditions of Theorem 3.1 be met. Then the discretization  $N_h p_h = 0$  cannot have extraneous solutions on meshes  $h$  with  $|h|$  sufficiently small.

*Proof.* We need only verify that a bound of the form (3.1) holds. For this purpose we shall make use of the following elementary inequality. For each given  $\hat{\epsilon} > 0$  there exists a constant  $M(\hat{\epsilon})$  such that

$$r^\alpha \leq M(\hat{\epsilon}) + \hat{\epsilon} r \quad \text{for all } r \geq 0.$$

Now let  $\epsilon > 0$  be given. Then

$$\begin{aligned} \|G_h p_h\|_\infty &= \max_{i,j} |f(z_{i,h}, p_j^*(z_{j,i}))| \\ &\leq C_1 + C_2 \max_{i,j} \left\{ \max_{0 \leq k \leq n-1} |p_j^{(k)}(z_{j,i})| \right\}^\alpha \\ &\leq C_1 + C_2 \|p_h\|_{n-1}^\alpha \\ &\leq C_1 + C_2 \|p_h\|_n^\alpha \leq K(\epsilon) + \epsilon \|p_h\|_n, \quad \text{with } K(\epsilon) \equiv C_1 + C_2 M\left(\frac{\epsilon}{C_2}\right). \quad \square \end{aligned}$$

Note that Corollary 3.3 applies in particular if  $\alpha = 0$ ; i.e., extraneous solutions must disappear as  $|h| \rightarrow 0$  when the lower order part of the operator is bounded.

**4. Examples.** One can easily construct examples of discretizations that have extraneous solutions on a given mesh. It is somewhat more difficult to find a problem where the discretization has extraneous solutions on each of an infinite sequence of meshes with mesh size tending to zero. For stable discretizations of the type treated in [2] and in this work, such examples cannot be found for operators with a sublinear lower order part. The following example is given in [7]. The boundary value problem

$$(4.1) \quad \begin{aligned} u'' + u'|u'| &= 0, \quad x \in [0, 1], \\ u(0) &= u(1) = 1 \end{aligned}$$

has  $u(x) \equiv 1$  as its unique solution, while the discretization

$$\begin{aligned} (u_{j+1} - 2u_j + u_{j-1})/h^2 + (u_j - u_{j-1})/h |(u_j - u_{j-1})/h| &= 0, \quad 1 \leq j \leq J-1, \\ u_0 &= u_J = 1 \end{aligned}$$

has in addition to  $u_j = 1$  ( $0 \leq j \leq J$ ) also the zig-zag solution  $u_j = (-1)^j$  for all uniform

meshes with  $J$  even. Of course this difference approximation is only first order accurate. But this is not essential, as the following example indicates.

*Example 4.1.* To approximately solve

$$(4.2) \quad \begin{aligned} u'' + 2u[u']^2 &= 0, & x \in [0, 1], \\ u(0) = u(1) &= 1, \end{aligned}$$

consider the second order finite difference scheme

$$(4.3) \quad \begin{aligned} (u_{j+1} - 2u_j + u_{j-1})/h^2 + 2u_j[(u_{j+1} - u_{j-1})/2h]^2 &= 0, & 1 \leq j \leq J-1, \\ u_0 = u_J &= 1. \end{aligned}$$

The discretization (4.3) has in addition to the constant solution  $u_j = 1$  ( $0 \leq j \leq J$ ) also the solution

$$u_j = \begin{cases} 1 & \text{if } j = 0 \text{ or } 3, \text{ mod } 4, \\ -1 & \text{if } j = 1 \text{ or } 2, \text{ mod } 4, \end{cases}$$

provided  $J = 3 \pmod{4}$ . This can be verified directly by substitution. However, the continuous problem (4.2) has  $u(x) \equiv 1$  as its unique solution. For if  $v(x)$  is another solution of (4.1) with  $v(s) \neq 1$  for some  $s \in [0, 1]$ , then there exists a  $t \in (0, 1)$  such that  $v(t) \neq 1$ ,  $v'(t) = 0$ . Hence  $v(x)$ , as well as the constant function  $u(x) \equiv v(t)$ ,  $x \in [0, 1]$ , solve the initial value problem

$$\begin{aligned} u'' + 2u[u']^2 &= 0, & x \in [0, 1], \\ u(t) = v(t), & \quad u'(t) = 0. \end{aligned}$$

Therefore  $v(x) \equiv v(t)$  for all  $x \in [0, 1]$ , which contradicts the fact that  $v(1) = 1$ . (A similar argument can be used for (4.1).) Thus the solution  $u(x) \equiv 1$  is unique. It is also clearly isolated.

For nonlinear problems containing a parameter, it can be instructive to compare the bifurcation diagram of a discretization to that of the continuous problem as is done, e.g., in [3]. As a simple numerical example consider the following:

*Example 4.2.* Discretize the nonlinear boundary value problem

$$\begin{aligned} u'' + \lambda \sin(u + u^2 + u^3) &= 0, & x \in [0, 1], \\ u(0) = u(1) &= 0 \end{aligned}$$

by the 4th order accurate, 3 point Numerov formula, with uniform mesh and mesh size  $h = \frac{1}{8}$ . Using the continuation and branch switching techniques of [10], one obtains the bifurcation diagram of Fig. 4.1. In the diagram the vertical axis represents a discrete  $L_2$ -norm of the approximate solution. Observe that  $u(x) \equiv 0$  is a solution for all  $\lambda$ . Bifurcation points have been circled. The branches originating from the secondary bifurcation points are actually double branches that coincide in the diagram due to some symmetry. If the mesh size is decreased to  $h = \frac{1}{70}$ , then the diagram changes to that of Fig. 4.2. This of course can be expected to be more like the actual diagram of the continuous problem. Extraneous solutions, which are abundant in Fig. 4.1, must disappear when the mesh size goes to zero, provided  $\lambda$  is such that the continuous problem has only isolated solutions. This follows from the boundedness of the lower order part of the continuous operator. However, we have observed that within the limits of the given bifurcation diagram, extraneous solutions do not disappear until the number of meshintervals  $J$  is taken greater than 50.



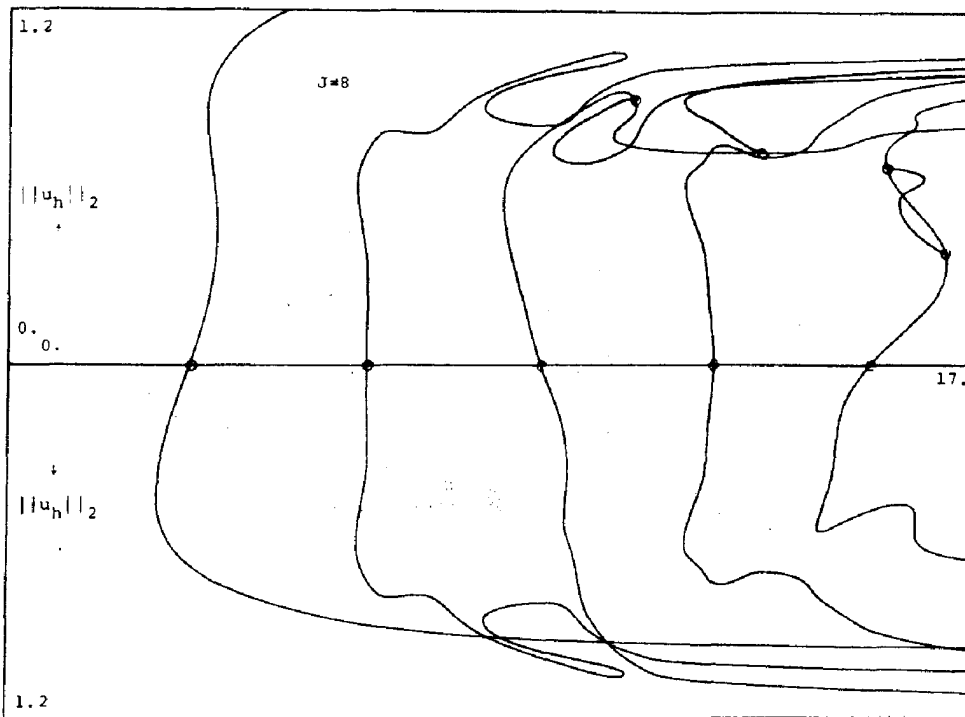


FIG. 4.1a. Bifurcation diagram of the discretization in Example 4.2 ( $J = 8$ ).

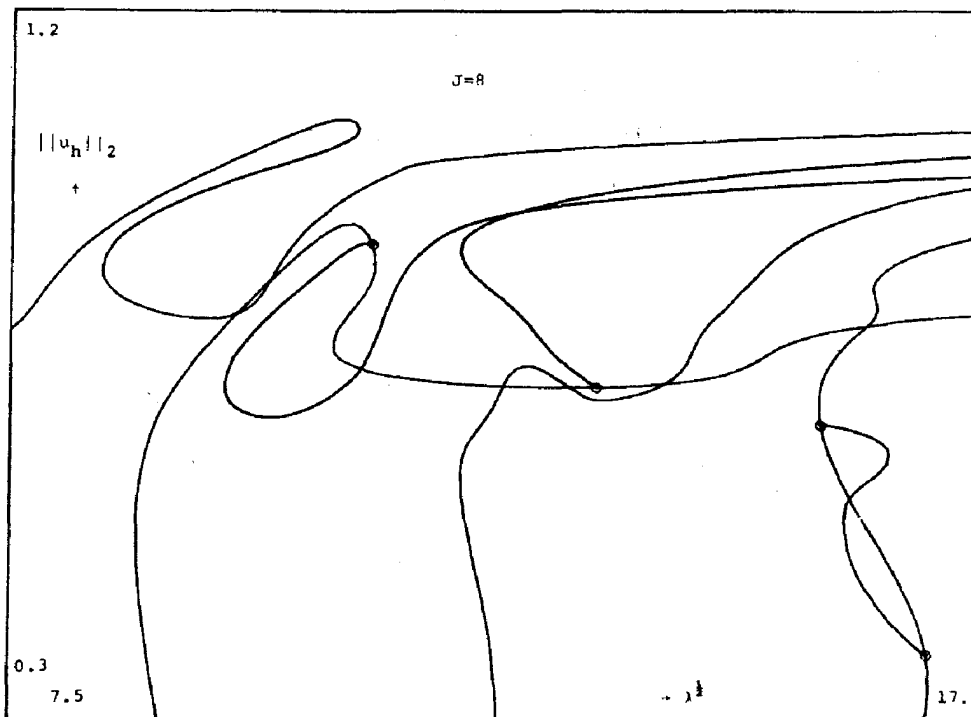


FIG. 4.1b. Local enlargement of the bifurcation diagram in Fig. 4.1a.

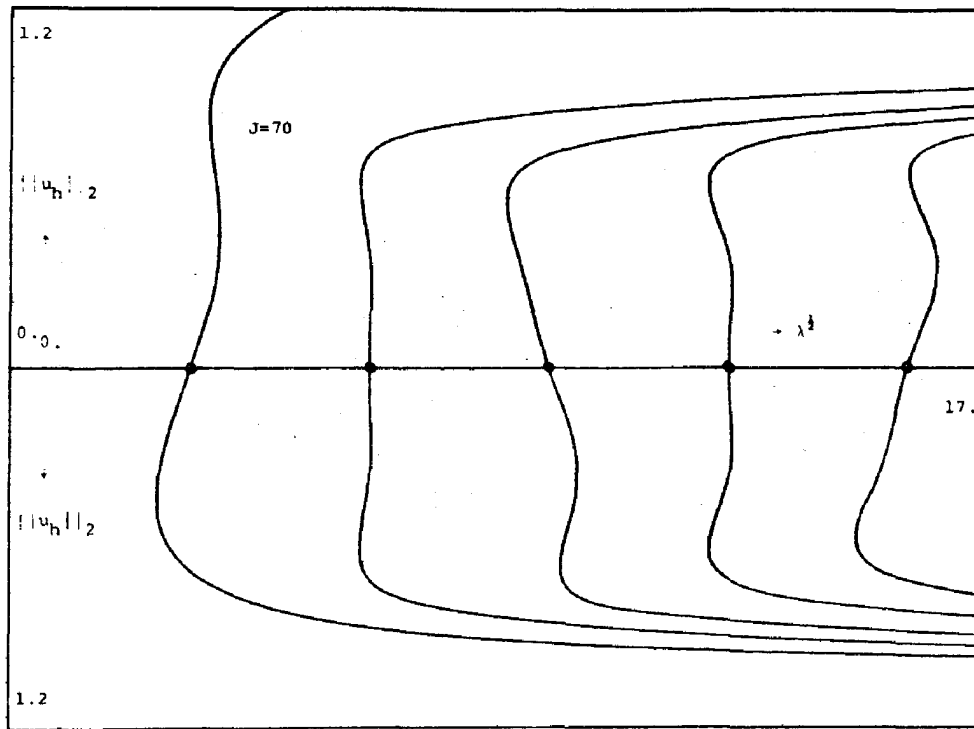


FIG. 4.2. Bifurcation diagram of the discretization in Example 4.2 ( $J = 70$ ).

Next we give an analytic example where the bifurcation diagram of a discretization has an extraneous bifurcation point and associated extraneous solutions, no matter how small the mesh size.

*Example 4.3.* The solutions of the one parameter nonlinear problem

$$(4.4) \quad \begin{aligned} v'' + (1 + [v(\frac{1}{2})]^2)\mu v &= 0, & x \in [0, 1], \\ v(0) = v(1) &= 0 \end{aligned}$$

are homeomorphically related to those of the linear problem

$$(4.5) \quad \begin{aligned} u'' + \lambda u &= 0, & x \in [0, 1], \\ u(0) = u(1) &= 0 \end{aligned}$$

through the transformation  $\lambda = \mu(1 + [v(\frac{1}{2})]^2)$ ,  $u = v$ . The same is true for their respective discretizations

$$(4.6) \quad \begin{aligned} (v_{j+1} - 2v_j + v_{j-1})/h^2 + (1 + v_{(1/2)J}^2)\mu v_j &= 0, & 1 \leq j \leq J-1, \\ v_0 = 0, & 4v_{J-1} - 3v_J = 0, & J \text{ even,} \end{aligned}$$

and

$$(4.7) \quad \begin{aligned} (u_{j+1} - 2u_j + u_{j-1})/h^2 + \lambda u_j &= 0, & 1 \leq j \leq J-1, \\ u_0 = 0, & 4u_{J-1} - 3u_J = 0, \end{aligned}$$

both of which are consistent and stable for isolated solutions. We claim that (4.7) has an eigenvalue  $\lambda$  that is asymptotically (as  $h \rightarrow 0$ ) given by  $\lambda_E \cong -1/12h^2$ . Thus, unlike the eigenvalues  $\lambda_k = (k\pi)^2$  of (4.5), which are all positive, the discrete problem (4.7) admits

a negative eigenvalue. To prove this, we note that an eigenvalue of (4.7) has the form

$$u_j = c_1 z_1^j + c_2 z_2^j,$$

where  $z_1(\lambda)$  and  $z_2(\lambda)$  are the roots of the characteristic equation

$$(4.8) \quad z^2 - 2z + 1 + \lambda h^2 z = 0.$$

Now if  $z_1$  is a root of (4.8) then so is  $z_1^{-1}$ . Moreover, the boundary condition  $u_0 = 0$  implies that  $c_1 = -c_2 \equiv c$ . Thus

$$u_j = c[z^j - z^{-j}].$$

The second boundary condition yields the equation

$$(4.9) \quad 4[z^{J-1} - z^{-(J-1)}] - 3[z^J - z^{-J}] = 0.$$

Notice that for large  $J$  (4.9) has a root  $z \equiv z(\lambda)$  with  $|z| > 1$ . This root asymptotically satisfies  $4z^{J-1} - 3z^J = 0$ ; i.e.,  $z = \frac{4}{3}$ . The corresponding eigenvalue  $\lambda_E \cong -1/12h^2$  is obtained from (4.8), and the associated eigenvector is  $u_j = c[(\frac{4}{3})^j - (\frac{3}{4})^j]$ . Note that  $u_j \neq 0$  if  $c \neq 0$  and  $j \neq 0$ .

The corresponding solution branch of the nonlinear discrete problem (4.6) is therefore asymptotically (as  $h \rightarrow 0$ ) given by

$$(\mu(c), v_j(c)), \quad -\infty < c < \infty,$$

where

$$\mu(c) = \frac{\lambda_E}{1 + c[(\frac{4}{3})^{J/2} - (\frac{3}{4})^{J/2}]^2}$$

and

$$v_j(c) = c[(\frac{4}{3})^j - (\frac{3}{4})^j].$$

(For related techniques see [8, 20].) The extraneous solution branch is shown in Fig. 4.3 for the case  $J = 16$ . This solution branch exists for all meshes with even  $J$  and is extraneous, since the continuous problem (4.4) has no nonzero solutions for negative  $\mu$ . In particular, for any  $\mu < 0$  one can choose  $h$  sufficiently small so that  $\lambda_E < \mu$ . For such  $\mu$  the discretization (4.6) evidently has three solutions, unlike the continuous problem (4.4).

Finally we give a numerical example, where the objective is to determine time periodic solutions to an autonomous system containing a parameter. It is generally necessary to reformulate the problem as a boundary value problem on a fixed interval, say  $[0, 2\pi]$ . This allows the numerical computation to proceed past turning points and makes the computation of asymptotically unstable solutions possible. A general code for the bifurcation analysis of autonomous systems has been developed by the second author and will be described more completely in a forthcoming paper.

*Example 4.4.* The dynamic behavior of a single first order chemical reaction in a continuously stirred tank reactor can be modeled by the ordinary differential equation

$$(4.10) \quad \begin{aligned} u_1' &= -u_1 + B \text{Da} (1 - u_2) e^{u_1} - \beta u_1, \\ u_2' &= -u_2 + \text{Da} (1 - u_2) e^{u_1}, \end{aligned}$$

where  $B$ ,  $\beta$  and  $\text{Da}$  are dimensionless parameters.

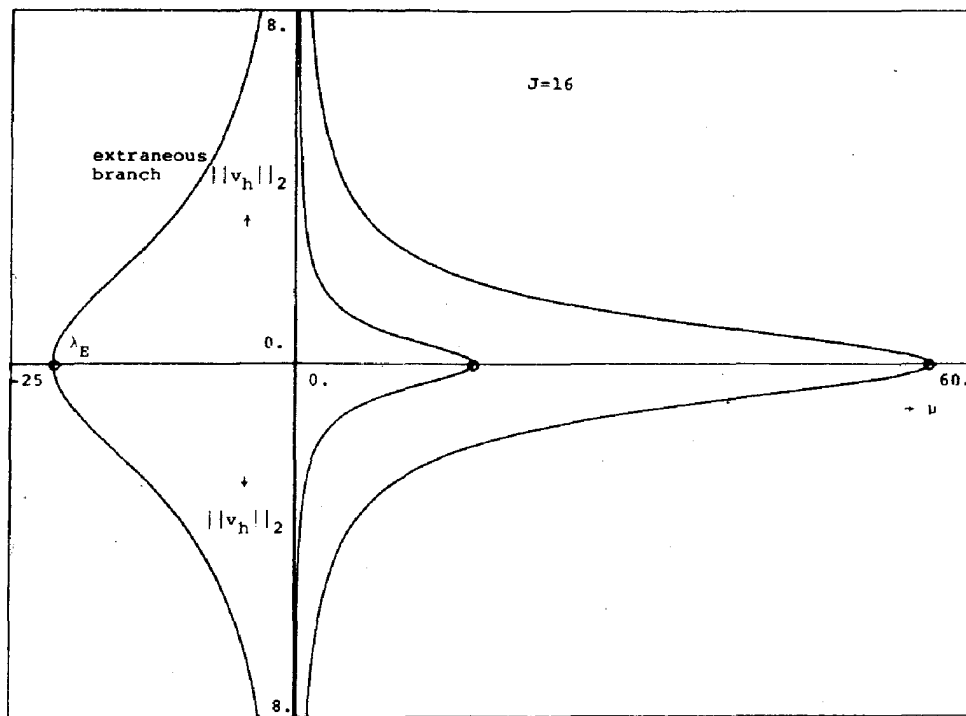


FIG. 4.3. Graph of the extraneous branch in Example 4.3.

For previous computations with initial value techniques, see [19]. Using vector notation and scaling the period to the interval  $[0, 2\pi]$ , we can write (4.10) as

$$(4.11) \quad u'(t) = \frac{\rho}{2\pi} f(u(t), \lambda), \quad \lambda \equiv Da,$$

$$(4.12) \quad u(0) = u(2\pi).$$

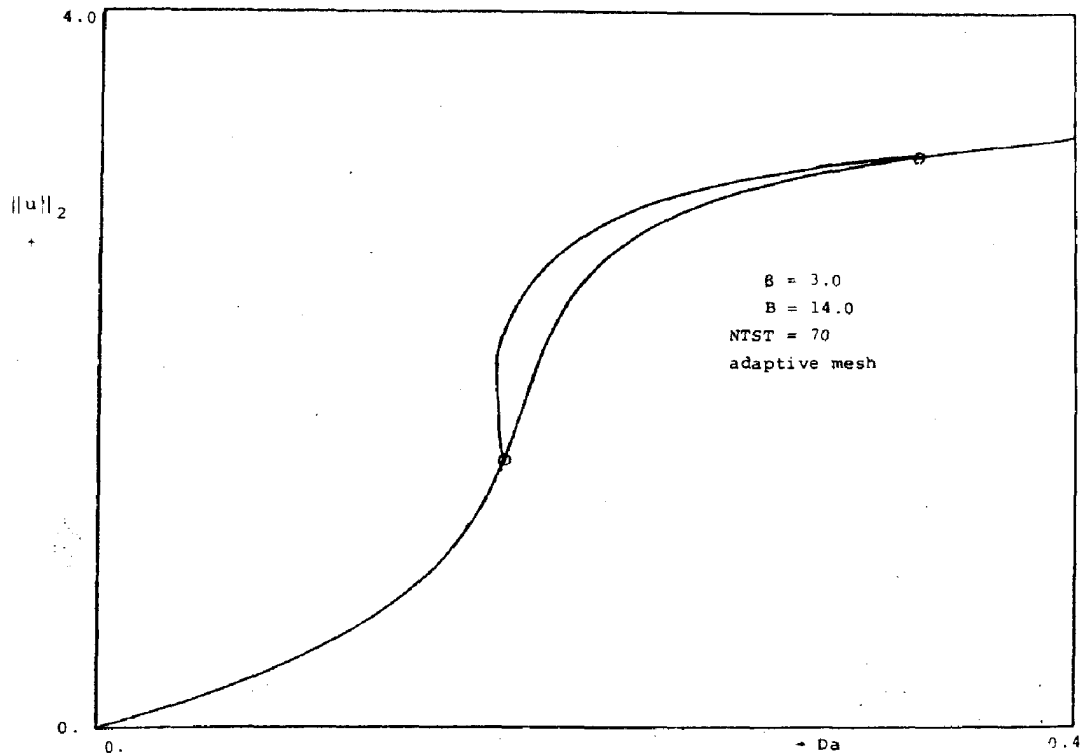
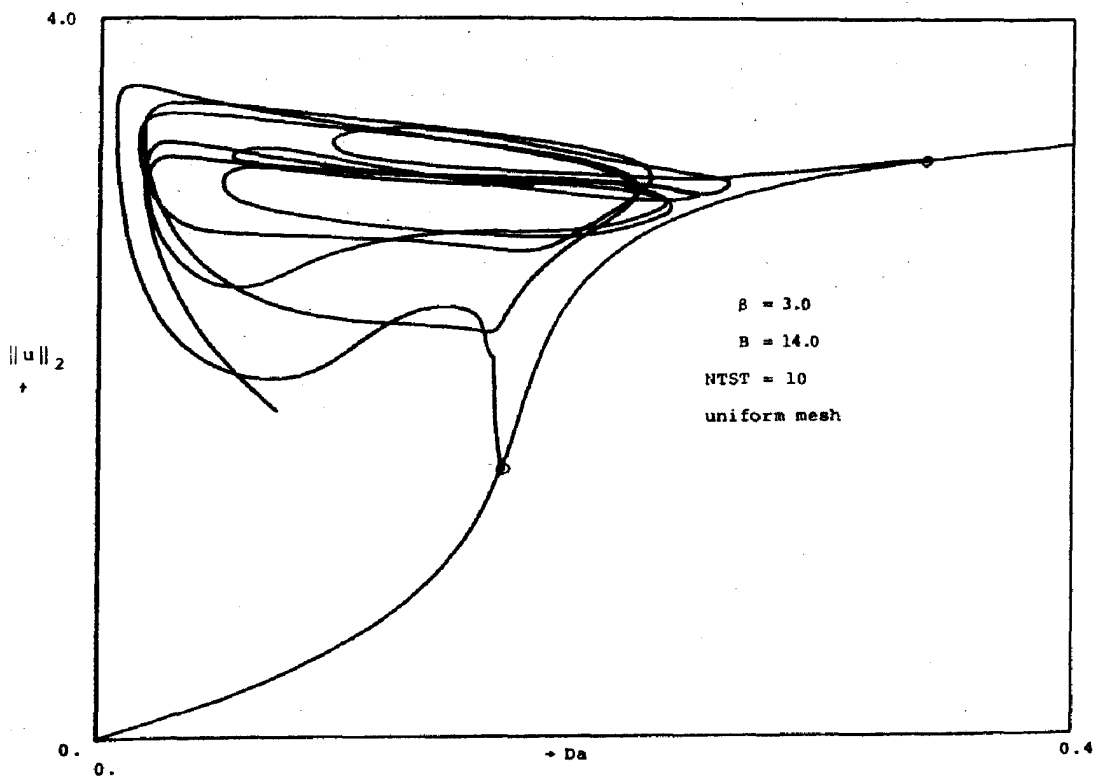
Further, we use the specific values  $\beta = 3$ ,  $B = 14$ . To remove the nonuniqueness, due to the fact that a periodic solution can be freely translated in time, we impose the orthogonality condition

$$(4.13) \quad (u(0) - u_0(0))^T f(u_0(0), \lambda_0) = 0.$$

Here  $(u_0, \lambda_0, \rho_0)$  denotes a given solution, while  $(u, \lambda, \rho)$  is the solution to be determined from (4.11), (4.12), (4.13) and

$$(4.14) \quad \|u - u_0\|^2 + (\lambda - \lambda_0)^2 + (\rho - \rho_0)^2 = \Delta s^2.$$

This procedure is repeated stepwise along a branch of periodic solutions. We discretize (4.11) by the method of collocation at Gauss points, using the piecewise polynomial space whose elements are globally continuous and quadratic polynomials in each mesh interval. With an adaptive mesh, 71 mesh points and two collocation points per mesh interval, the corresponding bifurcation diagram in [19] is recovered. (See Fig. 4.4.) The primary solution branch is a branch of steady states, while the secondary branch consists of periodic solutions. Along a significant portion of the periodic branch the solution changes very rapidly in a very small interval. Moreover, the location of this small interval does not remain fixed along the branch. The adaptive mesh is for this reason a necessity. For example, 71 uniformly spaced mesh points do not reproduce Fig. 4.4. A typical result is given in Fig. 4.5. Here an insufficient number of 10 uniformly distributed

FIG. 4.4. *Bifurcation diagram for (4.10).*FIG. 4.5. *Effect of an inaccurate discretization on the bifurcation diagram of (4.10).*

mesh points is used. Note the abundance of extraneous solutions. The ability of the method to compute past turning points is especially well illustrated, although the resulting diagram may have little physical significance.

*Remark.* Although the side condition (4.13) is theoretically sound, i.e., local existence theorems can be based upon it, for practical numerical computation the integrated form

$$\int_0^{2\pi} (u(t) - u_0(t)^T) f(u_0(t), \lambda_0) dt = 0$$

is preferable. The modified condition minimizes motion of regions of rapid change in  $u$  when progressing along a branch of periodic solutions. This property significantly aids the efficiency of automatic mesh selection. Details will be presented elsewhere.

**Acknowledgment.** The second author wishes to thank Dr. Rita Meyer-Spasche of the Max-Planck Institut für Plasmaphysik for some stimulating discussions.

#### REFERENCES

- [1] U. ASCHER, J. CHRISTIANSEN AND R. D. RUSSELL, *A collocation solver for mixed order systems of boundary value problems*, Math. Comp., 33 (1978), pp. 659–679.
- [2] W.-J. BEYN, *On the convergence of the finite difference method for nonlinear ordinary boundary value problems*, in Constructive Methods for Nonlinear Boundary Value Problems and Nonlinear Oscillations, Proc. Oberwolfach Research Inst., Nov. 19–25, 1978, J. Albrecht, L. Collatz, K. Kirchgässner, eds., Birkhäuser-Verlag, Basel, 1979, ISNM 48, pp. 9–19.
- [3] E. BOHL, *On the bifurcation diagram of discrete analogues for ordinary bifurcation problems*, Math. Meth. Appl. Sci., 1 (1979), pp. 566–571.
- [4] C. DE BOOR AND B. SWARTZ, *Collocation at Gaussian points*, SIAM J. Numer. Anal., 10 (1973), pp. 582–606.
- [5] E. J. DOEDEL, *Finite difference collocation methods for nonlinear two point boundary value problems*, SIAM J. Numer. Anal., 16 (1979), pp. 173–185.
- [6] E. J. DOEDEL, *Some stability theorems for finite difference collocation methods on nonuniform meshes*, BIT, 20 (1980), pp. 58–66.
- [7] R. GAINES, *Difference equations with boundary value problems for second order nonlinear ordinary differential equations*, SIAM J. Numer. Anal., 11 (1974), pp. 411–433.
- [8] R. D. GRIGORIEFF, *Diskrete Approximation von Eigenwertproblemen, II, Konvergenzordnung*, Numer. Math., 24 (1975), pp. 415–433.
- [9] H. B. KELLER, *Approximation methods for nonlinear problems with application to two point boundary value problems*, Math. Comp., 29 (1975), pp. 464–474.
- [10] H. B. KELLER, *Numerical solution of bifurcation and nonlinear eigenvalue problems*, in Applications of Bifurcation Theory, P. H. Rabinowitz, ed., Academic Press, New York, 1977.
- [11] H. O. KREISS, *Difference approximations for boundary and eigenvalue problems for ordinary differential equations*, Math. Comp., 26 (1972), pp. 605–624.
- [12] T. R. LUCAS AND G. W. REDDIEN, *Some collocation methods for nonlinear boundary value problems*, SIAM J. Numer. Anal., 9 (1972), pp. 341–356.
- [13] R. E. LYNCH AND J. R. RICE, *A high-order difference method for differential equations*, Math. Comp., 34 (1980), pp. 333–372.
- [14] M. R. OSBORNE, *A method for finite difference approximation to ordinary differential equations*, Computer J., 7 (1964), pp. 58–65.
- [15] M. R. OSBORNE, *Collocation, difference equations, and stitched function representations*, Proc. Dublin Conf. in Numerical Analysis, Dublin, 1974.
- [16] G. W. REDDIEN, *A survey of projection methods*, SIAM Rev., 22 (1980), pp. 156–171.
- [17] R. D. RUSSELL AND L. F. SHAMPINE, *A collocation method for boundary value problems*, Numer. Math., 19 (1972), pp. 1–28.
- [18] B. SWARTZ, *Compact implicit difference schemes for a differential equation's side conditions*, Math. Comp., 35 (1980), pp. 733–746.
- [19] A. UPPAL, W. H. RAY AND A. B. POORE, *On the dynamic behaviour of continuous stirred tank reactors*, Chem. Eng. Sci., 29 (1974), pp. 976–985.
- [20] J. M. VARAH, *On the stability of boundary conditions for separable difference approximations to parabolic equations*, SIAM J. Numer. Anal., 14 (1977), pp. 1114–1125.