SIAM J. NUMER. ANAL.
Vol. 15, No. 1, February 1978

# NUMERICAL METHODS FOR A CLASS OF FINITE DIMENSIONAL BIFURCATION PROBLEMS*

WERNER C. RHEINBOLDT†

**Abstract.** For an equation $H(y, t)=0$, where $H: D \subset R^{n+1} \to R^n$, let $p: J \subset R^1 \to R^n$ be a primary solution on which a simple bifurcation point $p^* = p(t^*)$ with rank $H_y(p^*, t^*) = n-1$ has been detected and a secondary solution is branching off. An iterative process is presented which starts at a point $p^0 = p(t_0)$ near $p^*$ and converges to a point on the secondary curve. It is similar in form to methods proposed by H. B. Keller and others but has considerably lower computational complexity. The process represents a chord iteration with singular iteration matrix and its convergence is derived from a general result for such singular chord iterations. Computational details for the implementation of the method and an informal program are given. Finally, some comments about extensions to the case rank $H_y(p^*, t^*) < n-1$ are made.

## 1. Introduction.

We consider equations of the form

$$H(y, t)=0 \tag{1.1}$$

with sufficiently differentiable mappings $H: D \subset R^{n+1} \to R^n$. Such equations arise in many applications where the principal task is the computation of certain solution curves of (1.1), that is, of smooth, simple, and connected paths $p: J \subset R^1 \to R^n$ for which

$$(p(t), t) \in D, \quad H(p(t), t)=0, \quad \forall t \in J. \tag{1.2}$$

In general, some tracing procedure is used to compute such curves (see, e.g., [8], [9]) and numerical difficulties can be expected to arise whenever the Jacobian $H_y(p(t), t)$ becomes singular.

We restrict ourselves here to the special case of simple bifurcation points where the rank-deficiency of $H_y$ at the singularity is one and a secondary solution is branching off the primary curve. This is a frequent case in practical problems, and we require a numerical method which produces on the secondary branch a point near the singularity from where a successful trace of that curve can be started.

There is a large literature on the theoretical analysis of bifurcation phenomena, but it includes only very few results about related computational methods. Several articles, notably [2], [3], [4], [5], [6], and [10], discussed iterative processes that are applicable in the above situation. However, the results were essentially theoretical in nature and for various reasons the processes are limited in computational efficiency.

The methods in [4], [5], and [6] involve at each step the solution of a nonlinear equation and hence are numerically cumbersome. In [2] and [3] this aspect of the methods was removed but the resulting iterations require the computation of certain second derivatives of $H$. In practice, $H$ often is a discrete analogue of an infinite-dimensional operator. Hence the dimension $n$ is large, and the evaluation of the first derivative of $H$ is costly while that of the second

* Received by the editors October 25, 1976.

† Computer Science Center, University of Maryland, College Park, Maryland 20742. This work was in part supported by the National Science Foundation under Grant GJ-35568X.

derivative is out of the question. All the mentioned processes presume the primary curve to be the zero solution. In other words, a local transformation is used implicitly under which the primary curve becomes a coordinate axis. During the iteration this manifests itself as a requirement for points on the primary curve arbitrarily near the singularity to be computable. This is not feasible because of the mentioned numerical difficulties encountered by most tracing procedures near the singularity.

In this article we present a new iterative process for determining the desired point on the secondary curve. It is similar in form to those in [2]–[6] but exhibits none of the mentioned numerical difficulties. In particular, we require only certain, readily available information about the primary curve at a single point near the singularity. The entire process involves exactly one evaluation of $H'$ and none of $H''$; otherwise, $H$ needs to be computed once per step and a linear system with the same matrix but different right sides has to be solved repeatedly. With this, the method has the same computational requirements as many tracing procedures and hence is easily usable in conjunction with them.

The process under discussion has the form of a chord method except that the iteration matrix is singular. The same is true for the procedures given in [2]–[6]. Therefore, we begin in § 2 with an extension of the standard semilocal convergence theorem for chord iterations (see, e.g., [7]) to this singular case. Then in § 3 a direct application of this general result leads readily to a convergence proof for the method considered here. The computational implementation of singular chord methods does not appear to have been addressed anywhere. For our process these computational aspects are discussed in § 4 and incorporated in an informal program. Analogous techniques should also apply to other singular chord iterations. Finally, in § 5 we end with some comments and open questions about the case of bifurcation points where the Jacobian has rank-deficiency larger than one.

**2. Singular chord methods.** The processes discussed here all have the form of a chord method

$$A(x^{k+1}-x^k)+Fx^k=0, \qquad k=0,1,\cdots, \tag{2.1a}$$

for determining a zero of a certain mapping $F: D_F \subset R^m \to R^m$. However they differ from the standard chord methods in that the matrix $A \in L(R^m)$ is singular with, say,

$$\operatorname{rank} A < m, \quad (\ker A)^{\perp} = X, \quad \operatorname{rge} A = Y. \tag{2.2}$$

Then, for (2.1a) to be solvable, we introduce the requirement

$$Fx \in Y, \quad \forall x \in D_F. \tag{2.3}$$

For any $b \in Y$, the equation $Ax = b$ has a unique (deflated) solution $x \in X$. Hence, if we add to (2.1a) the conditions

$$x^k \in X, \qquad k=0,1,\cdots, \tag{2.1b}$$

then the iterates $\{x^k\}$ are uniquely defined as long as they remain in $D_F$.

Under the assumption (2.2) the bound

$$\nu(A)=\left[\inf_{x\in X,x\neq 0}\frac{\|Ax\|}{\|x\|}\right]^{-1} \tag{2.4}$$

is finite and we have

$$\|x\|\leqq\nu(A)\|Ax\|,\quad \forall x\in X. \tag{2.5}$$

With this we may formulate the following convergence result.

THEOREM 1. *Let $F$: $D_F\subset R^m\to Y$ be of class $C^1$ on $D_F$ with*

$$\|F'(x)-F'(\bar{x})\|\leqq\gamma\|x-\bar{x}\|,\quad \forall x,\bar{x}\in D_F, \tag{2.6}$$

*and suppose that $A\in L(R^m)$ satisfies* (2.2). *If $x^0\in D_F\cap X$ was chosen such that*

$$\begin{aligned}&\text{(a)}\quad \nu(A)\|A-F'(x^0)\|\leqq\delta<1,\\ &\text{(b)}\quad \nu(A)\gamma\|x^1-x^0\|<\tfrac{1}{2}(1-\delta)^2,\\ &(c)\quad \bar{B}(x^0,\rho)\subset D_F\quad \text{for } \rho=\frac{2}{1-\delta}\|x^1-x^0\|,\end{aligned} \tag{2.7}$$

*then the iterates* (2.1a/b) *remain in the set $\mathcal{S}=\bar{B}(x^0,\rho)\cap X$ and converge to the unique solution of $Fx=0$ in $S$.*

*Proof.* For any $x\in S$ the solution $\bar{x}\in X$ of $A(\bar{x}-x)+Fx=0$ is unique and hence the association $x\to\bar{x}$ defines a map $G$ from $S$ into $X$. Since

$$A(G\bar{x}-Gx)=Fx-F\bar{x}-A(x-\bar{x}),\quad \forall x,\bar{x}\in S,$$

we obtain from (2.5), (2.6) and the mean value theorem that

$$\begin{aligned}\|G\bar{x}-Gx\|&\leqq\nu(A)\{\|Fx-F\bar{x}-F'(x^0)(x-\bar{x})\|+\|F'(x^0)-A\|\,\|x-\bar{x}\|\}\\ &\leqq\tfrac{1}{2}\nu(A)\gamma(\|x-x^0\|+\|\bar{x}-x^0\|)\|x-\bar{x}\|+\delta\|x-\bar{x}\|,\end{aligned} \tag{2.8}$$

whence

$$\|G\bar{x}-Gx\|\leqq(\nu(A)\gamma\rho+\delta)\|x-\bar{x}\|,\quad \forall x,\bar{x}\in S, \tag{2.9}$$

where, by (2.7), $\nu(A)\gamma\rho+\delta<1$. Thus $G$ is a contraction on $S$. For any $x\in S$ it follows from (2.8) that

$$\begin{aligned}\|Gx-x^0\|&\leqq\|Gx-Gx^0\|+\|x^1-x^0\|\\ &\leqq\tfrac{1}{2}\nu(A)\gamma\|x-x^0\|^2+\delta\|x-x^0\|+\|x^1-x^0\|\\ &\leqq\tfrac{1}{2}\nu(A)\gamma\rho^2+\delta\rho+\|x^1-x^0\|<\|x^1-x^0\|\left(1+\frac{2}{1-\delta}+1\right)=\rho.\end{aligned}$$

Therefore, $G$ maps $S$ into itself and the result is a direct consequence of the contraction mapping theorem.

The conditions (2.6) may be weakened along the lines of the Kantorovich theorem for Newton's method. The proof then has to rely on majorizing sequences (see [7]).

For later use we note that when two matrices $A_1, A_2 \in L(R^m)$ have the same nullspace $\mathcal{N} = \ker A_1 = \ker A_2$, then

$$\nu(A_2) \leqq \nu(A_1)(1 + \nu(A_2)\|A_1 - A_2\|). \tag{2.10}$$

In fact, for any $x \in \mathcal{N}^{\perp}$ it follows that

$$\begin{aligned}\|x\| &\leqq \nu(A_1)\|A_1 x\| \leqq \nu(A_1)[\|A_1 - A_2\|\,\|x\| + \|A_2 x\|] \\ &\leqq \nu(A_1)[\|A_1 - A_2\|\nu(A_2) + 1]\|A_2 x\|\end{aligned}$$

which implies (2.10).

**3. The basic method.** As in the introduction, we consider (1.1) with a mapping $H: D \subset R^{n+1} \to R^n$ which is now assumed to be of class $C^3$. Furthermore, let $p: J \subset R^n \to R^1$ be a simple path which is of class $C^2$ on the interval $J$ and satisfies there (1.2). For ease of notation, we write

$$z = \begin{pmatrix} y \\ t \end{pmatrix} \in R^{n+1}, \qquad q(t) = \begin{pmatrix} p(t) \\ t \end{pmatrix}, \quad \forall t \in J. \tag{3.1}$$

Suppose that the Jacobian $H_y(q(t))$ has an isolated singularity at $q^* = q(t^*)$, $t^* \in \operatorname{int} J$. As stated before, we restrict ourselves to the case where $L = H_y(q^*)$ has rank-deficiency one, that is,

$$\operatorname{rank} L = n - 1, \quad Lu = L^T v = 0, \quad u^T u = v^T v = 1. \tag{3.2}$$

The smoothness conditions for $H$ and $p$ evidently ensure that

$$H'(q(t))\dot{q}(t) = 0, \quad H''(q(t))(\dot{q}(t), \dot{q}(t)) + H'(q(t))\ddot{q}(t) = 0, \quad \forall t \in \operatorname{int} J. \tag{3.3}$$

The first of these relations shows that $H'(q^*)$ has rank $n-1$ and that the nullspace of $H'(q^*)$ is spanned by the vectors $r^* = \dot{q}(t^*)$ and $u^* = (u, 0)^T$. If

$$v^T c^* \neq 0 \quad \text{with } c^* = H''(q^*)(r^*, u^*), \tag{3.4}$$

then a theorem of Crandall and Rabinowitz [1] asserts that in some neighborhood of $q^*$ the totality of solutions of (1.1) consists of two continuous curves in $R^{n+1}$ intersecting at $q^*$, one of which is the primary curve.

Our aim is the construction of a computationally efficient iteration which converges to a point near $q^*$ on the secondary curve. The generic form of this type of bifurcation iteration derives from standard perturbation arguments (see [6]). For given (small) $\varepsilon \neq 0$ and $\eta \neq 0$ we introduce the notation

$$t_0 = t^* + \eta, \quad q^0 = q(t_0), \quad r^1 = \dot{q}(t_0), \quad r^2 = \ddot{q}(t_0), \tag{3.5a}$$

and the mappings

$$\begin{aligned}&\psi: R^1 \to R^{n+1}, \quad \psi(s) = q^0 + \varepsilon s r^1 + \tfrac{1}{2}\varepsilon^2 s^2 r^2, \\ &\Phi: R^{n+1} \to R^{n+1}, \quad \Phi(x) = \psi(s) + \varepsilon u^* + \varepsilon^2 \begin{pmatrix} w \\ 0 \end{pmatrix}, \quad \forall x = \begin{pmatrix} w \\ s \end{pmatrix} \in R^{n+1}, \\ &F: D_F \subset R^{n+1} \to R^{n+1}, \quad Fx = VH(\Phi(x)), \quad \forall x \in D_F = \Phi^{(-1)}(D), \\ &V = \begin{pmatrix} I \\ v^T \end{pmatrix}.\end{aligned} \tag{3.5b}$$

Then the process to be considered here has the form

$$A(x^{k+1}-x^k)+Fx^k=0, \qquad (u^*)^T x^{k+1}=0, \qquad k=0,1,\cdots, \quad x^0=0, \tag{3.6a}$$

where

$$A=\begin{pmatrix}\varepsilon^2 L & \varepsilon c\\ 0 & \varepsilon v^T c\end{pmatrix} \tag{3.6b}$$

and $c=c(\varepsilon,\eta)\in R^n$ is a suitably chosen vector such that $v^T c\neq 0$. We shall use the readily computable vector

$$c=H'(z^0)r^1, \qquad z^0=q^0+\varepsilon u^*. \tag{3.6c}$$

The methods in [2]–[6] are developed in terms of linearized equations. For the equation (1.1) they have also the general form (3.6) but with $\psi$ replaced by the primary curve $q(t^*+\varepsilon s)$. As indicated earlier, this leads to numerical difficulties since, in general, $q$ cannot be computed accurately near $t^*$. In [3] the vector $c=\varepsilon c^*$ is used which involves the second derivative of $H$ and hence is inaccessible. Formally, the method in [6] is obtained when $c=0$ and $s^{k+1}$ is a (specific) root of the nonlinear equation

$$v^T H(p(t^*+\varepsilon s)+\varepsilon u+\varepsilon^2 w^k, t^*+\varepsilon s)=0.$$

Let $\rho_0>0$, $\sigma_0>0$ be such that $\bar{B}(q^*,\rho_0)\subset D$ and $J_0=[t^*-\sigma_0, t^*+\sigma_0]\subset J$, and suppose that

$$\begin{aligned}&\|H''(z)\|\leqq\alpha_1, \quad \|H'''(z)\|\leqq\alpha_2, \quad \forall z\in\bar{B}(q^*,\rho_0),\\ &\|\dot{q}(t)\|\leqq\beta_1, \quad \|\ddot{q}(t)\|\leqq\beta_2, \quad \forall t\in J_0.\end{aligned} \tag{3.7}$$

In the following, we restrict $\varepsilon$ and $\eta$ to $0<|\eta|\leqq|\varepsilon|\leqq\varepsilon_0\leqq\sigma_0$. Then (3.5) and (3.7) imply that

$$\begin{aligned}&\|\Phi'(x)\|\leqq(\gamma_1+\gamma_2\|x\|)|\varepsilon|, \quad \|\Phi''(x)\|\leqq\beta_2\varepsilon^2, \quad \forall x\in R^{n+1},\\ &\|\Phi(x)-q^*\|\leqq(1+\beta_1+\gamma_1\|x\|+\tfrac{1}{2}\gamma_2\|x\|^2)|\varepsilon|, \quad \forall x\in R^{n+1},\end{aligned} \tag{3.8}$$

with $\gamma_1=\sigma_0+\beta_1$, $\gamma_2=\sigma_0\beta_2$. Hence for any given ball $B_\rho=\bar{B}(0,\rho)$, $\rho>0$, we have

$$\Phi(B_\rho)\subset\bar{B}(q^*,\rho_0), \quad \forall\varepsilon_0\leqq\varepsilon_1(\rho)=\rho_0/(1+\beta_1+\gamma_1\rho+\tfrac{1}{2}\gamma_2\rho^2). \tag{3.9}$$

Let $x, \bar{x}\in B_\rho$ and thus $z, \bar{z}\in B(q^*,\rho_0)$ for $z=\Phi(x)$, $z=\Phi(\bar{x})$. Then for any $h\in R^{n+1}$ we may write

$$\begin{aligned}(F'(x)-F'(\bar{x}))h = V\{&[H'(z)-H'(\bar{z})](\Phi'(x)-\Phi'(0))h\\ &+[H'(\bar{z})-H'(q^0)](\Phi'(x)-\Phi'(\bar{x}))h\\ &+[H'(z)-H'(\bar{z})-H''(q^0)(z-\bar{z})]\Phi'(0)h\\ &+H''(q^0)(z-\bar{z}-\Phi'(0)(x-\bar{x}),\Phi'(0)h)\\ &+[H''(q^0)(\phi'(0)(x-\bar{x}),\Phi'(0)h)\\ &\qquad +H'(q^0)\Phi''(x)(x-\bar{x},h)]\}.\end{aligned} \tag{3.10}$$

The first four terms on the right may be estimated in norm by application of the mean-value theorem and the inequalities (3.7), (3.8). It is easily seen that in each case we obtain a bound of the form

$$\text{const}\,|\varepsilon|^3\|x-\bar{x}\|\,\|h\|, \tag{3.11}$$

with a constant that depends at most cubically on $\rho$. For instance, for the first term (including the factor $V$) we find the constant $2\alpha_2\beta_2(\gamma_1+\gamma_2\rho)\rho$. In order to estimate the fifth term, let $h^i=(y^i,t_i)\in R^{n+1}$, $i=1,2$, be any two vectors. Then

$$\begin{aligned}\|H''(q^0)(\Phi'(0)h^1,\Phi'(0)h^2)+H'(q^0)\Phi''(0)(h^1,h^2)\|&\\ \leqq \varepsilon^4\|H''(q^0)(y^1,y^2)\|+|\varepsilon|^3\|H''(q^0)(y^1,t_2r^1)\|+|\varepsilon|^3\|H''(q^0)(y^2,t_1r^1)\|&\\ +\varepsilon^2\|t_1t_2[H(q^0)(r^1,r^1)+H'(q^0)r^2]\|,&\end{aligned}$$

and by (3.3) the last term vanishes. This shows that also the last term of (3.10) has a bound of the form (3.11). Hence, altogether we found that

$$\|F'(x)-F'(\bar{x})\|\leqq\gamma_3(\rho)|\varepsilon|^3\|x-\bar{x}\|,\quad \forall x,\bar{x}\in B_\rho, \tag{3.12}$$

where $\gamma_3$ is a cubic polynomial in $\rho$ with coefficients that depend only on the constants of (3.7) and (3.8).

From

$$\begin{aligned}\|c-\varepsilon c^*\|\leqq\|(H'(z^0)-H'(q^0)-H''(q^0)\varepsilon u^*)r^1\|&\\ +|\varepsilon|\,\|(H''(q^0)-H''(q^*))(r^1,u^*)\|+|\varepsilon|\,\|H''(q^*)(r^1-r^0,u^*)\|&\end{aligned}$$

we obtain that

$$\|c-\varepsilon c^*\|\leqq\gamma_4\varepsilon^2 \tag{3.13}$$

and hence, for sufficiently small $\varepsilon_0$, that

$$|v^Tc|\geqq|\varepsilon|\Big[|v^Tc^*|-\frac{1}{|\varepsilon|}|v^T(c-\varepsilon c^*)|\Big]\geqq\gamma_5|\varepsilon|,\qquad \gamma_5>0, \tag{3.14}$$

Thus $A(\varepsilon)$ has rank $n$ and

$$A(\varepsilon)u^*=A(\varepsilon)^Tv^*=0,\quad v^*=\frac{1}{\sqrt{2}}\begin{pmatrix}v\\-1\end{pmatrix},\quad (v^*)^TFz=0,\quad \forall z\in D_F.$$

This shows that (3.6a, b, c) is a singular chord method and it remains only to verify the conditions (2.7) of Theorem 1 for the starting vector $x^0=0$.

Let $A^*$ be the matrix (3.6b) with $c=\varepsilon c^*$. Then, clearly, $\nu(A^*)\leq\gamma_6/\varepsilon^2$ for some constant $\gamma_6$, and (3.11) implies that $\|A(\varepsilon)-A^*\|\leqq 2\gamma_4|\varepsilon|^3$. Thus, for small enough $\varepsilon_0$, it follows from (2.14) that

$$\nu(A(\varepsilon))\leqq\frac{\gamma_6/\varepsilon^2}{1-2\gamma_4\gamma_6|\varepsilon|}\leqq 2\frac{\gamma_6}{\varepsilon^2}.$$

By definition of $c$ we have

$$\|F'(0)-A(\varepsilon)\|=\varepsilon^2\left\|\begin{pmatrix}H_y(z^0)-L & 0\\ 0 & 0\end{pmatrix}\right\|\leqq\alpha_1(1+\beta_1)|\varepsilon|^3$$

and hence—after a further restriction of $\varepsilon_0$—

$$\nu(A(\varepsilon))\|F'(0)-A(\varepsilon)\| \leqq 2\alpha_1\gamma_6(1+\beta_1)|\varepsilon| \leqq \delta < 1$$

which is (2.6a). For simplicity, suppose that $\delta \leqq 1/2$. Then, with

$$\|x^1\| \leqq \nu(A(\varepsilon))\|F0\| \leqq 2\frac{\gamma_6}{\varepsilon^2}\|V\|\,\|Hz^0 - Hq^* - H'(q^*)(\varepsilon u^*)\| \leqq \gamma_7$$

the condition (2.6c) requires that $\rho = 8\gamma_7$. By (3.9) this is an allowable radius provided only that $\varepsilon_0 \leqq \varepsilon_1(\rho)$, and with it we find, after a possible further decrease of $\varepsilon_0$, that

$$\frac{1}{(1-\delta)^2}\nu(A(\varepsilon))\gamma_3(8\gamma_7)|\varepsilon|^3\gamma_7 \leqq 8\gamma_6\gamma_7\gamma_3(8\gamma_7)|\varepsilon| < \frac{1}{2}.$$

This completes the verification of (2.6) and therefore the proof of the following result:

THEOREM 2. *Let $H$ and $q$ satisfy the stated smoothness conditions and assume that* (3.2) *and* (3.4) *hold. Then there exist constants $\varepsilon_0 > 0$, $\rho > 0$ such that for any given $\varepsilon$, $\eta$ with $0 < |\eta| \leqq |\varepsilon| \leqq \varepsilon_0$, the iterates $\{x^k\}$ specified by* (3.5) *and* (3.6) *remain in the ball $\bar{B}(0,\rho) \subset \Phi^{(-1)}(D)$ and converge to the unique solution $x^* = x(\varepsilon)$ of $Fx = 0$ in that ball.*

Clearly, $z^* = \Phi(x^*)$ is a solution of (1.1) and

$$\|z^* - q(t^* + \varepsilon s^*)\| \geqq |\varepsilon|(1-(1+\beta_2)\rho^2\varepsilon_0), \qquad x^* = (w^*, s^*)^T,$$

implies that for small $\varepsilon_0$ the point $z^*$ must be on the secondary curve.

**4. Numerical aspects.** We turn now to the computational aspects of the process (3.6). In general, we expect to arrive in the neighborhood of the bifurcation point $q^*$ while tracing the primary curve $q = q(t)$, $t \in J$. Most tracing processes simply jump over the singular point, particularly when a steplength algorithm of the type in [8], [9] is employed. In our case (3.2) of rank-deficiency one the determinant $\mathcal{d}(t) = \det H_y(q(t))$ changes sign at $t = t^*$. Hence, in order to detect the singularity we monitor the sign of $\mathcal{d}$ on $J$. Many tracing algorithms involve the derivative $H'(q(t))$ and the tangent vector $\dot{q}(t)$. Hence $\mathcal{d}(t)$ is an easy byproduct of the trace calculations.

Let $t_1 < t^* < t_2$ be the $t$-step where the sign change of $\mathcal{d}$ has been detected. If desirable, we may reduce the steplength $\Delta t = t_2 - t_1$ by computing further points on the primary curve for $t$-values inside the interval. Once this leads to numerical difficulties or the steplength is sufficiently small, we choose $t_0$ as the endpoint of the interval where $|\mathcal{d}|$ is minimal. Numerical approximations $p^0$ and $\dot{p}^0$ of $p(t_0)$ and $\dot{p}(t_0)$, respectively, are direct results of the tracing process, and the difference quotient $(1/\Delta t)(\dot{p}(t_2) - \dot{p}(t_1))$ (based on the last interval) may be expected to provide a satisfactory estimate $\ddot{p}^0$ of $\ddot{p}(t_0)$.

The eigenvectors $u$ and $v$ of the (unknown) matrix $L = H_y(q^*)$ can be approximated by applying the inverse power method to $L_0 = H_y(q^0)$. More specifically, starting with some vector, say, $u = (1, 1, \cdots, 1)^T$, we iterate according to the formulas

(4.1) 1. Solve $L_o\tilde{u} = u$; 2. $u = \tilde{u}/\|u\|$; 3. Solve $L_0^T\tilde{v} = u$; 4. $v := \tilde{v}/\|v\|$.

The matrix $L_0$ was already used during the trace of the primary curve and is nonsingular by construction of $q^0$. By (3.7) we have $\|L_0 - L\| \leqq \alpha_1 \beta_1 |t_0 - t^*|$, and hence, for sufficiently small $\varepsilon_0$, a few iterations of (4.1) will usually suffice to produce the desired nullvectors $u$ and $v$ with reasonable accuracy. For details about the process we refer, for instance, to [11], [14], [15]. It is advisable to restart the method with a vector orthogonal to the one originally found; this provides a test for the basic rank-deficiency assumption (3.2).

Each step of the process (3.6) involves the numerical solution of

$$Lw = b, \qquad u^T w = 0 \tag{4.2}$$

with some vector $b \in \operatorname{rge} L$. For this we follow an approach proposed in [12]. With the orthogonal projections $P_u = I - uu^T$ and $P_v = I - vv^T$ onto $(\ker L)^\perp$ and $\operatorname{rge} L$, respectively, the solution of (4.2) is the unique vector $\tilde{w}$ for which

$$P_v L P_u \tilde{w} = P_v b, \qquad P_u \tilde{w} = \tilde{w}. \tag{4.3}$$

This suggests that we replace $L$ once again by the neighboring nonsingular matrix $L_0$ and then apply a form of iterative refinement. More specifically, starting from $w = 0$ we iterate according to the formulas

$$1.\ d := P_v(b - L_0 w); \quad 2.\ \text{Solve } L_0 \tilde{w} = d; \quad 3.\ w := w + P_u \tilde{w}. \tag{4.4}$$

Note that any vector $w$ generated by (4.4) satisfies $P_u w = w$. Hence, if the iterates have the limit $w^*$, then $P_v L_0 P_u w^* = P_v b$, $P_u w^* = w^*$, and a comparison with (4.3) shows that $\|w^* - \tilde{w}\|$ is bounded by $O(|\Delta t|)$. For a convergence analysis of the technique we refer to [12]. In general, the convergence may be expected to be fast.

Altogether then, we arrive at the following schematic algorithm. Here $w$ and $s$ correspond to the quantities $\varepsilon^2(w^{k+1} - w^k)$ and $\varepsilon(s_{k+1} - s_k)$ of (3.6), respectively.

1. Input $(t_0, p^0, \dot{p}^0, \ddot{p}^0, L_0, \varepsilon)$.
2. Use (4.1) to determine the approximate null vectors $u$, $v$ and to test (3.2).
3. Initialize: $y := p^0 + \varepsilon u$; $t := t_0$; $s := 0$.
4. $c := H_y(y, t)p^0 + H_t(y, t)$.
5. $\alpha := |v^T c|$; test whether $\alpha$ is too small.
6. Loop
   1. $h := H(y, t)$.
   2. $\tau := -v^T h/\alpha$.
   3. $b := -(\tau c + h)$.
   4. Use (4.4) to approximate the solution $w$ of (4.2).
   5. $\hat{s} := s + \frac{1}{2}\tau$.
   6. $y := y + \tau(\dot{p}^0 + \hat{s}\ddot{p}^0) + w$.
   7. $s := s + \tau$.
   8. $t := t^0 + s$.
7. Output $(y, t)$.

The process requires exactly one evaluation of $H'$, namely, for the calculation of $c$. Otherwise, we need to compute only $H$ itself and solve linear systems involving $L_0$; in other words, we have to decompose only one $n \times n$ matrix. If $\alpha$ is too small, either $\varepsilon$ was too small, or the problem is badly conditioned, in which case we may expect to encounter difficulties in steps 2 and 6.4 as well.

The method has been applied extensively with excellent success to a range of problems from structural engineering. In general, our experience has been that the convergence is very satisfactory (albeit linear) and less sensitive to the choice of $\varepsilon$ than to that of the length $|\Delta t|$ of the interval bracketing $t^*$. One of our examples was a finite element model of a clamped, thin, shallow, circular arch discussed in [13]. In that case the number of sweeps through loop 5 required to achieve a tolerance $\|w\|+|s| \leqq 10^{-5}$ are shown in Table 1.

TABLE 1

| $\|\Delta t\|$ \ $\varepsilon$ | .1 | .05 | .025 | .01 |
|---|---|---|---|---|
| .25 | 9 | 9 | 9 | 8 |
| .125 | 7 | 7 | 7 | 7 |
| .0625 | 5 | 5 | 5 | 5 |

The computations were performed on a UNIVAC 1108 in single precision and the dimension of the problem was $n = 45$. In practice the $\varepsilon$ and $\Delta t$ values are usually smaller; but the result is rather representative, at least, in the class of structural problems we considered.

**5. Comments on higher rank deficiencies.** Suppose that $H$ and $p$ satisfy the conditions of § 3 but that instead of (3.2) we have

$$\begin{aligned} &\operatorname{rank} L = n - l, \quad l > 1, \\ &LU = L^T V = 0 \quad \text{with orthogonal } U,\ V \in L(R^m, R^n). \end{aligned} \tag{5.1}$$

With this the problem has become considerably more complicated, and it is not surprising that the resulting numerical method has certain undesirable aspects.

In place of the condition (3.4) we need to consider now the algebraic bifurcation equation

$$V^T H''(q^*)(s\dot{q}(t^*)+\tfrac{1}{2}\bar{u}, \bar{u}) = 0, \quad \bar{u} = \begin{pmatrix} Ua \\ 0 \end{pmatrix} \in R^{n+1}, \quad a^T a = 1 \tag{5.2}$$

in the unknown vector $a \in R^l$ and scalar $s \in R^1$. In [6] it was shown that each isolated root $(a, s)$ of (5.2) corresponds to a complete "secondary" branch of solutions through the bifurcation point $q^*$. The branches for the two roots $(a, s)$ and $(-a, -s)$ are the same. In the case $l = 1$, the condition (3.4) ensures that there is only one such pair of roots and they are isolated.

Following again the ideas in [2]–[6], we may proceed as in § 3. Let $(a^0, s_0)$ be an isolated root of (5.2) and with some $\varepsilon \neq 0$ set $t_0 = t^* + \varepsilon s_0$ and

$$\begin{aligned} &\psi\colon R^1 \to R^{n+1}, \quad \psi(s) = q(t_0) + \varepsilon(s - s_0)\dot{q}(t_0) + \tfrac{1}{2}\varepsilon^2(s - s_0)^2\ddot{q}(t_0), \\ &\Phi\colon R^N \to R^{n+1}, \quad \Phi(x) = \psi(s) + \varepsilon\begin{pmatrix} Ua \\ 0 \end{pmatrix} + \varepsilon^2\begin{pmatrix} w \\ 0 \end{pmatrix}, \quad \forall x = \begin{pmatrix} w \\ a \\ s \end{pmatrix} \in R^N, \\ &\qquad\qquad N = n + l + 1, \\ &F\colon D_F \subset R^N \to R^N, \quad Fx = \begin{pmatrix} H(\Phi(x)) \\ V^T H(\Phi(x)) \\ \varepsilon^2(a^T a - 1) \end{pmatrix}, \quad \forall x \in D_F = \Phi^{(-1)}(D). \end{aligned} \tag{5.3}$$

Then the process to be considered is a singular chord method

$$A(x^{k+1}-x^k)+Fx^k=0, \qquad \tilde{U}^T x^{k+1}=0, \qquad k=0,1,\cdots, \tag{5.4}$$

where $\tilde{U}^T=(U^T,0,0)\in L(R^N,R^l)$ and the component $w^0\in R^n$ of the starting point $x^0=(w^0,a^0,s_0)^T$ is as yet unspecified. The matrix $A$ has the form

$$A=\begin{pmatrix}\varepsilon^2 L & \varepsilon KU & \varepsilon c\\ 0 & \varepsilon V^T KU & \varepsilon V^T c\\ 0 & 2\varepsilon^2(a^0)^T & 0\end{pmatrix} \tag{5.5}$$

with

$$K=H_y(\Phi(x^0)), \quad c=H'(\Phi(x^0))r^1. \tag{5.6}$$

If in (5.3), $\psi(s)$ is replaced by $q(t^*+\varepsilon s)$ and if

$$K=\varepsilon H''(q^*)(s_0\dot{q}(t^*)+\bar{u}^0), \quad c=\varepsilon H''(q^*)(\dot{q}(t^*),\bar{u}^0), \quad \bar{u}^0=\begin{pmatrix}Ua^0\\ 0\end{pmatrix},$$

are used, then (5.4/5) is essentially the process in [3]. The methods in [4]–[6] set $K=0$, $c=0$ and define $a^{k+1}$, $s^{k+1}$ as a certain exact solution of the full bifurcation equation $V^TH(\Phi(x))=0$.

The proof procedure of § 3 carries over easily to the iteration (5.4)–(5.6); that is, Theorem 1 applies for sufficiently small $|\varepsilon|$. We shall not enter into details here. The limit point will be on the secondary branch corresponding to the chosen root $(a^0,s_0)$ of (5.2). However, for the convergence proof we need to choose $w^0$ such that

$$Lw^0=H''(q^*)(s_0\dot{q}(t^*)+\tfrac{1}{2}\bar{u}^0,\bar{u}^0), \qquad U^Tw^0=0. \tag{5.7}$$

Hence, it turns out that all components of the starting point $x^0$ are defined in terms of the second derivative of $H$ at $q^*$. Thereafter, the computational requirements of the process correspond to those of the method (3.6a, b, c), and the implementation techniques of § 4 are readily adapted to it.

Clearly, any such process may be expected to utilize information about the particular secondary branch to which it is to converge. In our case this information is contained principally in $x^0$. At the same time, numerical experiments show that, in general, the convergence of (5.4)–(5.6) is hardly affected by use of an approximation of $x^0$ provided only that the submatrix

$$\begin{pmatrix}V^TKU & V^Tc\\ 2\varepsilon(a^0)^T & 0\end{pmatrix}$$

of $A$ remains nonsingular. For $x^0$ itself this is correct since the root $(a^0,s_0)$ of (5.2) is assumed to be isolated. The question remains open how to compute efficiently such allowable approximations of $x^0$ without using second derivatives of $H$.

## REFERENCES

[1] M. G. CRANDALL AND P. H. RABINOWITZ, *Bifurcation from simple eigenvalues*, J. Functional Analysis, 8 (1971), pp. 321–340.

[2] Y. M. J. DEMOULIN AND Y. M. CHEN, *An iteration method for studying the bifurcation solutions of the nonlinear equations, $L(\lambda)u+\varepsilon R(\lambda,u)=0$*, Numer. Math., 23 (1974), pp. 47–61.

[3] ———, *An iteration method for solving nonlinear eigenvalue problems*, SIAM J. Appl. Math., 28 (1975), pp. 588–595.
[4] J. P. KEENER AND H. B. KELLER, *Perturbed bifurcation theory*, Arch. Rational Mech. Anal., 50 (1973), pp. 159–175.
[5] H. B. KELLER, *Nonlinear bifurcation*, J. Differential Equations, 7 (1970), pp. 417–434.
[6] H. B. KELLER AND W. F. LANGFORD, *Iterations, perturbations and multiplicities for nonlinear bifurcation problems*, Arch. Rational Mech. Anal., 48, (1972), pp. 83–108.
[7] J.M. ORTEGA AND W. C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
[8] W. C. RHEINBOLDT, *On the solution of some nonlinear equations arising in the application of finite element methods*, Mathematics of Finite Elements and Applications, J. Whiteman, ed., Academic Press, London, 1976, pp. 465–482.
[9] ———, *Numerical continuation methods for finite element applications*, U.S.–German Symposium on Finite Elements, J. Bathe, ed., MIT Press, Cambridge, MA., in press.
[10] J. SCHRÖDER, *Störungsrechnung bei Eigenwert und Verzweigungsaufgaben*, Arch. Rational Mech. Anal., 1 (1958), pp. 436–468.
[11] G. W. STEWART, *Introduction to Matrix Computations*, Academic Press, New York, 1973.
[12] ———, *On the implicit deflation of nearly singular systems of linear equations*, Univ. of Maryland, Computer Science Tech. Rep. TR-437, 1976.
[13] A. C. WALKER, *A nonlinear finite element analysis of shallow circular arches*, Int. J. Solids and Structures, 5 (1969), pp. 97–107.
[14] J. H. WILKINSON, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, England, 1965.
[15] ———, *Note on inverse iterations and ill-conditioned eigensystems*, Acta Univ. Carolinae-Math. et Phys., 112 (1974), pp. 173–177.