SIAM J. SCI. COMPUT.
Vol. 21, No. 5, pp. 1835–1850

MESH INDEPENDENCE OF MATRIX-FREE METHODS FOR PATH FOLLOWING*

W. R. FERNG† AND C. T. KELLEY‡

Abstract. In this paper we consider a matrix-free path following algorithm for nonlinear parameter-dependent compact fixed point problems. We show that if these problems are discretized so that certain collective compactness and strong convergence properties hold, then this algorithm can follow smooth folds and capture simple bifurcations in a mesh-independent way.

Key words. path following, bifurcation, collective compactness, fold point, mesh independence, matrix-free method, GMRES, Arnoldi method, singularity

AMS subject classifications. 45G10, 65F10, 65H10, 65H17, 65J15

PII. S1064827598339360

1. Introduction.

The purpose of this paper is to extend the results in [4], [5], and [19] on mesh-independent convergence of the GMRES [24] iterative method for linear equations to a class of matrix-free methods for solution of parameter-dependent nonlinear equations of the form

$$G(u, \lambda) = 0. \tag{1.1}$$

In (1.1) $G : X \times \mathbb{R} \to X$, where X is a Banach space.

We present an algorithm for numerical path following and detection of simple bifurcations together with conditions on a sequence of approximate problems,

$$G^h(u, \lambda) = 0, \tag{1.2}$$

(with $G^0 = G$ for consistency of notation) that imply that the performance of the algorithm is independent of the level h of the discretization. Hence, for such problems, methods, and discretizations, the difficulties raised in [26] and [27] will not arise.

In this section we set the notation and specify the kinds of singularities that we will consider. This discussion, and that of algorithms for path following in section 2.1 and detection of simple bifurcation and branch switching in section 3, does not depend on the discretization, and we use the notation G for both G^0 and G^h for $h > 0$. When we describe our assumptions on the discretization in section 4 and present an example in section 5, the distinction between G^0 and G^h becomes important and the notation in those sections reflects that difference.

1.1. Notation and simple singularities.

We let

$$\Gamma = \{(u, \lambda) \mid G(u, \lambda) = 0, \lambda \in \mathbb{R}\} \tag{1.3}$$

*Received by the editors May 26, 1998; accepted for publication (in revised form) January 4, 1999; published electronically April 28, 2000.
http://www.siam.org/journals/sisc/21-5/33936.html

†Department of Applied Mathematics, National Chiao-Tung University, Hsin-Chu 30050, Taiwan (ferng@math.nctu.edu.tw). The research of this author was supported by National Science Council, Taiwan.

‡North Carolina State University, Department of Mathematics and Center for Research in Scientific Computation, Box 8205, Raleigh, NC 27695-8205 (Tim_Kelley@ncsu.edu). The research of this author was supported by National Science Foundation grant DMS-9700569.

denote the solution path in $X \times \mathbb{R}$ of pairs (u, λ) that satisfy (1.1), G_u denote the Fréchet derivative with respect to u, and G_λ denote the derivative with respect to the scalar λ.

Throughout this paper we make the following assumption.

ASSUMPTION 1.1. *G is continuously Fréchet differentiable in (u, λ). G_u is a Fredholm operator of index zero, and*

$$I - G_u(u, \lambda), \qquad G_\lambda \in \mathcal{COM}(X) \quad \textit{for all } (u, \lambda) \in X \times \mathbb{R}. \tag{1.4}$$

Here $\mathcal{COM}$ denotes the space of compact operators.

For $A \in \mathcal{L}(X)$, the space of bounded operators on X, we let $\mathcal{N}(A)$ be the null space of A and $\mathcal{R}(A)$ be the range of A. Assumption 1.1 implies that $\mathcal{R}(G_u(u, \lambda))$ is closed in X and the dimension of $\mathcal{N}(G_u(u, \lambda))$ is the codimension of $\mathcal{R}(G_u(u, \lambda))$.

Following [16], we say that a point $(u, \lambda) \in \Gamma$

- is a regular point if $G_u(u, \lambda)$ is nonsingular;
- is a simple singularity if 0 is an eigenvalue of G_u with algebraic and geometric multiplicity one,
  - is a simple fold if it is a simple singularity and $G_\lambda \notin \mathcal{R}(G_u)$, and
  - is a simple bifurcation point if it is a simple singularity and $G_\lambda \in \mathcal{R}(G_u)$.

2. Algorithms for path following and branch switching. In this section we describe the iterative methods for path following, detection of bifurcation, and branch switching that we analyze in the subsequent sections and discuss some alternative approaches. Our approach is matrix-free, which means that we use no matrix storage or matrix factorizations at all. However, the results that we prove on mesh-independent well-conditioning of the linear equations that must be solved to compute the Newton step in the corrector equation and the eigenvalue problems that are solved in detection of bifurcation are applicable to other methods. The analysis of the performance of the nonlinear corrector iteration, which depends on the conditioning of the linearized problem, is also relevant to methods that solve the linear system for the step directly, such as those discussed in [15] and [8]. While we focus on a variation of the approach in [12] for detection of bifurcation, our results on mesh-independent well-conditioning of eigenvalue problems apply equally well to the earlier methods described in [23] and [28].

2.1. Path following and arclength continuation. For path following in the absence of singularities the standard approach is a predictor-corrector method. The methods differ in the manner in which $u'(\lambda) = du/d\lambda$ is computed. Assume that (u_0, λ_0) is a known solution point and we want to compute (u_1, λ_1) at a nearby $\lambda = \lambda_1$.

If the Jacobian $G_u(u_0, \lambda_0)$ is nonsingular, then the implicit function theorem insures the existence of a unique smooth arc of solutions $(u(\lambda), \lambda)$, through (u_0, λ_0) with $u(\lambda_0) = u_0$. Furthermore, with the smoothness assumption on G, it follows that $u'(\lambda) = \frac{d}{d\lambda} u(\lambda)$ exists and differentiation of (1.1) with respect to λ gives the equation for u':

$$G_u(u_0, \lambda_0) u'(\lambda_0) = -G_\lambda(u_0, \lambda_0). \tag{2.1}$$

The Euler predictor uses u' directly and approximates $u_1 = u(\lambda_1)$ with

$$u_1^{(0)} = u_0 + (\lambda_1 - \lambda_0) u'(\lambda_0). \tag{2.2}$$

While the results in this paper can be applied to the solution of (2.1) by GMRES, we choose to avoid the additional solve and approximate u' with a difference. The secant predictor is

$$u_1^{(0)} = u_0 + \frac{\lambda_1 - \lambda_0}{\lambda_0 - \lambda_{-1}}(u_0 - u_{-1}), \tag{2.3}$$

where (u_{-1}, λ_{-1}) is a second solution pair with $\lambda_{-1} < \lambda_0$.

The nonlinear iteration is an inexact Newton iteration [7] in which

$$u_1^{k+1} = u_1^k + d^k,$$

where d^k satisfies the inexact Newton condition

$$\|G_u(u_1^k, \lambda_1)d^k + G(u_1^k, \lambda_1)\| < \eta_k \|G(u_1^k, \lambda_1)\|. \tag{2.4}$$

The inexact Newton condition can be viewed as a relative residual termination criterion for an iterative linear solver applied to the equation for the Newton step

$$G_u(u_1^k, \lambda_1)d = -G(u_1^k, \lambda_1).$$

When GMRES is used as the linear iteration, as it is in this paper, the nonlinear solver is referred to as Newton-GMRES.

The same procedure can be used if there are simple folds if the problem is expanded by using pseudo-arclength continuation [14], [15]. Here we introduce a new parameter s and solve

$$F(x, s) = \begin{bmatrix} G(x) \\ N(x, s) \end{bmatrix} = 0, \tag{2.5}$$

where $x = (u, \lambda)$ and $N(u, \lambda, s) = 0$ is a normalization equation.

We will use the secant normalization

$$\begin{aligned} N(u, \lambda, s) = \theta \left(\frac{u(s_0) - u(s_{-1})}{s_0 - s_{-1}} \right)^T (u(s) - u(s_0)) \\ + (1 - \theta) \left(\frac{\lambda(s_0) - \lambda(s_{-1})}{s_0 - s_{-1}} \right) (\lambda(s) - \lambda(s_0)) - (s - s_0) = 0 \end{aligned} \tag{2.6}$$

when two points on the path are available and the norm-based normalization

$$N(u, \lambda, s) = \theta \|u - u(s_0)\|^2 + (1 - \theta)|\lambda - \lambda(s_0)|^2 - (s - s_0)^2 \tag{2.7}$$

begins the path following. Both normalizations are independent of the discretization. This independence plays a role in the analysis that follows.

It is known [16] that arclength continuation turns simple folds in (u, λ) to regular points in (u, λ, s).

3. Singular point detection and branch switching. Although a continuation procedure incorporated with a pseudo-arclength normalization can circumvent the computational difficulties caused by turning points and usually jump over bifurcation points, it is usually desirable and important to be able to detect and locate the singularities in both cases. In the case where direct methods are used for linear algebra a sign change in the determinant of F can be used to detect simple bifurcation

and a change in sign of $d\lambda/ds$ to detect simple folds. We use a variation of the approach in [12], which requires only the largest (in magnitude) solution of a generalized eigenvalue problem.

Suppose that $(x(s_a), s_a)$ and $(x(s_b), s_b)$ are two regular points and the path following procedure is going from s_a to s_b. Let $A(s_a) = F_x(x(s_a), s_a)$ and $A(s_b) = F_x(x(s_b), s_b)$. Then both $A(s_a)$ and $A(s_b)$ are nonsingular. Recall that if $(x(s), s)$ is a bifurcation point on the solution path, then the null space $\mathcal{N}(F_x(x(s), s))$ is one-dimensional and there exists a nonzero vector v such that

$$F_x(x(s), s)v = 0. \tag{3.1}$$

Applying the Lagrange interpolation to $A(s) = F_x(x(s), s)$ we have

$$A(s) = \frac{s - s_b}{s_a - s_b}A(s_a) + \frac{s - s_a}{s_b - s_a}A(s_b) + \frac{1}{2}(s - s_a)(s - s_b)E(s), \tag{3.2}$$

where the matrix $E(s)$ denotes the perturbation matrix in the interpolation. Then instead of solving the usually nonlinear eigenvalue problem (3.1) we make a linear approximation,

$$\left[\frac{s - s_b}{s_a - s_b}A(s_a) + \frac{s - s_a}{s_b - s_a}A(s_b)\right] w = 0,$$

which leads to a generalized eigenvalue problem

$$A(s_a)w = \sigma A(s_b)w \tag{3.3}$$

with

$$\sigma = \frac{s - s_a}{s - s_b} = 1 + \frac{s_b - s_a}{s - s_b}. \tag{3.4}$$

Therefore, an s value which causes the Jacobian $F_x(x(s), s)$ to be singular can be approximated by

$$\hat{s} = s_b + \frac{1}{\sigma - 1}(s_b - s_a). \tag{3.5}$$

Our approach differs from that in [12] in that the roles of s_a and s_b are interchanged. The method of [12] solves

$$A(s_b)\hat{w} = \hat{\sigma}A(s_a)\hat{w}, \tag{3.6}$$

where

$$\hat{\sigma} = \frac{s - s_b}{s - s_a} = 1 - \frac{s_b - s_a}{s - s_a}. \tag{3.7}$$

In the approach of [12], $F_x(x, s)$ is factored at s_a in order to compute dx/ds for the predictor; that factorization is used again to precondition an Arnoldi iteration for the corrector and used one last time in the formulation of the eigenvalue problem to detect singularities. Since we do not factor F_x at all, the roles of s_a and s_b can be interchanged.

We found in our experiments that our approach, using (3.3), gave a better approximation of the location of the singularity than one using (3.6). A heuristic explanation

of this, suggested by an anonymous referee, is when (3.3) and (3.4) are used, the largest generalized eigenvalue in magnitude is desired. If, then, $|\sigma|$ is large, a small absolute error in the computation of σ would generate an even smaller error in the computation of s. To see this let $\sigma+\epsilon$ be the computed value of σ. By (3.5) the error in s is

$$|s_b - s_a|\left|\frac{1}{\sigma-1} - \frac{1}{\sigma+\epsilon-1}\right| = \frac{|s_b - s_a||\epsilon|}{|\sigma-1||\sigma+\epsilon-1|} \approx \frac{|s_b - s_a||\epsilon|}{\sigma^2} \ll |\epsilon|$$

if $|\sigma| \gg 1$. Conversely, if (3.6)–(3.7) are used, the generalized eigenvalue smallest in magnitude is sought. If $\hat{\sigma}+\epsilon$ is the computed value, and $|\hat{\sigma}|$ is small, then the error in $\hat{s}$ is

$$\frac{|s_b - s_a||\epsilon|}{|1-\hat{\sigma}||1-\hat{\sigma}+\epsilon|} \approx |s_b - s_a||\epsilon|$$

if $|\hat{\sigma}| \ll 1$.

Equation (3.4) implies that a bifurcation point s closest to the interval $[s_a, s_b]$ corresponds to the largest eigenvalue in magnitude σ of (3.3). A negative σ signals that there is a bifurcation point between s_a and s_b, $0 < \sigma < 1$ indicates that there is a bifurcation point close behind s_a, and a "large" positive σ means that a bifurcation point is approaching. When $\sigma \approx 1$, it should be interpreted that no bifurcation point is nearby.

Since $(x(s_b), s_b)$ is a regular point and $A(s_b)$ is nonsingular, the generalized eigenvalue problem (3.3) can be solved via the equivalent linear eigenvalue problem

$$A(s_b)^{-1}A(s_a)w = \sigma w, \tag{3.8}$$

where only the largest eigenvalue and corresponding eigenvector are needed. We will solve (3.8) with the Arnoldi method and prove that if s_a and s_b are not too near a singular point, then the eigenvalue problem is well conditioned in a mesh-independent way.

A simple fold point can be predicted in a similar way with s replaced by λ and $A(s)$ replaced by $A(\lambda) = G_u(u(s), \lambda(s))$.

At a simple bifurcation point two branches of solutions intersect nontangentially. We next describe how the information obtained from the solution of the eigenvalue problem can be used for branch switching. Suppose a bifurcation point $x_0 = x(s_0) = (u(s_0), \lambda(s_0))$ on the primary branch is determined. Since the eigenvector w corresponding to the largest eigenvalue σ for the eigenvalue problem described above is an approximation for a null vector of $F_x(s_0)$, thus $[G_u(s_0), G_\lambda(s_0)]w \approx 0$. Conversely, the tangent vector $\dot{x}_0 = dx/ds(s_0)$, which is the solution of

$$\begin{bmatrix} G_u(s_0) & G_\lambda(s_0) \\ N_u(s_0) & N_\lambda(s_0) \end{bmatrix} \begin{bmatrix} \dot{u}(s_0) \\ \dot{\lambda}(s_0) \end{bmatrix} = \begin{bmatrix} 0 \\ -N_s(s_0) \end{bmatrix}, \tag{3.9}$$

is also a null vector of $[G_u(s_0), G_\lambda(s_0)]$. Since $\dim(\mathcal{N}[G_u(s_0), G_\lambda(s_0)]) = 2$ and w and $\dot{x}_0$ are linearly independent it is recommended in [12], where G_u is factored and dx/ds is computed using (3.9), that $\mathcal{N}([G_u(s_0), G_\lambda(s_0)])$ be approximated by $\text{span}\{\dot{x}_0, w\}$. In the matrix-free case considered in this paper, we approximate the tangent vector by a secant approximation

$$\delta_x = \frac{x(s_0) - x(s_{-1})}{s_0 - s_{-1}}.$$

We approximate the tangent direction of the new branch by orthogonalizing w against δ_x:

$$\hat{w} = w - \frac{\delta_x^T w}{\delta_x^T \delta_x}\delta_x, \qquad \tilde{w} = \hat{w}/\|\hat{w}\|_2. \tag{3.10}$$

To obtain a regular point on the secondary branch, we solve the following augmented nonlinear system

$$F_2(x) = \begin{bmatrix} G(x) \\ N_2(x) \end{bmatrix}, \tag{3.11}$$

where

$$N_2(x) = \tilde{w}^T(x - x_0 - \epsilon\tilde{w}) \tag{3.12}$$

and ϵ is a "switching factor." This switching factor is problem dependent and should be chosen large enough so that the solution of (3.11) does not fall back to the primary branch [16]. The nonlinear equation (3.11) can be solved by Newton-type method with initial iterate $x_0 + \epsilon\tilde{w}$. After moving onto the secondary branch, the continuation procedure for path following on the secondary branch is identical to the one on the primary branch.

In [12] the linear systems for computing tangent vectors (3.9) are solved with a preconditioned Arnoldi iteration and the eigenvalues of the Hessenberg matrix produced in the Arnoldi iteration are used to predict singular points. Since a factorization of $A(s_a)$ is used as the preconditioner in [12], the prediction comes with very little cost. In our approach, Jacobians are not factored and the tangent vector is not computed, instead, a secant vector is used as an approximation. The eigenvalue problem for singularity prediction is performed separately using some iterative algorithm. This leaves us the flexibility of choosing any robust iterative solver and the associated preconditioner for the continuation procedure.

Other methods for singularity detection based on solution of eigenvalue problems have been proposed in [28], [10], and [23].

4. Approximations and mesh independence. This is the only section in which the properties of the discrete problems are explicitly addressed. We assume, as is standard in the integral equations literature [1], [3], that the discretization used in (1.2) has been constructed, by interpolation if necessary, so that G^h has the same domain and range as G.

Recall [1] that a family of linear operators $\{T^\alpha\}_{\alpha\in\mathcal{A}}$ is collectively compact if the set

$$\cup_\alpha T^\alpha(\mathcal{B}) \tag{4.1}$$

is precompact in X. In (4.1) $\mathcal{B}$ is the unit ball in X. In the special case that the indexing set $\mathcal{A}$ is an interval $[0, h_0]$, we will denote the index by h. We say that $T^h \to T$ strongly as $h \to 0$ and write

$$T^h \xrightarrow{s} T$$

if $T^h u \to Tu$ as $h \to 0$ in the norm of X for all $u \in X$.

All of the results are based on the following result (Theorem 1.6 from [1]) and some simple consequences.

THEOREM 4.1. *Let $\{T^h\}$ be a family of collectively compact operators on X that converge strongly to $T \in \mathcal{COM}(X)$. Assume that $I - T$ is nonsingular. Then there is $h_0 > 0$ such that $I - T^h$ is nonsingular for all $h \le h_0$ and $(I - T^h)^{-1}$ converges strongly to $(I - T)^{-1}$.*

A simple compactness argument implies uniform bounds for parameter dependent families of collectively compact strongly convergent operators.

COROLLARY 4.2. *Let $a < b$ be given and assume that*

$$\{T^h(s)\}_{h\ge 0, s\in[a,b]}$$

is a collectively compact family of operators such that $T^h(s) \xrightarrow{s} T^0(s) = T(s)$ for each fixed s. Assume that $\{T^h(s)\}$ is uniformly Lipschitz continuous in s and that $I - T(s)$ is nonsingular for all $s \in [a, b]$. Then there are $M, h_0 > 0$ such that

$$\|I - T^h(s)\| \le M \qquad \text{and} \qquad \|(I - T^h(s))^{-1}\| \le M$$

for all $h \le h_0$ and $s \in [a, b]$.

We let

$$\Gamma^h = \{(u, \lambda) \mid G^h(u, \lambda) = 0, \lambda \in \mathbb{R}\} \tag{4.2}$$

and let $\Gamma = \Gamma^0$.

We assume that the following holds.

ASSUMPTION 4.1.

(1) *$G^h(u, \lambda) \to G(u, \lambda)$ for all $(u, \lambda) \in X \times \mathbb{R}$ as $h \to 0$.*
(2) *G^h is Lipschitz continuously Fréchet differentiable and the Lipschitz constants of G^h_u and G^h_λ are independent of h.*
(3) *For all $(u, \lambda) \in X \times \mathbb{R}$*

$$G^h_u(u, \lambda) \xrightarrow{s} G_u(u, \lambda) \quad \text{as } h \to 0. \tag{4.3}$$

(4) *There are $\delta_0, h_0 > 0$ such that if*

$$\mathcal{N}(\delta) = \{(v, \mu) | \|v - u\| + |\mu - \lambda| < \delta \text{ for some } (u, \lambda) \in \Gamma\}, \tag{4.4}$$

then the families of operators

$$\{I - G^h_u(v, \mu)\}_{\{(v,\mu)\in\mathcal{N}(\delta_0), h\in[0,h_0]\}} \quad \text{and} \quad \{G^h_\lambda(v, \mu)\}_{\{(v,\mu)\in\mathcal{N}(\delta_0), h\in[0,h_0]\}}$$

are collectively compact.

We augment $G^h = 0$ with a mesh-independent Lipschitz continuously differentiable arclength normalization $N(u, \lambda, s) = 0$ and define

$$F^h(x, s) = \begin{bmatrix} G^h(x) \\ N(x, s) \end{bmatrix} = 0. \tag{4.5}$$

Examples of normalizations N that do not depend on h are (2.6) and (2.7). We let

$$\Gamma(s) = \{(u(s), \lambda(s)) \mid F(u, \lambda, s) = 0\}. \tag{4.6}$$

Consistently with the notation in the previous sections, we let $u^h(\lambda)$ and $x^h(s)$ denote solutions to $G^h(u, \lambda) = 0$ and $F^h(x, s) = 0$.

4.1. The corrector equation and simple folds. At regular points when λ is used as the continuation parameter the equation for the Newton step is

$$G_u^h(u,\lambda)z = -G^h(u,\lambda).$$

Theorem 4.1 and our collective compactness assumptions imply that if $G_u(u,\lambda)$ is nonsingular, so is G_u^h for sufficiently small h and $G_u^h(u,\lambda)^{-1}$ is strongly convergent to $G_u(u,\lambda)^{-1}$. However, this convergence is not uniform and Theorem 4.1 is valid only if we remain away from singular points.

One can treat simple folds as regular points by means of arclength continuation. If we solve the augmented system, (2.5), and this system has only regular points, then $\|F_x^{-1}\|$ is bounded on finite segments of Γ. If, moreover, the normalization $N(u,\lambda,s)$ does not depend on the discretization, as it will not if (2.6) or (2.7) are used, then our assumptions imply that the finite-dimensional problems are as well conditioned as the infinite-dimensional problems. We let $\kappa(A) = \|A\|\|A^{-1}\|$ denote the condition number of an invertible operator A.

THEOREM 4.3. *Let Assumptions* 1.1 *and* 4.1 *hold. Let* $[s_a, s_b]$ *be such that*

$$\hat{\Gamma} = \{\hat{\Gamma}(s) : s \in [s_a, s_b]\} \subset \{\Gamma(s) : s \in [s_a, s_b]\}$$

is a single smooth arc (in particular, nonempty) and F_x *is nonsingular on* $\hat{\Gamma}(s)$*; then there are* δ_1*,* K*, and* h_0 *such that* $\kappa(F_x^h(x,s)) \le K$ *for all* $(x,s) \subset \mathcal{N}(\delta_1) \times [s_a, s_b]$ *and* $h \le h_0$*.*

Proof. Assumption 1.1 and the mesh independence of N imply that we may apply Corollary 4.2 to $T^h(s) = F_x^h(x(s),s)$. Hence, there is $h_0 > 0$ and M such that for all $h \le h_0$ and $s \in [s_a, s_b]$

$$\|F_x^h(x(s),s)\| \le M \qquad \text{and} \qquad \|F_x^h(x(s),s)^{-1}\| \le M.$$

Let L be the (h-independent) Lipschitz constant of F_x^h. Then if $\|x - x(s)\| \le \delta \le \delta_0$,

$$\|F_x^h(x,s)\| \le M + L\delta.$$

Moreover,

$$\|I - F_x^h(x(s),s)^{-1}F_x^h(x,s)\| \le ML\delta,$$

and, hence, if $ML\delta < 1$, the Banach lemma implies that $F_x^h(x,s)$ is nonsingular and

$$\|F_x^h(x,s)^{-1}\| \le \frac{M}{1 - ML\delta}. \tag{4.7}$$

So if $\delta_1 < 1/(2ML)$ the proof is complete with $K = 2M(M + L\delta_1)$. ∎

The bound, (4.7), part 4.1 of Assumption 4.1, and the Kantorovich theorem [13], [17], [20], imply convergence of x^h to x.

COROLLARY 4.4. *Let Assumptions* 1.1 *and* 4.1 *hold. Let* $[s_a, s_b]$ *be such that* $\{\hat{\Gamma}(s) : s \in [s_a, s_b]\}$ *is a single smooth arc with at most simple fold singularities. Then there is a unique solution arc for* F^h*,*

$$\hat{\Gamma}^h(s) \subset \mathcal{N}(\delta) \times [s_a, s_b],$$

and $x^h(s) \to x(s)$ *as* $h \to 0$ *uniformly for* $s \in [s_a, s_b]$*.*

So, for h sufficiently small, the secant predictor

$$(x^h)^{(0)}(s_1) = x^h(s_0) + \frac{s_1 - s_0}{s_0 - s_{-1}} \left(x^h(s_0) - x^h(s_{-1}) \right), \tag{4.8}$$

converges to $x^{(0)}(s_1)$ as $h \to 0$ uniformly in (s_{-1}, s_0, s_1). Hence if the steps in arclength $\{\delta_s^n\}$ are independent of h and the secant predictor is used, then the accuracy of the initial iterate to the corrector equation is independent of h. By Theorem 4.3, the condition of the linear equation for the Newton step is independent of h as well. Therefore, the performance of the *nonlinear iteration* is independent of h.

As for the GMRES iteration that computes the Newton step, the methods from [19], [4], and [5] may be extended to show that the GMRES iteration for the Newton step converges r-superlinearly in a manner that is independent of h, x, and s. All that one needs is an uniform clustering of the eigenvalues of $F_x^h(x, s)$ on $\mathcal{M}(\delta)$ that is independent of x, s, and h. This follows directly from Theorem 4.7 in [1]. In the results that follow, we count eigenvalues by multiplicity and order them by decreasing distance from zero.

THEOREM 4.5. *Let $\{T^h\}$ be a family of collectively compact operators on X that converge strongly to $T \in \mathcal{COM}(X)$. Let $\{\lambda_j^h\}_{j=1}^{\infty}$ be the eigenvalues of T^h and $\{\lambda_j\}_{j=1}^{\infty}$ be the eigenvalues of T. Let $\rho > 0$ and assume that $|\lambda_j| < \rho$ for all $j > M$. Then there is $h_0 > 0$ such that if $h \le h_0$, then*

$$\left|\lambda_j^h\right| < \rho \qquad \text{and} \qquad \lim_{h \to 0} \lambda_j^h = \lambda_j$$

for all $1 \le j \le M$. Moreover, if λ_1 has algebraic and geometric multiplicity of one, then so does λ_1^h for $h \le h_0$. Moreover, there is a sequence of eigenfunctions $\{w^h\}$ of T^h corresponding to the eigenvalue λ_1^h, so that $w^h \to w$, and an eigenfunction of T corresponding to the eigenvalue λ_1.

With this in hand, the eigenvalue clustering result can be obtained in the same way that Theorem 4.3 was derived from Theorem 4.1. We begin with the analog of Corollary 4.2.

COROLLARY 4.6. *Let the assumptions of Theorems* 4.3 *and* 4.5 *hold. Let $\{\mu_j^h(x, s)\}_{j=1}^{\infty}$ be the eigenvalues, counted by multiplicity, of $I - F_x(x, s)$. Let $\rho > 0$ be given. There are $h_0, M > 0$ such that if $h \le h_0$, then*

$$\left|\mu_j^h(x(s), s)\right| < \rho$$

for all $s \in [s_a, s_b]$ and $j > M$.

THEOREM 4.7. *Let the assumptions of Theorems* 4.3 *and* 4.5 *hold. Let $\{\mu_j^h(x, s)\}_{j=1}^{\infty}$ be the eigenvalues, counted by multiplicity, of $I - F_x(x, s)$. Let $\rho > 0$ be given. There are $\delta_2, h_0, M > 0$ such that if $h \le h_0$ and $(x, s) \in \mathcal{N}(\delta_2) \times [s_a, s_b]$, then $h_0, M > 0$ such that if $h \le h_0$, then*

$$\left|\mu_j^h(x, s)\right| < \rho$$

for all $j > M$.

Before we state our r-superlinear convergence theorem we must set some more notation. We write the equation for the Newton step for the corrector equation as

$$F_x^h(x, s) d^h = -F(x, s)^h. \tag{4.9}$$

We let $d_k^h(x, s)$ denote the kth GMRES iteration and let $r_k^h(x, s)$ the kth GMRES residual. As is standard in nonlinear equations, $d_0^h = 0$, and so $r_0^h = -F^h(x, s)$.

Aside from the dependence on the parameter s, which can be accounted for with a compactness argument, the statement and proof of Theorem 4.8 are the same as those of Theorem 1.1 of [5].

THEOREM 4.8. *Let Assumptions* 1.1 *and* 4.1 *hold. Let* $[s_a, s_b]$ *be such that* $\hat{\Gamma} = \{\hat{\Gamma}(s) : s \in [s_a, s_b]\}$ *is a single smooth arc and* F_x *is nonsingular on* $\hat{\Gamma}(s)$. *Then for all* $\rho \in (0,1)$ *there are* $M, \delta_3, h_1 > 0$ *such that*

$$\|r_k^h(x,s)\| \le M\rho^k \|F^h(x,s)\| \tag{4.10}$$

for all $(x,s) \subset \mathcal{N}(\delta_3) \times [s_a, s_b]$ *and* $h \le h_1$.

Theorem 4.8 states that any desired rate of linear convergence can be obtained in a mesh-independent way and hence, for (x,s) fixed, the convergence of the linear iteration is r-superlinear in a mesh-independent way.

Theorems 4.3 and 4.8 are related in that both imply mesh-independent convergence of the nonlinear Newton or inexact Newton iteration. Theorem 4.3 implies that the rate of quadratic convergence for Newton's method is independent of h. Theorem 4.8 implies that the number of GMRES iterations needed to reduce the relative residual in an inexact Newton method to a given tolerance can be bounded independently of h and s and that, therefore, the cost of a nonlinear iteration in terms of linear iterations is also independent of h. Similar conclusions about mesh independence results for inexact Newton methods have been made in other situations; see [18], for example.

4.2. Simple bifurcation. Now consider a path $\hat{\Gamma}_0 \subset \Gamma$,

$$\hat{\Gamma}_0 = \{\Gamma(s) : s \in [s_a, s_b]\}$$

which has a single simple bifurcation at $s = s_c \in (s_a, s_b)$. We let $\hat{\Gamma}_1$ be the branch of solutions that intersects $\hat{\Gamma}_0$ at $(x(s_c), s_c)$. We will denote solutions on $\hat{\Gamma}_1$ by y and the arclength parameter on $\hat{\Gamma}_1$ by t. Hence, for $t \in [t_a, t_b]$,

$$F(y(t), t) = 0.$$

We let $(y(t_c), t_c)$ be the bifurcation point on $\hat{\Gamma}_1$. Hence $y(t_c) = x(s_c)$.

For any $\delta > 0$, the results in section 4.1 hold for the paths

$$\hat{\Gamma}_0^- = \{\hat{\Gamma}_0(s) : s \in [s_a, s_c - \delta]\}, \qquad \hat{\Gamma}_0^+ = \{\hat{\Gamma}_0(s) : s \in [s_c + \delta, s_b]\},$$
$$\hat{\Gamma}_1^- = \{\hat{\Gamma}_1(t) : t \in [t_a, t_c - \delta]\}, \quad \text{and} \quad \hat{\Gamma}_1^+ = \{\hat{\Gamma}_1(t) : t \in [t_c + \delta, t_b]\}.$$

We define $\hat{\Gamma}_0^h$ in a similar way. Since $x^h(s) \to x(s)$ uniformly for $s \in [s_a, s_c - \delta] \cup [s_c + \delta, s_b]$ by the results in the previous section, simple bifurcation from $\hat{\Gamma}_0^h$ can only arise for $s \in (s_c - \delta, s_c + \delta)$.

In this section we describe how the prediction of the bifurcation point, the conditioning of the generalized eigenproblem, the solution of that eigenproblem, and the accurate tracking of the other branch depend on h. Since perturbation of G, even in finite dimension, can change the structure of $\hat{\Gamma}_0$ from two intersecting arcs to two disconnected arcs [21], [22], we cannot show that the bifurcation will be preserved (see [6] for results on this in the context of finite element discretizations). We can show that if s_c is far enough away from s_a and s_b, the other path will be detected even if it does not correspond to a bifurcation for the finite-dimensional problem.

For $(x(s), s) \in \hat{\Gamma}_0^{\pm}$, $F_x(x, s)$ will be nonsingular and we can consider the eigenvalue problems

$$F_x^h(x^h(s_b), s_b)^{-1} F_x^h(x^h(s_a), s_a) w^h = \sigma^h w^h, \tag{4.11}$$

where the largest eigenvalue in magnitude is sought. If (4.11) is solved via the Arnoldi method, the matrix-vector products of $F_x^h(x^h(s_b), s_b)^{-1} F_x^h(x^h(s_a), s_a)$ with a vector will each require a product of $F_x^h(x^h(s_b), s_b)^{-1}$ with a vector, i.e., a linear solve. If this solve is performed with GMRES, then Theorem 4.8 implies that the number of iterations required to approximate that matrix-vector product can be bounded independently of h, x, s.

Our assumptions imply, if $s_b - s_a$ is sufficiently small, that $\sigma = \sigma^0$ is a simple eigenvalue with geometric multiplicity 1. Since

$$\{I - F_x^h(x^h(s_b), s_b)^{-1} F_x^h(x^h(s_a), s_a)\}_{h \geq 0}$$

is a collectively compact and strongly convergent sequence, we can apply Theorem 4.5 to conclude that for h sufficiently small, σ^h is a simple eigenvalue with geometric multiplicity 1 as well. Moreover σ^h will be well separated from the next largest eigenvalue in magnitude, hence the conditioning of the eigenvalue problem is independent of h.

5. Numerical example. In this section we show how the results in this paper can be applied to a concrete instance of the methods discussed in the previous sections. We then report on numerical observations that illustrate the mesh independence of the convergence rates. In section 5.1 we describe the implementation, choice of algorithmic parameters, and termination tolerances. We describe our example problem, a nonlinear two point boundary value problem taken from [16], its discretization, and the family of maps in section 5.2. We present a bifurcation diagram and the numerical results in section 5.3.

5.1. Methodology. In the example presented in this section we use the normalization equation (2.6) with $\theta = .5$. We used a step of

$$\Delta_{path} = s_0 - s_{-1} = .02$$

in s for path following.

We apply the secant predictor (2.3) to generate the initial iterate. For the corrector, we choose a forward-difference Newton-GMRES algorithm [17]. The outer iteration that generates the sequence $\{(x^{(k)}, s)\}$ terminates when the nonlinear residual norm satisfies

$$\|F(x^{(k)}, s)\| \leq \tau_a + \tau_r \|F(x^{(0)}, s)\|,$$

where the absolute error tolerance $\tau_a = 10^{-7}$ and relative error tolerance $\tau_r = 10^{-7}$ were used. To avoid oversolving on the linear equation for the Newton step z_k the forcing terms η_k such that

$$\|F_x(x^{(k)}, s) z_k + F(x^{(k)}, s)\| \leq \eta_k \|F(x^{(k)}, s)\|$$

were adjusted with a method from [9]. We use the l^2 norm and scale the norm by a factor of $1/N$ for differential and integral equations so that the results are independent of the computational mesh.

The generalized eigenvalue problem (3.3) that characterizes the singularity prediction and branch switching need not be solved as each new point on the path is computed (i.e., $s_b = s_a + \Delta_{path}$ is too frequent). In the example considered in this section, we solved the eigenvalue problem to make a prediction of bifurcation after a step of

$$\Delta_{eig} = s_b - s_a = 20 * \Delta_{pred} = 4$$

had been taken. We used a simple version of the Arnoldi method [2] based on modified Gram–Schmidt orthogonalization to solve (3.3) with reorthogonalization at each iteration. The Jacobian of the inflated system is usually neither symmetric nor positive definite, and we treat the generalized eigenvalue problem as linear eigenvalue problem and solve the linear system involved explicitly with preconditioned GMRES.

After j steps the Arnoldi method produces

$$A(\lambda_b)^{-1}A(\lambda_a)Q_j = Q_j H_j + h_{j+1,j}q_{j+1}e_j^T,$$

where H_j is an upper Hessenberg matrix with the h_{ij}'s as its nonzero entries and Q_j is orthogonal with the q_j's as columns. If (θ, y) is an eigenpair of H_j and $w = Q_j y$, then

$$\|A(\lambda_b)^{-1}A(\lambda_a)w - \theta w\|_2 = |h_{j+1,j}||e_j^T y| \tag{5.1}$$

provides a computable error bound. The Ritz pair, (θ, w), is used to approximate the eigenpair of $A(\lambda_b)^{-1}A(\lambda_a)$.

We terminated the Arnoldi iteration when

$$|h_{j+1,j}||e_j^T y| < 10^{-4}$$

or $j > 5$. The prediction $\hat{\lambda}$ is accepted when $\hat{\lambda}$ lies in between λ_a and λ_b or

$$|\hat{\lambda} - \lambda_b| < \frac{1}{2}|\lambda_a - \lambda_b|.$$

We found this strategy sufficient for the example here and were able to use (3.10) to move onto the new branch. Of course, the values of Δ_{eig} is problem dependent (as it is for other methods) as is the criteria for accepting a prediction.

5.2. Example problem. As an example we consider the following two-point boundary value problem [16]:

$$\begin{aligned} u'' + u^3 + \lambda &= 0 \quad \text{in } (0,1), \\ u(0) = u(1) &= 0. \end{aligned} \tag{5.2}$$

We transform (5.2) to an equation satisfying Assumption 1.1 by multiplying both sides with the Green's operator $\mathcal{G}$ for u'', which is defined by

$$(\mathcal{G}f)'' = f; \qquad (\mathcal{G}f)(0) = (\mathcal{G}f)(1) = 0.$$

We obtain

$$G(u, \lambda) = u + \mathcal{G}(u^3 + \lambda) = 0. \tag{5.3}$$

The map G defined by (5.3) satisfies Assumption 1.1 with $X = C[0, 1]$.

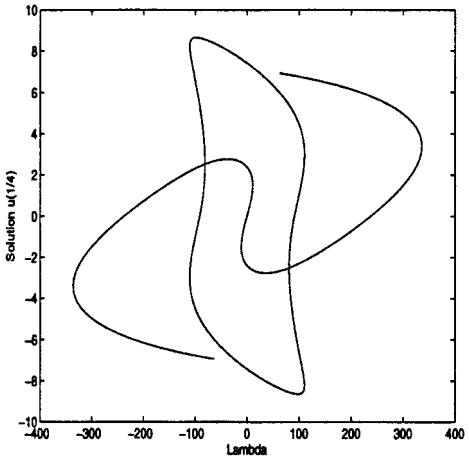

FIG. 5.1. *Solution* $u(\frac{1}{4})$ *versus* λ.

Our discretization of (5.2) and (5.3) uses a uniform mesh of spacing $h = \frac{1}{N}$. We denote the solution to the discrete system by $U \in R^{N+1}$. A fourth-order finite difference discretization leads to the discrete nonlinear system of order $N-1$,

$$\left(\frac{1}{h^2} + \frac{1}{12}U_{j-1}^2\right)U_{j-1} - \left(\frac{2}{h^2} - \frac{5}{6}U_j^2\right)U_j + \left(\frac{1}{h^2} + \frac{1}{12}U_{j+1}^2\right)U_{j+1} + \lambda = 0 \tag{5.4}$$

for $j = 1, \ldots, N-1$. $U_0 = U_N = 0$ at the boundaries. We multiply both sides of (5.4) with the discrete Green's operator $\mathcal{G}^{FD,h}$ for the mesh to obtain a fully discrete approximation to G:

$$G^{FD,h}(U,\lambda)_j = U_j + \mathcal{G}^{FD,h}\left(\frac{1}{12}U_{j-1}^3 - \frac{5}{6}U_j^3 + \frac{1}{12}U_{j+1}^3 + \lambda\right) = 0. \tag{5.5}$$

If we let $R_h : X \to R^{N+1}$ denote the evaluation map

$$(R_h u)_j = u(hj), \quad 0 \le j \le N,$$

and $P_h : R^{N+1} \to X$ denote piecewise linear interpolation, then we can define a sequence of maps G^h that satisfies Assumption 4.1 by

$$G^h(u,\lambda) = u + P_h(G^{FD}(R_h u, \lambda) - R_h u). \tag{5.6}$$

Now U is a solution of the fully discrete problem if and only if $P_h U$ is a solution of (5.6); similarly u is a solution of (5.6) if and only if $R_h u$ is a solution of the fully discrete problem. A similar statement holds for the linearized problems. Hence the results on conditioning for the linear equations and eigenproblems apply not only to the family $\{G^h\}$ but also to the fully discrete problems that are solved numerically. The convergence properties of G^h to G follow from the consistency of the discretizations. Collective compactness of $\{I - G_u^h\}$ follows from the fact that the matrix entries of $\mathcal{G}^{FD,h}$ converge to the values of the Green's function at the mesh points [11].

5.3. Numerical results. Starting from the trivial solution $u = 0$ at $\lambda = 0$, the algorithm traces the solution path, locates bifurcation points, switches branches, and

captures the secondary solution curves. Indeed there are a total of two bifurcation points at $\lambda \approx 81, -81$ and eight turning points at $\lambda \approx \pm 11, \pm 110, \pm 336$ in the region of interest. The primary solution branch represents the symmetric solutions and the secondary solution branch represents nonsymmetric periodic solutions bifurcating from the primary branch. Figure 5.1 plots the solution $u(\frac{1}{4})$ versus λ showing the folds and bifurcations on the curve.

TABLE 5.1
Prediction of bifurcation point along the primary branch going toward $\lambda \approx -81$, $N = 128$.

λ_a	λ_b	Prediction (3.3)	Prediction (3.6)
$1.0000e-02$	$6.0988e+00$	$9.1417e+01$	$1.2288e+01$
$6.0998e+00$	$1.0574e+01$	$1.2614e+02$	$-9.3188e+03$
$1.0574e+01$	$8.4208e+00$	$4.0988e-01$	$1.4026e+01$
$8.4208e+00$	$1.1361e+00$	$-2.2855e+01$	$-3.5346e+01$
$1.1361e+00$	$-7.1699e+00$	$-3.6599e+01$	$-4.7426e+01$
$-7.1699e+00$	$-1.5772e+01$	$-4.6063e+01$	$-6.1496e+01$
$-1.5772e+01$	$-2.4500e+01$	$-5.3739e+01$	$-8.3880e+01$
$-2.4500e+01$	$-3.3295e+01$	$-6.0284e+01$	$-1.0833e+02$
$-3.3295e+01$	$-4.2128e+01$	$-6.5987e+01$	$-1.3475e+02$
$-4.2128e+01$	$-5.0984e+01$	$-7.0930e+01$	$-1.6373e+02$
$-5.0984e+01$	$-5.9854e+01$	$-7.5103e+01$	$-1.9663e+02$
$-5.9854e+01$	$-6.8734e+01$	$-7.8384e+01$	$-2.3781e+02$
$-6.8734e+01$	$-7.7621e+01$	$-8.0558e+01$	$-3.0825e+02$

TABLE 5.2
Total number of Newton and preconditioned GMRES iterations on the primary branch.

Problem size	λ value	Newton	P-GMRES
$N = 64$	$\lambda = +1.0895e+01$	5	10
	$\lambda = -4.0017e+01$	5	12
	$\lambda = -8.0127e+01$	5	12
$N = 128$	$\lambda = +1.0894e+01$	4	7
	$\lambda = -4.0163e+01$	5	12
	$\lambda = -8.0386e+01$	5	13
$N = 256$	$\lambda = +1.0894e+01$	4	7
	$\lambda = -4.0223e+01$	5	13
	$\lambda = -8.0482e+01$	5	12

TABLE 5.3
Total number of Newton and preconditioned GMRES iterations at switching point and on the secondary branch.

Problem size	λ value	Newton	P-GMRES
$N = 64$	switching	7	21
	$\lambda = -8.1319e+01$	5	12
	$\lambda = -1.0020e+02$	5	11
$N = 128$	switching	7	22
	$\lambda = -8.1571e+01$	5	14
	$\lambda = -1.0015e+02$	5	13
$N = 256$	switching	6	20
	$\lambda = -8.1262e+01$	5	12
	$\lambda = -1.0001e+02$	5	11

In Table 5.1 we report the bifurcation predictions along the solution path going toward $\lambda \approx -81$. One can see that (3.3) is a more accurate predictor than (3.6). Tables 5.2 and 5.3 illustrate the mesh independence of the linear and nonlinear iterations

TABLE 5.4
Total number of preconditioned GMRES and Arnoldi iterations required at different sections on the path when a prediction procedure is performed.

Path	N	λ_a	λ_b	P-GMRES	Arnoldi	Residual
Turning	64	$7.0814e+00$	$1.0632e+01$	53	5	$2.9490e-02$
point	128	$6.0998e+00$	$1.0574e+01$	50	5	$1.9156e-05$
nearby	256	$9.1379e+00$	$1.0883e+00$	55	5	$4.6579e-03$
Regular	64	$-3.1304e+01$	$-4.0215e+01$	65	5	$8.0766e-06$
point	128	$-2.4500e+01$	$-3.3295e+01$	69	5	$3.3738e-05$
	256	$-3.1175e+01$	$-3.9836e+01$	71	5	$3.0663e-05$
Bifurcation	64	$-6.7012e+01$	$-7.5953e+01$	56	4	$9.5183e-07$
point	128	$-6.8734e+01$	$-7.7621e+01$	56	4	$1.3263e-06$
detected	256	$-7.4823e+01$	$-8.3605e+01$	62	4	$1.3251e-06$

at various points on the primary (5.2) and secondary ((5.3) branches. The iteration statistics remain virtually unchanged as the mesh is refined.

Table 5.4 lists the total number of preconditioned GMRES iterations and Arnoldi iterations corresponding to three representative prediction intervals. In Table 5.4 we show how both the GMRES iterations needed to approximate the product of $F_x^h(x^h(s_b), s_b)^{-1}$ with $F_x^h(x^h(s_a), s_a)$ (or G_u^{-1} with G_u in the case where we predict a turning point) and the overall number of Arnoldi iterations is independent of the mesh. Each Arnoldi step requires about 12–14 GMRES iterations. This is similar to the numbers listed in Tables 5.2 and 5.3. The residual in the table is the Arnoldi residual when the iteration terminates.

Acknowledgments. The authors wish to thank Gene Allgower for making us aware of reference [25], Klaus Böhmer for providing us with a copy of [25], and Alastair Spence for providing many other pointers to the literature. We are also grateful to Dan Sorensen for his council on the solution of eigenvalue problems and an anonymous referee for insights into the advantages of (3.3, 3.4) over (3.6, 3.7).

REFERENCES

[1] P. M. ANSELONE, *Collectively Compact Operator Approximation Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1971.
[2] W. ARNOLDI, *The principle of minimized iterations in the solution the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 152–163.
[3] K. E. ATKINSON, *A Survey of Numerical Methods for Fredholm Integral Equations of the Second Kind*, SIAM, Philadelphia, 1976.
[4] S. L. CAMPBELL, I. C. F. IPSEN, C. T. KELLEY, AND C. D. MEYER, *GMRES and the minimal polynomial*, BIT, 36 (1996), pp. 664–675.
[5] S. L. CAMPBELL, I. C. F. IPSEN, C. T. KELLEY, C. D. MEYER, AND Z. Q. XUE, *Convergence estimates for solution of integral equations with GMRES*, J. Integral Equations Appl., 8 (1996), pp. 19–34.
[6] M. CROUZEIX AND J. RAPPAZ, *On Numerical Approximation in Bifurcation Theory*, Recherches en Mathématiques Appliqeés 13, Springer-Verlag, Berlin, Heidelberg, New York, 1990.
[7] R. S. DEMBO, S. C. EISENSTAT, AND T. STEIHAUG, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408.
[8] E. J. DOEDEL, *Lecture Notes on Numerical Analysis of Bifurcation Problems*, Lecture notes from Sommerschule über Nichtlineare Gleichungssysteme, Hamburg, Germany, March 17–21, 1997. Available by anonmyous ftp to ftp.cs.concordia.ca in pub/doedel/doc/hamburg.ps.Z.
[9] S. C. EISENSTAT AND H. F. WALKER, *Choosing the forcing terms in an inexact Newton method*, SIAM J. Sci. Comput., 17 (1996), pp. 16–32.
[10] J.-S. GUO, W.-W. LIN, AND C.-S. WANG, *Nonequivalence deflation for the solution of matrix latent value problems*, Linear Algebra Appl., 231 (1995), pp. 15–45.

[11] P. Henrici, *Discrete Variable Methods in Ordinary Differential Equations*, Wiley, New York, 1962.
[12] J. Huitfeldt and A. Ruhe, *A new algorithm for numerical path following applied to an example from hydrodynamic flow*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 1181–1192.
[13] L. Kantorovich and G. Akilov, *Functional Analysis*, 2nd ed., Pergamon Press, New York, 1982.
[14] H. B. Keller, *Constructive methods for bifurcation and nonlinear eigenvalue problems*, in Computing Methods in Applied Science and Engineering, Lecture Notes in Math. 704, R. Glowinski and J. Lions, eds., Springer-Verlag, New York, 1977, pp. 241–251.
[15] 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, pp. 359–384.
[16] H. B. Keller, *Lectures on Numerical Methods in Bifurcation Theory*, Tata Institute of Fundamental Research, Lectures on Mathematics and Physics, Springer-Verlag, New York, 1987.
[17] C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, 1995.
[18] C. T. Kelley and E. W. Sachs, *Multilevel algorithms for constrained compact fixed point problems*, SIAM J. Sci. Comput., 15 (1994), pp. 645–667.
[19] C. T. Kelley and Z. Q. Xue, *GMRES and integral operators*, SIAM J. Sci. Comput., 17 (1996), pp. 217–226.
[20] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
[21] E. Riess, *Column buckling–an elementary example of bifurcation*, in Bifurcation Theory and Nonlinear Eigenvalue Problems, J. Keller and S. Antman, eds., W. A. Benjamin, New York, 1969, pp. 1–16.
[22] E. Riess, *Imperfect bifurcation*, in Applications of Bifurcation Theory, P. Rabinowitz, ed., Academic Press, New York, 1977, pp. 37–71.
[23] A. Ruhe, *Algorithms for the nonlinear eigenvalue problem*, SIAM J. Numer. Anal., 10 (1973), pp. 674–689.
[24] Y. Saad and M. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
[25] R. Sebastian, *Andwendung von Krylov-Verfahren auf Verzweigungs-und Fortsetzungsprobleme*, Ph.D. thesis, Philipps-Universität Marburg, Marburg, Germany, 1995.
[26] H. F. Walker, *An Adaption of Krylov Subspace Methods to Path Following Problems*, Tech. Rep. 12/96/88, Mathematics and Statistics Department, Utah State University, Logan, UT, December 1996.
[27] H. F. Walker, *An Approach to Continuation Using Krylov Subspace Methods*, Tech. Rep. 1/97/89, Mathematics and Statistics Department, Utah State University, Logan, UT, January 1997.
[28] Y. F. Zhou and A. Ruhe, *Numerical path following and eigenvalue criteria for branch switching*, in Large Scale Eigenvalue Problems, J. Cullum and R. Willoughby, eds., North-Holland Math. Stud. 127, North-Holland, Amsterdam, 1986, pp. 319–334.