Numerical Methods for Bifurcation Problems

Alastair Spence
Ivan G. Graham
Department of Mathematical Sciences, University of Bath
Claverton Down, Bath BA2 7AY, U.K.

February 4, 2002

1 Introduction

This set of lecture notes provides an introduction to the numerical solution of bifurcation problems. The theory is given for finite dimensional problems – so we shall require only matrix theory, finite dimensional calculus, etc. Only the basic principles for three of the most common bifurcations will be discussed, but the hope is that after reading these notes a student should be able to tackle the original journal papers. Almost all the results extend to infinite dimensional operators defined in an appropriate setting, e.g. Banach or Hilbert Spaces.

There are many books on bifurcation theory – for example Chow & Hale [2] gives an all-round treatment, Vanderbauwhede [48] gives an early account of bifurcation in the presence of symmetries, and the important books by Golubitsky & Schaeffer [9], and Golubitsky, Stewart & Schaeffer [10] look at multiparameter bifurcation problems using singularity theory. An early conference proceedings is Rabinowitz [37], which contains one of the first papers on the numerical *analysis* (as compared with numerical methods) for bifurcation problems written by H. B. Keller [25]. As might be expected, early books about the numerical analysis of bifurcations were conference proceedings, see Mittelmann & Weber [32], Küpper et.al. [28], Küpper et.al. [29], Roose et.al.[40] and Seydel et.al. [46].

H.B. Keller's book [26] is a published version of lectures on Numerical Methods in Bifurcation Problems delivered at the Indian Institute of Science, Bangalore. Rheinboldt's book [39] is a collection of his papers and also gives information and listing of the code PITCON for numerical continuation of parameter dependent nonlinear problems. The code AUTO, developed by Doedel

[7], but with recent extensions by several others, is now the leading piece of software for nonlinear systems, and can handle steady and time dependent problems, and discretized boundary value problems. Seydel [45] contains discussion of numerical methods and many interesting examples. A comprehensive treatment, including a full discussion of numerical methods using singularity theory, is in the recent book by Govaerts [13]. Beyn [1] gives a survey on numerical methods for dynamical systems, including methods for homoclinic and heteroclinic orbits. In fact, AUTO now has an option to compute and follow paths of these orbits.

The plan of these notes is as follows. Section 2 contains three generic examples, namely a fold bifurcation, a Hopf bifurcation, and bifurcation from the trivial solution. Section 3 contains an account of Newton's method and the Implicit Function Theorem. Section 4 discusses the ideas behind Keller's pseudo-arclength numerical continuation algorithm [25]. Sections 5, 6 and 8 provide an introduction to the numerical analysis of the three types of bifurcation phenomena introduced in Section 2. Section 7 discusses bifurcation theory in nonlinear ODEs using results in Sections 5 and 6.

There are many phenomena not considered here, for example, bifurcation in the presence of symmetry (see for example [51] and the books [9], [10], [48]) and high order singularities in multiparameter problems (see [9], [10], [13], [23]).

2 Examples

Bifurcation is the study of nonlinear problems with parameters, with the main interest being the determination of changes in solution behaviour as a parameter varies. In particular, interest centres on how to detect, calculate and classify points where there is a change in the type of solution of the nonlinear problem. This section contains some examples of some typical bifurcation phenomena.

In these notes we shall consider systems of the form

$$\boldsymbol{F}(\boldsymbol{x}, \lambda) = \boldsymbol{0} \tag{2.1}$$

where $\boldsymbol{F} : \mathbb{R}^{n+1} \to \mathbb{R}^n$, $\boldsymbol{x} \in \mathbb{R}^n$ is the *state* variable, and $\lambda \in \mathbb{R}$ is a *parameter*. We shall study the behaviour of $\boldsymbol{x}$ as λ varies, in fact, loosely speaking, λ may be thought of as the independent variable and $\boldsymbol{x}$ as the dependent variable.

Problems like (2.1) arise when studying autonomous systems of ordinary differential equa-

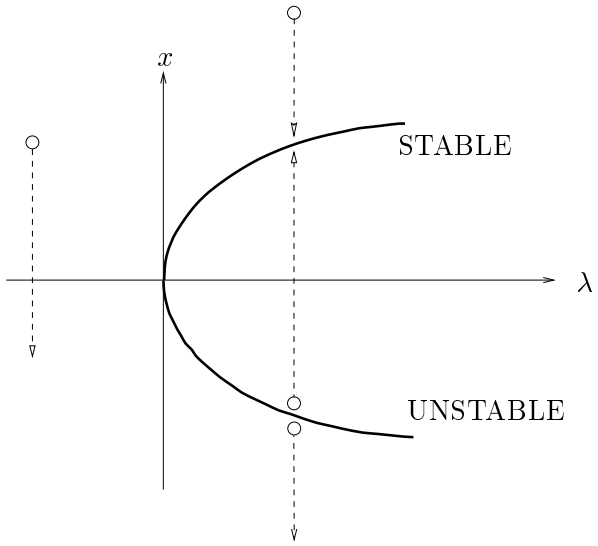

Figure 2.1:

tions

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}, \lambda), \quad \boldsymbol{x}(0) \text{ given}, \boldsymbol{x}(t) \in \mathbb{R}^n. \tag{2.2}$$

Steady states (equilibria) of (2.2) are given by $\frac{d\boldsymbol{x}}{dt} = \mathbf{0}$, and hence satisfy (2.1). An important topic in the study of systems like (2.2) is the analysis of how the solutions change as λ varies and the determination of changes in stability. Often a first step is to find the steady states by solving (2.1) for a range of λ values and then determine any changes in stability of these steady states. This theme is described in the first example.

Example 2.1 *Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined by*

$$f(x, \lambda) = \lambda - x^2 = 0.$$

For $\lambda < 0$ there are no real solutions; for $\lambda = 0$, $x = 0$ (twice); and for $\lambda > 0$ there are two solutions, $x = \pm\sqrt{\lambda}$. The steady solutions are shown by the solid line in figure 2.1. Consider now the solutions of the ODE

$$x_t = \lambda - x^2, \ x(t) \in \mathbb{R}, \ x(0) \text{ given.}$$

If $\lambda - x(0)^2 < 0$, then initially $x_t < 0$ and x decreases in time. If $\lambda - x(0)^2 > 0$ then $x_t > 0$ and x increases in time. The trajectories (dashed lines) for 4 different initial values (denoted $\circ$) are shown in Figure 2.1. Thus for $\lambda > 0, x = \sqrt{\lambda}$ is a stable equilibrium, and for $x = \sqrt{-\lambda}$ is an unstable equilibrium. □

It is clear that even in this simple example, knowledge of the zeros of $f(x,\lambda)=0$ helps us understand the behaviour of solutions of $x_t = f(x,\lambda)$. It is instructive before reading on to carry out a similar analysis for $x_t = \lambda x - x^3$.

The type of solution behaviour exhibited in Example 2.1 occurs in many physical examples. The point $(x,\lambda)=(0,0)$ is called a fold point (turning point or, in the dynamical systems literature, a saddle node) and we return to this kind of phenomenon in Section 5.

A very clear account of the stability of nonlinear systems is given in Chapter 9 of [20], where it is proved (p.187) that if $\boldsymbol{x}$ is a stable steady state for a given λ, then the Jacobian matrix $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x},\lambda)$ (i.e. the matrix with the (i,j)th component $\frac{\partial F_i}{\partial x_j}(\boldsymbol{x},\lambda)$) has no eigenvalues with positive real part. Hence stability of a steady state of (2.2) is lost when one or more eigenvalues of $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x},\lambda)$ moves into the right half-plane as λ varies. It is an instructive exercise to see in Example 2.1 how the eigenvalue of the (1×1) Jacobian matrix changes along the path of steady state solutions.

In Example 2.1 it was trivial to find the path of steady states analytically. In general the solutions to a nonlinear problem $\boldsymbol{F}(\boldsymbol{x},\lambda)=\boldsymbol{0}$ will not be known analytically. In the following sections we shall describe how to compute such solution paths and recognise the parameter values at which the number of solutions changes.

The following example is two dimensional, and it is more convenient to use (x,y) rather than $\boldsymbol{x}$.

Example 2.2 *Consider the pair of ODEs*

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \lambda & 1 \\ -1 & \lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - (x^2+y^2) \begin{pmatrix} x \\ y \end{pmatrix}.$$

Clearly $(x(t),y(t))=\boldsymbol{0}$ *is a solution for any* $\lambda\in\mathbb{R}$. *For any* $\lambda>0$ *each* $(x(t),y(t))^T$ *satisfying* $x(t)=\sqrt{\lambda}\sin t, y(t)=\sqrt{\lambda}\cos t$, *is a nontrivial periodic solution. The Jacobian of the right hand side is*

$$\begin{pmatrix} \lambda & 1 \\ -1 & \lambda \end{pmatrix} - \begin{pmatrix} 3x^2+y^2 & 2xy \\ 2xy & x^2+3y^2 \end{pmatrix}.$$

At $(x,y)=(0,0)$ *the eigenvalues of this matrix are* $\lambda\pm i$ *and so the trivial solution is stable for* $\lambda<0$ *and unstable for* $\lambda>0$. *On the other hand a short calculation shows that the eigenvalues of this matrix on the above nontrivial solution are* $-\lambda\pm\sqrt{\lambda^2-1}$, *which always have negative real part when* $\lambda>0$ *and so the periodic solution is stable. (See also §1 in [26].)*

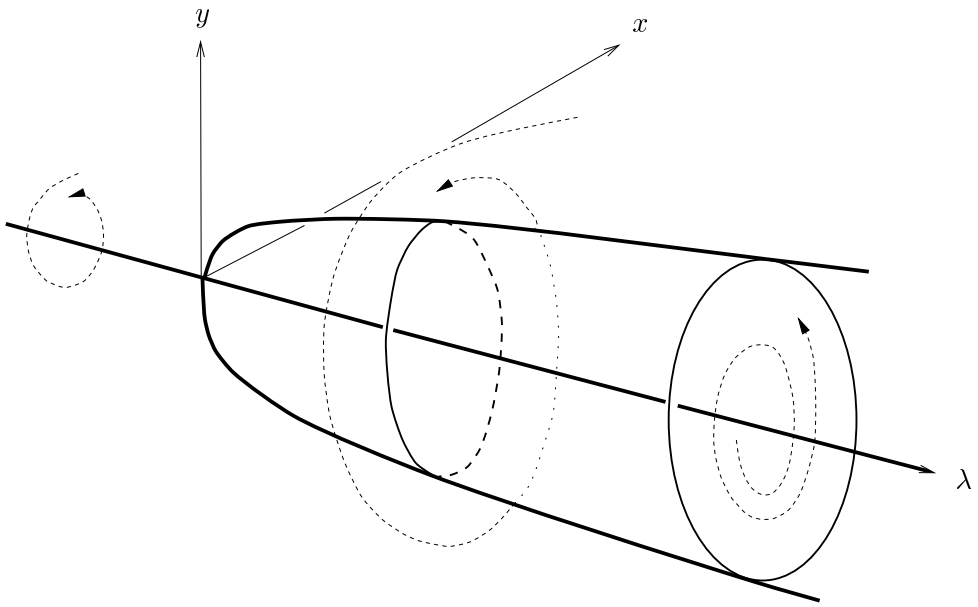

Figure 2.2:

In summary: As λ passes through zero there is a birth of periodic orbits in (2.2). The eigenvalues of $\begin{pmatrix} \lambda & 1 \\ -1 & \lambda \end{pmatrix}$ *are $\lambda \pm i$, and these are purely imaginary at $\lambda = 0$. This is a simple example of a Hopf bifurcation. We discuss this topic in Chapter 8.* □

Example 2.3 *Consider the differential equation*

$$\frac{d^2y}{dt^2} + \lambda \sin y = 0 \tag{2.3}$$

where $\lambda > 0$ is given and $y(t)$ is to be found on $t \in [0, l]$ subject to the boundary conditions

$$\frac{dy}{dt}(0) = 0 = \frac{dy}{dt}(l) \ . \tag{2.4}$$

This models the behaviour of an elastic rod occupying $0 \leq t \leq l$ which is fixed at each end and subject to a force λ in the direction of the rod (see [2], Chap. 1 for a figure and more detailed discussion).

Here $y(t)$ represents the angle the tangent to the rod at a distance t along the rod makes to the horizontal. Physically, as λ increases the rod can buckle. Obviously the trivial solution $y \equiv 0$ solves (2.3), (2.4) for all λ. The interesting solution is the buckled state $y \not\equiv 0$. The differential equation is tractable to theoretical analysis (the first step is to multiply (2.3) by $\frac{dy}{dt}$ and integrate) and it is shown in [2] that nontrivial solutions emanate from $(y, \lambda) = (0, m^2\pi^2/l^2)$, $m = 1, 2, 3, \ldots$ as shown in Figure 2.3. (The value of $y(0)$ is plotted on the vertical axis). The

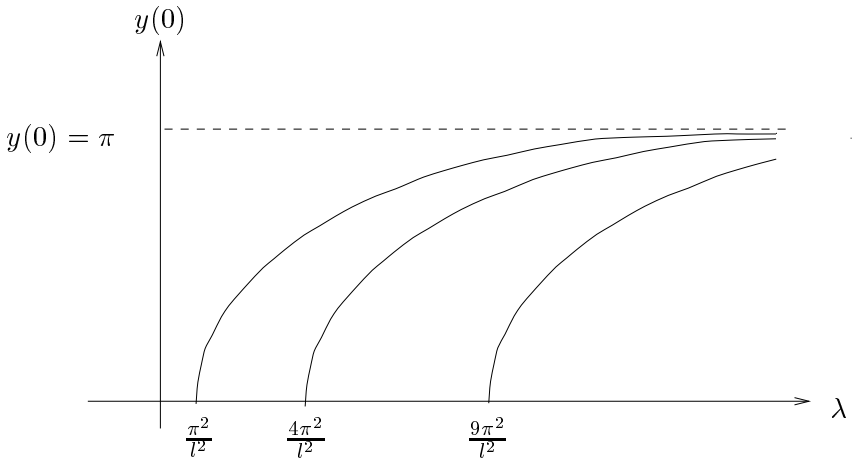

Figure 2.3:

nontrivial branches correspond to buckled states. This is classical "bifurcation from the trivial solution" and it was the analysis of this and similar buckling problems that prompted the initial studies in bifurcation theory.

In practise we may compute buckled states by approximating (2.3), (2.4). One way to do this is to set $h = l/n$ $(n \in \mathbb{Z})$ and introduce the mesh

$$t_i = ih \ , \quad i = 0, \cdots, n$$

of equally spaced points on $[0, l]$. Then set $y_i = y(t_i)$ and approximate

$$\frac{d^2y}{dt^2}(t_i) \quad \text{by} \quad \frac{1}{h^2}\left\{y_{i-1} - 2y_i + y_{i+1}\right\}.$$

Substituting in (2.3) and forcing equality gives the approximation

$$\frac{1}{h^2}\{Y_{i-1} - 2Y_i + Y_{i+1}\} + \lambda \sin Y_i = 0, \quad i = 1, \ldots, (n-1), \tag{2.5}$$

where $Y_i \simeq y(t_i)$. We can approximate (2.4) by

$$\frac{Y_1 - Y_0}{h} = 0 = \frac{Y_n - Y_{n-1}}{h} \ . \tag{2.6}$$

Now using (2.5) for $i = 1, .., (n-1)$ together with (2.6) yields the $(n-1)$ dimensional nonlinear

system

$$\begin{aligned}\boldsymbol{F}(\boldsymbol{Y},\lambda) &= A\boldsymbol{Y}+\lambda\sin(\boldsymbol{Y}),\\ &= \frac{1}{h^2}\begin{bmatrix} -1 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & & & & & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -1 \end{bmatrix}\begin{bmatrix} Y_1 \\ \\ \\ \\ Y_{n-1}\end{bmatrix}+\lambda\begin{bmatrix} \sin Y_1 \\ \\ \\ \\ \sin Y_{n-1}\end{bmatrix}=\mathbf{0}. \qquad (2.7)\end{aligned}$$

Clearly $\boldsymbol{Y}=\mathbf{0}$ is always a solution. It is of interest to find out (i) For what λ do nontrivial $\boldsymbol{Y}$ exist? (ii) Do they approximate the λ given by the ordinary differential equation theory outlined above? (iii) What is the corresponding $\boldsymbol{Y}$? (iv) How would we compute $\boldsymbol{Y}$ as λ varies? □

2.1 Some Multivariate Calculus

For $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$, $\boldsymbol{x}^T\boldsymbol{y}=\sum_{i=1^n} x_iy_i$, $\|\boldsymbol{x}\|=\{\boldsymbol{x}^T\boldsymbol{x}\}^{\frac{1}{2}}$. For $r>0$ define

$$\begin{aligned} B(\boldsymbol{x},r)&=\{\boldsymbol{y}\in\mathbb{R}^n:\|\boldsymbol{x}-\boldsymbol{y}\|<r\}\text{ open ball}\\ \overline{B}(\boldsymbol{x},r)&=\{\boldsymbol{y}\in\mathbb{R}^n:\|\boldsymbol{x}-\boldsymbol{y}\|\le r\}\text{ closed ball}\end{aligned}.$$

If D is an open subset of $\mathbb{R}^n$, then $\boldsymbol{F}:D\to\mathbb{R}^n$ is *continuous* at $\boldsymbol{x}\in D$ if $\forall\varepsilon>0\ \exists\delta>0$ s.t. $\boldsymbol{y}\in B(\boldsymbol{x},\delta)\Longrightarrow\boldsymbol{F}(\boldsymbol{y})\in B(\boldsymbol{F}(\boldsymbol{x}),\varepsilon)$.

A function $f:D\subseteq\mathbb{R}^n\to\mathbb{R}$ is called *continuously differentiable* at $\boldsymbol{x}$ if $\frac{\partial f}{\partial x_i}$ exists and is continuous at $\boldsymbol{x}$ for each $i=1,\cdots,n$.

If $\boldsymbol{F}:D\subseteq\mathbb{R}^n\to\mathbb{R}^n$ and $\boldsymbol{F}(\boldsymbol{x})=(f_1(\boldsymbol{x}),\cdots,f_n(\boldsymbol{x}))^T$ then the *Jacobian* of $\boldsymbol{F}$ is the matrix $\boldsymbol{F}_{\boldsymbol{x}}$ which is defined $\forall\boldsymbol{x}\in D$ by $(\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}))_{ij}=\frac{\partial f_i}{\partial x_j}(\boldsymbol{x})$.

($\boldsymbol{F}$ is called continuously differentiable at $\boldsymbol{x}$ if $\frac{\partial f_i}{\partial x_j}$ exists and is continuous at $\boldsymbol{x}$ for each i,j.

Definition A sequence of vectors $\{\boldsymbol{x}^k\}_{k=1}^{\infty}\subset\mathbb{R}^n$ is said to *converge* to $\boldsymbol{x}^*\in\mathbb{R}^n$ if $\|\boldsymbol{x}^k-\boldsymbol{x}^*\|\to 0$ for some vector norm $\|\cdot\|$ on $\mathbb{R}^n$.

Remark Since all norms on $\mathbb{R}^n$ are equivalent, the choice of norm in this definition is arbitrary.

Definition If a sequence of approximate solutions $\{\boldsymbol{x}^k\}\subset\mathbb{R}^n$ are converging to a solution $\boldsymbol{x}^*\in\mathbb{R}^n$ we say the convergence is *order* p if

$$\|\boldsymbol{x}^*-\boldsymbol{x}^{k+1}\|\le C\|\boldsymbol{x}^*-\boldsymbol{x}^k\|^p,\ \forall k\ge o\ ,$$

with C independent of k.

If $p = 1$ convergence is called *linear*.

If $p = 2$ convergence is called *quadratic*.

3 Newton's Method and the Implicit Function Theorem

The main computational tool to solve systems like (2.1) is *Newton's method*, which we discuss in §3.1. The main theoretical tool, which also has important numerical implications is the *Implicit Function Theorem* which we discuss in §3.2. Two applications of the Implicit Function Theorem are discussed in §3.3.

Recall that if $\boldsymbol{G} : D \subset \mathbb{R}^n \to \mathbb{R}^n$, and $\|\cdot\|$ denotes a norm on $\mathbb{R}^n$ then $\boldsymbol{G}$ is called *Lipschitz continuous* (with respect to $\|\cdot\|$) if there exists $\gamma \in \mathbb{R}$, such that for all $\boldsymbol{x}, \boldsymbol{y} \in D$

$$\|\boldsymbol{G}(\boldsymbol{x}) - \boldsymbol{G}(\boldsymbol{y})\| \leq \gamma \|\boldsymbol{x} - \boldsymbol{y}\|, \tag{3.1}$$

and we write $\boldsymbol{G} \in \mathrm{Lip}_\gamma(D)$. Throughout these lectures $\|\cdot\|$ will denote the Euclidean norm $\|\boldsymbol{x}\| = \{\boldsymbol{x}^T\boldsymbol{x}\}^{1/2}$ on $\mathbb{R}^n$ and also the matrix norm induced by the Euclidean norm. With respect to this norm, $B(\boldsymbol{x}, r)$ will denote the open ball in $\mathbb{R}^n$ with centre $\boldsymbol{x}$ and radius r, while $\bar{B}(\boldsymbol{x}, r)$ denotes its closure.

3.1 Newton's Method for Systems

A very nice treatment of Newton's method for systems of nonlinear equations is given in [6]. To find a root, $\boldsymbol{x}_0$ say, of $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{0}$, Newton's method for a given starting guess $\boldsymbol{x}^0$ is

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \boldsymbol{d}^k, \quad \text{where } \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}^k)\boldsymbol{d}^k = -\boldsymbol{F}(\boldsymbol{x}^k), \qquad k \geq 0. \tag{3.2}$$

Theorem 3.1 *Assume*

(a) $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ *is continuously differentiable in an open convex set* $D \subset \mathbb{R}^n$, *and* $\boldsymbol{F}_{\boldsymbol{x}} \in \mathrm{Lip}_\gamma(B(\boldsymbol{x}_0, r))$, *for some* $r > 0$,

(b) $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}_0)$ *is nonsingular.*

Then provided $\boldsymbol{x}^0$ *satisfies* $\boldsymbol{x}^0 \in B(\boldsymbol{x}_0, \epsilon)$, *for small enough* $\epsilon > 0$, *Newton's method is well defined and* $\boldsymbol{x}^k \to \boldsymbol{x}_0$ *quadratically. (See [6] for a fuller account.)*

In fact ϵ can be given explicitly as $\min\{r, 1/2\beta\gamma\}$, where $\beta = \|(\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}_0))^{-1}\|$, and this approaches 0 as $r \to 0$ or $\beta \to \infty$ or $\gamma \to \infty$.

It is interesting to formulate the matrix eigenvalue problem as a system of $(n+1)$ equations in $(n+1)$ unknowns, as is done in the following example.

Example 3.2 *Let A be a real symmetric matrix, with simple eigenvalue μ_0 and corresponding eigenvector $\boldsymbol{\phi}_0$ satisfying $\boldsymbol{\phi}_0^T\boldsymbol{\phi}_0 = 1$. Consider the problem of computing $(\boldsymbol{\phi}_0, \mu_0)$ by Newton's method. Define $\boldsymbol{F} : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ by*

$$\boldsymbol{F}(\boldsymbol{y}) = \begin{pmatrix} A\boldsymbol{\phi} - \mu\boldsymbol{\phi} \\ \boldsymbol{\phi}^T\boldsymbol{\phi} - 1 \end{pmatrix} = \boldsymbol{0}, \qquad \text{where } \boldsymbol{y} = \begin{pmatrix} \boldsymbol{\phi} \\ \mu \end{pmatrix} \in \mathbb{R}^{n+1}. \tag{3.3}$$

If we apply Newton's method to compute the eigenpair $(\boldsymbol{\phi}^T, \mu)^T$ then the first step in verifying the convergence would be to check that the Jacobian matrix $\boldsymbol{F}_{\boldsymbol{y}}$ is nonsingular. A simple calculation shows

$$\boldsymbol{F}_{\boldsymbol{y}}(\boldsymbol{y}) = \begin{pmatrix} A - \mu I & -\boldsymbol{\phi} \\ 2\boldsymbol{\phi}^T & 0 \end{pmatrix}.$$

The proof that this is nonsingular at $\boldsymbol{y}_0^T = (\boldsymbol{\phi}_0^T, \mu_0)$ may be obtained directly or by application of case (ii) *of the ABCD Lemma, which we now state.*

Lemma 3.3 ("ABCD Lemma" (Keller [25])) *Given an $n \times n$ real matrix A, $\boldsymbol{c}$, $\boldsymbol{b} \in \mathbb{R}^n$, $d \in \mathbb{R}$, consider the $(n+1) \times (n+1)$ bordered matrix*

$$M = \begin{pmatrix} A & \boldsymbol{b} \\ \boldsymbol{c}^T & d \end{pmatrix}.$$

(i) If A is nonsingular then M is nonsingular if and only if $d - \boldsymbol{c}^T A^{-1}\boldsymbol{b} \neq 0$.

(ii) If $\operatorname{rank}(A) = n-1$, M is nonsingular if and only if
$\boldsymbol{\psi}^T\boldsymbol{b} \neq 0$ for all $\boldsymbol{\psi} \in \ker(A^T) \setminus \{\boldsymbol{0}\}$,
and
$\boldsymbol{c}^T\boldsymbol{\phi} \neq 0$ for all $\boldsymbol{\phi} \in \ker(A) \setminus \{\boldsymbol{0}\}$.

(iii) If $\operatorname{rank}(A) \leq n-2$, then M is singular.

Clearly different normalisations for the eigenvectors are possible. Replacing $\boldsymbol{\phi}^T\boldsymbol{\phi} = 1$ in (3.3) with $\boldsymbol{e}_r^T\boldsymbol{\phi} = 1$, where $\boldsymbol{e}_r$ is the unit vector with $(\boldsymbol{e}_r)_i = \delta_{ir}$, one can show that Newton's method applied to the eigenvalue problem can be interpreted as a version of inverse iteration (see [47] for more details).

3.2 The Implicit Function Theorem

The Implicit Function Theorem is obtained as an application of the Contraction Mapping Theorem to a nonlinear system with a parameter. So, let us first recall the Contraction Mapping Theorem.

Theorem 3.4 (Contraction Mapping Theorem) *Suppose*

(i) $\boldsymbol{G} \in \mathrm{Lip}_\alpha(\bar{B}(\boldsymbol{x}_0, r))$ *for some* $r > 0$, *with* $0 \leq \alpha < 1$;

(ii) $\|\boldsymbol{x}_0 - \boldsymbol{G}(\boldsymbol{x}_0)\| \leq (1-\alpha)r$.

Then

(a) For all $\boldsymbol{x}^0 \in \bar{B}(\boldsymbol{x}_0, r)$, *the sequence* $\boldsymbol{x}^k$ *defined by* $\boldsymbol{x}^{k+1} = \boldsymbol{G}(\boldsymbol{x}^k)$ *converges to a limit* $\boldsymbol{x}^\star \in \bar{B}(\boldsymbol{x}_0, r)$;

(b) $\boldsymbol{x}^\star$ *is the unique fixed point of* $\boldsymbol{G}$ *in* $\bar{B}(\boldsymbol{x}_0, r)$.

The proof is in most books on nonlinear equations. Other versions of this theorem replace the assumption *(ii)* with the requirement that $\boldsymbol{G}(\bar{B}(\boldsymbol{x}_0, r)) \subseteq \bar{B}(\boldsymbol{x}_0, r)$. For numerical analysis purposes the present version is better since checking *(ii)* requires only checking that $\boldsymbol{G}(\boldsymbol{x}_0)$ should not be too far from $\boldsymbol{x}_0$. The proof of this version of the Contraction Mapping Theorem is in [38].

The contraction mapping theorem has many uses. One example of its use is in the analysis of the modified Newton method, given by the fixed point iteration $\boldsymbol{x}^{k+1} = \boldsymbol{G}(\boldsymbol{x}^k)$, where

$$\boldsymbol{G}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}^0)^{-1}\boldsymbol{F}(\boldsymbol{x}). \tag{3.4}$$

Using the Contraction Mapping Theorem it can be shown that if $\boldsymbol{x}^0$ is sufficiently close to a solution $\boldsymbol{x}_0$ of $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{0}$ then $\boldsymbol{x}^k \to \boldsymbol{x}_0$ linearly as $k \to \infty$.

Another example of its use is in the proof of the Implicit Function Theorem. Consider the parameter dependent problem

$$\boldsymbol{F}(\boldsymbol{x}, \lambda) = \boldsymbol{0},$$

for $(\boldsymbol{x}, \lambda) \in D$, where $\boldsymbol{F} : D \subset \mathbb{R}^{n+1} \to \mathbb{R}^n$. Let S be the solution set

$$S = \{(\boldsymbol{x}, \lambda) \in D : \quad \boldsymbol{F}(\boldsymbol{x}, \lambda) = \boldsymbol{0}\}.$$

It is natural to ask the following question. If $(\boldsymbol{x}_0, \lambda_0) \in S$ and λ is near λ_0, is there a corresponding unique $\boldsymbol{x}(\lambda)$ such that $(\boldsymbol{x}(\lambda), \lambda) \in S$ and $\boldsymbol{x}(\lambda_0) = \boldsymbol{x}_0$? If so, we say $\boldsymbol{x}$ is *parametrised* by λ near $(\boldsymbol{x}_0, \lambda_0)$, written "$\boldsymbol{x} = \boldsymbol{x}(\lambda)$ near $(\boldsymbol{x}_0, \lambda_0)$". The Implicit Function Theorem provides the answer, but first consider a simple example.

Example 3.5 *Consider*

$$f(x, \lambda) = x^2 + \lambda^2 - 1.$$

Clearly if $|x_0| < 1$ *and* $(x_0, \lambda_0) \in S$ *then* $x = x(\lambda)$ *near* (x_0, λ_0).

Since $\boldsymbol{F}$ now depends on $\boldsymbol{x}$ and λ, we use the notation $\boldsymbol{F}_{\boldsymbol{x}}$ to mean the $n \times n$ matrix with (i,j)th element $\dfrac{\partial F_i}{\partial x_j}$, and by $\boldsymbol{F}_\lambda$ we mean the $n \times 1$ vector with elements $\dfrac{\partial F_i}{\partial \lambda}$. If $(\boldsymbol{x}_0, \lambda_0) \in D$ we write

$$\begin{aligned} \boldsymbol{F}^0 &= \boldsymbol{F}(\boldsymbol{x}_0, \lambda_0), \quad \boldsymbol{F}^0_{\boldsymbol{x}} = \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}_0, \lambda_0), \\ \boldsymbol{F}^0_\lambda &= \boldsymbol{F}_\lambda(\boldsymbol{x}_0, \lambda_0), \quad \text{etc .} \end{aligned}$$

The Implicit Function Theorem

In the proof of this theorem we assume that for all $(\boldsymbol{x}, \lambda), (\boldsymbol{x}, \mu), (\boldsymbol{y}, \lambda) \in D$,

$$\begin{aligned} &\text{(A1)} \qquad \|\boldsymbol{F}(\boldsymbol{x}, \lambda) - \boldsymbol{F}(\boldsymbol{x}, \mu)\| \le \sigma_2 |\lambda - \mu| \\ &\text{(A2)} \qquad \|\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}, \lambda) - \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{y}, \lambda)\| \le \gamma_1 \|\boldsymbol{x} - \boldsymbol{y}\| \\ &\text{(A3)} \qquad \|\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}, \lambda) - \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}, \mu)\| \le \gamma_2 |\lambda - \mu|. \end{aligned}$$

Clearly (A1–A3) hold if $\boldsymbol{F}$ has two continuous derivatives with respect to $(\boldsymbol{x}, \lambda) \in D$. In many applications in fact $\boldsymbol{F}$ will be infinitely continuously differentiable on D, which we write as $\boldsymbol{F} \in C^\infty(D)$.

Theorem 3.6 *(***Implicit Function Theorem***) Suppose (A1–A3) hold and suppose there exists* $(\boldsymbol{x}_0, \lambda_0) \in D$ *such that*

(A4) $\boldsymbol{F}(\boldsymbol{x}_0, \lambda_0) = \boldsymbol{0}$,

(A5) $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}_0, \lambda_0)$ *is nonsingular.*

Then there exist neighbourhoods $B(\lambda_0, \varepsilon_\lambda)$, $B(\boldsymbol{x}_0, \varepsilon_x)$ *of* λ_0, $\boldsymbol{x}_0$ *such that for all* $\lambda \in B(\lambda_0, \varepsilon_\lambda)$ *there exists* $\boldsymbol{x}(\lambda) \in B(\boldsymbol{x}_0, \varepsilon_x)$ *with*

(a) $\boldsymbol{F}(\boldsymbol{x}(\lambda), \lambda) = \boldsymbol{0}$,

(b) $\boldsymbol{x}(\lambda)$ *is the unique solution of* $\boldsymbol{F}(\boldsymbol{x}, \lambda) = \boldsymbol{0}$ *in* $B(\boldsymbol{x}_0, \varepsilon_x)$,

(c) $\boldsymbol{x}(\lambda_0) = \boldsymbol{x}_0$,

(d) $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}(\lambda), \lambda)$ *is nonsingular for all* $\lambda \in B(\lambda_0, \varepsilon_\lambda)$,

(e) $\boldsymbol{x}(\lambda)$ *is continuous with respect to* $\lambda \in B(\lambda_0, \varepsilon_\lambda)$.

Remark If $\boldsymbol{F} \in \mathcal{C}^\infty(D)$ then (e) can be replaced by $\boldsymbol{x} \in \mathcal{C}^\infty(B(\lambda_0, \varepsilon_\lambda))$.

Proof The proof of *(a),(b)* and *(c)* uses the Contraction Mapping Theorem applied to $\boldsymbol{K}(\boldsymbol{x}, \lambda) := \boldsymbol{x} - \left(\boldsymbol{F}^0_{\boldsymbol{x}}\right)^{-1} \boldsymbol{F}(\boldsymbol{x}, \lambda)$, which is very like the form of the mapping $\boldsymbol{G}$ used in the theory of the modified Newton method (3.4). *(d)* follows using 3.1.4 in [6], and *(e)* by standard manipulation. □

With respect to the parameter dependent problem we make the following definition.

Definition 3.7 $(\boldsymbol{x}_0, \lambda_0) \in S$ *is called a regular point of* S *if* $\boldsymbol{F}^0_{\boldsymbol{x}}$ *is nonsingular. The Implicit Function Theorem can then be applied to show* $\boldsymbol{x} = \boldsymbol{x}(\lambda)$ *near* $(\boldsymbol{x}_0, \lambda_0)$. *If a point* $(\boldsymbol{x}_0, \lambda_0) \in S$ *is not regular it is called a singular point.*

Example 3.8 *Consider*

$$\boldsymbol{F}(\boldsymbol{x}, \lambda) = \begin{bmatrix} x_1^2 + x_2^2 - \lambda \\ x_2^2 - 2x_1 + 1 \end{bmatrix}.$$

It is clear that the solution set S *is the intersection of the circle centred on the origin with radius* $\sqrt{\lambda}$, *and a parabola. (It is helpful to draw a sketch.) Clearly* $\boldsymbol{F} \in \mathcal{C}^\infty(\mathbb{R}^{n+1})$ *and* $\boldsymbol{F}_{\boldsymbol{x}} = \begin{bmatrix} 2x_1 & 2x_2 \\ -2 & 2x_2 \end{bmatrix}$. *Consider* $(\boldsymbol{x}_0, \lambda_0) := (0.73, 0.68, 1)$. *Then* $(\boldsymbol{x}_0, \lambda_0) \in S$ *and* $\det(\boldsymbol{F}^0_{\boldsymbol{x}}) = 4(x_0)_1(x_0)_2 + 4(x_0)_2 \neq 0$. *So* $(\boldsymbol{x}_0, \lambda_0)$ *is a regular point and the Implicit Function Theorem shows that* $\boldsymbol{x} = \boldsymbol{x}(\lambda)$ *near* $(\boldsymbol{x}_0, \lambda_0)$. *Consider instead* $(\boldsymbol{x}_0, \lambda_0) = (1/2, 0, 1/4) \in S$. *Then* $\boldsymbol{F}^0_{\boldsymbol{x}}$ *is found to be singular. So* $(\boldsymbol{x}_0, \lambda_0)$ *is a singular point and we cannot conclude that* $\boldsymbol{x} = \boldsymbol{x}(\lambda)$ *near* $(\boldsymbol{x}_0, \lambda_0)$. *(Plotting the path of solutions* $\boldsymbol{x}(\lambda)$ *against* λ *shows why not.) The difficulty here is simply that the solution set turns around at* $\lambda = \lambda_0 = \frac{1}{4}$.

This is a special type of singular point called a fold or turning point.

Definition 3.9 *If* $(\boldsymbol{x}_0, \lambda_0) \in S$ *is a singular point and if Rank*$(\boldsymbol{F}^0_{\boldsymbol{x}}) = n - 1$ *then* $(\boldsymbol{x}_0, \lambda)$ *is called a fold point (or turning point) if* $\boldsymbol{F}^0_\lambda \notin$ *Image* $(\boldsymbol{F}^0_{\boldsymbol{x}})$. *In this case the* $n \times (n+1)$ *augmented*

Jacobian $[\boldsymbol{F}^0_{\boldsymbol{x}}|\boldsymbol{F}^0_\lambda]$ *must have rank* n *and hence has a subset of* n *linearly independent columns. By selecting the variables corresponding to these columns as the dependent variables we can still apply the Implicit Function Theorem.*

Example 3.10 *Consider again Example 3.8.* $(\boldsymbol{x}_0, \lambda_0) = (1/2, 0, 1/4)$*, and hence*

$$[\boldsymbol{F}^0_{\boldsymbol{x}}|\boldsymbol{F}^0_\lambda] = \begin{bmatrix} 1 & 0 & -1 \\ -2 & 0 & 0 \end{bmatrix},$$

which has full rank. The first and third columns are linearly independent so if we write

$$\boldsymbol{G}(\boldsymbol{y}, x_2) = \begin{bmatrix} x_1^2 - \lambda + x_2^2 \\ -2x_1 + 1 + x_2^2 \end{bmatrix}.$$

The solution set for $\boldsymbol{G} = \boldsymbol{0}$ *is identical to the solution of* $\boldsymbol{F} = \boldsymbol{0}$ *but* x_2 *is now considered to be a parameter and* $\boldsymbol{y} = (x_1, \lambda)$*. Then* $\boldsymbol{G}_{\boldsymbol{y}} = \begin{bmatrix} 2x_1 & -1 \\ -2 & 0 \end{bmatrix}$*, which is nonsingular at* $(\boldsymbol{y}_0, (x_2)_0) = (\frac{1}{2}, \frac{1}{4}, 0)$ *so the Implicit Function Theorem shows that* $\boldsymbol{y} = \boldsymbol{y}(x_2)$ *near* $(\boldsymbol{y}_0, (x_2)_0)$*.* □

This example shows that change of parametrisation can remove the problems of a fold point. If a singular point is not a fold point, further analysis is required (see §6).

3.3 Two Examples

We now give two examples

Example 3.11 *(See Example 2.3) Consider the* $n-1$ *dimensional nonlinear system with* $\boldsymbol{F}$: $\mathbb{R}^n \to \mathbb{R}^{n-1}$ *given in (2.7). Clearly* $(\boldsymbol{Y}_0, \lambda_0) := (\boldsymbol{0}, \lambda_0) \in S$*, for all* $\lambda_0 \in \mathbb{R}$*, and*

$$\begin{aligned} [\boldsymbol{F}_{\boldsymbol{Y}}|\boldsymbol{F}_\lambda] &= [A - \lambda \mathrm{diag}(\cos \boldsymbol{Y})|\sin \boldsymbol{Y}] \\ &= [A - \lambda_0 I|\boldsymbol{0}] \text{ at } (\boldsymbol{Y}_0, \lambda_0). \end{aligned}$$

Now A *has the* $(n-1)$ *eigenvalues* $\mu_k = \frac{n^2}{l^2}\left(2 - 2\cos\frac{k\pi}{(n-1)}\right)$*, (with corresponding eigenfunctions* $\boldsymbol{x}^k$*, with* $x^k_j = \cos(k\pi(2j-1)/2(n-1))$*) for* $k = 0, ..., (n-2)$*, which are distinct and hence simple. So if* $\lambda_0 \neq -\mu_k$ *for any* k *then* $(\boldsymbol{Y}_0, \lambda_0)$ *is a regular point, and so near* λ_0 *we have* $\boldsymbol{Y} = \boldsymbol{Y}(\lambda)$*. But if* $\lambda_0 = \mu_k$*, then* $\mathrm{Rank}(A - \lambda_0 I) = n - 2$*, so* $(\boldsymbol{Y}_0, \lambda_0)$ *is a singular point. In addition* $\boldsymbol{F}^0_\lambda = \boldsymbol{0} \in \mathrm{Image}(A - \lambda_0 I) = \mathrm{Image}(\boldsymbol{F}^0_{\boldsymbol{Y}})$*, so* $(\boldsymbol{Y}_0, \lambda_0)$ *is not a fold point either.*

Example 3.12 *(Perturbation theory for algebraically simple eigenvalues.)*
Let A be a real $n \times n$ matrix with a simple eigenvalue μ_0 (i.e. algebraic multiplicity is 1) and corresponding eigenvector $\boldsymbol{\phi}_0$. If A is perturbed to $A + \epsilon B$, one question is to find the dominant term in the perturbation of μ_0.

Start the perturbation theory by considering the nonlinear system (cf Example 3.2 but without the assumption that A is symmetric)

$$F(\boldsymbol{y}, \epsilon) := \begin{pmatrix} (A + \epsilon B)\boldsymbol{\phi} - \mu\boldsymbol{\phi} \\ \boldsymbol{\phi}^T\boldsymbol{\phi} - 1 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} \boldsymbol{\phi} \\ \mu \end{pmatrix}. \tag{3.5}$$

Clearly with $\boldsymbol{y}_0 = (\boldsymbol{\phi}_0^T, \mu_0)^T$, $\boldsymbol{F}(\boldsymbol{y}_0, 0) = \mathbf{0}$, and $\boldsymbol{F}_{\boldsymbol{y}}(\boldsymbol{y}_0, 0)$ is nonsingular. (This is proved using part (ii) of Lemma 3.3, though note that the condition of algebraic simplicity is needed since A is no longer assumed symmetric.) Thus using the Implicit Function Theorem, for small $|\epsilon|$ there exists a unique $y(\epsilon)$ such that $\boldsymbol{F}(\boldsymbol{y}(\epsilon), \epsilon) = 0$ and $\boldsymbol{F}_{\boldsymbol{y}}(\boldsymbol{y}(\epsilon), \epsilon)$ is nonsingular. The latter result ensures that $\mu(\epsilon)$ is simple. Since $\mu(\epsilon) \in C^\infty(\mathbb{R})$ we can write $\mu(\epsilon) = \mu_0 + \epsilon\mu'(0) + \mathcal{O}(\epsilon^2)$. To find the dominant term in the error we need to find $\mu'(0)$. To do this we differentiate $(A + \epsilon B)\boldsymbol{\phi}(\epsilon) = \mu(\epsilon)\boldsymbol{\phi}(\epsilon)$ with respect to ϵ, set $\epsilon = 0$, and multiply on the left by $\boldsymbol{\psi}_0 \in \ker((A - \mu_0 I)^T) \setminus \{\mathbf{0}\}$. This leads to

$$\mu'(0) = \boldsymbol{\psi}_0^T B\boldsymbol{\phi}_0 / \boldsymbol{\psi}_0^T\boldsymbol{\phi}_0.$$

(Note that $\boldsymbol{\psi}_0^T\boldsymbol{\phi}_0 \neq 0$. If $\boldsymbol{\psi}_0^T\boldsymbol{\phi}_0 = 0$, then $\boldsymbol{\phi}_0 \in \mathrm{Image(A - \mu_0 I)}$ and $\dim(\ker(\mathrm{A - \mu_0 I})^2) > 1$, contradicting the assumption of algebraic simplicity.) If A is symmetric then $\mu'(0) = \boldsymbol{\phi}_0^T B\boldsymbol{\phi}_0 / \boldsymbol{\phi}_0^T\boldsymbol{\phi}_0$.

4 Computation of solution paths

In this section we consider the general problem

$$\boldsymbol{F}(\boldsymbol{x}, \lambda) = \mathbf{0}, \tag{4.1}$$

where $\boldsymbol{F} : \mathbb{R}^{n+1} \to \mathbb{R}^n$, $\boldsymbol{F} \in \mathcal{C}^\infty(\mathbb{R}^{n+1})$. Set

$$S = \{(\boldsymbol{x}, \lambda) \in \mathbb{R}^{n+1} : \boldsymbol{F}(\boldsymbol{x}, \lambda) = \mathbf{0}\}. \tag{4.2}$$

Often in applications one is interested in computing the whole set S or a continuous portion of it. For example in fluid dynamics $\boldsymbol{x}$ may represent the velocity and pressure of a flow, whereas λ is some physical parameter such as the Reynolds number. In practice S is computed by finding

a discrete set of points on S and then using some graphics package to interpolate. So the basic numerical question to consider is: Given a point $(\boldsymbol{x}_0, \lambda_0) \in S$ how would we compute a nearby point on S? Throughout we use the notation $\boldsymbol{F}^0 = \boldsymbol{F}(\boldsymbol{x}_0, \lambda_0)$, $\boldsymbol{F}^0_{\boldsymbol{x}} = \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}_0, \lambda_0)$, etc.

If $\boldsymbol{F}^0_{\boldsymbol{x}}$ is nonsingular then the Implicit Function Theorem implies that for λ near λ_0 the solutions of $\boldsymbol{F}(\boldsymbol{x}, \lambda) = \boldsymbol{0}$ satisfy $\boldsymbol{x} = \boldsymbol{x}(\lambda)$ with $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}(\lambda), \lambda)$ nonsingular. Hence Theorem 3.1 implies that Newton's method for finding the solution $\boldsymbol{x}(\lambda)$ of $\boldsymbol{F}(\boldsymbol{x}(\lambda), \lambda) = \boldsymbol{0}$ with starting value $\boldsymbol{x}_0$ will converge in some ball centred on $\boldsymbol{x}(\lambda)$ for small enough $\lambda - \lambda_0$.

A simple strategy for computing a point of S near $(\boldsymbol{x}_0, \lambda_0)$ is to choose a steplength $\Delta\lambda$, set $\lambda_1 = \lambda_0 + \Delta\lambda$ and solve

$$\boldsymbol{F}(\boldsymbol{x}, \lambda_1) = \boldsymbol{0}$$

by Newton's method with starting guess $\boldsymbol{x}^0 = \boldsymbol{x}_0$. We then know this will work if $\Delta\lambda$ is sufficiently small. However this method will fail (or at best require repeated reduction of step $\Delta\lambda$) as a turning point is approached. For this reason the pseudo-arclength method described in the next section was introduced.

4.1 Keller's pseudo-arclength continuation [25]

Ideally we would like a method that has no difficulties near, or passing round, a fold point. This isn't unreasonable since at a fold point there is nothing geometrically "wrong" with the curve, though λ is the wrong parameter to use to describe the curve. In this section we shall assume that there is an arc of S such that at all points in the arc

$$\text{Rank } [\boldsymbol{F}_{\boldsymbol{x}} | \boldsymbol{F}_\lambda] = n, \tag{4.3}$$

and so any point in the arc is either a regular or fold point of S. The Implicit Function Theorem thus implies that the arc is a smooth curve in $\mathbb{R}^{n+1}$, and so there is a unique tangent direction at each point of the arc.

Let t denote any parameter used to describe the arc. Then along the arc $(\boldsymbol{x}, \lambda) = (\boldsymbol{x}(t), \lambda(t))$. Suppose $(\boldsymbol{x}_0, \lambda_0) = (\boldsymbol{x}(t_0), \lambda(t_0))$ and denote the tangent at $(\boldsymbol{x}_0, \lambda_0)$ by $\boldsymbol{\tau}_0 = (\dot{\boldsymbol{x}}_0, \dot{\lambda}_0)$ where

$$\dot{\boldsymbol{x}} = \tfrac{d\boldsymbol{x}}{dt}, \qquad \dot{\lambda} = \frac{d\lambda}{dt},$$
$$\dot{\boldsymbol{x}}_0 = \dot{\boldsymbol{x}}(t_0), \qquad \dot{\lambda}_0 = \dot{\lambda}(t_0).$$

The tangent $\boldsymbol{\tau}_0$ is well defined even if $(\boldsymbol{x}_0, \lambda_0)$ is a fold point and can be computed in practice using the following result.

Lemma 4.1 *Assume (4.3). Then the tangent at* $(\boldsymbol{x}_0, \lambda_0) \in S$ *satisfies*

$$\boldsymbol{\tau}_0 \in \ker\, [\boldsymbol{F}^0_{\boldsymbol{x}} | \boldsymbol{F}^0_\lambda]. \tag{4.4}$$

Proof Since $\boldsymbol{F}(\boldsymbol{x}(t), \lambda(t)) = \boldsymbol{0}$, differentiating with respect to t gives

$$\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}(t), \lambda(t))\dot{\boldsymbol{x}}(t) + \boldsymbol{F}_\lambda(\boldsymbol{x}(t), \lambda(t))\dot{\lambda}(t) = \boldsymbol{0}.$$

Put $t = t_0$ and we have $\begin{bmatrix} \dot{\boldsymbol{x}}_0 \\ \dot{\lambda}_0 \end{bmatrix} \in \ker\, [\boldsymbol{F}^0_x | \boldsymbol{F}^0_\lambda]$ and so the result follows. □

Suppose now that $\boldsymbol{\tau}_0 = [\boldsymbol{s}_0, \sigma_0]$ denotes the unit tangent i.e. $\boldsymbol{\tau}_0^T \boldsymbol{\tau}_0 = 1$. We can use this vector to devise an extended system which can be solved by Newton's method without fail for a point $(\boldsymbol{x}_1, \lambda_1)$ on S near $(\boldsymbol{x}_0, \lambda_0)$. The appropriate extended system is

$$\boldsymbol{H}(\boldsymbol{y}, t) = \boldsymbol{0} \tag{4.5}$$

where $\boldsymbol{y} = (\boldsymbol{x}, \lambda) \in \mathbb{R}^{n+1}$ and $\boldsymbol{H} : \mathbb{R}^{n+2} \to \mathbb{R}^{n+1}$. is given by

$$\boldsymbol{H}(\boldsymbol{y}, t) = \begin{bmatrix} \boldsymbol{F}(\boldsymbol{x}, \lambda) \\ \boldsymbol{s}_0^T(\boldsymbol{x} - \boldsymbol{x}_0) + \sigma_0(\lambda - \lambda_0) - (t - t_0) \end{bmatrix}. \tag{4.6}$$

The last equation in system (4.5) is the equation of the plane perpendicular to $\boldsymbol{\tau}_0$ a distance $\Delta t = (t - t_0)$ from t_0 (see Figure 4.1). So in (4.5) we in fact implement a specific parametrisation local to $(\boldsymbol{x}_0, \lambda_0)$, namely parametrisation by the length of the projection of $(\boldsymbol{x}, \lambda)$ onto the tangent direction at $(\boldsymbol{x}_0, \lambda_0)$.

With $\boldsymbol{y}_0 = (\boldsymbol{x}_0, \lambda_0)$, we have $H(\boldsymbol{y}_0, t_0) = \boldsymbol{0}$ and $\boldsymbol{H}\boldsymbol{y}(\boldsymbol{y}_0, t_0) = \begin{bmatrix} \boldsymbol{F}^0_{\boldsymbol{x}} & \boldsymbol{F}^0_\lambda \\ \boldsymbol{s}_0^T & \sigma_0 \end{bmatrix}$. Since $(\boldsymbol{s}_0^T, \sigma_0)^T$ is orthogonal to each of the rows of $[\boldsymbol{F}^0_{\boldsymbol{x}}, \boldsymbol{F}^0_\lambda]$, the matrix $\boldsymbol{H}\boldsymbol{y}(\boldsymbol{y}_0, t_0)$ is nonsingular and so by the Implicit Function Theorem solutions of (4.5) satisfy $\boldsymbol{y} = (\boldsymbol{x}, \lambda) = (\boldsymbol{x}(t), \lambda(t))$ for t near t_0. For $t_1 = t_0 + \Delta t$ and Δt sufficiently small we know that $\boldsymbol{F}(\boldsymbol{y}, t_1) = \boldsymbol{0}$ has a unique solution $\boldsymbol{y} = \boldsymbol{y}(t_1) = (\boldsymbol{x}_1, \lambda_1)$ and $\boldsymbol{H}\boldsymbol{y}(\boldsymbol{y}(t_1), t_1)$ is nonsingular. Thus Newton's method will converge for small enough Δt. If we take as starting guess $\boldsymbol{y}^0 = \boldsymbol{y}_0 = (\boldsymbol{x}_0, \lambda_0)$, it is a straightforward exercise to show, (i) $\boldsymbol{y}^1$, the first Newton iterate, is given by $\boldsymbol{y}^1 = (\boldsymbol{x}_0, \lambda_0) + \Delta t(\boldsymbol{s}_0, \sigma_0)$, that is, the first iterate "steps out" along the tangent, as one might expect, and (ii) $\boldsymbol{\tau}_0^T(\boldsymbol{y}^k - \boldsymbol{y}_0) = \Delta t \quad \forall\, k \geq 1$, which means that all the Newton iterates lie in the plane shown in figure 4.1.

Since length along the tangent at $(\boldsymbol{x}_0, \lambda_0)$ is used as parameter this technique is called *pseudo-arclength continuation* ([25]).

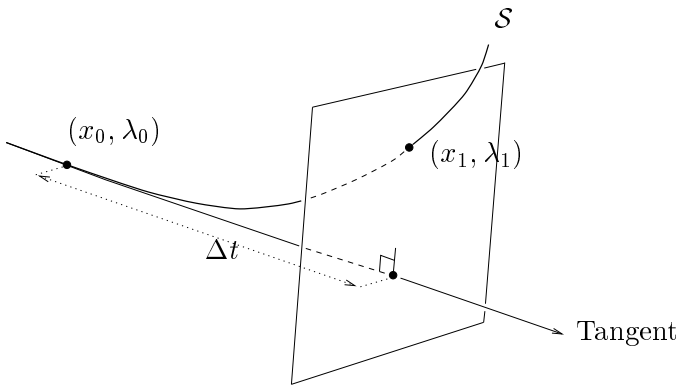

Figure 4.1:

Another interpretation Since we know that solutions of (4.5) satisfy $\boldsymbol{y} = \boldsymbol{y}(t)$, t near t_0, where t is pseudo-arclength, we have $\boldsymbol{H}(\boldsymbol{y}(t), t) = \boldsymbol{0}$ and we can differentiate with respect to t to get $\boldsymbol{H}_{\boldsymbol{y}}(\boldsymbol{y}(t), t)\dot{\boldsymbol{y}}(t) + \boldsymbol{H}_t(\boldsymbol{y}(t), t) = \boldsymbol{0}$. Since $\boldsymbol{H}_{\boldsymbol{y}}$ is nonsingular for t near t_0,

$$\dot{\boldsymbol{y}}(t) = -\boldsymbol{H}_{\boldsymbol{y}}(\boldsymbol{y}(t), t)^{-1}\boldsymbol{H}_t(\boldsymbol{y}(t), t), \tag{4.7}$$

which is an ordinary differential equation for $\boldsymbol{y}(t)$ with initial condition

$$\boldsymbol{y}(t_0) = \boldsymbol{y}_0. \tag{4.8}$$

We can use Euler's method to solve (4.7), (4.8) in which case the first step is

$$\boldsymbol{y}^1_{\text{Euler}} = \boldsymbol{y}_0 - \Delta t[\boldsymbol{H}^0_{\boldsymbol{y}}]^{-1}\boldsymbol{H}^0_t. \tag{4.9}$$

Since $\boldsymbol{H}^0_t = \begin{bmatrix} \boldsymbol{0} \\ -1 \end{bmatrix}$, (4.9) is equivalent to

$$\begin{aligned} \boldsymbol{y}^1_{\text{Euler}} &= \boldsymbol{y}_0 + \begin{bmatrix} \boldsymbol{F}^0_{\boldsymbol{x}} & \boldsymbol{F}^0_{\lambda} \\ \boldsymbol{s}_0^T & \sigma_0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \Delta t \end{bmatrix} \\ &= \boldsymbol{y}^0 + \Delta t \begin{bmatrix} \boldsymbol{s}_0 \\ \sigma_0 \end{bmatrix}. \end{aligned}$$

So the first step of Newton's method for (4.5) is equivalent to one step of Euler's method applied to (4.7), (4.8). We can think of this as using Euler's method to provide a "predicted guess" for $\boldsymbol{y}(t_1) = (\boldsymbol{x}_1, \lambda_1)$. Then continuing with Newton's method can be thought of as "correcting" this initial guess.

The choice of an appropriate step control strategy for Δt seems to be harder than in the ODE case, perhaps because the real problem is $\boldsymbol{F}(\boldsymbol{x}(t), \lambda(t)) = \boldsymbol{0}$ and not the differential equation derived from it. This topic is discussed in §4.6 of [45] or §7.4 of [39] but experience indicates that simple techniques often work just as well as sophisticated approaches.

Practical implementation of pseudo-arclength continuation

The following is a suggested algorithm for implementing the pseudo-arclength continuation method introduced above.

Step 1 Suppose $\boldsymbol{F}^0_{\boldsymbol{x}}$ is nonsingular, solve

$$\boldsymbol{F}^0_{\boldsymbol{x}} \boldsymbol{z}_0 = -\boldsymbol{F}^0_{\lambda} \tag{4.10}$$

for $\boldsymbol{z}_0$. Then set $\begin{bmatrix} \boldsymbol{s}_0 \\ \sigma_0 \end{bmatrix} = \dfrac{1}{(\boldsymbol{z}_0^T \boldsymbol{z}_0 + 1)^{1/2}} \begin{bmatrix} \boldsymbol{z}_0 \\ 1 \end{bmatrix}$.

Step 2 (Euler predictor) Choose a step length Δt and set

$$\begin{bmatrix} \boldsymbol{x}^1 \\ \lambda^1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_0 \\ \lambda_0 \end{bmatrix} + \Delta t \begin{bmatrix} \boldsymbol{s}_0 \\ \sigma_0 \end{bmatrix}. \tag{4.11}$$

Step 3 (Newton's method) For $k \geq 1$ iterate

$$\begin{bmatrix} \boldsymbol{x}^{k+1} \\ \lambda^{k+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}^k \\ \lambda^k \end{bmatrix} + \begin{bmatrix} \boldsymbol{d}^k \\ \delta^k \end{bmatrix}$$

with

$$\begin{bmatrix} \boldsymbol{F}^k_{\boldsymbol{x}} & \boldsymbol{F}^k_{\lambda} \\ \boldsymbol{s}_0^T & \sigma_0 \end{bmatrix} \begin{bmatrix} \boldsymbol{d}^k \\ \delta^k \end{bmatrix} = - \begin{bmatrix} \boldsymbol{F}^k \\ \boldsymbol{s}_0^T(\boldsymbol{x}^k - \boldsymbol{x}_0) + \sigma_0(\lambda^k - \lambda_0) - \Delta t \end{bmatrix} \tag{4.12}$$

Note that if $\boldsymbol{F}^0_{\boldsymbol{x}}$ is singular then the continuation method will still work provided the condition Rank $[\boldsymbol{F}^0_{\boldsymbol{x}} | \boldsymbol{F}^0_{\lambda}] = n$ holds, but Step 1 will fail to find the tangent vector. In practise this problem usually does not arise since $\boldsymbol{F}^0_{\boldsymbol{x}}$ only becomes singular at isolated points on the solution set S and effectively the probability of landing precisely on such a point is 0. However, care is needed when $\boldsymbol{F}^0_{\boldsymbol{x}}$ is nearly singular as is discussed in the next subsection. As a further precaution many continuation methods monitor the determinant of $\boldsymbol{F}^0_{\boldsymbol{x}}$.

If the tangent at a singular point is required then the null vector $\boldsymbol{z}_0$ of $\boldsymbol{F}^0_{\boldsymbol{x}}$ can be computed (say, by the inverse power method) and then the tangent vector can be taken as

$$\begin{bmatrix} \boldsymbol{s}_0 \\ \sigma_0 \end{bmatrix} = \frac{1}{(\boldsymbol{z}_0^T \boldsymbol{z}_0)^{1/2}} \begin{pmatrix} \boldsymbol{z}_0 \\ 0 \end{pmatrix}.$$

4.2 Block Elimination

As seen in (4.12) it is repeatedly necessary to solve "bordered systems" with coefficient matrix

$$M^k = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{x}}^k & \boldsymbol{F}_{\lambda}^k \\ \boldsymbol{s}_0^T & \sigma_0 \end{bmatrix}.$$

In many applications, where $\boldsymbol{F}(\boldsymbol{x}, \lambda)$ arises from a solution of a differential equation, $\boldsymbol{F}_{\boldsymbol{x}}^k$ may have some special structure (e.g. tridiagonal, banded, sparse) which makes systems with matrix $\boldsymbol{F}_{\boldsymbol{x}}^k$ easy to solve, but this structure is not present in M^k. Then the "block elimination method" (see [25]) is useful for quickly solving such systems.

"Block elimination" is merely Gaussian Elimination performed blockwise. If A is an $n \times n$ matrix, $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$ and $d \in \mathbb{R}$ then the block matrix

$$M := \begin{pmatrix} A & \boldsymbol{b} \\ \boldsymbol{c}^t & d \end{pmatrix} = \begin{pmatrix} I & \boldsymbol{0} \\ \boldsymbol{l}_n^T & 1 \end{pmatrix} \begin{pmatrix} A & \boldsymbol{b} \\ \boldsymbol{0}^T & u_{n+1} \end{pmatrix}, \tag{4.13}$$

where $\boldsymbol{l}_n = \boldsymbol{c}^T A^{-1}$, $u_{n+1} = d - \boldsymbol{c}^T A^{-1} \boldsymbol{b}$. If $\boldsymbol{l}_{n+1}$ and u_{n+1} are computed, then the system

$$\begin{pmatrix} A & \boldsymbol{b} \\ \boldsymbol{c}^T & d \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ y \end{pmatrix} = \begin{pmatrix} \boldsymbol{f} \\ g \end{pmatrix} \tag{4.14}$$

is readily solved using block forward and back substitution. One algorithm to accomplish this is

(i) Solve $A\boldsymbol{z} = \boldsymbol{b}$, and $A\boldsymbol{w} = \boldsymbol{f}$, and then set

(ii) $y = (g - \boldsymbol{c}^T \boldsymbol{w})/(d - \boldsymbol{c}^T \boldsymbol{z})$, $\qquad \boldsymbol{x} = \boldsymbol{w} - y\boldsymbol{z}$.

If A and M are both well conditioned then this algorithm for (4.14) works well, but if A is poorly conditioned, as occurs in pseudo-arclength continuation near a fold point, then it may fail to produce reliable results (in linear algebra terms, the algorithm is not "backward stable"). A complete analysis of why the algorithm fails in the latter case was first given by Moore [34] using a deflation argument. The account by Govaerts [11] avoids deflation but provides a stable algorithm based on combining the decomposition (4.13) with the alternative decomposition

$$\begin{pmatrix} A & \boldsymbol{b} \\ \boldsymbol{c}^T & d \end{pmatrix} = \begin{pmatrix} A & \boldsymbol{0} \\ \boldsymbol{c}^T & l_{n+1} \end{pmatrix} \begin{pmatrix} I_n & \boldsymbol{u}_n \\ \boldsymbol{0}^T & 1 \end{pmatrix}. \tag{4.15}$$

Roughly speaking the improved algorithm of Govaerts uses an iterative refinement approach. The idea is as follows. In actual calculation the step *(ii)* above produces a good approximation,

$\hat{y}$ say, for y, but the approximation for $\boldsymbol{x}$ is often worthless. So the $\boldsymbol{x}$ approximation is discarded. To compute the residual after the first solve the approximate solution $(\boldsymbol{x}_0, y_0) = (\mathbf{0}, \hat{y})$ is used. Then the approximation is corrected using the block LU decomposition (4.15) in a second solve. The analysis of why this works is fairly technical [11]. The main work in the resulting stable algorithm involves two solves with A and one solve with A^T. Moore [34] provides a stable algorithm also using only 3 solves (see quoted papers for analysis and algorithmic details).

5 The computation of fold (turning) points

Let $(\boldsymbol{x}_0, \lambda_0)$ be a point on S satisfying

$$\boldsymbol{F}^0_{\boldsymbol{x}} \text{ is singular}, \quad \text{rank}[\boldsymbol{F}^0_{\boldsymbol{x}} | \boldsymbol{F}^0_{\lambda}] = n. \tag{5.1}$$

Such a point is a fold point (see Example 2.1 and Definition 3.9). In a general one parameter problem $\boldsymbol{F}(\boldsymbol{x}, \lambda) = \mathbf{0}$, fold points are the generic singular points and there are many examples of their occurrence in applications. It is important to understand this type of nonlinear phenomenon in its own right, but also because the theoretical analysis and numerical methods for more complicated singularities are often extensions of fold point techniques.

Before attempting a discussion of the n-dimensional case in §5.1 it is a useful exercise to analyse first the scalar case: this is the subject of the following example.

Example 5.1 *Assume $f : \mathbb{R}^2 \to \mathbb{R}$ with $f \in C^\infty(\mathbb{R}^2)$ and set $S = \{(\boldsymbol{x}, \lambda) \in \mathbb{R}^2 : f(x, \lambda) = 0\}$. Assume* $\text{rank}[f_x, f_\lambda] = 1$ *on S (i.e. either $f_x \neq 0$ or $f_\lambda \neq 0$ at all points $(x, \lambda) \in S$). (i) Analyse the behaviour of S near a fold point (x_0, λ_0) i.e. where $f^0_x = 0$. (ii) Derive a 2×2 system for the accurate calculation of a fold point and determine when the fold point is a regular solution of this system.*

Sketch of solution: *Recall the usual notation $f^0_x = f_x(x_0, \lambda_0)$, etc. By the rank condition, at a fold point $f^0_\lambda \neq 0$ and so the Implicit Function Theorem implies $\lambda = \lambda(x)$ near x_0. Repeated differentiation of $f(x, \lambda(x)) = 0$ provides $\lambda'(x_0) = 0$, $\lambda''(x_0) = -f^0_{xx}/f^0_\lambda$, and so the first three terms of the Taylor expansion of $\lambda(x)$ can be found. (Sketch $\lambda(x)$ when $f^0_{xx} \neq 0$.) Thus if $f^0_{xx} \neq 0$ we see that $\lambda'(x)$ changes sign at $x = x_0$. To accurately compute (x_0, λ_0), consider the 2×2 system $\boldsymbol{G}(\boldsymbol{y}) = [f, f_x]^T = \mathbf{0}$ to be solved for $\boldsymbol{y} = (\boldsymbol{x}, \lambda)$. It is easy to show that (x_0, λ_0) is a regular point of $\boldsymbol{G}(\boldsymbol{y}) = \mathbf{0}$ if and only if $f^0_{xx} \neq 0$.*

We shall see in the following subsection that the theory for fold points of the n-dimensional system $\boldsymbol{F}(\boldsymbol{x}, \lambda) = \mathbf{0}$ is very similar to the one dimensional example above.

5.1 Analysis of Fold Points

Consider (4.1) with S as in (4.2). Let $(\boldsymbol{x}(t), \lambda(t))$ denote a smooth arc of S with $\operatorname{rank}[\boldsymbol{F}_{\boldsymbol{x}} | \boldsymbol{F}_\lambda] = n$ and let $(\boldsymbol{x}_0, \lambda_0) = (\boldsymbol{x}(t_0), \lambda(t_0))$ satisfy the fold point condition (5.1). Let

$$\boldsymbol{\phi}_0 \in \ker(\boldsymbol{F}^0_{\boldsymbol{x}}) \setminus \{\mathbf{0}\}, \quad \boldsymbol{\psi}_0 \in \ker({\boldsymbol{F}^0_{\boldsymbol{x}}}^T) \setminus \{\mathbf{0}\}. \tag{5.2}$$

Note that $\boldsymbol{\psi}_0^T \boldsymbol{F}^0_\lambda \neq 0$ (since otherwise $\boldsymbol{\psi}_0 \in \ker\, [\boldsymbol{F}^0_{\boldsymbol{x}}, \boldsymbol{F}^0_\lambda]^T = \{\mathbf{0}\}$). Differentiation of $\boldsymbol{F}(\boldsymbol{x}(t), \lambda(t)) = \mathbf{0}$ with respect to t, evaluation at $t = t_0$, and left multiplication by $\boldsymbol{\psi}_0^T$ yields

$$\dot\lambda(t_0) = 0, \ \dot{\boldsymbol{x}}(t_0) = \alpha \boldsymbol{\phi}_0, \ \ddot\lambda(t_0) = -\alpha^2 \boldsymbol{\psi}_0^T (\boldsymbol{F}^0_{\boldsymbol{x}\boldsymbol{x}} \boldsymbol{\phi}_0) \boldsymbol{\phi}_0 / \boldsymbol{\psi}_0^T \boldsymbol{F}^0_\lambda, \tag{5.3}$$

for some $\alpha \neq 0$, where $\boldsymbol{F}_{\boldsymbol{x}\boldsymbol{x}} \boldsymbol{\phi}_0$ denotes the Jacobian matrix of the n-vector $\boldsymbol{F}_{\boldsymbol{x}} \boldsymbol{\phi}_0$.

Definition 5.2 *We call* $(\boldsymbol{x}_0, \lambda_0)$ *a* quadratic *fold point if* $\ddot\lambda(t_0) \neq 0$.

To compute the fold point it is natural to set up the system

$$\boldsymbol{T}(\boldsymbol{y}) := \begin{pmatrix} \boldsymbol{F}(\boldsymbol{x}, \lambda) \\ \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}, \lambda) \boldsymbol{\phi} \\ \boldsymbol{\phi}^T \boldsymbol{\phi} - 1 \end{pmatrix} = \mathbf{0}, \quad \text{to be solved for } \boldsymbol{y} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\phi} \\ \lambda \end{pmatrix} \in \mathbb{R}^{2n+1}, \tag{5.4}$$

where $\boldsymbol{T} : \mathbb{R}^{2n+1} \to \mathbb{R}^{2n+1}$. Here the second and third equations say that $\boldsymbol{F}_{\boldsymbol{x}}$ has a zero eigenvalue with corresponding normalised eigenvector $\boldsymbol{\phi}$, (cf. Example 3.2). We shall see in §5.2 that there are alternative choices of system to compute a fold point, but (5.4) is very convenient for analysis. In fact, the following theorem shows that (5.4) characterises a quadratic fold point.

Theorem 5.3 *Let* $(\boldsymbol{x}_0, \lambda_0)$ *satisfy (5.1).*

(a) The point $(\boldsymbol{x}_0, \boldsymbol{\phi}_0, \lambda_0) \in \mathbb{R}^{2n+1}$ *is a regular solution of* $\boldsymbol{T}(\boldsymbol{y}) = \mathbf{0}$ *if and only if* $\ddot\lambda(t_0) \neq 0$, *i.e.* $(\boldsymbol{x}_0, \lambda_0)$ *is a quadratic fold point.*

(b) In addition, assume $\mu_0 = 0$ *is an algebraically simple eigenvalue of* $\boldsymbol{F}^0_{\boldsymbol{x}}$. *Let* $\mu(t)$ *denote the eigenvalue of* $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}(t), \lambda(t))$ *near* $t = t_0$ *with* $\mu(t_0) = 0$. *Then* $\dot\mu(t_0) \neq 0$ *if and only if* $(\boldsymbol{x}_0, \lambda_0)$ *is a quadratic fold point.*

The second result says that at a quadratic fold point a simple eigenvalue passes smoothly through zero. Thus there is a smooth change in sign of $\det(\boldsymbol{F}_{\boldsymbol{x}})$ through a quadratic fold point and this fact is often used in continuation codes.

Proof The proof of Theorem 5.3 is straightforward. One way is to consider $\boldsymbol{T}_{\boldsymbol{y}}(\boldsymbol{y}_0)$, given by

$$\boldsymbol{T}_{\boldsymbol{y}}(\boldsymbol{y}_0) = \begin{bmatrix} \boldsymbol{F}^0_{\boldsymbol{x}} & 0 & \boldsymbol{F}^0_{\lambda} \\ \boldsymbol{F}^0_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{\phi}_0 & \boldsymbol{F}^0_{\boldsymbol{x}} & \boldsymbol{F}^0_{\boldsymbol{x}\lambda}\boldsymbol{\phi}_0 \\ 0 & 2\boldsymbol{\phi}_0^T & 0 \end{bmatrix}$$

and use part *(ii)* of Keller's ABCD Lemma (Lemma 3.3). In the notation of the lemma, $A = \begin{pmatrix} \boldsymbol{F}^0_{\boldsymbol{x}} & 0 \\ \boldsymbol{F}^0_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{\phi}_0 & \boldsymbol{F}^0_{\boldsymbol{x}} \end{pmatrix}$ with corresponding choices for $\boldsymbol{b}$ and $\boldsymbol{c}$. It can then be shown that $\boldsymbol{T}_{\boldsymbol{y}}(\boldsymbol{y}_0)$ is nonsingular if and only if $\boldsymbol{\psi}_0^T(\boldsymbol{F}^0_{\boldsymbol{x}}\boldsymbol{\phi}_0)\boldsymbol{\phi}_0 \neq 0$, and part *(a)* follows from (5.3). The proof of part *(b)* follows by formulating the eigenvalue problem as

$$\boldsymbol{G}(\boldsymbol{z}, t) = \begin{pmatrix} \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}(t), \lambda(t))\boldsymbol{\phi} - \mu\boldsymbol{\phi} \\ \boldsymbol{\phi}^T\boldsymbol{\phi} - 1 \end{pmatrix} = \boldsymbol{0}, \quad \boldsymbol{z} = \begin{pmatrix} \boldsymbol{\phi} \\ \mu \end{pmatrix} \in \mathbb{R}^{n+1}. \tag{5.5}$$

Now $\boldsymbol{G}(\boldsymbol{z}_0, t_0) = \boldsymbol{0}$ and $\boldsymbol{G}_{\boldsymbol{z}}(\boldsymbol{z}_0, t_0)$ is nonsingular (cf. Example 3.2), and the Implicit Function Theorem shows that $\mu = \mu(t)$ is a simple eigenvalue of $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}(t), \lambda(t))$ near $t = t_0$. Differentiation of $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}(t), \lambda(t))\boldsymbol{\phi}(t) = \mu(t)\boldsymbol{\phi}(t)$, evaluation at $t = t_0$, and left multiplication by $\boldsymbol{\psi}_0^T$ provides $\dot{\mu}(t_0) \neq 0$ iff $\ddot{\lambda}(t_0) \neq 0$. (Note that $\boldsymbol{\psi}_0^T\boldsymbol{\phi}_0 \neq 0$ since zero is an algebraically simple eigenvalue of $\boldsymbol{F}^0_{\boldsymbol{x}}$.) □

5.2 Numerical Calculation of Fold points

The system (5.4) can easily be used to compute fold points. It was so used by Seydel [44], [43] and by Moore and Spence [35]. It is important to realise that when solving (5.4) by Newton's method one need not solve the $(2n+1) \times (2n+1)$ linear systems directly. In [35] an efficient solution procedure is described using only solves with an $n \times n$ nonsingular matrix, which is formed from $\boldsymbol{F}_{\lambda}$ and $n-1$ linearly independent columns of $\boldsymbol{F}_{\boldsymbol{x}}$. The details are not given here, but the main work involves 4 linear solves with the *same* matrix i.e. only one LU factorisation of an $n \times n$ matrix is needed per Newton step to solve (5.4).

The fact that the solution of the $(2n+1) \times (2n+1)$ linear Jacobian systems is accomplished using solves of $n \times n$ systems has been used many times since, and another example of this technique is in the calculation of Hopf bifurcation points (see §8 and [16]).

Griewank and Reddien [17, 18] (and with improvements Govaerts [12]) suggested an alternative way of calculating fold points (and other higher order singularities). This involves setting up a "minimal" defining system

$$\boldsymbol{T}(\boldsymbol{y}) = \begin{bmatrix} \boldsymbol{F}(\boldsymbol{x},\lambda) \\ g(\boldsymbol{x},\lambda) \end{bmatrix} = \boldsymbol{0}, \qquad \boldsymbol{y} \in \mathbb{R}^{n+1}, \tag{5.6}$$

where $g(\boldsymbol{x},\lambda) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is implicitly defined through the equations

$$M(\boldsymbol{x},\lambda) \begin{bmatrix} \boldsymbol{v}(\boldsymbol{x},\lambda) \\ g(\boldsymbol{x},\lambda) \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ 1 \end{bmatrix}, \tag{5.7}$$

and

$$(\boldsymbol{w}^T(\boldsymbol{x},\lambda), g(\boldsymbol{x},\lambda))M(\boldsymbol{x},\lambda) = (\boldsymbol{0}^T, 1), \tag{5.8}$$

where

$$M(\boldsymbol{x},\lambda) = \begin{bmatrix} \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x},\lambda) & \boldsymbol{b} \\ \boldsymbol{c}^T & d \end{bmatrix}, \tag{5.9}$$

for some $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$, $d \in \mathbb{R}$. (The fact that $g(\boldsymbol{x},\lambda)$ is defined uniquely by *both* (5.7) and (5.8) may be seen since both equations imply that $g(\boldsymbol{x},\lambda) = [M^{-1}(\boldsymbol{x},\lambda)]_{n+1,n+1}$.) Note that $M(\boldsymbol{x},\lambda)$ is a bordering of $\boldsymbol{F}_{\boldsymbol{x}}$, as arises in the numerical continuation method (§4). Assuming $\boldsymbol{b}, \boldsymbol{c}$, and d are chosen so that $M(\boldsymbol{x},\lambda)$ is nonsingular (see the ABCD Lemma 3.3) then $g(\boldsymbol{x},\lambda)$ and $\boldsymbol{v}(\boldsymbol{x},\lambda)$ in (5.7) are uniquely defined. (Note, if S is parametrised by t near $(\boldsymbol{x}_0,\lambda_0)$, i.e. $(\boldsymbol{x}(t),\lambda(t))$ near $t = t_0$, then $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{x}(t),\lambda(t))$ and $g = g(\boldsymbol{x}(t),\lambda(t))$, and these functions may be differentiated with respect to t.) Also, if we apply Cramer's Rule in (5.7) (an idea due to Govaerts) we have (with $M(\boldsymbol{x},\lambda)$ nonsingular)

$$g(\boldsymbol{x},\lambda) = \det(\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x},\lambda))/\det(M(\boldsymbol{x},\lambda)) \tag{5.10}$$

and so

$$g(\boldsymbol{x},\lambda) = 0 \iff \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x},\lambda) \text{ is singular.}$$

It is easily shown that quadratic fold points are regular solutions of (5.6). To apply Newton's method to (5.6) derivatives of $g(\boldsymbol{x},\lambda)$ are required and these can be found by differentiation of (5.7). When the details of an efficient implementation of Newton's method applied to (5.6) are worked out then the main cost is two linear solves with M and one with M^T. A nice summary of these ideas is given in Beyn [1]. A complete account is in the forthcoming book by Govaerts [13].

6 Bifurcation from the trivial solution

As usual we consider the problem $\boldsymbol{F}(\boldsymbol{x}, \lambda) = \boldsymbol{0}$ and we assume that $\boldsymbol{F} \in C^\infty(\mathbb{R}^{n+1})$. We shall also assume that

$$\boldsymbol{F}(\boldsymbol{0}, \lambda) = \boldsymbol{0}, \quad \text{for all } \lambda \in \mathbb{R}, \tag{6.1}$$

with $(\boldsymbol{0}, \lambda)$ being the path of trivial solutions. (As an example, consider $f \in \mathbb{R}^2$, $f(x, \lambda) = x\lambda - x^3$.) In this section we use the $\boldsymbol{G}^0$ to denote the value of any function $\boldsymbol{G}$ at $(\boldsymbol{0}, \lambda_0)$.

A formal definition of bifurcation from the trivial solution is as follows.

Definition 6.1 *A point $(\boldsymbol{0}, \lambda) \in \mathbb{R}^{n+1}$ is said to be a bifurcation point for $\boldsymbol{F}(\boldsymbol{x}, \lambda) = \boldsymbol{0}$ if and only if there exists a sequence $(\boldsymbol{x}_k, \lambda_k)$, with $\boldsymbol{x}_k \to 0$, $\lambda_k \to \lambda_0$, as $k \to \infty$, such that $\boldsymbol{F}(\boldsymbol{x}_k, \lambda_k) = \boldsymbol{0}$ and $\boldsymbol{x}_k \neq 0$ for all k.*

Application of the Implicit Function Theorem shows immediately that if $(\boldsymbol{0}, \lambda_0)$ is a bifurcation point then $\boldsymbol{F}^0_{\boldsymbol{x}}$ must be singular. We shall assume that $\boldsymbol{F}^0_{\boldsymbol{x}}$ has an algebraically simple zero eigenvalue: that is, there exist $\boldsymbol{\phi}_0, \boldsymbol{\psi}_0 \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that

$$\begin{aligned} \ker(\boldsymbol{F}^0_{\boldsymbol{x}}) &= \operatorname{span}\{\boldsymbol{\phi}_0\}, \\ \ker({\boldsymbol{F}^0_{\boldsymbol{x}}}^T) &= \operatorname{span}\{\boldsymbol{\psi}_0\}, \\ \text{and} \quad \boldsymbol{\psi}_0^T \boldsymbol{\phi}_0 &\neq 0. \end{aligned} \tag{6.2}$$

In §6.1 we shall consider the scalar case since it is simpler and the scalar result is used in the n-dimensional case. We shall see in §7 that the scalar case is interesting in its own right. In §6.2 we then discuss the n-dimensional case.

6.1 Scalar case

Consider the problem $f(x, \lambda) = 0$, with $f(0, \lambda) = 0$ for all λ. Then we have the following theorem.

Theorem 6.2 *Suppose $f : D \subseteq \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ where D is an open subset of $\mathbb{R}^2$. Suppose $f \in C^\infty(D)$. Let $S = \{(x, \lambda) \in D : f(x, \lambda) = 0\}$, and assume $(0, \lambda) \in S$, for all $\lambda \in \mathbb{R}$. Also assume for some $\lambda_0 \in \mathbb{R}$*

(i) $f^0_x = 0$,

(ii) $f^0_{x\lambda} \neq 0$,

(where $f_x^0 = f_x(0,\lambda_0)$, etc). Then nontrivial solutions bifurcate from the trivial solution $x = 0$ at $(0,\lambda_0)$. Moreover, near λ_0,

$$\lambda(x) = \lambda_0 + xv(x) \tag{6.3}$$

where $v(x)$ is a smooth function of x with $v(0) = -f_{xx}^0/2f_{x\lambda}^0$.

Proof Note that $f_\lambda^0 = f_{\lambda\lambda}^0 = f_{\lambda\lambda\lambda}^0 = \cdots = 0$ since $(0,\lambda) \in S$ for all λ. Thus $\text{rank}[f_x^0, f_\lambda^0] = 0$ and the Implicit Function Theorem cannot be applied directly on $f(x,\lambda) = 0$. However, we can introduce a new problem on which the Implicit Function Theorem can be applied. Define $h(x,v) : \mathbb{R}^2 \to \mathbb{R}$ by, for $x \neq 0$,

$$h(x,v) = \frac{1}{x^2} f(x,\lambda), \quad v = (\lambda - \lambda_0)/x. \tag{6.4}$$

Expanding $f(x, \lambda_0 + xv)$ in powers of x using Taylor's theorem, and using $f^0 = f_x^0 = f_\lambda^0 = f_{\lambda\lambda}^0 = 0$, gives, for $x \neq 0$,

$$h(x,v) = \frac{1}{2!}[f_{xx}^0 + 2f_{x\lambda}^0 v] + x[\text{smooth function of } v].$$

Define $h(x,v)$ at $x = 0$ by

$$h(0,v) = \frac{1}{2!}[f_{xx}^0 + 2f_{x\lambda}^0 v]. \tag{6.5}$$

Then h is a smooth function of (x,v). Moreover by setting

$$v_0 = -f_{xx}^0/2f_{x\lambda}^0, \tag{6.6}$$

then, at $(x,v) = (0,v_0)$ we have, from (6.5), $h^0 = h(0,v_0) = 0$. Moreover, by assumption *(ii)*, $h_v^0 \neq 0$, and the Implicit Function Theorem gives that for $|x|$ sufficiently small there exists $v(x)$ such that $h(x,v(x)) = 0$. Using this function v, recall (6.4) and define

$$\lambda(x) = \lambda_0 + xv(x), \tag{6.7}$$

from which $f(x,\lambda(x)) = 0$ and the existence of nontrivial bifurcating solutions is proved.

Differentiating (6.7) and evaluation at $x = 0$ gives

$$\lambda'(0) = v(0) = v_0 = -f_{xx}^0/2f_{x\lambda}^0, \tag{6.8}$$

which tells us the slope of the tangent to the nontrivial solution branch at the bifurcation point (see Figure 6.1). □

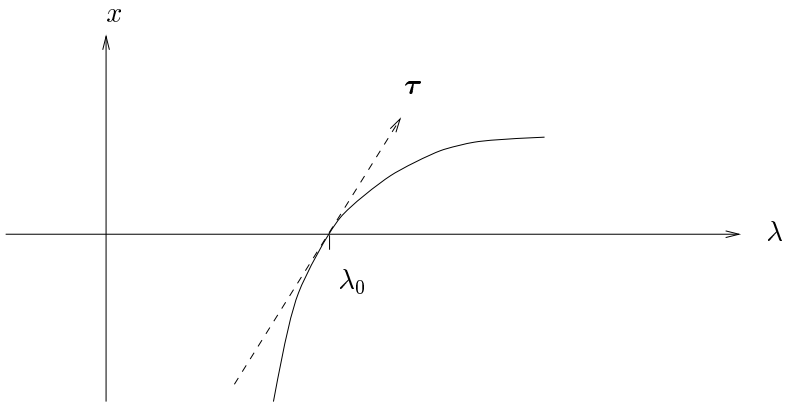

Figure 6.1: Here τ is the tangent to the bifurcating branches at $(0, \lambda_0)$. The slope of the tangent is $(-2f^0_{x\lambda}/f^0_{xx})$.

To compute $(x, \lambda(x))$ near $(0, \lambda_0)$ with $x \neq 0$, consider solving the system

$$\boldsymbol{G}(\boldsymbol{y}, t) := \begin{pmatrix} f(x, \lambda) \\ x - t \end{pmatrix} = \boldsymbol{0}, \qquad \boldsymbol{y} = \begin{pmatrix} x \\ \lambda \end{pmatrix} \in \mathbb{R}^2. \tag{6.9}$$

If $(\boldsymbol{y}, t)$ solves (6.9) with sufficiently small $t \neq 0$, then $x = t$ and $(x, \lambda) = (t, \lambda(t)) =: \boldsymbol{y}(t)$. Moreover,

$$\det\left(\boldsymbol{G}_{\boldsymbol{y}}(\boldsymbol{y}(t), t)\right) = -f_\lambda(t, \lambda(t)) = (-f^0_{x\lambda})t + \mathcal{O}(t^2)$$

and so the Newton theory (Theorem 5.2.1 in [6]) shows that convergence of Newton's method can only be guaranteed for starting guesses in a ball of radius $\mathcal{O}(t)$. If we take as starting guess the following point on the tangent $\boldsymbol{\tau}$ depicted in Figure 6.2:

$$\begin{pmatrix} x^0 \\ \lambda^0 \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda_0 \end{pmatrix} + t \begin{pmatrix} 1 \\ \lambda'(0) \end{pmatrix}$$

with $\lambda'(0)$ given by (6.8), then

$$\begin{pmatrix} t \\ \lambda(t) \end{pmatrix} - \begin{pmatrix} x^0 \\ \lambda^0 \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda(t) - \lambda(0) - t\lambda'(0) \end{pmatrix} = \mathcal{O}(t^2)$$

and one can show that Newton's method will converge for sufficiently small t. Notice that in this case t is *not* the pseudo-arclength parameter.

6.2 n-dimensional case

For the general case we have the classical theorem by Crandall and Rabinowitz on bifurcation from a simple eigenvalue [4] (which holds for general operators on Banach spaces under minimal

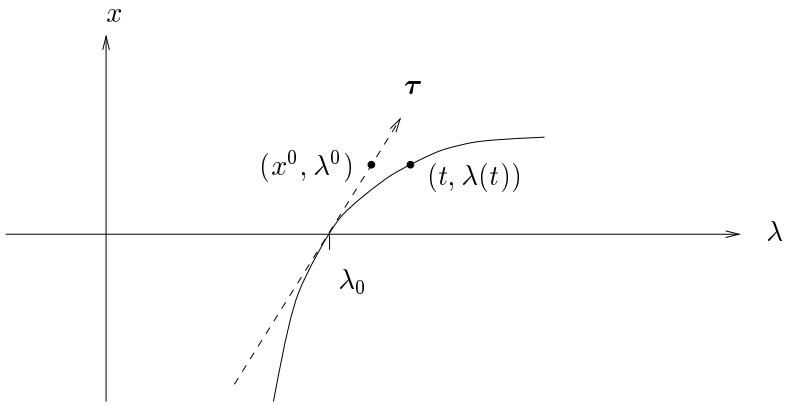

Figure 6.2: Numerical continuation away from the bifurcation point. Here $x^0 = t$ and in this case t is merely the value of the (non-zero) x component.

smoothness requirements).

Theorem 6.3 *Suppose* $\boldsymbol{F} \in C^\infty(\mathbb{R}^{n+1})$ *and (6.1), (6.2) hold. If*

$$\boldsymbol{\psi}_0^T \boldsymbol{F}^0_{\boldsymbol{x},\lambda} \boldsymbol{\phi}_0 \neq 0, \tag{6.10}$$

then $(\mathbf{0}, \lambda_0)$ *is a bifurcation point, and there exist smooth functions* $(\boldsymbol{x}(t), \lambda(t))$ *parametrised by* t *near* $t = 0$ *such that* $\boldsymbol{F}(\boldsymbol{x}(t), \lambda(t)) = \mathbf{0}$ *with* $\lambda(0) = \lambda_0$, $\boldsymbol{x}(0) = \mathbf{0}$, *and* $\boldsymbol{x}'(0) = \boldsymbol{\phi}_0$.

Proof We present a proof which we believe is within the scope of a typical UK research student just starting a PhD. More elegant proofs are given in most text books on bifurcation theory. The method of proof is an example of the "Lyapunov-Schmidt reduction" [2].

We give the proof only under the assumption that $\boldsymbol{F}^0_{\boldsymbol{x}}$ has distinct real eigenvalues, $\mu_0, \ldots, \mu_{n-1}$, with linearly independent eigenvectors $\{\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{n-1}\}$, and $\{\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_{n-1}\}$ the corresponding linearly independent eigenvectors of $(\boldsymbol{F}^0_{\boldsymbol{x}})^T$. Recall that $\mu_0 = 0$ and all the other eigenvalues of $\boldsymbol{F}^0_{\boldsymbol{x}}$ are nonzero. Also, because the eigenvalues are distinct and simple, $\boldsymbol{\psi}_i^T \boldsymbol{\phi}_j = 0$, $(i \neq j)$ and $\boldsymbol{\psi}_i^T \boldsymbol{\phi}_i \neq 0$.

Now since the $\boldsymbol{\psi}_i$ span $\mathbb{R}^n$, the equation $\boldsymbol{F}(\boldsymbol{x}, \lambda) = \mathbf{0}$ can be written as

$$\boldsymbol{\psi}_0^T \boldsymbol{F}(\boldsymbol{x}, \lambda) = 0 \tag{6.11}$$

$$\text{and} \quad \boldsymbol{\psi}_i^T \boldsymbol{F}(\boldsymbol{x}, \lambda) = 0, \quad i = 1, \ldots, (n-1). \tag{6.12}$$

(The proof of this is an elementary exercise.) Also we write $\boldsymbol{x} \in \mathbb{R}^n$ in the form

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{y}, t) = t\boldsymbol{\phi}_0 + V\boldsymbol{y}, \tag{6.13}$$

where V is the $n \times (n-1)$ matrix with ith column $\boldsymbol{\phi}_i$, for $i = 1, \ldots, (n-1)$. Thus we decompose $\mathbb{R}^n$ into $\mathbb{R}^n = \text{span}\{\boldsymbol{\phi}_0\} \oplus \mathcal{R}$, where $\mathcal{R} = \text{Image}(\boldsymbol{F}^0_{\boldsymbol{x}}) = \text{span}\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{n-1}\}$. Note that $\boldsymbol{\psi}_0^T \boldsymbol{v} = 0$ for all $\boldsymbol{v} \in \mathcal{R}$. Now consider the $n-1$ equations given by (6.12) in the form

$$\tilde{\boldsymbol{F}}(\boldsymbol{y}, t, \lambda) = \boldsymbol{0}, \quad \tilde{\boldsymbol{F}} : \mathbb{R}^{n-1} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^{n-1} \tag{6.14}$$

where $(\tilde{\boldsymbol{F}}(\boldsymbol{y}, t, \lambda))_i = \boldsymbol{\psi}_i^T \boldsymbol{F}(t\boldsymbol{\phi}_0 + V\boldsymbol{y}, \lambda)$. We shall use the Implicit Function Theorem to parametrise the solution $\boldsymbol{y}$ of (6.14) as a function of (t, λ). To do this observe that

$$\tilde{\boldsymbol{F}}(\boldsymbol{0}, 0, \lambda) = \boldsymbol{0} \in \mathbb{R}^{n-1},$$

and the matrix $\tilde{\boldsymbol{F}}\boldsymbol{y}(\boldsymbol{0}, 0, \lambda)$ is given by

$$\left[\tilde{\boldsymbol{F}}\boldsymbol{y}(\boldsymbol{0}, 0, \lambda)\right]_{i,j} = \boldsymbol{\psi}_i^T \boldsymbol{F}\boldsymbol{x}(\boldsymbol{0}, \lambda)\boldsymbol{\phi}_j, \quad i, j = 1, \ldots, (n-1).$$

Thus $[\tilde{\boldsymbol{F}}\boldsymbol{y}(\boldsymbol{0}, 0, \lambda_0)]_{ij} = \boldsymbol{\psi}_i^T \boldsymbol{F}^0_{\boldsymbol{x}} \boldsymbol{\phi}_j = \mu_i \boldsymbol{\psi}_i^T \boldsymbol{\phi}_i$, and since $\mu_i \neq 0$, $i = 1, \ldots, (n-1)$, $\tilde{\boldsymbol{F}}\boldsymbol{y}(\boldsymbol{0}, 0, \lambda_0)$ is nonsingular on $\mathbb{R}^{n-1}$, and so the Implicit Function Theorem (extended to the case of a two-dimensional parameter [2]) shows that solutions $\boldsymbol{y}$ of (6.14) may be parametrised by (t, λ) i.e. $\boldsymbol{y} = \boldsymbol{y}(t, \lambda)$ near $(0, \lambda_0)$. Thus $\tilde{\boldsymbol{F}}(\boldsymbol{y}(t, \lambda), t, \lambda) = \boldsymbol{0}$ for (t, λ) near $(0, \lambda_0)$ and (by uniqueness) $\boldsymbol{y}(0, \lambda) = \boldsymbol{0}$. In addition, $\boldsymbol{y}_t(0, \lambda_0) = \boldsymbol{0}$ since $\tilde{\boldsymbol{F}}_t(\boldsymbol{0}, 0, \lambda_0) = \boldsymbol{0}$, and $\boldsymbol{y}_\lambda(0, \lambda) = \boldsymbol{0}$ since $\tilde{\boldsymbol{F}}_\lambda(\boldsymbol{0}, 0, \lambda) = \boldsymbol{0}$.

Having solved (6.14) (i.e. the equations (6.12)), we return to equation (6.11) and write it in the form

$$f(t, \lambda) := \boldsymbol{\psi}_0^T \boldsymbol{F}(t\boldsymbol{\phi}_0 + V\boldsymbol{y}(t, \lambda), \lambda) = 0. \tag{6.15}$$

This is in the form of a scalar nonlinear problem, and all that remains to prove the existence of nontrivial solutions is to check the conditions of Theorem 6.2. Not surprisingly (6.15) (or (6.11)) is called the *bifurcation equation* in the Lyapunov-Schmidt reduction of $\boldsymbol{F}(\boldsymbol{x}, \lambda) = \boldsymbol{0}$. It is essentially a projection of the original n dimensional system into the one dimensional space spanned by the null eigenvector. Observe that

$$\begin{aligned} f(0, \lambda) &= \boldsymbol{\psi}_0^T \boldsymbol{F}(V\boldsymbol{y}(0, \lambda), \lambda) \\ &= \boldsymbol{\psi}_0^T \boldsymbol{F}(\boldsymbol{0}, \lambda) \\ &= 0 \end{aligned}$$

and

$$f_t(0, \lambda_0) = \boldsymbol{\psi}_0^T \boldsymbol{F}^0_{\boldsymbol{x}} \boldsymbol{\phi}_0 = 0,$$

since $\phi_0 \in \ker(\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{0}, \lambda_0))$. Finally

$$\begin{aligned} f^0_{t\lambda} &= \psi_0^T(\boldsymbol{F}^0_{\boldsymbol{x}\boldsymbol{x}}(\phi_0 + V\boldsymbol{y}^0_t)(V\boldsymbol{y}^0_\lambda) + \boldsymbol{F}^0_{\boldsymbol{x}\lambda}(\phi_0 + V\boldsymbol{y}^0_t) + \boldsymbol{F}^0_{\boldsymbol{x}} V\boldsymbol{y}^0_{t\lambda}) \\ &= \psi_0^T \boldsymbol{F}^0_{\boldsymbol{x}\lambda}\phi_0 \neq 0 \end{aligned}$$

by (6.10). Here we have used $\boldsymbol{y}^0_t = \boldsymbol{0}, \boldsymbol{y}^0_\lambda = \boldsymbol{0}$, and $V\boldsymbol{y}^0_{t\lambda} \in \mathcal{R}$ to simplify the expression for $f^0_{t\lambda}$. So the conclusions of Theorem 6.2 apply. Hence nontrivial solutions of $f(t, \lambda) = 0$ bifurcate at $\lambda = \lambda_0$. The corresponding $\boldsymbol{x}(t) = \boldsymbol{x}(\boldsymbol{y}(t, \lambda), t) = t\phi_0 + V\boldsymbol{y}(t, \lambda)$, with $\boldsymbol{x}(0) = \boldsymbol{0}$ and $\boldsymbol{x}'(0) = \phi_0$, provides the nontrivial solution of $\boldsymbol{F}(\boldsymbol{x}, \lambda) = \boldsymbol{0}$. □

Detection of bifurcation points is relatively easy for this case. We seek points where $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{0}, \lambda)$ is singular. If $\mu(\lambda)$ denotes the eigenvalue of $\boldsymbol{F}_{\boldsymbol{x}}(0, \lambda)$ along the trivial solution path with $\mu(\lambda_0) = 0$ then it is a simple exercise to show that $\mu'(\lambda_0) \neq 0$ if and only if (6.10) holds. Hence (6.10) is another example of a nondegeneracy condition which can be interpreted as an eigenvalue going through zero with nonzero speed. Note that $\det(\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{0}, \lambda))$ changes sign at the bifurcation point as λ passes through λ_0.

Example 6.4 *In Example 2.3, equation (2.7) is*

$$\boldsymbol{F}(\boldsymbol{Y}, \lambda) = A\boldsymbol{Y} + \lambda \sin(\boldsymbol{Y}).$$

(It is an instructive exercise to go through the proof of Theorem 6.3 for this example.) In this case $\boldsymbol{F}(\boldsymbol{0}, \lambda) = \boldsymbol{0}$*, for all* $\lambda \in \mathbb{R}$*,* $\boldsymbol{F}_{\boldsymbol{Y}}(\boldsymbol{0}, \lambda) = (A + \lambda I)$*, and* $\boldsymbol{F}_{\boldsymbol{Y}\lambda}(\boldsymbol{0}, \lambda) = I$*. Now* A *has* $(n-1)$ *algebraically simple eigenvalues*

$$\mu_k = -h^{-2}(2 - 2\cos k\pi/(n-1)), \quad k = 0, 1, \ldots, n-2.$$

The zero eigenvalue is ruled out of the buckling example on physical grounds, so Theorem 6.3 shows that bifurcation from the trivial solution occurs at $\lambda = -\mu_k$*,* $k = 0, \ldots, (n-2)$ *since the nondegeneracy condition (6.5) reduces to the condition that the eigenvalue be algebraically simple.* □

If the n-dimensional problem arises from a discretization of an ODE or PDE then the eigenvalues and eigenfunctions of the linearisation of the continuous problem might be known analytically. To find λ_0 and ϕ_0 in the discretized problem a simple inverse iteration approach applied to $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{0}, \lambda)$ with the exact value for λ at the bifurcation point used would almost certainly work very quickly.

To move off the trivial branch a technique similar to that in §6.1 may be used. At $(0, \lambda_0)$ $[\boldsymbol{F}^0_{\boldsymbol{x}} | \boldsymbol{F}^0_{\lambda}]$ has a two dimensional kernel spanned by $(\boldsymbol{\phi}_0^T, 0)^T$ and $(\mathbf{0}^T, 1)^T$. It is straightforward to show that the tangent to the bifurcating nontrivial branch has the form

$$\boldsymbol{\tau}_0^T = ((-2\boldsymbol{\psi}_0^T \boldsymbol{F}^0_{\boldsymbol{x}\lambda}\boldsymbol{\phi}_0)\boldsymbol{\phi}_0^T, \boldsymbol{\psi}_0^T(\boldsymbol{F}^0_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{\phi}_0)\boldsymbol{\phi}_0)$$

(cf. the scalar case in the previous section). The fact that the ordinary pseudo-arclength method works in these circumstances is proved in [22]. However, the direct analogue of the approach in §6.1 is merely to set equal to t one component of $\boldsymbol{x}$. The best component to choose is the rth, where $(\boldsymbol{\phi}_0)_r$ is the component of maximum modulus of $\boldsymbol{\phi}_0$ (see [39]).

Remarks

(i) It is important to note that bifurcation from the trivial solution is a rather special case but nonetheless a very important case in applications. The trivial solution forms an invariant subspace under the action of $\boldsymbol{F}$ in $\mathbb{R}^{n+1}$ and bifurcating nontrivial solutions break the subspace. Werner [49] gives a general theory of subspace-breaking bifurcation. A different case but with comparable results arises when $\boldsymbol{F}(\boldsymbol{x}, \lambda)$ satisfies a symmetry condition (see [48], [51]).

(ii) In the absence of any special features (for example, symmetry) a bifurcation where two (nontrivial) solution curves intersect will not typically arise in a one parameter problem $\boldsymbol{F}(\boldsymbol{x}, \lambda) = \mathbf{0}$. In this case one needs *two* parameters to detect and compute bifurcation points (see [33], [18],[23]).

7 Bifurcation in Nonlinear ODEs

The bifurcation theory in these notes is given for finite dimensional problems. However several of the theoretical results on bifurcation can be applied to infinite dimensional problems involving nonlinear ODEs by use of the shooting method, which also provides a computational tool. In fact Poincaré's analysis of bifurcation of periodic orbits in ODEs using the Poincaré section is the first example of the use of a shooting method to prove analytical results. The use of shooting to study steady bifurcations in nonlinear boundary value problems (BVPs) seems to have been considered first by J.B.Keller in 1960. The treatment here is based on Keller's article [27]. An account of the numerical analysis of shooting methods for nonlinear BVPs is given in [24].

First recall a standard theorem of existence, uniqueness and continuity with respect to initial data for systems of ordinary differential equations (ODEs) of the form

$$\boldsymbol{u}' = \boldsymbol{f}(t, \boldsymbol{u}), \quad t > a \tag{7.1}$$

with initial condition

$$\boldsymbol{u}(a) = \boldsymbol{\alpha} \tag{7.2}$$

where $\boldsymbol{u}(t) \in \mathbb{R}^n$ is to be found for $t > a$, $\boldsymbol{f} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ is given, $\boldsymbol{\alpha} \in \mathbb{R}^n$ is given, and $a \in \mathbb{R}$ is given.

Theorem 7.1 *Suppose $\boldsymbol{f}$ is continuous on $[a, b] \times \mathbb{R}^n$, and suppose*

$$\|\boldsymbol{f}(t, \boldsymbol{u}) - \boldsymbol{f}(t, \boldsymbol{v})\| \leq L\|\boldsymbol{u} - \boldsymbol{v}\| \tag{7.3}$$

for some $L > 0$ and for $t \in [a, b]$ and all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. Then for any $\boldsymbol{\alpha} \in \mathbb{R}^n$ the IVP (7.1), (7.2) has a unique solution $\boldsymbol{u} = \boldsymbol{u}(t, \boldsymbol{\alpha})$ defined for $t \in [a, b]$. Moreover $\boldsymbol{u}$ is Lipschitz continuous in $\boldsymbol{\alpha}$, and in fact

$$\|\boldsymbol{u}(t, \boldsymbol{\alpha}) - \boldsymbol{u}(t, \boldsymbol{\beta})\|_2 \leq e^{L(t-a)}\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2$$

for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^n$.

Remarks

1. In many problems of interest (7.3) will not be true over *all* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, but rather over all $\boldsymbol{u}, \boldsymbol{v} \in B(\boldsymbol{\alpha}, r)$ for some $\boldsymbol{\alpha} \in \mathbb{R}^n$, $r > 0$ fixed. In this case a similar theorem holds, but the solution may exist only for $t \in [a, b_0]$ with $b_0 = \min\{b, r/L\}$.

2. The numerical solution of (7.1), (7.2) over finite ranges of $t > a$ is now well understood. Many codes exist in which a user specifies $\boldsymbol{f}, a, b$ and $\boldsymbol{\alpha}$ and a required tolerance, and the program returns the value of $\boldsymbol{u}(b)$ at b or at any intermediate points between a and b. We will assume that (7.1), (7.2) have a unique solution which can be found numerically for $t \in [a, b]$, with $b > a$.

7.1 The shooting method for ODEs

Consider the second order ODE

$$-y'' - g(t, y, y') = 0, \quad t \in [a, b] \tag{7.4}$$

subject to the boundary conditions

$$a_0 y(a) - a_1 y'(a) = \alpha \tag{7.5}$$

$$b_0 y(b) + b_1 y'(b) = \beta \tag{7.6}$$

with $|a_0| + |a_1| \neq 0, \quad |b_0| + |b_1| \neq 0$. Here it is assumed that $g(t, y_1, y_2)$ is continuous on $D := \{(t, y_1, y_2) : \ t \in [a, b], \ y_1^2 + y_2^2 < \infty\}$ and satisfies a uniform Lipschitz condition in y_1 and y_2.

To solve this BVP consider the associated initial value problem (IVP)

$$-u'' - g(t, u, u') = 0 \quad \text{(as in (7.4))} \tag{7.7}$$

subject to the initial conditions

$$a_0 u(a) - a_1 u'(a) = \alpha \quad \text{(as in (7.5))} \tag{7.8}$$

and

$$c_0 u(a) - c_1 u'(a) = s, \tag{7.9}$$

where s is a parameter which will be determined below. We choose c_0, c_1 s.t.

$$d := a_1 c_0 - a_0 c_1 \neq 0. \tag{7.10}$$

Then (7.8), (7.9) are independent initial conditions and the matrix $\begin{bmatrix} a_0 & -a_1 \\ c_0 & -c_1 \end{bmatrix}$ is invertible. From Theorem 7.1 we know that (7.7)–(7.9) has a unique solution, which we denote by $u(t; s), \ t > 0, \ s \in \mathbb{R}$. To solve (7.4)–(7.6) we need to find s such that

$$f(s) = 0 \tag{7.11}$$

where $f(s)$ is defined by the right-hand boundary condition (7.6)

$$f(s) = \{b_0 u(b; s) + b_1 \frac{\partial u}{\partial t}(b; s) - \beta\}. \tag{7.12}$$

Thus the solution of (7.4)–(7.6) is reduced to solving the nonlinear problem (7.11) where f is implicitly defined in terms of solutions of (7.7)–(7.9).

The numerical analysis of shooting methods for solving (7.4)–(7.6) where solutions of (7.7)–(7.9) are evaluated numerically is given in [24].

The equivalence of (7.4)–(7.6) to (7.11) is given in the following lemma (see [24],[27]).

Lemma 7.2 (i) *If s_0 solves (7.11) then $y(t) = u(t; s_0)$ solves (7.4)–(7.6). If $y(t)$ solves (7.4)–(7.6) then $s_0 = c_0 y(a) - c_1 y'(a)$ solves (7.11).*

(ii) *s_0 is the unique solution of (7.11) if and only if $y(t)$ is the unique solution of (7.4)–(7.6).*

Proof (i) Suppose $f(s_0) = 0$ then $y(t) = u(t; s_0)$ solves (7.4)–(7.6). Conversely if $y(t)$ solves (7.4)–(7.6) then set $s_0 = c_0 y(a) - c_1 y'(a)$. By uniqueness of solutions to initial value problems $u(t; s_0) = y(t)$ and $f(s_0) = 0$.

(ii) See Theorem 2 [27] where a more general problem is considered. □

To analyse and solve (7.7)–(7.9) we reduce to a first order system. To do this we set $u_1 = u$, $u_2 = u' = u_1'$. Then (7.7) becomes the 2×2 system

$$\begin{bmatrix} u_1' \\ u_2' \end{bmatrix} = \begin{bmatrix} u_2 \\ -g(t, u_1, u_2) \end{bmatrix} =: \boldsymbol{f}(t, \boldsymbol{u}) \tag{7.13}$$

and (7.8), (7.9) become (using (7.10)),

$$\begin{bmatrix} u_1(a) \\ u_1(a) \end{bmatrix} = \frac{1}{d} \begin{bmatrix} -c_1 & a_1 \\ -c_0 & a_0 \end{bmatrix} \begin{bmatrix} \alpha \\ s \end{bmatrix}, \tag{7.14}$$

and so (7.13), (7.14) has a unique solution using Theorem 7.1.

To implement Newton's method for (7.11) we need to be able to evaluate not only $f(s)$ but also $f_s(s) = \left\{ b_0 w(b; s) + b_1 \dfrac{\partial w}{\partial t}(b; s) \right\}$ where $w = \dfrac{\partial u}{\partial s}$. We obtain an equation for w by differentiating (7.7)–(7.9) with respect to s to get the IVP

$$\left.\begin{aligned} -w'' - g_u(t, u, u')w - g_{u'}(t, u, u')w' &= 0, \\ a_0 w(a) - a_1 w'(a) &= 0, \\ c_0 w(a) - c_1 w'(a) &= 1. \end{aligned}\right\} \tag{7.15}$$

(Note $'$ always means differentiation with respect to t.) This system together with (7.7)–(7.9) can be reduced to a first order system of dimension 4 which we can solve using standard ODE software. We illustrate this by means of an example.

Example 7.3 *Consider the boundary value problem*

$$\left.\begin{aligned} -y'' + e^y &= 0, \\ y(0) = 0,\ y(1) &= 0, \end{aligned}\right\} \tag{7.16}$$

and the corresponding IVP

$$\left.\begin{aligned} -u'' + e^u &= 0, \\ u(0) = 0,\ u'(0) &= s. \end{aligned}\right\} \tag{7.17}$$

Denote the solution by $u(t;s)$. The nonlinear problem (7.11) is

$$f(s) := u(1;s) = 0. \tag{7.18}$$

To find f_s, set $w = \dfrac{\partial u}{\partial s}$, then differentiate (7.17) with respect to s

$$\left.\begin{aligned} -w'' + e^u w &= 0, \\ w(0) &= 0, \\ w'(0) &= 1. \end{aligned}\right\} \tag{7.19}$$

From the solution of this we obtain $f_s(s) = w(1;s)$.

We solve (7.17), (7.19) simultaneously by the substitutions

$$u_1 = u, \qquad u_2 = u' = u_1',$$
$$u_3 = w, \qquad u_4 = w' = u_3',$$

to obtain a first order system of four equations:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}' = \begin{bmatrix} u_2 \\ \exp(u_1) \\ u_4 \\ \exp(u_1)u_3 \end{bmatrix}$$

with initial condition

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}_{t=0} = \begin{bmatrix} 0 \\ s \\ 0 \\ 1 \end{bmatrix}.$$

We solve this system up to $t = 1$, from which we obtain

$$f(s) = u_1(1), \quad f_s(s) = u_3(1).$$

If s is a guess to the solution of (7.11) then the values of $f(s)$ and $f_s(s)$ can be used to generate a new guess for s using Newton's method, and this process can be iterated.

7.2 Analysis of parameter dependent ODEs

As mentioned at the start of this section, the shooting method can be used as a technique for theoretical analysis as well as numerically solving problems.

Example 7.4 *(Recall Example 2.3). Prove that nontrivial solutions exist for the following BVP*

$$\left.\begin{aligned} -y'' &= \lambda \sin y = 0, \\ y'(0) &= 0,\ y'(l) = 0. \end{aligned}\right\} \tag{7.20}$$

To prove this we use again the shooting approach and consider the IVP

$$\left.\begin{aligned} -u'' &= \lambda \sin u = 0, \\ u'(0) &= 0,\ u(0) = s. \end{aligned}\right\} \tag{7.21}$$

Note that the solution u depends on t, s and also λ, ie. $u = u(t; s, \lambda)$. In view of the right hand boundary conditions in (7.20), we consider

$$f(s, \lambda) = \frac{\partial u}{\partial t}(l; s, \lambda) = 0. \tag{7.22}$$

By uniqueness for the IVP (7.21) we have

$$f(0, \lambda) = 0 \qquad \text{for all } \lambda \in \mathbb{R}.$$

By Lemma 7.2, (7.22) is equivalent to (7.20), so nontrivial solutions y of (7.20) bifurcate at $\lambda = \lambda_0$ if and only if nontrivial solutions s of (7.22) bifurcate at $\lambda = \lambda_0$. To prove the latter assertion we use Theorem 6.2, for which we need f_s and $f_{s\lambda}$. So set $w = \dfrac{\partial u}{\partial s}$ and differentiate (7.21) with respect to s to obtain

$$\left.\begin{aligned} -w'' - \lambda(\cos u)w = 0, \\ w'(0) = 0,\ w(0) = 1, \end{aligned}\right\} \tag{7.23}$$

with solution $w = w(t; s, \lambda)$. Now when $s = 0$, $u = 0$ (by uniqueness for (7.21)) and (7.23) implies that $w(t) = w(t; 0, \lambda)$ satisfies the linear 2nd order ODE

$$-w'' - \lambda w = 0. \tag{7.24}$$

Thus $w(t) = A \sin \sqrt{\lambda} t + B \cos \sqrt{\lambda} t$, and to satisfy the boundary conditions, we have $A = 0,\ B = 1$. Then, by (7.22)

$$f_s(0, \lambda) = \frac{\partial w}{\partial t}(l; 0, \lambda) = -\sqrt{\lambda} \sin(\sqrt{\lambda} l). \tag{7.25}$$

This vanishes for $\lambda = \lambda_0 = \frac{m^2\pi^2}{l^2}, \quad m = 0, 1, 2....$ (In the application of the buckling of a rod the case $\lambda = 0$ is ruled out on physical grounds.) To check if bifurcation occurs at $\lambda = \lambda_0$, we have to compute $f_{s\lambda}(0, \lambda_0)$. To do this set

$$v = \frac{\partial w}{\partial \lambda},$$

and differentiate (7.23) with respect to λ to get

$$\left.\begin{aligned} &-v'' - (\cos u)w + \lambda(\sin u)\tfrac{\partial u}{\partial \lambda}w - \lambda(\cos u)v = 0 \\ &v'(0) = 0, \ v(0) = 0. \end{aligned}\right\} \tag{7.26}$$

At $s = 0, \quad u = 0$, and so $v(t) = v(t; 0, \lambda)$ satisfies

$$\left.\begin{aligned} &-v'' - \lambda v = w \\ &v'(0) = 0, \ v(0) = 0. \end{aligned}\right\} \tag{7.27}$$

Bifurcation occurs at λ_0 if

$$f_{s\lambda}(0, \lambda_0) = \frac{\partial v}{\partial t}(l, 0, \lambda_0) \neq 0. \tag{7.28}$$

Suppose (7.28) does not hold. Then $v(t) = v(t; 0, \lambda_0)$ satisfies (7.27) with $\lambda = \lambda_0$, together with

$$v'(l) = 0. \tag{7.29}$$

Then (7.27) implies

$$\begin{aligned} \int_0^l w^2 &= -\int_0^l v''w - \lambda_0 \int_0^l vw \\ &= +\int_0^l v'w' - \lambda_0 \int_0^l vw \quad \text{by (7.29)} \\ &= -\int_0^l vw'' - \lambda_0 \int_0^l vw \quad \text{since } w'(0) = 0 = w'(l) \\ &= \int_0^l v(-w'' - \lambda_0 w) = 0 \quad \text{by (7.24)}, \end{aligned}$$

which is impossible since $w(t) = w(t; 0, \lambda_0) = \sqrt{\lambda_0}\cos\sqrt{\lambda_0}t$ and $\lambda_0 \neq 0$. So bifurcation from the trivial solution occurs at $\lambda = \lambda_0 = \frac{m^2\pi^2}{l^2}, \quad m = 1, 2, ...$

7.3 Calculation of fold points in ODEs using shooting

We consider this technique via an example. See also Seydel [45].

Example 7.5 *Consider the following nonlinear ODE:*

$$-y'' - \lambda \exp(y) = 0, \quad y(0) = 0 = y(1).$$

Using the development in Example 7.4 we set up an associated IVP and $f(s) := u(1; s, \lambda)$. *To calculate the values of* f *and its derivatives in the shooting method, we have the three initial value problems:*

$$\boxed{\begin{array}{rcl} -u'' &=& \lambda e^u \\ u(0) &=& 0 \\ u'(0) &=& s \end{array}} \qquad \boxed{\begin{array}{rcl} -w'' &=& \lambda e^u w \\ w(0) &=& 0 \\ w'(0) &=& 1 \end{array}} \qquad \boxed{\begin{array}{rcl} -v'' &=& \lambda e^u v + e^u \\ v(0) &=& 0 \\ v'(0) &=& 0 \end{array}}$$

Then

$$f(s, \lambda) = u(1; s, \lambda); \quad f_s(s, \lambda) = w(1; s, \lambda); \quad f_\lambda(s, \lambda) = v(1; s, \lambda).$$

We may follow the solution curve of $f(s, \lambda) = 0$ *numerically by continuation with respect to* s *or* λ, *with a check on size of* $|f_s|,\ |f_\lambda|$. *If one of these becomes small then we use the other as a parameter. If they are not both zero then curve has only turning points. Observe also that when* $s = 0,\ \lambda = 0$ *we have*

$$\begin{aligned} u &= 0, \text{ so} \quad f(0,0) = 0; \\ w &= t, \text{ so} \quad f_s(0,0) = 1; \\ v &= -\frac{1}{2}t^2, \text{ so} \quad f_\lambda(0,0) = -\frac{1}{2}. \end{aligned}$$

We may use $(0, 0)$ *as the starting point for continuation. To solve for* u, v, w *simultaneously, set* $u_1 = u,\ u_2 = u',\ u_3 = w,\ u_4 = w',\ u_5 = v,\ u_6 = v'$. *Then the three problems become:*

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{bmatrix}' = \begin{bmatrix} u_2 \\ -u_7 \exp(u_1) \\ u_4 \\ -u_7 \exp(u_1) u_3 \\ u_6 \\ -u_7 \exp(u_1) u_5 - \exp(u_1) \\ 0 \end{bmatrix}, \text{ with } \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{bmatrix} = \begin{bmatrix} 0 \\ s \\ 0 \\ 1 \\ 0 \\ 0 \\ \lambda \end{bmatrix} \text{ at } t = 0.$$

For any (s, λ) *we solve this system by any numerical routine, and then*

$$f(s, \lambda) = u_1(1),\ f_s(s, \lambda) = u_3(1),\ f_\lambda(s, \lambda) = u_5(1). \tag{7.30}$$

These may be used in a continuation method as described in §4.

8 Hopf Bifurcation

As was seen in Example 2.2, one way a steady state of $\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}, \lambda)$ can lose stability as λ varies is when a complex pair of eigenvalues of $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}, \lambda)$ crosses the imaginary axis. This situation is described by the classical Hopf bifurcation theorem [19].

Theorem 8.1 (Hopf Bifurcation) *Let $\boldsymbol{F} \in C^2(\mathbb{R}^{n+1})$ and assume*

(i) $\boldsymbol{F}(\boldsymbol{x}_0, \lambda_0) = \boldsymbol{0}$,

(ii) $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}_0, \lambda_0)$ *has a simple, purely imaginary eigenvalue* $\mu(\lambda_0) = +i\beta_0$, $\beta_0 \neq 0$, *with eigenvector* $\boldsymbol{\phi}_0 + i\boldsymbol{\psi}_0$, *and no other eigenvalues on the imaginary axis apart from at* $-i\beta_0$,

(iii) $\{\frac{d}{d\lambda}\mathrm{Re}(\mu)\}\Big|_{\lambda=\lambda_0} \neq 0$.

Then there exists an $a_0 > 0$ and a parameter a such that $\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}, \lambda)$ has a smooth branch of $T(a)$-periodic solutions $(\boldsymbol{x}(t, a; \lambda(a)), \lambda(a))$ for $0 \leq t \leq T(a)$, for all $|a| < a_0$ with the following properties

$$\begin{aligned} \boldsymbol{x}(t, a; \lambda(a)) &= \boldsymbol{x}^s(\lambda(a)) + a(\cos(\beta_0 t)\boldsymbol{\phi}_0 - \sin(\beta_0 t)\boldsymbol{\psi}_0) + \mathcal{O}(a^2), \\ \lambda(a) &= \lambda_0 + \mathcal{O}(a^2), \\ T(a) &= \frac{2\pi}{\beta_0} + \mathcal{O}(a^2), \end{aligned}$$

where $\boldsymbol{x}^s(\lambda(a))$ denotes the steady solution at $\lambda = \lambda(a)$.

The theorem states that at $(\boldsymbol{x}_0, \lambda_0)$ there is a birth of periodic solutions that may be parametrised by the amplitude a. If conditions *(i),(ii)* and *(iii)* hold then $(\boldsymbol{x}_0, \lambda_0)$ is called a *Hopf bifurcation point*.

Note that since $\boldsymbol{F}^0_{\boldsymbol{x}}$ is nonsingular, the Implicit Function Theorem ensures that S (the solution set of the steady problem) may be parametrised by λ near $\lambda = \lambda_0$. Using the approach in the proof of part (b) of Theorem 5.3 extended to $\mathbb{C}^{n+1}$ shows that a (real or complex) simple eigenvalue of $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}(\lambda), \lambda)$ is a smooth function of λ near λ_0. Hence we can write $\mathrm{Re}(\mu) = \mathrm{Re}(\mu(\lambda))$.

Condition *(iii)* in Theorem 8.1 is another example of a nondegeneracy condition where an eigenvalue smoothly crosses the imaginary axis.

Theorem 8.1 is due to Hopf in a famous paper in 1942, though in 1929 Andronov was the first to formulate a theorem and Poincaré's work in 1892 contained examples of this type of

bifurcation. So this phenomenon is now often called Poincaré/Andronov/Hopf bifurcation. A nice treatment of the theory of Hopf bifurcation with references to the work of Andronov and Poincaré is given in Wiggins [52].

8.1 Calculation of a Hopf Bifurcation Point

If a good estimate of the Hopf bifurcation point is known then it may be computed exactly by setting up and solving an appropriate extended system (cf. the fold point system in (5.4).) Consider the nonlinear system

$$\boldsymbol{H}(\boldsymbol{y}) = \boldsymbol{0} \tag{8.1}$$

where

$$\boldsymbol{H}(\boldsymbol{y}) := \begin{pmatrix} \boldsymbol{F}(\boldsymbol{x},\lambda) \\ \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x},\lambda)\boldsymbol{\phi} - \beta\boldsymbol{\psi} \\ \boldsymbol{c}^T\boldsymbol{\phi} - 1 \\ \boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x},\lambda)\boldsymbol{\psi} + \beta\boldsymbol{\phi} \\ \boldsymbol{c}^T\boldsymbol{\psi} \end{pmatrix}, \quad \boldsymbol{y} := \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\phi} \\ \lambda \\ \boldsymbol{\psi} \\ \beta \end{pmatrix} \in \mathbb{R}^{3n+2} \tag{8.2}$$

with $\boldsymbol{H} : \mathbb{R}^{3n+2} \to \mathbb{R}^{3n+2}$. This is the obvious system to write down as can be seen from conditions *(i)* and *(ii)* in Theorem 8.1. There are two conditions on the eigenvector $\boldsymbol{\phi} + i\boldsymbol{\psi}$ since a complex vector requires two real normalisations.

The following theorem is readily proved (see [16]).

Theorem 8.2 *Let $(\boldsymbol{x}_0, \lambda_0)$ be a Hopf bifurcation point (i.e. (i), (ii) and (iii) of Theorem 8.1 hold) and assume $\boldsymbol{c}$ has non-zero projection on* span $\{\boldsymbol{\phi}_0\}$. *Then $\boldsymbol{y}_0 := (\boldsymbol{x}_0^T, \boldsymbol{\phi}_0^T, \lambda_0, \boldsymbol{\psi}_0^T, \beta_0)^T \in \mathbb{R}^{3n+2}$ is a regular solution of (8.1).*

Note that fold points also satisfy (8.1) since if $(\boldsymbol{x}_0, \lambda_0)$ is a fold point and $\boldsymbol{\phi}_0 \in \ker(\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}_0, \lambda_0))$ then $\boldsymbol{y}_0 = (\boldsymbol{x}_0, \boldsymbol{\phi}_0, \lambda_0, \boldsymbol{0}, 0)$ satisfies $\boldsymbol{H}(\boldsymbol{y}_0) = \boldsymbol{0}$. In fact $\boldsymbol{y}_0$ is a regular solution if the conditions of Theorem 5.3 hold.

System (8.1) was first introduced by Jepson [21] and independently by Griewank and Reddien [16] who showed that the linearisation of (8.1) could be reduced to solving systems with a bordered form of $\boldsymbol{F}_{\boldsymbol{x}}^2(\boldsymbol{x},\lambda) + \beta^2 I$. This is natural since an alternative system for a Hopf bifurcation can be derived by using the fact that the second and fourth equations of (8.1) can be written as $(\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x},\lambda) + \beta^2 I)\boldsymbol{v} = 0$ with $\boldsymbol{v} = \boldsymbol{\phi}$ or $\boldsymbol{\psi}$.

To eliminate the possibility of computing a fold point rather than a Hopf bifurcation point, Werner and Janovsky [50] used the system

$$\mathbf{R}(\boldsymbol{y}) = \mathbf{0} \tag{8.3}$$

where

$$\mathbf{R}(\boldsymbol{y}) = \begin{pmatrix} \boldsymbol{F}(\boldsymbol{x}, \lambda) \\ (\boldsymbol{F}_{\boldsymbol{x}}^2(\boldsymbol{x}, \lambda) + \nu I)\boldsymbol{\phi} \\ \boldsymbol{c}^T\boldsymbol{\phi} \\ \boldsymbol{c}^T\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}, \lambda)\boldsymbol{\phi} - 1 \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{\phi} \\ \lambda \\ \nu \end{pmatrix} \in \mathbb{R}^{2n+2} \tag{8.4}$$

in which $\mathbf{R} : \mathbb{R}^{2n+2} \to \mathbb{R}^{2n+2}$, and where $\boldsymbol{c}$ is a constant vector. The last equation in (8.3) ensures that the solution cannot be a fold point. The system $\mathbf{R}(\boldsymbol{y}) = \mathbf{0}$ is closely related to a system derived by Roose and Hlavacek [41], but (8.3) has several advantages when computing paths of Hopf bifurcations if a second parameter is varying (see [50]).

The fact that (8.1) or (8.3) is regular at a Hopf bifurcation is important since Newton's method (or some variant) will probably be used to solve the system. Just as is the case in §5 for the computation of a fold point, there are efficient ways of solving the Jacobian systems in Newton's method. In [16] an efficient procedure is described for the solution of the $(3n+2)\times(3n+2)$ Jacobian systems arising from (8.1) by solving systems with a bordering of $(\boldsymbol{F}_{\boldsymbol{x}}^2(\boldsymbol{x}, \lambda) + \beta^2 I)$. We do not give the details here. A nice summary is given in [1].

8.2 The Detection of Hopf Bifurcations in Large Systems

The extended systems in §8.1 can only be used when we know we are near a Hopf point. The following section describes how this might be determined in practice.

When computing a path of steady solutions of $\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}, \lambda)$ using a numerical continuation method it is easy to pass over a Hopf bifurcation point without "noticing" it, since when a complex pair of eigenvalues crosses the imaginary axis there is no easy detection test based on the linear algebra of the continuation method. In particular the sign of the determinant of $\boldsymbol{F}_{\boldsymbol{x}}$ does not change. If n is small then the simplest test is merely to compute all the eigenvalues of $\boldsymbol{F}_{\boldsymbol{x}}$ during the continuation. For large n, say when $\boldsymbol{F}$ arises from a discretized PDE, such an approach will usually be out of the question. The efficient detection of Hopf bifurcations in large systems is an important and, as yet, unsolved problem. The review article [8] discusses in detail both classical techniques from complex analysis and linear algebra-based methods. It

is natural to try to use classical ideas from complex analysis for this problem since then one seeks an *integer*, namely the number of eigenvalues in the unstable half-plane, and counting algorithms are applicable. This is explored for large systems in [14] but there is still work to be done in this area.

The *rightmost* eigenvalues of $\boldsymbol{F}_{\boldsymbol{x}}(\boldsymbol{x}, \lambda)$ determine the (linearised) stability of the steady solutions of $\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}, \lambda)$ and one strategy for the detection of Hopf bifurcation points is to monitor a few of the rightmost eigenvalues as the path of steady state solutions is computed. (Note that the rightmost eigenvalue is not a continuous function of λ, see [36].) Standard iterative methods, e.g. Arnoldi's method and simultaneous iteration, compute extremal or dominant eigenvalues, and there is no guarantee that the rightmost eigenvalue will be computed by direct application of these methods to $\boldsymbol{F}_{\boldsymbol{x}}$. The approach in [8] and [3] is to first transform the eigenvalue problem using the Generalised Cayley Transform

$$C(A) = (A - \alpha_1 I)^{-1}(A - \alpha_2 I), \quad \alpha_1, \alpha_2 \in \mathbb{R},$$

which has the key property that if $\mu \neq \alpha_1$ is an eigenvalue of A then $\theta := (\mu - \alpha_1)^{-1}(\mu - \alpha_2)$ is an eigenvalue of $C(A)$. Also, $\mathrm{Re}(\mu) \leq (\geq)(\alpha_1 + \alpha_2)/2$ if and only if $|\theta| \leq (\geq)1$. Thus eigenvalues to the right of the line $\mathrm{Re}(\mu) = (\alpha_1 + \alpha_2)/2$ are mapped outside the unit circle and eigenvalues to the left of the line mapped inside the unit circle. In [8] and [3] algorithms based on computing dominant eigenvalues of $C(\boldsymbol{F}_{\boldsymbol{x}})$ using Arnoldi or simultaneous iteration are presented, with consequent calculation of rightmost eigenvalues of $\boldsymbol{F}_{\boldsymbol{x}}$. These algorithms were tested on a variety of problems, including systems arising from mixed finite element discretizations of the Navier-Stokes equations. Quite large problems can in fact be tackled. Indeed, in [15] the problem of the stability of flow over a backward facing step is discussed in detail and the rightmost eigenvalues of a system with over 3×10^5 degrees of freedom are found using the Generalised Cayley transform allied with simultaneous iteration.

However it was later noted (see [31]) that since

$$C(A) = I + (\alpha_1 - \alpha_2)(A - \alpha_1 I)^{-1},$$

Arnoldi's method applied to $C(A)$ builds the same Krylov subspace as Arnoldi's method applied to the shift-invert transformation $(A - \alpha_1 I)^{-1}$. Thus if Arnoldi's method is the eigenvalue solver there is no advantage in using the Cayley transform, which needs two parameters, over the standard shift-invert transformation (see [31]).

One can think of the approach in [3] as the computation of the subspace containing the eigenvectors corresponding to the rightmost eigenvalues of $\boldsymbol{F_x}$. A similar theme, derived using a completely different approach, is described by [42] and refined by [5]. In these papers the subspace corresponding to a set of (say rightmost) eigenvalues is computed using a hybrid iterative process based on a splitting technique. Roughly speaking a small subspace is computed using a Newton-type method and the solution in the larger complementary space is found using a Picard (contraction mapping) approach. One advantage is that the Jacobian matrix $\boldsymbol{F_x}$ need never be evaluated.

When detecting Hopf bifurcations in the Navier-Stokes equations using mixed finite elements, a generalised eigenvalue problem of the form $A\phi = \mu B\phi$ arises where B is singular. A common method is to apply Arnoldi's method to the shifted-inverted matrix $(A - \alpha B)^{-1}B$, which is singular since B is singular. In [30] it is noted that great care is needed here when using Arnoldi's method because of the generation of spurious eigenvalues due to perturbation of the zero eigenvalue. The details are quite technical and are omitted.

Finally we note that chapter 5 of [45] contains an overview of Hopf detection techniques.

References

[1] W. J. Beyn. Numerical methods for dynamical systems. In W. Light, editor, *Advances in Numerical Analysis*, pages 175–227. Clarendon Press, Oxford, 1991.

[2] S-N. Chow and J. K. Hale. *Methods of Bifurcation Theory.* Springer-Verlag, New York, 1982.

[3] K. A. Cliffe, T. J. Garratt, and A. Spence. Eigenvalues of the discretized Navier-Stokes equation with application to the detection of Hopf bifurcations. *Advances in Computational Maths.*, **1**:337–356, 1993.

[4] M. G. Crandall and P. H. Rabinowitz. Bifurcation from a simple eigenvalue. *J. Functional Analysis*, **8**:321–340, 1971.

[5] Bryan D. Davidson. Large-scale continuation and numerical bifurcation for partial differential equations. *SIAM J. Numer. Anal.*, **34**:2008–2027, October 1997.

[6] J.E Dennis Jr and Robert B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice Hall, New Jersey, 1983.

[7] E. J. Doedel and J. P. Kernevez. AUTO: Software for continuation and bifurcation problems in ordinary differential equations. Technical report, Caltech, Pasadena, 1986.

[8] T. J. Garratt, G. Moore, and A. Spence. Two methods for the numerical detection of Hopf bifurcations. In T. Küpper R. Seydel, F. W. Schneider and H. Troger, editors, *Bifurcation and Chaos; Analysis, Algorithms, Applications*, volume **97**. Birkhäuser, Basel, 1991.

[9] M. Golubitsky and D. G. Schaeffer. *Singularities and Groups in Bifurcation Theory, Vol I*, volume **51**. Springer-Verlag, New York, 1985.

[10] M. Golubitsky, I. Stewart, and D. G. Schaeffer. *Singularities and Groups in Bifurcation Theory, Vol II*, volume **69**. Springer-Verlag, New York, 1988.

[11] W. Govaerts. Stable solvers and block elimination for bordered systems. *SIAM J Matrix Anal. and Applications*, **12**:469–483, 1991.

[12] W. Govaerts. Bordered matrices and singularities of large nonlinear systems. *Int. J. of Bifurcation and Chaos*, **5**:243–250, 1995.

[13] W. Govaerts. *Numerical Methods for Bifurcations of Dynamic Equilibria*. SIAM, To be published, 1999.

[14] W. Govaerts and A. Spence. Detection of Hopf points by counting sectors in the complex plane. *Numer. Math.*, **75**:43–58, 1996.

[15] P. M. Gresho, D. K. Gartling, J. R. Torczynski, K. A. Cliffe, K. H. Winters, T. J. Garratt, A. Spence, and J. W. Goodrich. Is the steady viscous incompressible 2D flow over a backward facing step ar Re=800 stable? *Int. J. Numer. Meth. Fluids*, **17**:501–541, 1993.

[16] A. Griewank and G. Reddien. The calculation of Hopf points by a direct method. *IMA J. Numer. Anal.*, **3**:295–303, 1983.

[17] A. Griewank and G. Reddien. Characterization and computation of generalized turning points. *SIAM J. Numer. Anal.*, **21**:176–185, 1984.

[18] A. Griewank and G. Reddien. Computation of cusp singularities for operator equations and their discretizations. *J. Comp. Appl Maths*, **16**:133–153, 1989.

[19] B. D. Habard, N. D. Kazarinoff, and Y-H Wan. Theory and applications of Hopf bifurcation. *London Math. Soc. Lecture Note Series*, **41**, 1981.

[20] M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra.* Academic Press, New York, 1974.

[21] A. D. Jepson. *Numerical Hopf Bifurcation.* PhD thesis, Caltech, Pasadena, 1981.

[22] A. D. Jepson and D. W. Decker. Convergence near bifurcation. *SIAM J. Numer. Anal.*, **23**:959–975, 1986.

[23] A. D. Jepson and A. Spence. The numerical solution of nonlinear equations having several parameters, I: scalar equations. *SIAM J. Numer. Anal.*, **22**:736–759, 1985.

[24] H. B. Keller. *Numerical methods for two-point boundary value problems.* Ginn-Blaisdell, Waltham, 1968.

[25] H. B. Keller. *Numerical solution of bifurcation and nonlinear eigenvalue problems*, pages 359–384. Academic Press, New York, 1977.

[26] H. B. Keller. *Numerical methods in bifurcation problems.* Springer-Verlag, 1987.

[27] J. B. Keller. Bifurcation Theory for Ordinary Differential Equations. In J. B. Keller and S. Antmann, editors, *Bifurcation Theory and Nonlinear Eigenvalue Problems.* Benjamin, New York, 1969.

[28] T. Küpper, H. D. Mittelmann, and H. Weber. *Numerical Methods for Bifurcation Problems*, volume **70**. Birkhäuser, Basel, 1984.

[29] T. Küpper, R. Seydel, and H. Troger. Bifurcation: Analysis, Algorithms, Applications. *Proceedings of a conference in Dortmund*, **79**, 1987.

[30] K. Meerbergen and A. Spence. Implicitly Restarted Arnoldi with purification for the shift-invert transformation. *Mathematics of Computation*, **67**:667–689, 1997.

[31] K. Meerbergen, A. Spence, and D. Roose. Shift-Invert and Cayley transforms for detection of rightmost eigenvalues of nonsymmetric matrices. *BIT*, **34**:409–423, 1994.

[32] H. D. Mittelmann and H. Weber. *Bifurcation problems and their Numerical Solution*, volume **54**. Birkhaüser, Basel, 1980.

[33] G. Moore. Numerical Treatment of nontrivial bifurcation points. *Numer. Funct. Anal. and Optimiz.*, **2**:441–472, 1980.

[34] G. Moore. Some remarks on the deflated block elimination method. In T. Küpper, R. Seydel, and H. Troger, editors, *Bifurcation: Analysis, Algorithms and Applications*, International Series in Numerical Mathematics, pages 222–234. Birkhäuser, 1987.

[35] G. Moore and A. Spence. The calculation of Turning Points of nonlinear equations. *SIAM J. Numer. Anal.*, **17**:567–576, 1980.

[36] R. Neubert. Predictor-Corrector techniques for detecting Hopf bifurcation points. *International J. Bifurcation and Chaos*, **3**:1311–1318, 1993.

[37] P. H. Rabinowitz. *Applications of Bifurcation Theory.* Academic Press, New York, 1977.

[38] L. B. Rall. *Computational solution of nonlinear operator equations.* Wiley, New York, 1969.

[39] W. C. Rheinboldt. *Numerical Analysis of parametrized nonlinear equations.* Wiley-Interscience, New York, 1986.

[40] D. Roose, B. De Dier, and A. Spence. *Continuation and Bifurcations: Numerical Techniques and Applications*, volume **313**. Kluwer, Dordrecht, 1990.

[41] D. Roose and V. Hlavacek. A direct method for the computation of Hopf bifurcation points. *SIAM J. Appl. Math.*, **45**:879–894, 1985.

[42] G. Schroff and H. B. Keller. Stabilization of unstable procedures: The recursive projection method. *SIAM J. Numer. Analysis*, **30**:1099–1120, 1993.

[43] R. Seydel. Numerical Computation of branch points in nonlinear equations. *Numer. Math.*, **32**:339–352, 1979.

[44] R. Seydel. Numerical Computation of branch points in ordinary differential equations. *Numer. Math.*, **32**:51–68, 1979.

[45] R. Seydel. *Practical Bifurcation and Stability Analysis: From equilibrium to Chaos.* Springer-Verlag, New York, 1994.

[46] R. Seydel, T. Küpper, F. W. Schneider, and H. Troger. *Bifurcation and Chaos: Analysis, Algorithms, Applications*, volume **97** of *International Series in Numerical Mathematics*. Birkhäuser, Basel, 1991.

[47] G. Symm and J. H. Wilkinson. Realistic Error Bounds for a Simple Eigenvalue and its Associated Eigenvector. *Numer. Math.*, **35**:113–126, 1980.

[48] A. Vanderbauwhede. Local bifurcation and symmetry. In *Research notes in mathematics*, volume **75**. Pitman, London, 1982.

[49] B. Werner. Regular systems for bifurcation points with underlying symmetries. In T. Küpper, H. D. Mittelmann, and H. Weber, editors, *Numerical Methods for Bifurcation Problems*, volume **70**, pages 562–574. Birkhäuser, Basel, 1984.

[50] B. Werner and V. Janovsky. Computation of Hopf branches bifurcating from Takens-Bogdanov points for problems with symmetry. In R. Seydel, T. Küpper, F. W. Schneider, and H. Troger, editors, *Bifurcation and Chaos: Analysis, Algorithms, Applications*, pages 377–388. Birkhauser, 1991.

[51] B. Werner and A. Spence. The computation of symmetry breaking bifurcation points. *SIAM J. Numer. Anal.*, **21**:388–399, 1984.

[52] S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer-Verlag, New York, 1990.