Lectures on
Numerical Methods In Bifurcation Problems

By

H.B. Keller

Lectures delivered at the
Indian Institute Of Science, Bangalore
under the
T.I.F.R.-I.I.Sc. Programme In Applications Of Mathematics

Notes by

A.K.Nandakumaran and Mythily Ramaswamy

Published for the
Tata Institute Of Fundamental Research
Springer-Verlag
Berlin Heidelberg New York Tokyo

Author
H.B. Keller
Applied Mathematics 217-50
California Institute of Technology
Pasadena, California 91125
U.S.A.

ISBN 3-540-20228-5 Springer-verlag, Berlin, Heidelberg,
New York. Tokyo
ISBN 0-387-20228-5 Springer-verlag, New York. Heidelberg.
Berlin. Tokyo

Printed by
INSDOC Regional Centre.
Indian Institute of Science Campus,
Bangalore-560 012
and published by H.Goctzc, Springer-Verlag,
Heidelberg, West Germany
Printed In India

Preface

These lectures introduce the modern theory of continuation or path following in scientific computing. Almost all problem in science and technology contain parameters. Families or manifolds of solutions of such problems, for a domain of parameter variation, are of prime interest. Modern continuation methods are concerned with generating these solution manifolds. This is usually done by varying one parameter at a time - thus following a parameter path curve of solutions.

We present a familiar, interesting and simple example in Chapter 1 which displays most of the basic phenomena that occur in more complex problems. In Chapter 2 we examine some local continuation methods, bases mainly on the implicit function theorem. We go on to introduce concepts of global continuation, degree theory and homotopy invariance with several important applications in Chapter 3. In Chapter 4, we discuss practical path following procedures, and introduce folds or limit point singularities. Pseudo-arclength continuation is also introduced here to circumvent the simple fold difficulties. General singular points and bifurcations are briefly studied in Chapter 5 where branch switching and (multiparameter) fold following are discussed. We also very briefly indicate how periodic solutions path continuation and Hopf bifurcations are Incorporated into our methods. Finally in Chapter 6, we discuss two computational examples and some details of general methods employed in carrying out such computations.

This material is based on a series of lectures I presented at the Tata institute of Fundamental Research in Bangalore, India during December 1985, and January 1986. It was a most stimulating and enjoyable experienced for me, and the response and interaction with the audience was unusually rewarding. The lecture notes were diligently recorded and written up by Mr. A.K. Nandakumar of T.I.F.R., Bangalore. The final chapter was mainly worked out with Dr. Mythily Ramaswamy of T.I.F.R. Ramaswamy also completely reviewed the entire manuscript, corrected many of the more glaring errors and made many other improvements. Any remaining errors are due to me. The iteration to converge on the final manuscript was allowed only one step due to the distance involved. The result, however, is surprisingly close to parts of my original notes which are being independently prepared for publication in a more extend form.

I am most appreciative to the Tata Institute of Fundamental Research for the opportunity to participate in their program. I also wish to thank the U.S. Department of Energy and the U.S.Army Research Office who have for years sponsored much of the research that has resulted in these lectures under the grants: 0E-AS03-7603-00767 and DAAG29-80-C-0041.

Herbert B. Keller
Pasadena, California
December 29, 1986

Contents

Chapter 1
Basic Examples

1.1 Introduction

1

Our aim in these lectures is to study constructive methods for solving nonlinear systems of the form :

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

where λ is a possibly multidimensional parameter and G is a smooth function or operator from a suitable function space into itself. Frequently we will work in finite dimensional spaces. In this introductory chapter we present two examples from population dynamics and study the behaviour of solutions regarding bifurcation, stability and exchange of stability. Second chapter describes a local continuation method and using this we will try to obtain global solution of (1.1). The important tool for studying this method is the *implicit function theorem*. We will also present various predictor - solver methods. The third chapter deals with global continuation theory. Degree theory and some of its applications are presented there. Later in that chapter we will study global homotopies and Newton methods for obtaining solutions of the equation (1.1).

In the fourth chapter, we describe a practical procedure for computing paths and introduce the method of arclength continuation. Using the bordering algorithm presented in that chapter, we can compute paths in

an efficient way. In chapter 5, we will study singular points and bifur-
cation. A clear description of various methods of continuation past sin-
2 gular points, folds, branch switching at bifurcations points is presented
in that chapter. We also study multiparameter problems, Hopfbifurca-
tions later in that chapter. The final chapter presents some numerical
results obtained using some of the techniques presented in the previous
chapters.

1.2 Examples (population dynamics)

We start with a simple, but important example from population dynamics (see [11], [24]). Let $u(t)$ denote the populations density in a particular region, at time t. The simplest growth law states that the *rate of change of population is proportional to the existing density, that is:*

$$\frac{du}{dt} = \beta u, \tag{1.3}$$

where β is the reproduction rate. The solution for this problem is given by

$$u(t) = u_\circ e^{\beta(t-t_\circ)}, u_\circ = u(t_\circ). \tag{1.4}$$

Note that $u(t) \to \infty$ as $t \to \infty$. This means that population grows indefinitely with time. Obviously, we know that such situation is not possible in practice. Hence β cannot be a constant. It must decrease as u increases. Thus as a more realistic model, we let β be a function of u, say linear in u :

$$\beta = \beta_1(1 - \frac{u}{u_1}). \tag{1.5}$$

Here $\beta_1 > 0$, is the ideal reproduction rate, and u_1 is the maximal density that can be supported in the given environment.

3 Note that if $u > u_1$ then $\beta < 0$ so that $u(t)$ decays. On the other hand,
if $u < u_1$ then $u(t)$ grows as time increases. The solutions curve with β
as above is sketched in figure 1.1; it is:

$$u(t) = \frac{u_\circ u_1 e^{\beta_1 t}}{(u_1 - u_\circ) + u_\circ e^{\beta_1 t}}, \quad \text{where} \quad u_\circ = u(0). \tag{1.6}$$

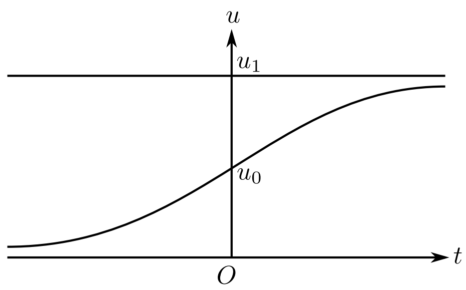


Figure 1.1:

We now consider the more general case, in which coupling with an external population density, say $u_\circ$, is allowed. Specifically we take

$$\frac{du}{dt} = \beta_1(1 - \frac{u}{u_1})u + \alpha_o(u_\circ - u) \tag{1.7}$$

where,

$$\begin{aligned}
\beta_1 &= \text{ ideal reproduction rate}(> 0),\\
u_1 &= \text{ maximal density},\\
u_\circ &= \text{ exterior density},\\
\alpha_\circ &= \text{ flow rate } (\alpha_\circ \geq 0 \text{ or } \alpha_\circ \leq 0).
\end{aligned}$$

Naturally, the population of a particular region may depend upon the population of the neighbouring regions. If the populations of the exterior is less, then species may move to that region ($\alpha_\circ > 0$) and vice **4**
versa ($\alpha_\circ < 0$). The term $\alpha_\circ(u_\circ - u)$ accounts for this behaviour.

We scale (1.7) by setting

$$\begin{aligned}
\lambda_1 &= \frac{\beta_1 - \alpha_\circ}{\beta_1} u_1,\\
\lambda_1 &= \frac{\alpha_\circ}{\beta_1} u_1 u_\circ,\\
t &= \bar{t}\frac{u_1}{\beta_1}.
\end{aligned}$$

Then we have

$$\frac{du}{dt} = G(u, \lambda) \equiv -u^2 + \lambda_1 u + \lambda_2. \tag{1.8}$$

Here $\lambda = (\lambda_1, \lambda_2)$ denotes the two independent parameters λ_1 and λ_2.

STEADY STATES: The steady state are solutions of

$$G(u, \lambda) = 0.$$

These solutions are :

$$u_{\pm} = \frac{\lambda_1}{2} + \sqrt{(\frac{\lambda_1}{2})^2 + \lambda_2} \tag{1.9}$$

These solutions are distinct unless

$$\lambda_2 = -(\frac{\lambda_1}{2})^2. \tag{1.10}$$

Along this curve in the (λ_1, λ_2) plane, $G_u(u, \lambda) = 0$. This curve is known as a fold (sometimes it is called the bifurcation set) and the number of solutions changes as one crosses this curve (See Fig. 1.2).

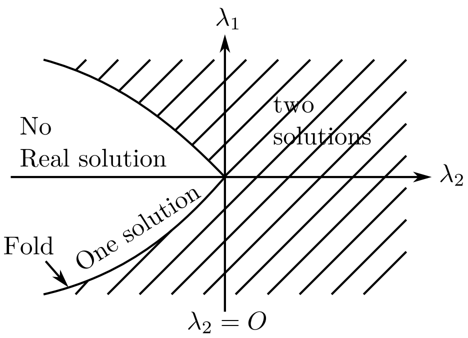


Figure 1.2:

5

We examine three distinct cases :

$$\text{(i) } \lambda_2 = 0 \quad \text{(ii) } \lambda_2 > 0 \quad \text{(iii) } \lambda_2 < 0.$$

(i) $\lambda_2 = 0$: The solutions are two straight lines given by $u \equiv 0$ and $u = \lambda_1$. They intersect at the origin. The origin is thus a **bifurcation point**, where two distinct solution branches intersect.

(ii) $\lambda_2 > 0$: In this case, there are two solutions u_+ and u_-, arcs of the hyperbola, whose asymptotes are given by $u = 0$ and $u = \lambda_1$.

(iii) $\lambda_2 < 0$: In this case, a real solution exists only if $|\lambda_1| > \sqrt{-4\lambda_2}$. They are the hyperbolae conjugate to those of case (ii) (See Fig. 1.3).

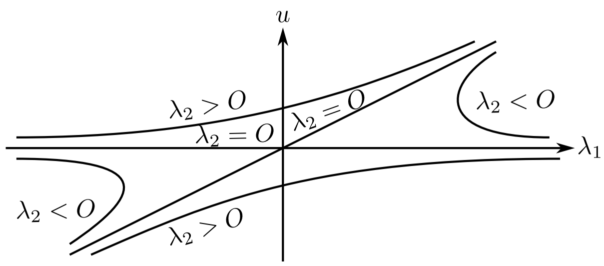


Figure 1.3:

The solution surfaces in the space $(u, \lambda_1, \lambda_2)$ is sketched is Fig. 1.4. **6**
It is clearly a saddle surface with saddle point at the origin. As we see later on, the cases (ii) and (iii) are example of perturbed bifurcation.

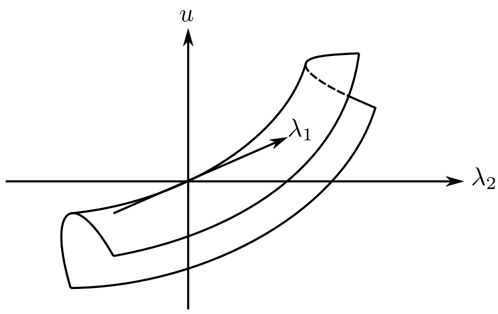


Figure 1.4:

STABILITY : To examine the stability of these states we note that (1.8) can be written as:

$$\frac{du}{dt} = -(u - u_+)(u - u_-). \tag{1.11}$$

Then it clearly follows that :

$$\frac{du}{dt} \begin{cases} < 0 \text{ if } u < u_- < u_+ \\ > 0 \text{ if } u_- < u < u_+ \\ < 0 \text{ if } u_- < u_+ < u. \end{cases} \tag{1.12}$$

This means that u decreases when it is in the semi-infinite intervals $u < u_-$ and $u_+ < u$. It increases in $u_- < u < u_+$. Hence it follows that
7 u_+ is always stable and that u_- is always unstable. Note that the trivial solution $u \equiv 0$ consists of u_+ for $\lambda_1 < 0$. and u_- for $\lambda_1 > 0$. Thus in the bifurcation case, $\lambda_2 = 0$, the phenomenon of exchange of stability occurs as λ_1 changes sign. That is the branch which is stable changes as λ_1 passes though the bifurcation value $\lambda_1 = 0$.

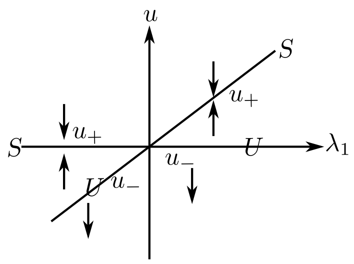


Figure 1.5: ($\lambda_2 = 0$)

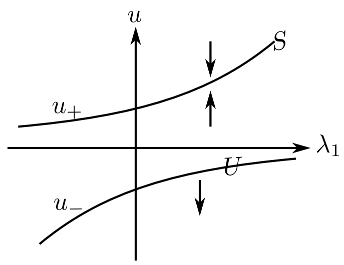


Figure 1.6: ($\lambda_2 > 0$)

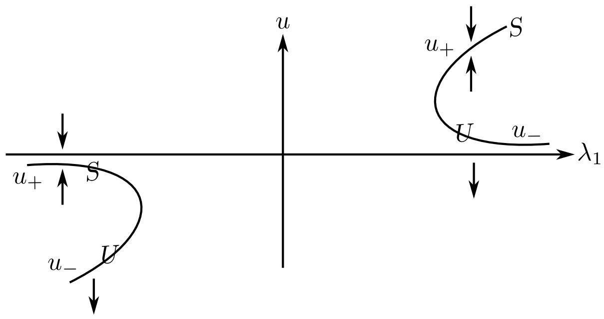


Figure 1.7: ($\lambda_2 < 0$)

We consider another model of reproduction rate β, say, quadratic in u:

$$\beta = \beta_1[1 - (\frac{u}{u_1})^2]. \tag{1.13}$$

Then equation (1.8) reduces to : **8**

$$\frac{du}{dt} = G(u, \lambda) \equiv -u^3 + \lambda_1 u + \lambda_2, \tag{1.14}$$

where now :

$$\lambda_1 \equiv (1 - \frac{\alpha}{\beta_1})u_1^2,$$
$$\lambda_2 \equiv \frac{\alpha}{\beta_1}u_1^2 u_\circ,$$
$$t \equiv t\frac{u_1^2}{\beta_1}.$$

As before we find steady states and then examine their stability. The study states are given by the roots of :

$$G(u, \lambda) \equiv -u^3 + \lambda_1 u + \lambda_2 = 0 \tag{1.15}$$

The three roots are given by :

$$u_\circ = a + b, u_\pm \equiv -(\frac{a+b}{2}) \pm i\sqrt{3}(\frac{a-b}{2}) \tag{1.16}$$

where,

$$a = (\frac{\lambda_2}{2} + (\frac{\lambda_2^2}{4} - \frac{\lambda_1^3}{27})^{1/2})^{1/3}$$
$$b = (\frac{\lambda_2}{2} + (\frac{\lambda_2^2}{4} - \frac{\lambda_1^3}{27})^{1/2})^{1/3}$$

In particular either we have 3 real roots or one real root. Clearly, if $\lambda_2^2 - \frac{4}{27}\lambda_1^3 > 0$, the there will one real root and two conjugate complex roots. If $\lambda_2^2 - \frac{4}{27}\lambda_1^3 = 0$ then there will be three real roots of which two are equal. If $\lambda_2^2 - \frac{4}{27}\lambda_1^3 < 0$ then there will be three distinct real roots. This can be seen as follows. Put

$$a = (a_1 + ib_1)^{1/3}, \text{ then } b = (a_1 - ib_1)^{1/3}.$$

By changing $a_1 + ib_1$ to polar co-ordinates, we can easily see that
9 $a + b$ is a real number and $a - b$ is purely imaginary. Now

$$G_u(u, \lambda) = -3u^2 + \lambda_1.$$

Combining (1.15) together with $G_u(u, \lambda) = 0$ we get

$$\lambda_1^3 = \frac{27}{4}\lambda_2^2. \tag{1.17}$$

This curve in the (λ_1, λ_2) plane represents a **fold** where the two real roots become equal; across the fold they become complex. Again note that across the fold (Bifurcation Set) the number of solutions changes. Observe that at $\lambda_2 = 0$, the solution contains the trivial branch $u = 0$ and the parabola whose branches are $u_\pm = \pm\sqrt{\lambda_1}$ which passes through the origin. Hence the origin is a **bifurcation point.** We call this configuration a **pitchfork bifurcation.** For $\lambda_2 > 0$ or $\lambda_2 < 0$ there is no bifurcation. The fold has a cusp at the origin and is sketched in Fig. 1.8.

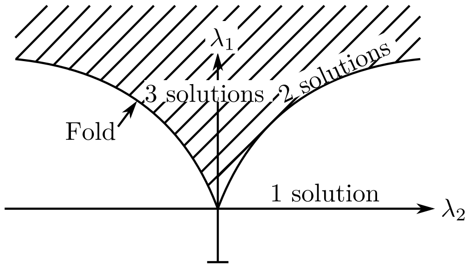


Figure 1.8:

Now we will analyse the stability results in different cases. (1.14) can be written as :

$$\frac{du}{dt} = -(u - u_o)(u - u_+)(u - u_-). \tag{1.18}$$

10

(i) $\lambda_2 = 0$: The dynamic in this case are simply generated by,

$$\frac{du}{dt} = -u(u - \sqrt{\lambda_1})(u + \sqrt{\lambda_1}), \tag{1.19}$$

and hence we see that:

$$\frac{du}{dt} \begin{cases} > 0 \text{ if } u < -\sqrt{\lambda_1} < 0, \\ < 0 \text{ if } -\sqrt{\lambda_1} < u < 0, \\ > 0 \text{ if } 0 < u < \sqrt{\lambda_1}, \\ < 0 \text{ if } 0 < \sqrt{\lambda_1} < u. \end{cases} \tag{1.20}$$

Thus u_o is stable for $\lambda_1 < 0$ and it becomes unstable as λ_1 changes sign to $\lambda_1 > 0$. In this latter range both $u_\pm$ are stable.

(ii) $\lambda_2 > 0$: Then

$$\frac{du}{dt} \begin{cases} > O \text{ if } u < u_o, \\ > 0 \text{ if } u_o < u < u_-, \\ > 0 \text{ if } u_- < u < u_+, \\ < 0 \text{ if } u_+ < u. \end{cases} \tag{1.21}$$

(iii) $\lambda_2 < 0$: In a similar way the stability results can be obtained here also.

The stability results are indicated in figures 1.9, 1.10, 1.11. The solution surface is sketched in figure 1.12.

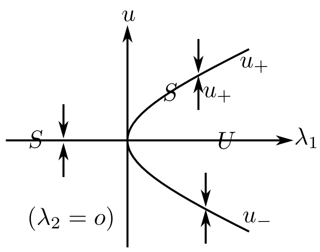


Figure 1.9: ($\lambda_2 = 0$)

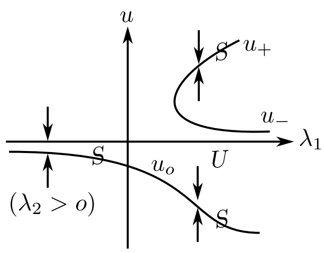


Figure 1.10: ($\lambda_2 > 0$)

11

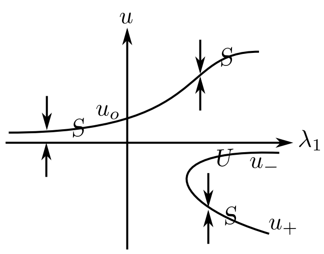


Figure 1.11: ($\lambda_2 < 0$)

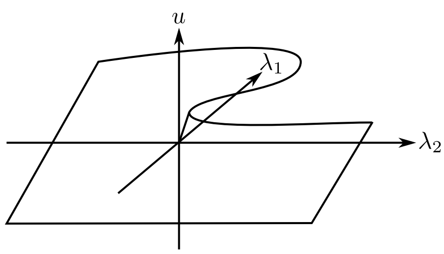


Figure 1.12:

Now we present one more example from population dynamics, in **12**
which there are two species in the same region. We know that there is a constant struggle for survival among different species of animals living in the same environment. For example, consider a region inhabited by foxes and rabbits. The foxes eat rabbits and hence this population grows as the rabbits diminish. But when the population of rabbits decreases sufficiently the fox population must decrease (due to hunger). As a result the rabbits are relatively safe and their population starts to increase again. Thus we can expect a repeated cycle of increase and decrease of the two species, i.e. a periodic solution. To model this phenomenon, we take the coupled system :

$$
\begin{aligned}
\tilde{u}_t &= \beta_u[1 - (\frac{\tilde{u}}{\tilde{u}_1})^2 - (\frac{\tilde{v}}{\tilde{v}_1})^2]\ \tilde{u} + \alpha_u(\tilde{u}_o - \tilde{u}) - \tilde{\gamma}_u\tilde{v},\\
\tilde{v}_t &= \beta_v[1 - (\frac{\tilde{u}}{\tilde{u}_1})^2 - (\frac{\tilde{v}}{\tilde{v}_1})^2]\ \tilde{v} + \alpha_v(\tilde{v}_o - \tilde{v}) + \tilde{\gamma}_v\tilde{u}.
\end{aligned}
\tag{1.22}
$$

Here $\tilde{u}$ is prey density and $\tilde{v}$ is predator density and $\tilde{\gamma}_u > 0$, $\tilde{\gamma}_v > 0$. We have also assumed that both species compete for the same "vegetation" so that each of their effective reproduction rates are diminished together. To scale we now introduce:

$$
\begin{aligned}
u &\equiv \frac{\tilde{u}}{\tilde{u}_1},\ v \ \equiv \frac{\tilde{v}}{\tilde{v}_1},\\
u_o &\equiv \frac{\tilde{u}_o}{\tilde{u}_1},\ v_o \ \equiv \frac{\tilde{v}_o}{\tilde{v}_1}\\
\gamma_u &\equiv \tilde{\gamma}_u\frac{\tilde{v}_1}{\tilde{u}_1},\ \gamma_v \ \equiv \tilde{\gamma}_v\frac{\tilde{u}_1}{\tilde{v}_1}.
\end{aligned}
\tag{1.23}
$$

But to simplify we will consider only a special case in which the eight parameters are reduced to one by setting : **13**

$$
\begin{aligned}
&\beta_u = \beta_v = \gamma_u = \gamma_v = 1,\\
&u_o = v_o = 0,\\
&\ \lambda = 1 - \alpha_u = 1 - \alpha_v.
\end{aligned}
\tag{1.24}
$$

Note that here λ is similar to the parameter λ_1 of our previous example.

Now (1.24) reduces to the system :

$$\begin{pmatrix} u \\ v \end{pmatrix}_t = \begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} - (u^2 + v^2) \begin{pmatrix} u \\ v \end{pmatrix}. \tag{1.25}$$

Note that $\binom{u}{v} = \binom{0}{0}$, the trivial solution is valid for all λ. First we consider a small perturbation $\binom{\delta_u}{\delta_v}$ about the trivial solution $\binom{0}{0}$. This satisfies the linearized problem :

$$\begin{pmatrix} \delta_u \\ \delta_v \end{pmatrix}_t = A \begin{pmatrix} \delta_u \\ \delta_v \end{pmatrix}, \quad \text{where} \quad A = \begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix}. \tag{1.26}$$

The eigenvalues of A are given by :

$$n_\pm = \lambda \pm i. \tag{1.27}$$

These eigenvalues are distinct and hence the solutions of the linearized problem must have the form :

$$\begin{aligned} \delta u &= a_1 e^{n_+ t} + a_2 e^{n_- t} \\ \delta v &= b_1 e^{n_+ t} + b_2 e^{n_- t} \end{aligned} \tag{1.28}$$

For all $\lambda < 0$, we thus have that $\delta u, \delta v \to 0$ as $t \to \infty$. Hence the trivial state is a stable solution. On the other hand, if $\lambda > 0$, then δu, δv grow exponentially as $t \to \infty$ and the trivial state is unstable for $\lambda > 0$.
14 Observe that there is an exchange of stability as λ crosses the origin. For $\lambda = 0$ we get periodic solution of the linearized problem. Indeed we now show that the exact nonlinear solutions have similar feature, but that stable periodic solutions exist for all $\lambda > 0$.

Introduce polar co-ordinates in the (u, v) plane :

$$\begin{aligned} &u = \rho \cos\theta, v = \rho \sin\theta, \\ &u^2 + v^2 = \rho^2, \tan\theta = \frac{v}{u}. \end{aligned} \tag{1.29}$$

The system (1.25) becomes, with $\rho \geq 0$:

$$\begin{aligned} \dot\rho \cos\theta - \rho\dot\theta \sin\theta &= \lambda\rho \cos\theta - \rho \sin\theta - \rho^3 \cos\theta, \\ \dot\rho \sin\theta + \rho\dot\theta \cos\theta &= \rho \cos\theta + \lambda\rho \sin\theta - \rho^3 \cos\theta. \end{aligned} \tag{1.30}$$

Appropriate combinations of these equations yield :

(1.31) $$\dot{\rho} = \lambda\rho - \rho^3 = \rho(\lambda - \rho^2),$$

(1.32) $$\dot{\theta} = +1.$$

Thus $\theta(t) = \theta_0 + t$, where θ_0 is an arbitrary constant. From (1.31), for $\lambda > 0$, $\rho(t)$ is given by $\rho(t)(\dfrac{\tau + \rho(t)}{\tau - \rho(t)})^{1/2} = Ce^{\lambda t}$, where $\tau = \sqrt{\lambda}$ and C is an arbitrary constant. For $\lambda < 0, \rho(t)$ is given by $\dfrac{\rho(t)}{\sqrt{\rho^2(t) - \lambda}} = Ce^{\lambda t}$. We now examine two cases $\lambda > 0$ and $\lambda < 0$.

(i) $\lambda < 0$: Then $\dot{\rho} < 0$ for all t. This implies $\rho(t)$ decreases to 0 as t increases.

(ii) $\lambda > 0$: Here we have to consider 3 possibilities for the state at any 15
time t_o:

(iia) $\rho(t_o) < \rho_{cr} \equiv \sqrt{\lambda}$

Now $\rho(t) < 0$ and $\rho(t)$ increases towards ρ_{cr} as $t \uparrow \infty$

(iib) $\rho(t_o) > \rho_{cr}$

Now $\rho(t) < 0$ and $\rho(t)$ decreases towards ρ_{cr} as $t \uparrow \infty$

(iic) $\rho(t_o) = \rho_{cr}$

Now $\rho(t) \equiv 0$ and we have the periodic solution :

$$u = \rho_{cr} \cos(\theta_o + t), v = \rho_{cr} \sin(\theta_o + t).$$

These solutions are unique upto a phase, θ_o. The "solution surface" in the (u, v, λ) space contains the trivial branch and a paraboloid, which is sketched in Fig. 1.13

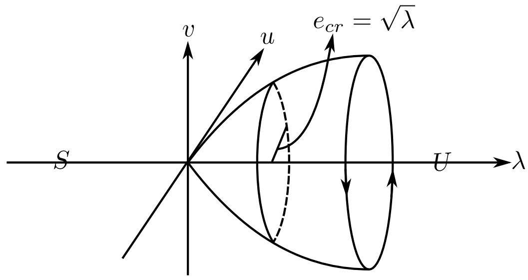


Figure 1.13:

This is a standard example of Hopf bifurcation - a periodic solution branches off from the trivial steady state as λ crosses a critical value,
16 in this case $\lambda = 0$. The important fact is that for the complex pair of eigenvalues $\rho_{\pm}(\lambda)$, the real part crosses zero and as we shall see in general, it is this behaviour which leads to Hopf bifurcation.

We have seen the change in the number of solutions across folds and two types of steady state bifurcations and one type of periodic bifurcation in our special population dynamics example. We shall now study much more general steady states and time dependant problems in which essentially these same phenomena occur. Indeed they are in a sense typical of the singular behaviour which occurs as parameters are varied in nonlinear problems. Of course even more varied behaviour is possible as we shall see (see [4], [13], [23], for example). One of the basic tools in studying singular behaviour is to reduce the abstract case to simple algebraic forms similar to those we have already studied (or generalizations).

Chapter 2

Local continuation methods

2.1 Introduction

We will use a local continuation method to get global solutions of the 17
general problem:

$$G(u, \lambda) = 0 \tag{2.1}$$

Suppose we are given a solution (u^o, λ^o) of (2.1). The idea of local continuation is to find a solution at $(\lambda^o + \delta\lambda)$ for a small perturbation $\delta\lambda$. Then perhaps, we can proceed step by step, to get a global solution. The basic tool for this study is the implicit function theorem.The continuation method may fail at some step because of the existence of singularities on the curve (for example folds or bifurcation points). Near these points there exist more than one solution and the implicit function theorem is *not* valid.

First we recall a basic theorem which is the main tool in proving the implicit function theorem.

2.2 Contraction Mapping Theorem

Let B be a Banach space and $F : B \to B$ satisfy:

(2.2)
(a) $F(\bar{\Omega}) \subset \bar{\Omega}$ *for some closed subset* $\bar{\Omega} \subset B$.
(b) $\|F(u) - F(v)\| \leq \theta\|u - v\|$ *for some* $\theta \in (0, 1)$ *and for all* $u, v \in \bar{\Omega}$.

18 *Then the equation*

$$u = F(u)$$

has one and only one solution $u^* \in \bar{\Omega}$. *This solution is the limit of any sequence* $\{u_U\}$, $U = 0, 1, 2, \ldots$ *generated by*

$$\text{(2.3)} \qquad \begin{array}{ll} \text{(a)} & u_o \in \bar{\Omega}, \quad \text{arbitrary;} \\ \text{(b)} & u_{U+1} = F(u_U), U = 0, 1, 2, \ldots\ldots \end{array}$$

The convergence is such that:

$$\text{(2.3)} \qquad \text{(c) } \|u^* - u_U\| \leq \frac{\theta^U}{1-\theta}\|u_o - F(u_0)\| \equiv \frac{\theta^U}{1-\theta}\|u_0 - u_1\|.$$

Proof. Well known □

We will state another lemma which shows how to find a set $\bar{\Omega}$, in which the conditions of theorem (2.2) hold.

2.4 Contracting Ball Lemma

Let $\rho > 0$, $\theta \in (0, 1)$ *be such that for some* $u_o \in B$, F *satisfies:*

$$\text{(2.4)} \qquad \begin{array}{ll} \text{(a)} & \|u_o - F(u_o)\| \leq (1-\theta)\rho. \\ \text{(b)} & \|F(u) - F(v)\| \leq \theta\|u - v\| \ \textit{for all} \ , u, v \in \bar{B}_\rho(u_o). \end{array}$$

where

$$\bar{B}_\rho(u_o) = \{u \in B : \|u - u_o\| \leq \rho\}.$$

Then the conditions (2.2) *of theorem* (2.2) *hold with* $\bar{\Omega} = \bar{B}_\rho(u_o)$.

Proof. For any $u \in \bar{B}_\rho(u_o)$,

$$\|F(u) - u_o\| \leq \|F(u) - F(u_o)\| + \|F(u_o) - u_o\| \leq \theta\rho + (1-\theta)\rho = \rho,$$

which proves (a) of (2.2). Part (b) of (2.2) is trivial □

2.5 Fundamental Theorem of Calculus

19

Let F be a differentiable operator defined on a Banach space B. Then for all $u, v \in B$:

$$F(u) - F(v) = \int_0^1 \frac{d}{dt} F(tu + (1-t)v)dt \tag{2.5}$$

We have on differentiating

$$\frac{d}{dt} F(tu + (1-t)v) = F_u(tu + (1-t)v)(u - v),$$

and if we see

$$\tilde{F}_u(u, v) \equiv \int_0^1 F_u(tu + (1-t)v)\, dt$$

then we obtain the mean value formula :

$$F(u) - F(v) = \tilde{F}_u(u, v)(u - v). \tag{2.6}$$

This is valid provided F is differentiable on the "line" joining u and v.

Now we will state and prove:

2.7 Implicit Function Theorem

Let $G : B_1 \times B_2 \to B_1$ satisfy for some $\rho_1 > 0$, $\rho_2 > 0$, sufficiently small, (B_1 is a Banach space and B_2 is the parameter space, either it 20
is Euclidean space or more generally it can be a Banach space) the following conditions:

(2.7)

(a) $G(u^0, \lambda^0) = 0$ for some $u^0 \in B_1, \lambda^0 \in B_2$.

(b) $G_u^0 \equiv G_u(u^0, \lambda^0)$ is nonsingular with a bounded inverse, i.e. : $\|(G_u^0)^{-1}\| \leq M_0$, for some constant M_0.

(c) $G(u, \lambda)$ and $G_u(u, \lambda)$ are continuous on $B_{\rho_1}(u^0) \times B_{\rho_2}(\lambda^0)$.

Then for all $\lambda \in B_{\rho_2}(\lambda^0)$, *there exists* $u(\lambda) \in B_{\rho_1}(u^0)$ *such that:*

$$(2.7)\quad \begin{array}{ll} \text{(d)} & u(\lambda^0) = u^0. \\ \text{(e)} & \text{Existence}: G(u(\lambda), \lambda) = 0. \\ \text{(f)} & \text{Uniqueness : For } \lambda \in B_{\rho_2}(\lambda^0) \text{ there is no solution of} \\ & G(u, \lambda) = 0 \text{ in } B_{\rho_1}(u^0) \text{ other than } u(\lambda). \\ \text{(g)} & \text{Continuous dependence : } u(\lambda) \text{ is continuous on} \\ & B_{\rho_2}(\lambda^0) \text{ and has, upto a factor, the same modulus} \\ & \text{of continuity with respect to } \lambda, \text{ as } G(u, \lambda). \end{array}$$

Proof. Since G_u^0 is nonsingular, it follows that:

$$G_u^0 u = G_u^0 u - G(u, \lambda) \text{ if and only if } G(u, \lambda) = 0.$$

□

Hence $G(u, \lambda) = 0$ is equivalent to :

$$(2.8)\qquad u = (G_u^0 u)^{-1}[G_u^0 u - G(u, \lambda)] \equiv F(u, \lambda)$$

Thus the problem of solving $G(u, \lambda) = 0$ reduces to finding the fixed point of the operator $F(u, \lambda)$ for a given λ. Note that u^0 is a fixed point for $\lambda = \lambda^0$.

Next, we will check the conditions (a) and (b) of the contracting
21 ball lemma (2.4) so that we can apply the contraction mapping theorem (2.2). Take any $\lambda \in B_{p_2}(\lambda^0)$ and use $F(u^0, \lambda^0) = u^0$, to get:

$$\begin{aligned} \|u^0 - F(u^0, \lambda)\| &= \|F(u^0, \lambda^0) - F(u^0, \lambda)\| \\ &\leq M_0 \|G(u^0, \lambda) - G(u^0, \lambda^0)\| \\ &\leq M_0 \omega_0(|\lambda^0 - \lambda|). \end{aligned}$$

Thus we have

$$(2.9)\qquad \|u^0 - F(u^0, \lambda)\| \leq M_0 \omega_0(p_2).$$

Here we have introduced the modulus of continuity ω_0 defined as:

$$(2.10)\qquad \omega_0(\rho) = \sup_{\left\{\begin{array}{c} \|\lambda - \tilde{\lambda}\| \leq \rho, \\ \lambda, \tilde{\lambda} \in B_{\rho_2}(\lambda^0), \\ u, v \in B_{\rho_1}(u^0). \end{array}\right\}} \|G(u, \lambda) - G(v, \tilde{\lambda})\|.$$

ω_0 is nonnegative and nondecreasing and also $\omega_0(\rho) \to 0$ as $\rho \to 0$ by (2.7c). Next fo $u, v \in B_{\rho_1}(u^0)$, we have:

$$F(u, \lambda) - F(v, \lambda) = (G_u^0)^{-1}[G_u^0(u - v) - (G(u, \lambda) - G(v, \lambda))],$$

$$= (G_u^0)^{-1}[G_u^0 - \tilde{G}_u(u, v, \lambda)](u - v), \tag{2.11a}$$

where,

$$\tilde{G}_u(u, v, \lambda) = \int_0^1 G_u(tu + (1 - t)v, \lambda)dt.$$

Thus we have:

$$G_u^0 - \tilde{G}_u(u, v, \lambda) = \int_0^1 [G_u - G_u(tu + (1 - t)v, \lambda)]dt,$$

$$= \int_0^1 [G_u(u^0, \lambda^0) - [G_u(u^0, \lambda) + [G_u(u^0, \lambda)$$

$$- G_u(tu + (1 - t)v, \lambda)]dt. \tag{2.11b}$$

But 22

$$||G_u(u^0, \lambda^0 - G_u(u^0, \lambda)|| \leq \omega_2(\rho_2), \tag{2.11c}$$

and

$$||G_u(u^0, \lambda - G_u(tu + (1 - t)v, \lambda)|| \leq \omega_1(\rho_1), \tag{2.11d}$$

where we have introduced the moduli of continuity ω_1 and ω_2 as :

$$\omega_1(\rho_1) = \sup_{\begin{Bmatrix} ||\lambda \in B_{\rho_2}(\lambda^0) \\ w, \in B_{\rho_1}(u^0), \end{Bmatrix}} ||G(u^0, \lambda) - G_u(w, \lambda)||.$$

and

$$\omega_2(\rho_2) = \sup_{\begin{Bmatrix} u, v \in B_{\rho_1}(u^0), \\ \lambda, \overline{\lambda} \in B_{\rho_2}(\lambda^0), \\ |\lambda - \overline{\lambda}| \leq \rho_2. \end{Bmatrix}} ||G_u(u, \lambda) - G_u(v, \overline{\lambda})||.$$

Again note that ω_1 and ω_2 are nonnegative and nondecreasing. Also $\omega_1(\rho_1) \to 0$ as $\rho_1 \to 0$ and $\omega_2(\rho_2) \to 0$ as $\rho_2 \to 0$ by (2.11c). Now using the results (2.11) it follows that :

$$\|F(u,\lambda) - F(v,\lambda)\| \leq M_0(\omega_1(\rho_1) + \omega_2(\rho_2))\|u - v\|. \tag{2.12}$$

For any fixed $\theta \in (0,1)$ we can choose ρ_1 and ρ_2 small enough to make :

$$M_0(\omega_1(\rho_1) + \omega_2(\rho_2)) \leq \theta < 1. \tag{2.13a}$$

Now fix ρ_1 and reduce ρ_2 (if necessary) so that:

$$M_0\omega_1(\rho_2) \leq (1-\theta)\rho_1. \tag{2.13b}$$

Then (2.9) and (2.12) together show that the conditions of lemma
23 (2.4) hold for some ρ_1 and ρ_2. Hence we can apply the theorem (2.2) which prove the results (2.7d,e,f).

Now we prove (2.7g). For $\lambda, \bar{\lambda} \in B_{\rho_2}(\lambda^0)$, we have:

$$\begin{aligned}
\|u(\lambda) - u(\bar{\lambda})\| &= \|F(u(\lambda),\lambda) - u(\bar{\lambda})\| \\
&\leq \|F(u(\lambda),\lambda) - \|F(u(\bar{\lambda}),\lambda)\| \\
&\qquad + \|F(u(\bar{\lambda}),\lambda) - F(u(\bar{\lambda}),\bar{\lambda})\|, \\
&\leq \theta\|u(\lambda) - u(\bar{\lambda})\| + M_0\omega_0(|\lambda - \bar{\lambda}|),
\end{aligned}$$

by (2.10) and (2.12). Thus we get :

$$\|u(\lambda) - u(\bar{\lambda})\| \leq \frac{M_0}{1-\theta}\omega_0(|\lambda - \bar{\lambda}|)$$

This shows that $u(\lambda)$ is continuous in $B_{\rho_2}(\lambda^0)$ and has the same modulus of continuity with respect to λ, as $G(u,\lambda)$, upto a scalar multiple $M_0/(1-\theta)$.

2.14 Step Length Bound

The Implicit function theorem simply gives conditions under which one can solve the equation $G(u,\lambda) = 0$ in a certain neighbourhood of u^0, λ^0.

In other words, if (u^0, λ^0) is a solution, we can solve $G(u, \lambda) = 0$, for each $\lambda \in B_{\rho_2}(\lambda^0)$, for some $\rho_2 > 0$. It is interesting to know (especially for applying continuation method) how large the neighbourhood $B_{\rho_2}(\lambda^0)$ may be. In fact actually we want to find the maximum ρ_2 for which (2.13a,b) hold.

We assume G and G_u satisfy Lipschitz conditions. Thus

$$\omega_U(\rho) \equiv K_U \rho, U = 0, 1, 2. \tag{2.15}$$

To get an idea of the magnitude of these Lipschitz constants we note that for smooth G: **24**

$$K_0 \approx \|G_\lambda\|, K_1 \approx \|G_{uu}\|, K_2 \approx \|G_{u\lambda}\|.$$

Using (2.15) in (2.13b) gives :

$$M_0 K_0 \rho_2 \leq (1 - \theta)\rho_1.$$

Thus if we take :

$$\rho_2 = \frac{(1 - \theta)\rho_1}{M_0 K_0}, \tag{2.16}$$

then (2.13b) holds. In addition we require:

$$M_0(K_1 \rho_1 + K_2 \rho_2) \leq \theta < 1.$$

Thus if we take

$$\rho_1 = \frac{\theta - M_0 K_2 \rho_2}{M_0 K_1},$$

then (2.16) yields,

$$\rho_2(\theta) = \frac{(1 - \theta)\theta}{A - B\theta},$$

where,

$$A = M_0^2 K_0 K_1 + B, \quad B = M_0 K_2.$$

We want to maximize ρ_2 as a function of θ over $(0, 1)$. The following properties hold (see Fig. 2.1):

(i) $\rho_2(0) = \rho_2(1) = 0$.

(ii) $\rho_2(0) > 0$ for $o < \theta < 1$ and $\rho_2(0) < 0$ for $\theta < 0$.

(iii) $\rho_2(0) \to -\infty$ as $\theta \uparrow A/B$.

(iv) for $\theta > A/B$, $\rho_2(0) > 0$ and $\rho_2(0) \to \infty$ as $\theta \downarrow A/B$ and also as $\theta \to \infty$

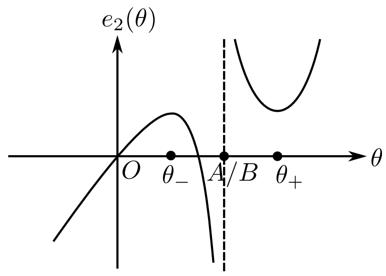


Figure 2.1:

25

Thus $\rho_2(\theta)$ has a maximum at $\theta = \theta_- \varepsilon(0.1)$ and a minimum at $\theta_+ > A/B$ which are easily determined to be :

$$\theta_\pm = \frac{A}{B}\{1 \pm (1 - B/A)^{1/2}\}$$

The maximum is thus:

$$\rho_{2_{\max}} = \rho_2(\theta_-) = \frac{[1 - \frac{A}{B}(1 - (1 - B/A)^{1/2})][1 - (1 - B/A)^{1/2}]}{B(1 - B/A)^{1/2}}$$

We write:

$$\frac{A}{B} = 1 + \frac{1}{\varepsilon}, \quad \text{where} \quad \frac{1}{\epsilon} = \frac{M_0 K_0 K_1}{K_2}$$

Then if $\in$ is small, say $0 <\in<< 1$, we have the expansion:

(2.17a) $$\theta_- = \frac{1}{2} + \frac{\varepsilon}{8} + \cdots\cdots ,$$

(2.17b) $$\rho_{2_{\max}} = \frac{1 - (\varepsilon^2/16) + \cdots\cdots}{(4M_0^2 K_0 K_1 + 2M_0 K_2) - (M_0 K_2/2)\varepsilon + \cdots\cdots}$$

26 We recall that $M_0 = ||(G_u^0)^{-1}||$. Then if G_u^0 becomes singular during continuation, M_0 becomes infinite and the continuation procedure explained here must fail. We notice this phenomenon first by the required step sizes getting smaller. Also note that small steps result from larger K_0, K_1 and K_2 as well.

In the implicit function theorem, we assumed that there is a solution (u^0, λ^0). Next we prove a similar theorem in which we assume only that $||G(u^0, \lambda^0)||$ is small.

2.18 Approximate Implicit Function Theorem

Let $G : B_1 \times B_2 \to B_1$ *satisfy for some* $\delta > 0$, $\rho_1 > 0$, $\rho_2 > 0$ *sufficiently small*:

(2.18)
(a) $||G(u^0, \lambda^0)|| \leq \delta$ for some $u^0 \in B_1, \lambda^0 \in B_2$.
(b) G_u^0 is invertible and $||(G_u^0)|| \leq M_0$ for some constant M_0.
(c) G and G_u are continuous on $B_{\rho_1}(u^0) \times B_{\rho_2}(\lambda^0)$.

Then there exist $u(\lambda)$, *for all* $\lambda \in B_{\rho_1}(u^0)$, *such that*:

(d) $G(u(\lambda), \lambda) = 0$ *and* $u(\lambda) \in (U^0)B_{\rho_1}(u^0)$ [Existence].
(e) For a given $\lambda \in B_{\rho_2}(\lambda^0)$ there is no solution of $G(u, \lambda) = 0$ in $B_{\rho_1}(u^0)$ other than $u(\lambda)$ [uniqueness].
(f) $u(\lambda)$ is continuous on $B_{\rho_2}(\lambda^0)$ and has the same modulus of continuity as $G_u(u, \lambda)$ [continuity].

Proof. We can prove the result in the same way as we have proved the implicit function theorem. Since G_u^0 is nonsingular, the problem $G(u, \lambda) = 0$ reduces to the fixed point problem

$$u = F(u, \lambda).$$

where 27

$$F(u, \lambda) = (G_u^0)^{-1}[G_u^0 u - G(u, \lambda)].$$

New for $\lambda \in B_{\rho_2}(\lambda^0)$, we have:

$$||u^0 - F(u^0, \lambda)|| \leq ||u^0 - F(u^0, \lambda^0)|| + ||F(u^0, \lambda^0) - F(u^0, \lambda)||,$$

and so:

$$\|u^0 - F(u^0, \lambda)\| \leq M_0(\delta + \omega_0(\rho_2)). \tag{2.19a}$$

Here ω_0 is given by (2.10)(2.10). Next, for $u, v \in B_{\rho_1}(u^0)$:

$$F(u, \lambda) - F(v, \lambda) = (G_u^0)^{-1}[G_u^0 - \widetilde{G}_u(u, v, \lambda)](u - v),$$

where $\widetilde{G}_u(u, v, \lambda)$ is defined as in (2.11a). Again using (2.11b,c,d) we obtain:

$$\|F(u, \lambda) - F(v, \lambda)\| \leq M_0(\omega_1(\rho_1) + \omega_2(\rho_2))\|u - v\|. \tag{2.19b}$$

Now for any $\theta \in (0, 1)$, choose ρ_1 and ρ_2, small enough, such that :

$$M_0(\omega_1(\rho_1) + (\omega_2(\rho_2)) \leq \theta < 1, \tag{2.19c}$$

and

$$M_0(\delta + \omega_0(\rho_2)) \leq (1 - \theta)\rho_1. \tag{2.19d}$$

Now we can apply the contraction mapping theorem to obtain the results □

2.20 Step Length Bound

We can also obtain step length bounds as before. Thus we assume (2.15) to hold and use it in (2.19d) to get

$$M_0(\delta + K_0\rho_2) \leq (1 - \theta)\rho_1$$

28

If we take :

$$\rho_2 = \frac{(1 - \theta)\rho_1 - M_0\delta}{M_0K_0} \tag{2.21a}$$

then (2.19d) holds. Again from (2.19c), using (2.15) we require:

$$M_0(K_1\rho_1 + K_2\rho_2) = \theta < 1. \tag{2.21b}$$

Substituting the value of ρ_1 from (2.21a,b) we get:

$$\widetilde{\rho}_2(\theta) = \frac{(1-\theta)\theta - C}{A - B\theta},$$

where,

$$A = M_0^2 K_0 K_1 + B = M_0 K_2, C = M_0^2 K_1 \delta.$$

$\widetilde{\rho}_2(\theta)$ is sketched in fig. 2.2. We see that $\widetilde{\rho}_2(\theta)$ has a maximum at $\theta_- \in (0,1)$ and a minimum at $\theta_+ > A/B$ which are given by :

$$\theta_\pm = \frac{A}{B}\{1 \pm (1 - B/A - (B^2/A^2)C)^{1/2}\}.$$

The maximum is thus:

$$\widetilde{\rho}_{2\,\max} = \widetilde{\rho}_2(\theta_-).$$

We have $C > 0$ and $A - B\theta \geq 0$ if $\theta \leq A/B$. Hence $\widetilde{\rho}_2$ max in this theorem is less than that in the original implicit function theorem. We write:

$$\frac{A}{B} = 1 + \frac{1}{\varepsilon}, \quad \text{where} \quad \varepsilon = \frac{K_2}{M_0 K_0 K_1},$$

and

$$C = M_0^2 K_1 \delta = D \cdot \frac{1}{\varepsilon},$$

where $D = \dfrac{M_0 K_0 \delta}{K_0}$. If ε is small say, $0 < \varepsilon << 1$, then we have the 29
expansions:

$$\theta_- = \frac{1}{2}(1 + D) + \frac{\varepsilon}{8}(1 - D)^2 + \cdots\cdots,$$

and

$$\widetilde{\rho}_{2\,\max} = \frac{(1-D)^2 + \frac{D}{2}(1-D)^2\varepsilon - \frac{1}{8}(1-D)^4\varepsilon^2 + \cdots\cdots}{(4M_0^2K_0K_1 + 2M_0K_2 - 2M_0K_2D) - \frac{M_0K_2}{2}(1-D)^2\varepsilon + \cdots\cdots}$$

Again as before, we can see that $\widetilde{\rho}_{2\,\max}$ is small if M_0 is large. Note that $\widetilde{\rho}_{2\,\max}$ is always less than $\rho_{2\,\max}$ in the implicit function theorem.

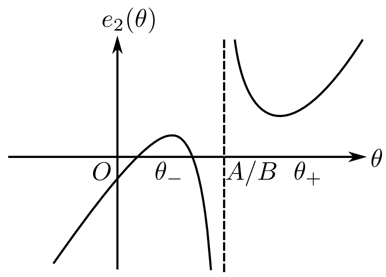


Figure 2.2:

2.22 Other Local Rootfinders

Let $g : \mathbb{R} \to \mathbb{R}$ be a smooth function. We will examine various methods for finding the roots of $g(x) = 0$. In particular, we consider the chord method, Newton's method etc. In the chord method, we choose an initial estimate x_0 of the solution and a slope a. Then the line with slope 'a' through $(x_0, g(x_0))$ will intersect the x-axis at x_1. This is the next approximation to the solution. We continue the process with x_1 replacing x_0, using the same slope 'a' (See Fig. 2.3). More precisely,

$$a(x_1 - x_0) = -g(x_0)$$

and in general

$$a(x_{k+1} - x_k) = -g(x_k)$$

The sequences $\{x_k\}$ thus generated, converges to a zero $g(x)$ if,

$$\max_{x \in H} \left| 1 - \frac{g'(x)}{a} \right| < 1,$$

where H is an interval containing both the initial value and a root of the equation.

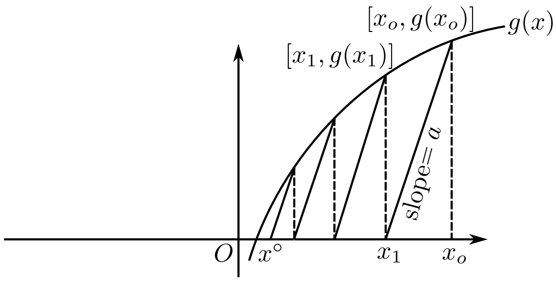


Figure 2.3:

In the higher dimensional case, we generalize the above to get the sequence $\{u_U\}$, form:

$$A[u_{U+1} - u_U] = -G(u_U),$$

where $G : \mathbb{R}^n \to \mathbb{R}^n$ and A is an $n \times n$ matrix. In particular if we use:

$$a = g'(x_0) \text{ and } A = G_u(u_0),$$

we get the special Newton method. 31

In Newton's method we vary the slope at each iterate and take :

$$\begin{aligned} a_U &= g'(x_U) \\ A_U &= G_U(u_U). \end{aligned}$$

The Newton-Kantorovich theorem gives conditions under which the Newton iterate converge.

2.23 Newton-Kantorovich Theorem

Let $G : B \to B$ (B is a Banach space) *satisfy for some* $u^0 \in B$ *and* $\rho_0^- > 0$*:*

(2.23)

(a) $G_u(u^0)$ is nonsingular with $\|(G_u^0)^{-1}\| \leq \beta$;

(b) $\|G_u^- 1(u^0)G(u^0)\| \leq \alpha$;

(c) $\|G_u(u) - G_u(v)\| \leq \gamma\|u - v\|$, in for all $u, v \in B_{\rho_0^-}(u^0) \setminus \{(u^0)\}$

(d) $\alpha\beta\gamma \leq \dfrac{1}{2}$ and $\rho_0^- \leq \dfrac{1 - \sqrt{1 - 2\alpha\beta\gamma}}{\beta\gamma}$.

Then the Newton iterates $\{u_U\}$ *defined by Newton's method*:

$$G_u(u_U)[u_{U+1} - u_U] = -G(u_U), U = 0, 1, 2, \ldots\ldots$$

with $u_0 = u^0$ *satisfy*:

(2.23)
(e) $u_U \in B_{\rho_0^-}(u^0)$.
(f) $\{u_U\}$ converges to u^*, a root of $G(u) = 0$ in $B_{\rho_0^+}(u^0)$.

32 *In addition* u^* *is the is the unique root of* G *in* $B_{\rho_0^+}(u^0)$ *where*

$$\rho_0^+ = (1 + (1 - 2\alpha\beta\gamma)^{1/2})/\beta\gamma.$$

Proof. See the reference [16] □

Now we prove a theorem, assuming the existence of a root, to show the basic idea of how the above method works.

2.24 Newton Convergence Theorem

Let $G : B \to B$ *and* $G(u) = 0$ *have a root* $u = u^*$. *For some* $\rho_* > 0$ *let* G *satisfy*:

(2.24)
(a) $\|G_u^{-1}(u^*)\| \leq \beta$.
(b) $\|G_u(u) - G_u(v)\| \leq \gamma\|u - v\|$, *for all* $u, v \in B_{\rho_*}(u^*)$.
(c) $\rho_*\beta\gamma < \dfrac{2}{3}$.

Then for every $u_0 \in B_{\rho_*}(u^*)$ *the Newton iterates satisfy*:

(2.24)
(d) $u_U \in B_{\rho_*}(u^*)$;
(e) $\|u_{U+1} - u^*\| \leq a\|u_U - u^*\|^2$;

where,

$$a \equiv \frac{B\gamma}{2(1 - \rho_*\beta\gamma)} < \frac{1}{\rho_*}.$$

Proof. For any $u \in B_{\rho_*}(u^*)$ we have the identity :

$$G_u(u) = G_u(u^*)\{I + G_u^{-1}(u^*)[G_u(u) - G_u(u^*)]\}.$$

Then (2.24a,b,c) imply that:

$$\|G_u^{-1}(u^*)[G_u(u) - G_u(u^*)]\| \leq \rho_* \beta\gamma < \frac{2}{3}.$$

Hence, by the Banach lemma, $\{I + G_u^{-1}(u^*)[G_u(u) - G_u(u^*)]\}$ is invertible and so is $G_u(u)$. From the same lemma we get the estimate : **33**

$$\|G_u^{-1}(u)\| \leq \frac{\beta}{1 - \rho_* \beta\gamma}. \tag{2.25}$$

Now we will prove by induction that $u_U \in B_{\rho_*}(u^*)$, $u = 0, 1, 2, \ldots$. Suppose $u_U \in B_{\rho_*}(u^*)$. Then

$$u_{U+1} - u^* = (u_U - u^*) - G_u^{-1}(u_U)[G(u_U) - G(u^*)].$$

By the mean value result (2.6), we have :

$$G(u_U) - G(u^*) = \widetilde{G}_u(u_U, u^*)(u_U - u^*),$$

and so:

$$u_{U+1} - u^* = G_u^{-1}(u_U)[G_u(u_U) - \widetilde{G}_u(u_U, u^*)](u_U - u^*).$$

But

$$\begin{aligned}
\|G_u(u_U) - \widetilde{G}_u(u_U, u^*)\| &\leq \int_0^1 \|G_u(u_U) - G_u(tu_U + (1-t)u^*)\|\, dt \\
&\leq \gamma \int_0^1 \|(1-t)(u_U - u^*)\|\, dt \\
&= \frac{\gamma}{2}\|u_U - u^*\|.
\end{aligned}$$

Hence

$$
(2.26) \qquad
\begin{aligned}
\|u_{U+1} - u^*\| &\leq \|G_u^{-1}(u_U)\|.\frac{\gamma}{2}\|u_U - u^*\|^2,\\
&\leq \frac{\beta}{1-\rho_*\beta\gamma}\frac{\gamma}{2}\|u_U - u^*\|^2,
\end{aligned}
$$

by (2.25). Note that

$$\frac{\beta\gamma}{2(1-\rho_*\beta\gamma)} \leq \frac{1}{\rho_*} \text{and } \|u_U - u^*\| \leq \rho_*.$$

34 Hence $u_{U+1} \in B_{\rho_*}(u^*)$ and by induction (d) follows. Then (e) follows from (2.26). □

Note. The convergence here, is quadratic. Thus if

$$a\|u_U - u^*\| \leq 10^{-p_U}$$

for some positive p_U, then

$$a\|u_{U+r} - u^*\| \leq 10^{-2^u \rho_U}\|u_U - u^*\| \text{ for any } r = 0, 1, 2, \ldots\ldots.$$

The choice of the initial iterate, u_0, is important in using Newton's method. It is not uncommon to spend 90% or more of the effort in finding a food approximate value of the root. Our study will show many ways in which such difficulties can be overcome. Another problem with Newton's method can be the time it takes to solve the linear system for the new iterate. This occurs, for example, if $B = \mathbb{R}^N$ for very large N, say $N \approx 10^3$ or larger (when approximating nonlinear P.D.E. problems). The so called **quasi-Newton methods** are designed to avoid the linear system problem by some device (for example updating secant method, which we are going to describe in the next section).

2.27 Predictor-Solver Methods

Let $G : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$. At the root (u^0, λ^0) of $G(u, \lambda) = 0$, let G_u^0 be nonsingular. Then the implicit function theorem shows the existence of

the unique branch of solutions $u = u(\lambda)$ in a neighbourhood of λ^0. We will briefly describe various **predictor-sovler methods** for determining $u(\lambda)$. These proceed by using the solution (u^0, λ). to construct a better approximation $u_0(\lambda)$ to the actual solution $u(\lambda)$. This is the predictor.
After obtaining the initial approximation, we apply an iteration scheme **35**
for which the sequence of iterates converges to the solution $u(\lambda)$. This is the solver.

Various Predictors

(i) **Trivial Predictor** : Here we take the initial approximation as (see (Fig. 2.4a)):

$$u_0(\lambda) = u^0 = u(\lambda^0)$$

i.e. the initial guess at λ is equal to the solution at λ^0. The error estimate is given by :

$$\begin{aligned} \|u(\lambda) - u_0(\lambda)\| &= \|u(\lambda) - u(\lambda^0)\|, \\ &\leq \frac{M_0}{1-\theta}\omega_0(|\lambda - \lambda^0|). \end{aligned}$$

Here $u(\lambda)$ is the actual solution and ω_0 is given by (2.10). If $G(u, \lambda)$ is Lipshitz continuous in u then

$$\|u(\lambda) - u_0(\lambda)\| \leq \frac{m_0}{1-\theta} K_0 |\lambda - \lambda^0|,$$

where K_0 is the corresponding Lipshitz constant.

(ii) **Secant-Predictor**: Here we assume that there are two known solutions (u^0, λ^0) and (u^1, λ^1). Then consider the line segment joining (u^0, λ^0) and (u^1, λ^1) in the (u, λ) space. Take $u_0(\lambda)$ as the point on this line with given λ-coordinate value λ. (See Fig. 2.4b) i.e.

$$u_0(\lambda) = u^1 + (\lambda - \lambda^1)\frac{u^1 - u^0}{\lambda^1 - \lambda^0}.$$

Then **36**

$$u(\lambda) - u_0(\lambda) = [u(\lambda) - u(\lambda^1)] - \frac{\lambda - \lambda^1}{\lambda^1 - \lambda^0}[u(\lambda^1) - u(\lambda^0)],$$
$$= \tilde{u}_\lambda(\lambda, \lambda^1)(\lambda - \lambda^1) - (\lambda - \lambda^1)\tilde{u}_\lambda(\lambda^1, \lambda^0).$$

By the mean value formula (2.6):

$$u(\lambda) - u_0(\lambda) = \int_0^1 dt[u_\lambda + (1-t)\lambda^1) - u_\lambda(t\lambda^1 + (1-t)\lambda^0)](\lambda - \lambda^1),$$
$$= \int_0^1 dt \int_0^1 \{u_{\lambda\lambda}(s(t\lambda + (1-t)\lambda^1) + (1-s)(t\lambda^1 + (1-t)\lambda^0))$$
$$(t(\lambda - \lambda^1) + (1-t)(\lambda^1 - \lambda^0))(\lambda - \lambda^1)\}ds.$$

Again we have use (2.6). Thus we get:

$$\|u(\lambda) - u_0(\lambda)\| \leq \frac{1}{2}K_2[(\lambda - \lambda^1) + (\lambda^1 - \lambda^0)](\lambda - \lambda^1),$$
$$= \frac{1}{2}K_2(\lambda - \lambda^0)(\lambda - \lambda^1), \qquad \lambda^0 < \lambda^1 < \lambda,$$

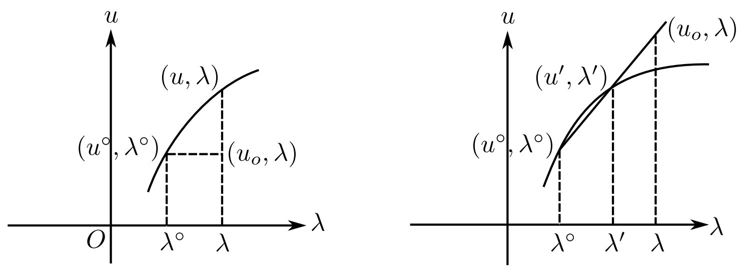
where,

$$K_2 = K_2(\lambda) \equiv \max_{\tilde{\lambda} \in [\lambda^0, \lambda]} \|u_{\lambda\lambda}(\tilde{\lambda})\|.$$

Since $\lambda^0 < \lambda^1 < \lambda$, this is an extrapolation.

In the interpolation case, that is , $\lambda^0 < \lambda < \lambda^1$, we have :

$$\|u(\lambda) - u_0(\lambda)\| \leq \frac{1}{2}K_2(\lambda - \lambda^0)(\lambda^1 - \lambda).$$



(a) (b)

Figure 2.4:

(iii) **Higher-order Predictor (Lagrange)**: If we know more than two solutions, then we can take higher order approximations to the next solution by using order interpolation formulae. **37**

(iv) **Tangent Predictor (Euler method)**: In the secant predictor method we assumed the existence of two solutions and then used the line segment joining these two solutions. Now we will consider only the solution at a single point and use the tangent to the solution curve at that point. (see Fig. 2.4). From the implicit function theorem we have:

$$G(u(\lambda), \lambda) = 0 \text{ for all } \lambda \in B_{\rho_2}(\lambda^0).$$

Differentiating with respect to λ, we get:

$$G_u(u(\lambda), \lambda)\dot{u}(\lambda) = -G_\lambda(u(\lambda), (\lambda),$$

and thus

$$\dot{u}^0 = \dot{u}(\lambda^0) = -(G_u^0)^{-1}G_\lambda^0.$$

Then we take the approximation as :

$$u_0(\lambda) = u^0 + (\lambda - \lambda^0)\dot{u}^0.$$

The error is given by:

$$u(\lambda) - u_0(\lambda) = \frac{1}{2}\ddot{u}^0(\lambda - \lambda^0)^2 + 0((\lambda - \lambda_0)^3).$$

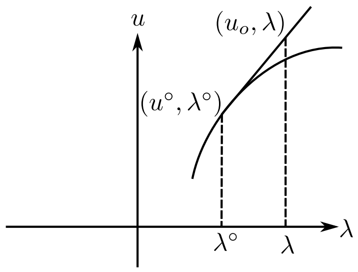


Figure 2.4: (c)

38 **Solvers**: Now using predictor $u_0(\lambda)$, our aim is to construct an iteration scheme of the form

$$A_U(u_{U+1} - u_U) = -G(u_U, \lambda), U = 0, 1, 2, \cdots\cdots ,$$

where $\{A_U\}$ are suitable matrices which assure the convergence of $\{u_U\}$ to a root

(i) **Special Newton Method**

In the special special Newton method, A_U is given by the constant operator G_u^0.

(ii) **Newton's Method**

In Newton's method, $A_U \equiv G_u^U = G_u(u_U, \lambda)$. One of the important advantages of Newton's method is that it can have quadratic convergence i.e. the sequence (u_U) satisfies :

$$\|u_{U+1} - u^*\| \leq \beta \|u_U - u^*\|^2,$$

for some constant β, where u^* is the actual solution.

(iii) **Updating-Secant Method**

As mentioned above a major disadvantage of Newton's method is that we need to solve a linear system with the coefficient matrix G_u^U step $U = 0, 1, 2, \ldots\ldots.$. This costs $0(N^3)$ operations at each step, where N is the dimension of the space.

Now we will introduce **Updating-Secant Method.** The idea of this
39 iteration scheme is to obtain a suitable approximation A_U to G_u^U so that the system at $(U+1)^{st}$ stage can easily be solved if we can solve the system at U^{th} stage. Further A_{U+1} is to satisfy a relation also satisfied by G_u^{U+1}. We will take A_{U+1} in the form:

$$A_{U+1} = A_U + C_U R_U^T, \tag{2.28}$$

where C_U and R_U are column vectors of dimension N and A_U is an $N\times N$ matrix. In particular we choose:

$$C_U \equiv \frac{(y_U - A_U S_U)}{< S_U, S_U >},$$

and

$$R_U^T \equiv S_U^T,$$

where

$$y_U = G(x_{U+1}) - G(x_U) \text{ and } S_U = x_{U+1} - x_U.$$

Now we have to solve the linear system $A_U S_U = -G(x_U)$ at each step. But this can be easily achieved using the Shermann-Morrison formula [for more details see [12]).

2.29 Lemma

Let A be an $N\times N$ invertible matrix and $u, v \in \mathbb{R}^N$. Then $A + uv^T$ is nonsingular if and only if $\sigma = 1 + (v, A^{-1}u) \neq 0$, and then the inverse of $(A + uv^T)$ is given by

$$(A + uv^T)^{-1} = A^{-1} - \frac{1}{\sigma} A^{-1} uv^T A^{-1} \tag{2.29a}$$

Using this result and (2.27), the inverse of A_{U+1} is given by

$$A_{U+1}^{-1} = A_U^{-1} + \frac{(s_U - A_U^{-1} y_U) s_U^T A_U^{-1}}{(s_U, A_U^{-1} y_u)} \tag{2.29b}$$

40

This method uses only $0(N^2)$ arithmetic operations. Naturally, we can expect a loss in the rate of convergence compared to Newton's method here we get only *superlinear convergence* instead of quadratic convergence. i.e.

$$\|u_{U+1} - u^*\| \leq \alpha_U \|u_U - u^*\|$$

where

$$\alpha_U \rightarrow 0 \text{ as } U \rightarrow \infty$$

This is generally known as a Quasi-Newton method (see ref: [9]).

Chapter 3

Global continuation theory

3.1 Introduction

41

In this section we will discuss some preliminary aspects which are useful in the study of **global continuation results**. We will start with some definitions regarding the function G from $\mathbb{R}^M$ into $\mathbb{R}^N$. The Jacobian of $G(x)$ is:

$$G'(x) = \frac{\partial G}{\partial x}(x) = \left(\frac{g_i}{x_j}(x_1, \ldots, x_M)\right)_{\substack{1\le j\le M\\ 1\le i\le N}},$$

where $x = (x_1, \ldots x_M)T$ and $G = (g_1, \ldots g_N)^T \cdot G'(x)$ is an $N \times M$ matrix and its rank must be less than or equal ti min (N, M).

Definitions. (1) Let G be continuously differentiable. Let

$$C = \{x \in \mathbb{R}^M : \text{ Rank } G'(x) < \min(N, M)\}.$$

The points in C are called critical points of G.

(2) The regular points are the points in the complement of C, denoted by $\mathbb{R}^M \setminus C$.

Note: The domain $\mathbb{R}^M$ can always be replace by the closure of an open set in $\mathbb{R}^M$.

(3) The image of the set C under G, i.e. $G(C)$ which is a subset of $\mathbb{R}^N$ 42
is the set of critical values of G.

(4) The complement of $G(C)$ in $\mathbb{R}^N$ is the set of regular values of G. i.e. $\mathbb{R}^N \setminus G(C)$.

Now we will state and prove *Sard's Theorem* (See [26]).

3.2 Sard's Theorem

Let $G \in C^1(\mathbb{R}^M) \cap C^{M-N+1}(\mathbb{R}^M)$. Then $G(C)$ has $\mathbb{R}^N$ - measure zero.

Proof. We will prove the result for $M = N$. The case $M < N$ is trivial. In this case there exist a nontrivial null vector of G'. This is precisely the idea we are going to use in the case $M = N$. For $M > N$ see [1].

So assume $M = N$. Let Q_0 be an arbitrary cube in $\mathbb{R}^N$ of side L. (Therefore vol $(Q_0) = L^N$). We will show that volume $(G(C \cap Q_0))$ is zero which implies that the measure of $G(C)$ is zero, since L is arbitrary.

With $\ell = \frac{L}{n}$, let $\{q_j\}$ be the set cubes of side ℓ, for $j = 1, 2, \ldots n^N$, which form a partition of Q_0. i.e.

$$Q_0 = \bigcup_{j=1}^{n^N} q_j.$$

Assume that each q_j contains at least one critical point, say x^j. Let $x \in q_j$, then

$$\begin{aligned} G(x) &= G(x^j + (x - x^j)) \\ &= G(x^j) + G'(x^j)(x - x^j) + o(\ell), \end{aligned}$$

where $\dfrac{o(\ell)}{\ell} \to 0$ as $\ell \to 0$.

43 Using the fact that, rank $G'(x^j) \leq N - 1$, we will show that for sufficiently small ℓ, $z^j(x) = G'(x^j)(x - x^j)$ lies in $N - 1$ dimensional subspace of $\mathbb{R}^N$ as x varies. To see this, let ξ^j be a unit null vector of $G'(x^j)$. Put

$$y^j(x) = x - x^j = [y^j - < y^j, \xi^j > \xi^j] + < y^j, \xi^j > \xi^j.$$

Note that the vector $y^j - < y^j, \xi^j > \xi^j$ has no component in the ξ^j direction and hence it lies in an $N-1$ dimensional subspace of $\mathbb{R}^N$. As x varies over q_j, all these vectors $\{z^j(x)\}$ lie in the same $N-1$ dimensional subspace. Since $G'(x^j)$ is independent of x, the measure of the set

$$\{G'(x^j)y^j(x) : x \in q_j\}$$

is less than or equal to $(K\ell)^{N-1}$, where K is a constant (maximum elongation constant for G on Q_0). Hence the volume of $\{G'(x^j)y^j(x) + o(\ell) : x \in q_j\}$ is less than or equal to $(K\ell)^{N-1} \times o(\ell)$. Thus

$$\begin{aligned}\mathrm{Vol}\{G(C \cap Q_0)\} &\le n^N (C\ell)^{N-1} o(\ell)\\ &= (C^{N-1}L^N)\frac{o(\ell)}{\ell} \to 0 \text{ as } \ell \to 0.\end{aligned}$$

This proves Sard's theorem for the case $M \le N$. □

3.3 Examples

(1) Consider the example from population dynamics:

$$G(u, \lambda) = u^2 - \lambda_1 u - \lambda_2.$$

Take

$$x \equiv \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} u \\ \lambda_2 \\ \lambda_3 \end{bmatrix}.$$

Then $G(u, \lambda)$ and its derivative can be written as **44**

$$\begin{aligned} G(x) &= x_1^2 - x_2 x_1 - x_3,\\ G'(x) &= [2x_1 - x_2, -x_1, -1]. \end{aligned}$$

Rank $(G'(x)) = 1$, which is the maximum rank. Therefore there are no critical points for this problem.

(2) Define $G : \mathbb{R}^2 \to \mathbb{R}$ by:

$$G(x) = x_1^2 - x_2 x_1 - k \text{ where } k \text{ is a constant.}$$

Then

$$G'(x) = [2x_1 - x_2, -x_1].$$

and it has the maximum rank 1, except at $x = (0, 0)$ where $G'(x)$ has rank zero. This is a critical point and it is the only critical point. The critical value is $G(0) = -k$.

Observe that the solution curves of $G(u) = 0$ are two disjoint curves for $k \neq 0$. For $k = 0$ they are two lines which intersect at the origin, which is a bifurcation point. See Fig. 3.1.

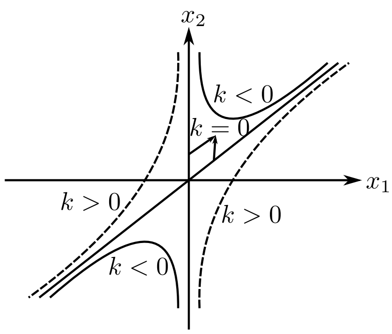


Figure 3.1:

3.4 Solution sets for Regular Values

45 We study now the solutions of $G(x) = p$, Where p is a regular value.

3.5 Lemma (Finite Set of Solutions)

Let $G : \mathbb{R}^N \to \mathbb{R}^N$ with $G \in C^1(\overline{\Omega})$, for some bounded open set $\Omega \subset \mathbb{R}^N$. Let $p \in \mathbb{R}^N$ be a regular value of G such that:

$$G(x) \neq p \text{ for } x \in \partial\Omega.$$

Then the set $G^{-1}(p) \cap \overline{\Omega}$ is finite, where:

$$G^{-1}(p) = \{x \in \mathbb{R}^N : G(x) = p\}.$$

Proof. We give a proof by contradiction. Let $\{x_j\}_{j=1}^{\infty} \in G^{-1}(p) \cap \overline{\Omega}$ and assume all x_j's are distinct. Then some subsequence of $\{x_j\}$ converges to some $x^* \in \overline{\Omega}$. By the continuity of G, we have:

$$G(x^*) = p.$$

This implies $x^* \notin \partial\Omega$. However by the implicit function theorem, there is one and only one solution $x = x(p)$ of $G(x) = p$ in some open neighbourhood of $x(p)$. This is a contradiction, since every neighbourhood of x^* contains infinitely may x_j's. This complete the proof. □

Next, we will consider the case $G : \mathbb{R}^{N+1} \to \mathbb{R}^N$ and study the solution set $G^{-1}(p)$, if p is a regular value.

3.6 Lemma (Global Solution Paths)

Let $G : \mathbb{R}^{N+1} \to \mathbb{R}^N$, and $G \in C^2(\mathbb{R}^{N+1})$. Let p be a regular value for G. Then $G^{-1}(p)$ is a C^1, one dimensional manifold. That is each of its connected components is diffeomorphic to a circle or infinite interval. In more detail, each component of $G^{-1}(p)$ can be represented by a differentiable function $X(s)$, $s \in \mathbb{R}$ and it satisfies one of the following: **46**

(3.6)
(a) $\|X(s)\| \to \infty$ as $|s| \to \infty$.
(b) $X(s)$ is a simple closed curve and $X(s)$ is periodic.
(c) $G^{-1}(p) = \phi$.

Proof. Assume that $G^{-1}(p) \neq \phi$, or else we are done. Let x^0 be such that:

$$G(x^0) - p = 0.$$

Consider the equation:

$$G(x) - p = 0.$$

Since p is regular value, the Jacobin of $G(x) - p$ at x^0 viz. $G'(x^0)$, must have the maximum rank N. Therefore, it has a minor of order N, which is nonsingular. Assume that $\dfrac{\partial G}{\partial(x_1,,,x_N)}(x^0)$ is nonsingular. Let $(x_i,..x_N) = u$ and $x_{N+1} = \lambda$. Thus we have:

$$F(u^0, \lambda^0) \equiv G(u^0, \lambda^0) - p = 0,$$

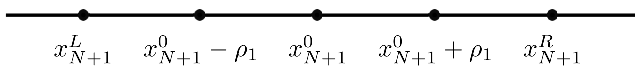
and

$$F_u(u^0, \lambda^0) = G_u(u^0, \lambda^0)$$

is nonsingular. Hence by the implicit function theorem there exists a unique solution $u = u(\lambda)$, for all λ such that $|\lambda - \lambda^0| < p_1(u^0, \lambda^0)$. i.e.
47 there exits a solution are:

$$\gamma^1 = \{x_1(x_{N+1}), \ldots, x_N(x_{N+1}), x_{N+1}\},$$

with X_{N+1} in some iterval $|x_{N+1} - x^0_{N+1}| < \rho_1(x^0)$. Extend this arc over a maximal interval (the interval beyond which we cannot extend the solution), say, $x^L_{N+1} < x_{N+1} < x^R_{N+1}$.



At the points, the Jacobian $\dfrac{\partial G}{\partial(x_1,..x_N)}(x^0)$ will be singular, otherwise we can again apply implicit function theorem to obtain the solution in a larger interval which will contradict the maximality of the interval. Now consider the point $x^R = (x_1(x^R_{N+1}), \ldots (x_N(x^R_{N+1}), x^R_{N+1})$. By the continuity of G we have:

$$G(x^R) = p.$$

Again since p is a regular value, there exists a minor of rank N, which is nonsingular. Assume that this is obtained from the matrix G_x by removing the j^{th} column of the matrix. As before, there exists an are γ^2, say,

$$\gamma^2 = \{x_1(x_j), \ldots x_{j-1}(x_j), x_j, x_{j+1}(x_j), \ldots x_{N+1}(x_j)\},$$

over a maximal interval. We can continue this procedure indefinitely. This family (γ^i) will overlap because the implicit function theorem gives a unique curve. Thus $\{\gamma^i\}$ form a C^1 curve. This curve can be globally parametrized. Suppose it is given by:

$$\Gamma = \{X(s); X(s_a) = x^0, s_a \leq s \leq s_b\}$$

and 48

$$G(X(s)) = p.$$

If $X(s)$ is not bounded, then the first part of the theorem follows. So assume that $X(s)$ is bounded. Then choose a sequence $\{s_j\} \to \infty$ such that $\{X(s_j)\} \to x^*$. By the continuity of G we get:

$$G(x^*) = p.$$

If the curve is not closed, x^* will be a limit point and also a solution of $G(x) - p = 0$. Since p is a regular value, we can apply the implicit function theorem to conclude that the curve has to be closed. This closed curve is simple, because of the uniqueness of the solution path. Hence in the case when $X(s)$ is bounded, we have a simple closed periodic solution which proves part (b) of the theorem. □

Remark. The main idea in the proof is the change of parameter from one component γ^j to another. This can be used as a practical procedure for carrying out global continuation; see [27].

3.7 Degree Theory

In this section we will assign an integer, called the '**degree**' to a continuous function F, defined on an open subset Ω, of $\mathbb{R}^N$ and a point $p \in \mathbb{R}^N$. This is an important tool in proving fixed point theorems and existence results for nonlinear problems. The degree of a continuous function F over a domain at a point p has the important property that if it is non - zero, then there exists a solution of $F(x) = p$ in the given domain. An- 49
other important property is that it depends on the values of the function only on the boundary and not in the interior.

3.8 Definition

Let $F : \mathbb{R}^N \to \mathbb{R}^N$ and $p \in \mathbb{R}^N$ satisfy:

(i) $F \in C^1(\bar{\Omega})$, where Ω is an open bounded subset of $\mathbb{R}^N$,

(ii) $p \notin F(\partial\Omega)$,

(iii) p is a regular value of F on Ω.

Then the degree of F on Ω for p is defined as:

$$\deg(F, \Omega, p) = \sum_{x \in [F^{-1}(p) \cap \bar{\Omega}]} \mathrm{Sgn}[\det F'(x)]. \tag{3.8}$$

Note. The degree is well defined, since by lemma 3.5, $[F^{-1}(p) \cap \bar{\Omega}]$ is a finite set and hence the right hand side is a finite sum.

Example. Consider $f : \mathbb{R} \to \mathbb{R}$ such that $f'(x)$ is continuous on $[a, b]$ and $f(a) \neq p$, $f(b) \neq p$ and $f(x) = p$ has only simple roots on $[a, b]$. Then

$$\deg(f, (a, b), p) = \sum_{j=1}^{k} \frac{f'(x_j)}{|f'(x_j)|},$$

where $\{x_j\}_{j=1}^{k}$ are the consecutive roots of $f(x) - p = 0$ contained in $[a, b]$. In this particular case, if $f'(x_j) > 0$, then $f'(x_{j+1}) < 0$ and vice versa. Therefore $\dfrac{f'(x_j)}{|f'(x_j)|} = +1$ and -1 alternatively. Hence

$$\deg(f, (a, b), p) \in \{1, 0, -1\}.$$

Now if p does not lie between $f(a)$ and $f(b)$, then there are either
50 no roots or an even number of roots for the equation $f(x) = p$ in (a, b). Hence the degree is zero. In the other case, there will be an odd number of roots, say $(2k+1)$, $k \geq 0$. Thus in the summation in (3.8), the first $(2k)$ terms cancel out and the last term is $+1$ or -1, depending on $f(b) > p$ or $f(b) < p$. Hence we can write:

$$\deg(f, (a, b), p) = \frac{1}{2}\left[\frac{f(b) - p}{|f(b) - p|} - \frac{f(a) - p}{|f(a) - p|}\right].$$

Note that in this case the degree depends only on the values of f at the boundary points a and b. This is true in the general case also.

Next we will relax condition (iii) of definition 3.8.

3.9 Definition

Let F satisfy (i) and (ii) of (3.8). Then,

$$\deg(F, \Omega, p) \equiv \deg(F, \Omega, q), \tag{3.9a}$$

where q satisfies:

$$\text{(3.9b)} \qquad \begin{array}{rl} \text{(i)} & q \text{ is a regular value of } F, \\ \text{(ii)} & \|q - p\| < \gamma \equiv \operatorname{dist}(F(\partial\Omega), p) \equiv \inf_{x \in \partial\Omega} \|F(x) - p\|. \end{array}$$

By Sard's theorem, the set of all regular values of F are dense in $\mathbb{R}^N$. Hence we can find regular values satisfying (3.9b(ii)). Also if q_1, q_2 are two regular values satisfying (3.9b(ii)), then they belong to the same connected component of $\mathbb{R}^N \setminus F(\partial\Omega)$ and hence (for a proof which is too lengthy to include here see [29]):

$$\deg(F, \Omega, q_1) = \deg(F, \Omega, q_2).$$

Therefore the above degree is well defined. We can also relax condition (i) of definition 3.8 and define the degree for $F \in C(\overline{\Omega})$. **51**

3.10 Definition

Assume F satisfies:

(i) $F \in C(\overline{\Omega})$.

(ii) $p \notin F(\partial\Omega)$.

Then

$$\deg(F, \Omega, p) \equiv \deg(\tilde{F}, \Omega, p) \tag{3.10}$$

where $\tilde{F}$ satisfies, using y of (3.9b(ii)):

(iii) $\tilde{F} \in C^1(\overline{\Omega})$,

(iv) $\|F - \tilde{F}\|_\infty < \dfrac{\gamma}{2}$.

Since F is continuous, we can approximate F as closely as desired, by differentiable function $\overline{F}$. (Take polynomials, for example). Conditions (ii) and (iv) imply that $p \notin \tilde{F}(\partial\Omega)$. Thus deg $(\tilde{F}, \Omega, p)$ is well defined by definition 3.9. If $\hat{F}$ is another smooth function, satisfying condition (iv), then by considering the homotopy

$$G(x,t) = t\tilde{F}(x) + (1-t)\hat{F}(x)$$

and using the homotopy invariance property of the degree in definition 3.8 (which we prove next), it follows that

$$\deg(\tilde{F}, \Omega, p) = \deg(\hat{F}, \Omega, p).$$

Thus definition 3.10 is independent of the choice of $\tilde{F}$. Thus the degree is well defined even for a function $F \in C(\overline{\Omega})$.

3.11 Theorem (Homotopy Invariance of the Degree)

52 *Let $G : \mathbb{R}^{N+1} \to \mathbb{R}^N$, Ω bounded open set in $\mathbb{R}^N$ and $p \in \mathbb{R}^N$ satisfy*:

$$(3.11) \qquad \begin{array}{ll} \text{(a)} & G \in C^2(\overline{\Omega} \times [0,1]), \\ \text{(b)} & G(u,\lambda) \neq p \ \text{ on } \partial\Omega \times [0,1], \\ \text{(c)} & G(u,0) \equiv F_0(u),\ G(u,1) \equiv F_1(u), \\ \text{(d)} & \underline{p} \ \text{ is a regular value for } G \text{ on} \\ & \overline{\Omega} \times [0,1] \ \text{ and for } F_0 \text{ and } F_1 \text{ on } \overline{\Omega}. \end{array}$$

Then $\deg(G(.,\lambda), \Omega, p)$ *is independent of* $\lambda \in [0,1]$. *In particular,*

$$\deg(F_0, \Omega, p) = \deg(F_1, \Omega, p).$$

Proof. The proof uses lemma 3.5. We will prove

$$\deg(F_0, \Omega, p) = \deg(F_1, \Omega, p).$$

The case for any $\lambda \in (0,1)$ is included. Since p is a regular value, we have:

$$\{F_0^{-1}(p) \cap \bar{\Omega}\} = \{u_i^0\},$$

and

$$\{F_0^{-1}(p) \cap \bar{\Omega}\} = \{u_j^1\},$$

where $\{u_i^0\}$ and $\{u_j^1\}$ are finite sets by lemma 3.5. Again, since p is a regular value, $\{G^{-1}(p) \cap (\bar{\Omega} \times [0,1])\}$ is a finite collection of arcs, denoted by $\{\Gamma_i(s)\}$. Let us parametrize any $\Gamma(s)$ as $(u(s), \lambda(s))$, where s denotes the arc length. On $\Gamma(s)$ we have:

$$G(s) \equiv G(u(s), \lambda(s)) = p.$$

Differentiating with respect to s, we get : **53**

$$G_u(s) \cdot \dot{u}(s) + G_\lambda(s)\dot{\lambda}(s) = 0.$$

Since s is the arc length, we have:

$$\dot{u}^T(s)\dot{u}(s) + \dot{\lambda}(s)\dot{\lambda}(s) = 1.$$

These two equations together can be written as:

$$A(s) \begin{pmatrix} \dot{u}(s) \\ \dot{\lambda}(S) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{3.12a}$$

where

$$A(s) = \begin{bmatrix} G_u(s) & G_\lambda(s) \\ \dot{u}^T(s) & \dot{\lambda}(s). \end{bmatrix}$$

We shall show that $A(s)$ is nonsingular on $\Gamma(s)$. First observe that at each point on $\Gamma(s)$, there is a unique tangent to the path. The matrix $[G_u(s), G_\lambda(s)]$ has rank N, since p is a regular value. Thus the null space of this matrix is one dimensional. Let is be spanned by the vector $\{\xi, n\}$. Since $(\dot{u}, \dot{\lambda})$ is also a null vector, we have:

$$(\dot{u}(s), \dot{\lambda}(s)) = c_1(\xi, \eta), \quad \text{for some } c_1 \neq 0.$$

Now if $A(s)$ is singular, then it must have a nontrivial null vector of the form $c_2(\xi, \eta)$ for $c_2 \neq 0$. Since

$$c_2 A(s)(\frac{\xi}{\eta}) = 0,$$

the last equation gives:

$$c_1 c_2(|\underset{\sim}{\xi}|^2 + \eta^2) = 0.$$

54 This implies that $c_1 c_2 = 0$, a contradiction. Hence $A(s)$ is nonsingular.

Now apply Cramer's rule to solve the above system for $\lambda(s)$ to get:

$$\dot{\lambda}(s) = \frac{\det G_u(s)}{detA(s)} \tag{3.12b}$$

Note that on $\Gamma(s)$, det $A(s) \neq 0$. The above result shows that $\dot{\lambda}(s)$ and det $G_u(s)$ change sign simultaneously.

We have by (3.8)

$$\deg(F_0, \Omega, p) = \sum_{\{u_i^0\}} \operatorname{sgn}(\det F_0'(u_i^0)),$$
$$\deg(F_0, \Omega, p) = \sum_{\{u_i^1\}} \operatorname{sgn}(\det F_1'(u_i^1)).$$

Observe that the arcs $\Gamma(s)$ can be classified into four different types :

(i) arcs joining two points from $\{u_i^0\}$

(ii) arcs joining two points from $\{u_j'\}$

(iii) arcs joining a point u_i^0 to a point u_j^1

(iv) arcs with no end points.

We shall use the arcs of type (i), (ii) and (iii) to relate the above two degrees.

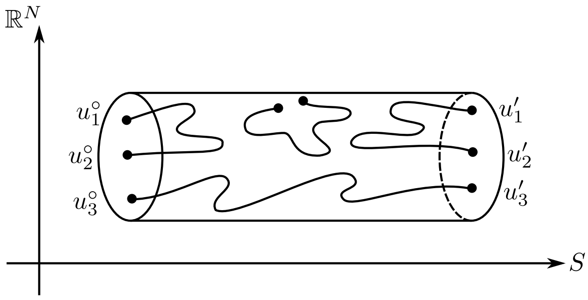


Figure 3.2:

In case (i), $\dot{\lambda}(s)$ has different signs at the end points, since along the $\Gamma(s)$, $\dot{\lambda}(s)$ changes sign an odd number of times. Hence det $G_u(s)$ has different signs at the end points. Therefore no contribution to deg (F_0, Ω, p) results from points. Similar result is true for case (ii). Obviously there is no contribution to the degree from the (iv)th case. So the only contribution comes from case (iii). Here note that $\dot{\lambda}(s)$ and hence det $G_u(s)$ have the same sign at both the end points, because $\dot{\lambda}(s)$ changes sign an even number of times. This shows that: **55**

$$\deg(F_0, \Omega, p) = \deg(F_1, \Omega, p)$$

□

The theorem is true for the other two definitions of degree also. Indeed when these definitions have been justified, our above proof gives the desired result with the hypothesis relaxed to continuous mappings and without the restriction (3.11d).

Some Important Properties of the Degree

(1) $f, g : \overline{\Omega} \to \mathbb{R}^N$ be such that:

(i) $f(x) = g(x)$ for all $x \in \partial\Omega$;

(ii) $f(x), g(x) \in C(\overline{\Omega})$;

(iii) $f(x) \neq p$ for all $x \in \partial\Omega$;

Then
$$\deg(f, \Omega, p) = \deg(g, \Omega, p).$$

(2) If p, q are close to each other, say,
$$\|p - q\| < \gamma = \operatorname{dist}(f(\partial\Omega), p).$$

56 Then
$$\deg(f_0, \Omega, p) = \deg(f_1, \Omega, q).$$

(3) Let $f, g : \overline{\Omega} \to \mathbb{R}^N$ be continuous and $p \notin f(\partial\Omega)Ug(\partial\Omega)$. Let f and g satisfy
$$\sup_{x\in\partial\Omega} |f(x) - g(x)| < \frac{\gamma}{2}.$$

Then
$$\deg(f, \Omega, p) = \deg(g, \Omega, p).$$

(4) Homotopy invariance property.

(5) (Excision) If K is a closed set contained in Ω and $p \notin f(K)Uf(\partial\Omega)$, then
$$\deg(f, \Omega, p) = \deg(f, \Omega\backslash K, p).$$

(6) Let $p \notin f(\overline{\Omega})$, then
$$\deg(f, \Omega, p) = 0.$$

(7) If $\deg(f, \Omega, p) \neq 0$, then $f(x) = p$ has a solution in Ω. (For reference, See : [26], [29])

Application 1. Fixed points and roots

3.13 Brouwer Fixed Point Theorem

Let $F : \mathbb{R}^N \to \mathbb{R}^N$ satisfy:

(3.13) (a) $F \in C(\bar{\Omega}), \Omega$ is a convex open bounded subset of $\mathbb{R}^N$.
(b) $F(\bar{\Omega}) \subset \Omega$.

Then $F(u) = u$ *for some* $u \in \Omega$. 57

Proof. Take any point $a \in \Omega$. Define

$$G(u, \lambda) = \lambda(u - F(u)) + (1 - \lambda)(u - a).$$

We shall that G, a homotopy between $(u - F(u))$ and $(u - a)$, does not vanish on $\partial\Omega \times [0, 1]$.

If $G(u, \lambda) = 0$, then we have :

$$u = \lambda F(u) + (1 - \lambda)a$$

By (3.13b) we get, $F(u) \in \Omega$. Since $a \in \Omega$, by convexity $u \in \Omega$ for $\lambda \in [0, 1]$. Therefore $G(u, \lambda) \neq 0$ on $\partial\Omega\times[0, 1]$. So homotopy invariance theorem implies that:

$$\begin{aligned} \deg(u - F(u), \Omega, 0) &= \deg(u - a, \Omega, 0) \\ &= [sgn \det\{I\}]_{u=a} \\ &= 1 \end{aligned}$$

Hence $u - F(u) = 0$ has a solution in Ω □

Note. If $F \in C^2(\bar{\Omega})$ the above proof is constructive. That is we can find arbitrarily good approximations to the fixed point by computation. To do this we consider:

$$G(u, \lambda) = \lambda(u - F(u)) + (1 - \lambda)(u - a) = 0$$

and follow the path $\Gamma(s)$ from $(u(0), \lambda(0)) = (a, 0)$, to $(u(s_F), \lambda(s_F)) = (F(u(s_F)), 1)$. If 0 is not a regular value, we can take δ arbitrarily small and a regular value. Then we can construct solutions for $u - F(u) = \delta$. It can be shown that 0 is a regular value for almost all $a \in \Omega$. 58

Another result on global existence of a solution is given as:

3.14 Theorem

Let $F \in C(\overline{\Omega})$, where Ω is open in $\mathbb{R}^N$ and satisfy, for some $x_0 \in \Omega$:

(3.14) $$< x - x_0, F(x) >> 0 \text{ for all } x \in \partial\Omega.$$

Then $F(x) = 0$ for some $x \varepsilon \Omega$.

Proof. Consider the homotopy

$$G(x, \lambda) = \lambda f(x) + (1 - \lambda)(x - x_0), 0 \leq \lambda \leq 1.$$

It is easy to prove that $G(x, \lambda) \neq 0$ on $\partial\Omega$ and for all $0 \leq \lambda \leq 1$. For if not, taking the inner product with $(x - x_0)$, we get:

$$< x - x_0, F(x) > +(1 - \lambda)||x - x_0||^2 = 0.$$

If $x \in \partial\Omega$ then this gives a contradiction by (3.14). Hence $G(x, \lambda) \neq 0$ for $x \in \partial\Omega$ and $0 \leq \lambda \leq 1$. Therefore

$$\deg(F, \Omega, 0) = \deg(x - x_0, \Omega, 0) = 1$$

□

Application II. Periodic solutions of O.D.E.

We shall use the Brouwer theorem to show the existence of periodic solutions of systems of ordinary differential equations.

3.15 Periodic Solution Theorem

59

Let $f(t, y) : \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}^N$ satisfy for some $T > 0$ and some convex open $\Omega \subset \mathbb{R}^n$.

(3.15)

(a) $f \in C([0, T] \times \overline{\Omega})$.
(b) $f(t + T, y) = f(t, y)$ for all $y \in \overline{\Omega}$
(c) $||f(t, y) - f(t, x)|| \leq K||x - y||$ for all $x, y \in \overline{\Omega}, 0 \leq t \leq T, K > 0$.
(d) $f(t, y)$ is directed into Ω for all $y \in \partial\Omega$ and $t_{\in}[0, T]$,

i.e. for all $y \in \partial\Omega$, $y + \varepsilon f(t, y) \in \Omega$, for all $\varepsilon > 0$ sufficiently small. Then the equation

$$\frac{dy}{dt} = f(t, y) \tag{3.15e}$$

has periodic solutions $y(t)$ with period T and $y(t) \in \Omega$, for all t.

Proof. Pick any $u \in \Omega$ and solve the initial value problem :

$$\frac{dy}{dt} = f(t, y),$$
$$y(0) = u.$$

Let the unique solution in the interval $[0, T]$ be denoted by $y(t, u)$. Then $y(t, u) \in \Omega$ for all $t > 0$. Otherwise, let t_1 be the first time it crosses the boundary of Ω. That is $y(t_1, u) \in \partial\Omega$, $t_1 > 0$, $y(t, u) \in \Omega$, for all $0 \leq t < t_1$ and,

$$\dot{y}(t_1, u) = f(t_1, y).$$

But condition (3.15d) says that $f(t_1, y)$ is directed into Ω which is not possible. Now consider $F(u) \equiv y)(T, u)$, for $u\varepsilon\Omega$; this F satisfies all the hypothesis of the Brouwer fixed point Theorem. Hence we have:

$$y(T, u) = u, \quad \text{for some } u\varepsilon\Omega.$$

i.e. we have a periodic solution passing through u. □ **60**

Application III. Bifurcation by degree theory

Definition. Let $G : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$ be C^1. Consider the problem:

$$G(u, \lambda) = 0. \tag{3.16}$$

Given a smooth path (or arc) of solutions, say,

$$\Gamma = \{(u(\lambda), \lambda) : \lambda_a < \lambda < \lambda_b\},$$

a point $(u^0, \lambda^0) \in \Gamma$ is said to be a bifurcation point for (3.16) if every ball $B_\rho(u^0, \lambda^0) \subset \mathbb{R}^{N+1}$, of radius $\rho > 0$, contains solutions of (3.16) not on Γ.

The following theorem shows that if the sign of det $G_u(u(\lambda), \lambda)$ changes at some point along Γ, then it is a bifurcation point. This is an important result in testing for bifurcation.

3.17 Theorem (Bifurcation Test)

Let Γ be a smooth solution arc of (3.16) *parametrized by λ. Let det $G_u(u(\lambda), \lambda)$ change sign at $\lambda^0 \in (\lambda_a, \lambda_b)$. Then $(u(\lambda^0)\lambda^0)$ is a bifurcation point of* (3.16).

Proof. We prove by contradiction. Assume that,

$$(u^0, \lambda^0) = (u(\lambda^0)\lambda^0),$$

is not a bifurcation point. Hence there exists a ball of radius ρ, which
61 does not contain any root of (3.16) other on the arc Γ. Choose η, $\delta > 0$ small enough so that the cylinder:

$$C_{\eta,\delta} = \{(u, \lambda) : u \in \bar{K}_\eta(u^0), \lambda \in [\lambda^0 - \delta, \lambda^0 + \delta]\},$$

where,

$$\bar{K}_\eta(u^0) = \{u \in \mathbb{R}^N : \|u - u^0\| \leq \eta\}.$$

is such that :

(i) $C_{\eta,\delta} \subset B_\rho(u^0, \lambda^0)$.

(ii) $G(u, \lambda) \neq 0$, for all $(u, \lambda) \in \partial K_\eta \times [\lambda^0 - \delta, \lambda^0 + \delta]\}$.

(iii) $u(\lambda^0 \pm \delta) \in K_\eta(u^0)$.

(iv) det $G_u(u(\lambda), \lambda)$ does not change sing in the intervals $(\lambda^0 - \delta, \lambda^0)$ and $(\lambda^0, \lambda^0 + \delta)$.

We can easily choose η and δ so that (i) is satisfied. Since the ball $B_\rho(u^0, \lambda^0)$ contains no roots other than on the are Γ and det $G_u(u(\lambda), \lambda)$ changes sing at λ^0 and is continuous we can shrink δ (if necessary) so that (ii), (iii) and (iv) are satisfied (see Fig. 3.3).

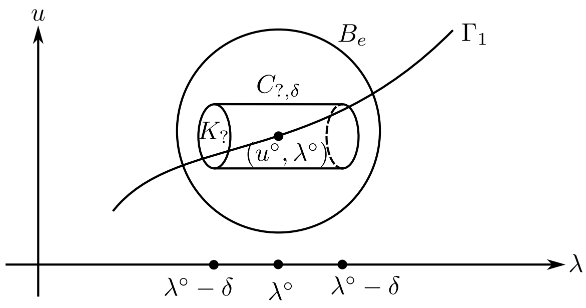


Figure 3.3:

Applying the homotopy theorem to the function $G(., \lambda^0 - \delta)$ and **62**
$G(., \lambda^0 + \delta)$, we get:

$$\deg(G(., \lambda^0 - \delta), K_\eta(u^0), 0) = \deg(G(., \lambda^0 + \delta), K_\eta(u^0), 0).$$

But

$$\deg(G(., \lambda^0 - \delta), K_\eta(u^0), 0) = \operatorname{Sgn} \deg G_u(u(\lambda^0 - \delta), \lambda^0 - \delta),$$

and

$$\deg(G(., \lambda^0 + \delta), K_\eta(u^0), 0) = \operatorname{Sgn} \deg G_u(u(\lambda^0 + \delta), \lambda^0 + \delta).$$

This is a contradiction, because det $G_u(u(\lambda), \lambda)$ has different signs at $(u(\lambda^0 + \delta), (\lambda^0 - \delta)$ and $(u(\lambda^0 + \delta), \lambda^0 + \delta)$. $\square$

Note. Here $\lambda^0 \in (\lambda_a, \lambda_b)$ is an interior point of the interval. We cannot apply this theorem if λ_0 is a boundary point. See Fig. 3.4.

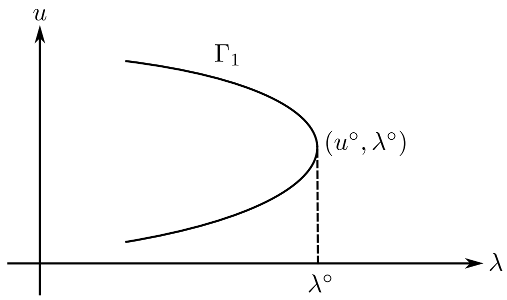


Figure 3.4:

Example. Let

$$G(u, \lambda) = Au - \lambda u = 0,$$

63 where A is an $n \times n$ symmetric matrix, $\lambda \in \mathbb{R} \cdot \Gamma = \{(0, \lambda) : -\infty < \lambda < \infty\}$ is the trivial path of solutions. We have:

$$\det G_u(u, \lambda) = \det(A - \lambda I), \qquad\qquad = p_n(\lambda),$$

where $p_n(\lambda)$ is a polynomial in λ of degree n. Note that $p_n(\lambda)$ changes sign at eigenvalues of odd multiplicity. Therefore every eigenvalue of odd multiplicity corresponds to a bifurcation point. We cannot predict anything (using the above theorem) about the bifurcation at eigenvalues of even multiplicity.

3.18 Global Homotopies and Newton's Method

To solve

$$F(u) = 0, \tag{3.19}$$

where $F : \mathbb{R}^N \to \mathbb{R}^N$ is a smooth function we consider for $u^0 \in \mathbb{R}$, the homotopy:

$$G(u, t) = F(u) - e^{-\alpha t} F(u^0) \equiv 0. \tag{3.20}$$

Here $\alpha > 0$ and $0 \leq t < \infty$, so that when $t \to \infty$, solution $u = u(t)$ of (3.20), if it exists for all $t > 0$, must approach a solution of (3.19). i.e.

$$\lim_{t\to\infty} u(t) = u^* \in F^{-1}(0).$$

Differentiating (3.19) we get:

$$F'(u)\frac{du}{dt} + \alpha F(u) = 0. \tag{3.21a}$$

The solution of this nonlinear differential equation together with the initial condition: **64**

$$u(0) = u^0 \tag{3.21b}$$

gives the homotopy path $u(t)$ from u^0 to a solution of (3.19)

$$u^* = \lim_{t\to\infty} u(t),$$

provided $u(t)$ exists on $[0, \infty)$

If we use Euler's method on (3.21a), to approximate this path we get the sequence $\{u^U\}$ defined by :

$$F'(u^U)[u^{U+1} - u^U] + \alpha \Delta t_U F(u^U) = 0,$$

where $\Delta t_U = t_{U+1} - t_U$.

If we take $\Delta t_U = \Delta t$ (uniform spacing) and $\alpha = (\Delta t)^{-1}$, this gives Newton's method to approximate a root of (3.19) starting with the initial guess u^0. Such a path does not exist always. But if $F'(u^*)$ is nonsingular and $\|u^0 - u^*\|$ is sufficiently small, it does exist. If $F'(u)$ is singular along the path defined by (3.21a), the method need not converge. This is one of the basic difficulties in devising global Newton methods.

A key to devising global Newton methods is to give up the monotone convergence implied by (3.20) (i.e. each component of F goes to 0 monotonically in t) and consider more general homotopies by allowing $\alpha = \alpha(u)$ in (3.21a). Branin [2] and Smale [30] used these techniques. The former used $\alpha(u)$ of the form: **65**

$$\alpha(u) = \operatorname{Sgn} \det F'(u),$$

and the latter used

$$\operatorname{Sgn}\alpha(u) = \operatorname{Sgn}\det F'(U).$$

Smale shows the if $F(u)$ satisfies the boundary conditions (3.22) (see below) for some bounded open set $\Omega \subset \mathbb{R}^N$, then for almost all $u^0 \varepsilon \partial\Omega$, the homotopy path defined by (3.21a) and (3.21b) is such that:

$$\lim_{t\to t_1} u(t) = u^*,$$

where $F(u^*) = 0$ and $0 < t_1 \leq \infty$. Note that with such choices for $\alpha(u)$, the corresponding schemes need not always proceed in the 'Newton direction', viz.-$[F'(u)]^{-1}F(u)$, but frequently go in just the *opposite* direction. The change in direction occurs whenever the Jacobian $\det F'(u(t))$ changes sign . Hence the singular matrices $F'(u)$ on the path $u(t)$ cause no difficulties in the proof of Smale's result. But there practical difficulties in computing near such points, where "small steps" must be taken. We shall indicate in theorem 3.24 how these difficulties can be avoided in principle, by using a different homotopy, namely the one appearing in (3.24a) below. To prove the theorem, we need the following :

3.22 Boundary conditions (Smale)

Let $\Omega \subset \mathbb{R}^N$ be and open bounded set with smooth connected boundary $\partial\Omega$. Let $F : \mathbb{R}^N \to \mathbb{R}^N$ be in $C^1(\Omega)$ and satisfy:

$$\text{(3.22)} \qquad \begin{array}{rl} \text{(a)} & F'(u) \text{ is nonsingular for all } u \in \partial\Omega, \text{ and} \\ \text{(b}_+\text{)} & [F'(u)]^{-1}F(u) \text{ is directed out of } \Omega, \\ & \text{for all } u \in \partial\Omega, \text{ or} \\ \text{(b}_-\text{)} & [F'(u)]^{-1}F(u) \text{ is directed into } \Omega, \text{ for all} \\ & u \in \partial\Omega. \end{array}$$

66

Suppose u^* is an isolated (i.e. $F'(u^*)$ is nonsingular) root of $F(u) = 0$, then the boundary conditions (3.22) are satisfied on the ball $B_\rho(u^*)$

provided the radius ρ is sufficiently small. Condition (3.22a) follows from :

$$F'(u) = F'(u^*) + 0(u - u*), \text{ for all } u \in B_\rho(u^*),$$

and the Banach lemma, provide ρ is sufficiently small. To show (b_+), use Taylor expansion:

$$F(u) = F(u^*) + F'(u^*) \cdot (u - u*) + 0((u - u^*))^2$$

Since $F(u^*) = 0$ and $F'(u^*)$ is nonsingular, using (3.22a) we get

$$F'(u)^{-1} F'(u^*) = (u - u*) + 0((u - u^*)^2).$$

The right hand side is directed out of $B_\rho(u^*)$ for ρ sufficiently small and so (b_+) holds.

We consider the equation

$$G(u, \lambda) = F(u(s)) - \lambda(s)F(u^0) = 0,$$

for some fixed u^0. The smooth homotopy path $\{u(s), \lambda(s)\}$ must satisfy the differential equation:

$$F'(u)\dot{u} - \dot{\lambda} F(u^0) = 0, \tag{3.23a}$$

In addition we impose the arclength condition: **67**

$$\|\dot{u}(s)\|^2 + \dot{\lambda}(s)^2 = 1. \tag{3.23b}$$

This has the effect of making s the arc length parameter along the path $(u(s), \lambda(s))$. If $\lambda(s^*) = 0$ at some point $s = s^*$ on the path, then $u(s^*) = u^*$ is a root of (3.19). Further several roots may be obtained, if $\lambda(s)$ vanishes several times on a path.

We shall show that if Smale's boundary conditions (3.22) are satisfied then for almost all $u^0 \in \partial\Omega$, the initial data $(u(0), \lambda(0)) = (u^0, 1)$ and (3.23a,b) define a path on which $\lambda(s)$ vanishes an odd numbers of times. (see [20]). The main Theorem is as follows:

3.24 Theorem

Let $F : \bar{\Omega} \to \mathbb{R}^N$ *be* C^2 *and satisfy the boundary conditions* (3.22)*. Then for any* $u^0 \in \partial\Omega$ *for which 0 is a regular value of:*

(3.24a) $$G(u, \lambda) \equiv F(u) - \lambda F(u^0)$$

there is a C^1 *solution* $(u(s), \lambda(s))$ *of the system:*

(b) $F'(u)\dot{u}(s) - \dot{\lambda}(s)F(u^0) = 0,$

(c) $\|\dot{u}(s)\|^2 + |\dot{\lambda}(s)|^2 = 1,$

over $0 \le s \le s_F$*, starting at:*

(3.24d) $$(u(0), \lambda(0)) = (u^0, 1),$$

and terminating at $(u(s_F), \lambda(s_F))$ *where:*

68 (e) $u(s_F) \in \partial\Omega$, $\lambda(s_F) < 0$, *and*

(3.24) $$|\lambda(s_F)| < L \equiv \max_{x \in \bar{\Omega}} \|f(x)\| \;/ \min_{y \in \partial\Omega} \|f(y)\|.$$

For an odd number of points $s_U \in (0, s_F)$*,*

(3.24f) $$\lambda(s_U) = 0 \quad \text{and} \quad F(u(s_U)) = 0.$$

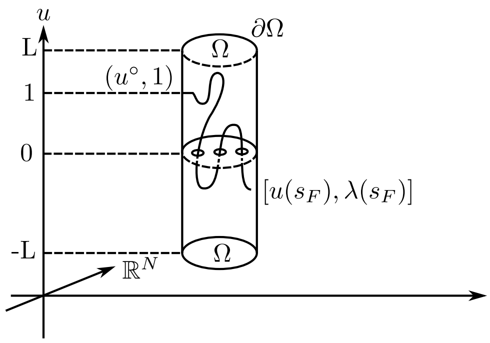


Figure 3.5:

Proof. In $\mathbb{R}^{N+1}$, we consider the cylinder $K \equiv \bar{\Omega} \times [-L, L]$, where L is defined as in (3.24d). See Fig. 3.5. Then for any fixed $u^0 \in \partial\Omega$, we have $G(u, \lambda) \neq 0$ on the bases of $K : \lambda = \pm L$ and $u \in \bar{\Omega}$. But on the cylindrical surface of K there is at least one solution of (3.24a), Viz. at $(u, \lambda) = (u^0, 1)$. Now 0 is a regular value and

$$\frac{\partial G(u, \lambda)}{\partial(u, \lambda)}|_{(u^0,1)} = [F'(u^0), -F(u^0)].$$

By assumption (3.22a), $F'(u^0)$ is nonsingular. Hence there is a C^1 arc $\Gamma(s) \equiv [u(s), \lambda(s)]$ which satisfies (3.24a,b,d). Taking s as the arclength, we obtain (3.24c) also. Choose the sign of $(\dot{u}(0), \dot{\lambda}(0))$ so that **69**
$\Gamma(s) \in K$ for $s > 0$; that is $\dot{u}(0)$ at u^0 points into Ω. Continuity along $\Gamma(s)$ determines the orientation of the tangent $(\dot{u}(s), \dot{\lambda}(s))$ satisfying (3.24c) in the interior of K. Since L is so large that G does not vanish for $|\lambda| = L$, the path $\Gamma(s)$ for $s > 0$ cannot meet the bases of K. The path $\Gamma(s)$ cannot terminate in the interior since if it had an interior point must lie on Γ. Then the implicit function theorem gives a contradiction, since 0 is a regular value. Thus $\Gamma(s)$ must meet the cylindrical surface of K for some $s = s_F > 0$. Since the tangent $(\dot{u}(s_F), \dot{\lambda}(s_F))$ to $\Gamma(s)$ at s_F cannot point into K, it follows that $\dot{u}(s_F)$ cannot point into Ω at $u(s_F) \in \partial\Omega$.

Multiplying (3.24b) by $\lambda(s)$ and using (3.24a):

$$\lambda(s).F'(u)\dot{u}(s) - \dot{\lambda}(s)F(u(s)) = 0 \text{ on } \Gamma.$$

$F'(u)$ is nonsingular at $u = u^0$ and $u = u(s_F)$. Therefore at points, we have:

$$\lambda(s)\dot{u}(s) = \dot{\lambda}(s)(F'(u(s)))^{-1}F(u(s))$$

Note that $\dot{\lambda}(0)$ and $\dot{\lambda}(s_F)$ are not zero, since $F'(u(s))$ is nonsingular for $s = 0$ and s_F. Now using the boundary condition (3.22b), We can deduce that

$$\frac{\lambda(0)}{\dot{\lambda}(0)} \frac{\lambda(s_F)}{\dot{\lambda}(s_F)} < 0. \tag{3.25}$$

Both $\lambda(0)\dfrac{\dot{u}(0)}{\dot{\lambda}(0)}$ and $\lambda(s_F)\dfrac{\dot{u}(s_F)}{\dot{\lambda}(s_F)}$ point out of (or into) Ω. But $\dot{u}(0)$ points into Ω and $\dot{u}(s_F)$ does not, so (3.25) follows.

Now we will show that $\dot{\lambda}(0)\dot{\lambda}(s_F) > 0$ which implies $\lambda(0)\lambda(s_F) < 0$. **70**
Hence $\lambda(s_F) < 0$ and so the theorem follows. We have:

$$G_u \cdot \dot{u} + G_\lambda \cdot \dot{\lambda} = 0$$
$$\dot{u}^T \cdot \dot{u} + \dot{\lambda}\dot{\lambda} = 1,$$

and from (3.12a,b):

$$\dot{\lambda}(s) = \frac{\det F'(u(s))}{\det A(s)}.$$

Now u^0 and $u(s_F)$ are $\partial\Omega$. By assumption (3.22a),$F'(u(s))$ is nonsingular for all $u \in \partial\Omega$. Since $\partial\Omega$ is connected, $\det F'(u(s))$ has the same sign at $s = 0$ and $s = S_F$. Also $A(s)$ is nonsingular all along Γ, as we had seen in the proof of theorem 3.11. Hence $\dot{\lambda}(0)$ and $\dot{\lambda}(s_F)$ have the same sign and the proof is complete $\square$

Chapter 4

Practical procedures in computing paths

4.1 Introduction

71

We will consider the problem either in uniform formulation (4.1a) or in formulation (4.1b):

(4.1a) $$G(x) = p \text{ where } G : \mathbb{R}^{N+1} \to \mathbb{R}^N;$$

(4.1b) $$G(u, \lambda) = P \text{ where, } G : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N.$$

Definition. (a) A path $\Gamma = \{x(s) : x(s) \in \Omega, G(x(s)) = p \text{ for all } s \in I\}$ is said to be regular if Rank $G_x(x(s)) = N$ for all $s \varepsilon I(\Omega \subset \mathbb{R}^{N+1})$.

(b) A path $\Gamma = \{(u(s), \lambda(s)) : u(s) \in \Omega, G(u(s), \lambda(s)) \underline{=} p \text{ for all } s \in I\}$ is regular if Rank $[G_u(s), G_\lambda(s)] = N$ for all $s \in I(\bar{\lambda} \subset \mathbb{R}^N)$.

Lemma 4.2. Rank$[G_u, G_\lambda] = N$ *if and only if either*

(4.2) (i) G_u *is nonsingular,*

or

(4.2) (ii) (a) $\dim N(g_u) = 1$,
(b) $G_\lambda \notin \text{Range}(G_u) \equiv R(G_u)$.

Proof. If (i) is true, then Rank $[G_\lambda, G_\lambda] = N$. In the other case, G_λ is not a linear combination of columns of G_u and since $N(G_u)$ has dimension $1, (N-1)$ columns of G_u are linearly independent and hence N columns **72**
of $[G_u, G_\lambda]$ are linearly independent.

Conversely, let Rank $[G_u, G_\lambda] = N$. Then either (i) holds or not. If it does we are done. If not, then $\dim N(G_u) = 1$ or else we cannot have Rank $[G_u, G_\lambda] = N$. But then we must also have $G_\lambda \not\in R(G_u)$. Hence the lemma □

Lemma 4.3. *Let* Γ *be a regular path of* (4.1a). *Then there exists an open tube* $T_\Gamma \subset \Omega$ *such that* p *is a regular value for* G *on* T_Γ *and* $\Gamma \subset T_\Gamma$.

Proof. Take any point $x \in \Gamma$, then there exists a minor matrix of rank N, say $M(x)$ of $G_x(x(s))$, such that $\det M(x) \neq 0$. Recall that

$$\det M(x) = \prod_{j=1}^{N} k_j(s)$$

where $\{k_j(s)\}$ are all ten eigenvalues of the matrix $M(x)$. Also $\{k_j(s)\}$ are smooth functions and $k_j(s) \neq 0$ for $j = 1, \ldots N$. Hence there will exists a neighbourhood of $x(s)$ in which the eigenvalues do not vanish. Therefore in this neighbourhood, the minor is nonvanishing. Such neighbourhoods exist at each point on Γ and so the tube T_Γ can be constructed □

Definition. A point $(u(s), \lambda(s))$ on Γ is said to be simple limit point (or simple fold) if (ii a,b) of lemma 4.2 holds.

We have on Γ:

$$G(u(s), \lambda(s)) - p = 0, \tag{4.4a}$$

73 so that:

$$G_u(s)\dot{u}(s) + G_\lambda(s)\dot{\lambda}(s) = 0. \tag{4.4b}$$

Note that at a fold point $s = s^0$, $\dot{\lambda}(s^0) = 0$ because

$$G_\lambda(s^0) \notin R(G_u(s^0)).$$

Hence $\dot{u}(s^0) \in N(G_u(s^0))$ at a fold point. Since $\dim N(G_u^0) = 1$ at simple fold point (u^0, λ^0), we can take:

$$N(G_u^0) = \text{ span } \{\phi\},$$

and

$$N(G_u^{oT}) = \text{ span } \{\psi\}.$$

From (4.4b) we have:

$$G_u^0 \ddot{u} + G_\lambda^0 \ddot{\lambda}^0 + G_{uu}^0 \dot{u}^0 \dot{u}^0 + 2G_{u\lambda}^0 \dot{u}^0 \dot{\lambda}^0 + G_{\lambda\lambda}^0 \dot{\lambda}^0 \dot{\lambda}^0 = 0.$$

Multiplying throughout by ψ^T and using the fact that $\xi \in R(G_u^0)$ if and only if $\psi^T \xi = 0$, we get:

$$\psi^T G_\lambda^0 \ddot{\lambda}^0 + \psi^T G_{uu}^0 \dot{u}^0 \dot{u}^0 = 0,$$

as $\dot{\lambda}(s_0) = 0$ at a fold point. Since $G_\lambda^0 \notin R(G_u^0)$, we have:

$$\psi^T G_\lambda^0 \neq 0,$$

and so

$$\ddot{\lambda}(s^0) = \frac{-\psi^T G_{uu}^0 \dot{u}^0 \dot{u}^0}{\psi^T G_\lambda^0}.$$

But $\dot{u}(s^0) = \alpha\phi$ for some scalar α. Hence: **74**

$$\ddot{\lambda}(s^0) = -\alpha^2 \frac{-\psi^T G_{uu}^0 \phi\phi}{\psi^T G_\lambda^0}.$$

If $\dfrac{-\psi^T G_{uu}^0 \phi\phi}{\psi^T G_\lambda^0} \neq 0$, then we say that (u^0, λ^0) is a simple quadratic fold. We similarly define a 'fold of order m' if $\lambda^{(k)}(s^0) = 0$ for all $k = 1, \ldots m-1$ and $\lambda^m(s^0) \neq 0$.

Usual methods of computing paths fail at a fold point. So in this section we present an algorithm for computing paths, that does not fail at simple fold point.

4.5 Pseudoarclength Continuation

We already mentioned that at fold points on a regular path, Newton's method fails during natural parameter continuation. The main idea in pseudoarclength continuation is to drop the natural parametrization by λ and use some other parameterization. Consider the equation:

$$G(u, \lambda) = 0, \tag{4.6}$$

where $G : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$. If (u^0, λ^0) is any point on a regular path and $(\dot{u}, \dot{\lambda}^0)$ is the unit tangent to the path, then we adjoin to (4.6) the scalar normalization:

$$N(u, \lambda, \Delta s) \equiv \dot{u}^{0^T}(u - u^0) + \dot{\lambda}^0(\lambda - \lambda^0) - \Delta S = 0 \tag{4.7}$$

This is the equation of a plane, which is perpendicular to the tangent $(\dot{u}^0, \dot{\lambda}^0)$ at a distance Δs from (u^0, λ^0) (see Fig. 4.1). This plane will
75 intersect the curve Γ if Δs and the curvature of Γ are not too large. That is we solve (4.6) and (4.7) simultaneously for $(u(s), \lambda(s))$. Using Newton's method this leads to the linear system:

$$\begin{bmatrix} G_u^U & G_\lambda^U \\ \dot{u}^{o^T} & \dot{\lambda}^0 \end{bmatrix} \begin{bmatrix} \Delta u^U \\ \Delta \lambda^U \end{bmatrix} = - \begin{bmatrix} G^U \\ N^U \end{bmatrix} \tag{4.8}$$

Here $G_u^U = G_u(u^U, \lambda^U)$, $G_\lambda^U = G_\lambda(u^U, \lambda^U)$, and the iterates are $u^{U+1} = u^U + \Delta u^U$ and $\lambda^{U+1} = \lambda^U + \Delta \lambda^U$.

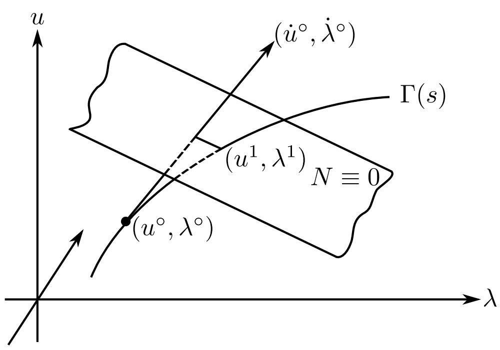


Figure 4.1:

We will give practical procedures for solving the system (4.8) on a regular path. Recall that on such a path G_u^0 may be nonsingular or singular but the $(N+1)$ order coefficient matrix should be nonsingular. One proof of the nonsingular of the coefficient matrices in (4.8) can be based on the following result (see: [7]).

Lemma 4.9. *Let B be a Banach space and let the linear operator $\tilde{A}$:* 76
$B \times \mathbb{R}^U \to B \times \mathbb{R}^U$ have the form:

$$\tilde{A} \equiv \begin{bmatrix} A & B \\ C^* & D \end{bmatrix}, \quad \textit{where} \quad \begin{cases} A : B \to B, B : \mathbb{R}^U \to B, \\ C^* : B \to \mathbb{R}^U, D : \mathbb{R}^U \to \mathbb{R}^U. \end{cases}$$

(i) *If A is nonsingular then $\tilde{A}$ is nonsingular if and only if:*

$$D - C^* A^{-1} B \textit{ is nonsingular.} \tag{4.9a}$$

(ii) *If A is singular with*

$$\dim N(A) = \operatorname{Codim} R(A) = u, \tag{4.9b}$$

then $\tilde{A}$ is nonsingular if and only if:

$$\begin{aligned} &(c_1) \dim R(B) = u, (c_2) R(B) \bigcap R(A) = 0, \\ &(c_3) \dim R(C^*) = u, (c_4) N(A) \bigcap N(C^*) = 0. \end{aligned} \tag{4.9c}$$

(iii) *If A is singular with* $\dim N(A) > u$*, then $\tilde{A}$ is singular*

Here C^* denotes the adjoin of C. In our analysis we use only the cases $u = 1$ and $B \equiv \mathbb{R}^N$. Then the conditions (4.9c) reduce to

$$B \notin R(A) \text{ and } C^T \notin R(A^T). \tag{4.10}$$

where A^T is the transpose of A.

Note. Instead of using the earlier mentioned normalization (4.7), we can use other relation. One obvious generalization of (4.7) is:

$$N \equiv \theta \dot{u}^{0^T}(u - u^0) + (1-\theta)\lambda^0(\lambda - \lambda^0) - \Delta s = 0, 0 < \theta < 1;$$

Another type of normalization is: 77

$$N(u, \lambda, s) = \theta \|u(s) - u(s^0)\|^2 + (1-\theta)|\lambda(s) - \lambda(s^0)|^2 - (s - s^0)^2 = 0.$$

Alternatively if we know nearby point on Γ say at $s = s_0$ and $s = s_{-1}$ then we can use:

$$N(u, \lambda, s) = \theta \left[\frac{u(s_0) - u(s_{-1})}{s_0 - s_{-1}} \right]^T (u(s) - u(s_0)) \\ + (1-\theta) \left[\frac{\dot{\lambda}(s_0) - \dot{\lambda}(s_{-1})}{s_0 - s_{-1}} \right] (\lambda(s) - \lambda(s_0)) - (s - s_0) = 0.$$

This employs a scant rather a tangent. We call all of the above pseudo arclength normalizations. For a references, see [18], [19].

4.11 Bordering Algorithm

We write the coefficient matrix of (4.8) in the form:

(4.12) $$\tilde{A} = \begin{bmatrix} A & b \\ c^T & d, \end{bmatrix}$$

where, A is an $N \times N$ matrix, b, $c \in \mathbb{R}^N$ and $d \in \mathbb{R}$. Then we consider the linear system:

(4.13) $$\tilde{A} \binom{x}{\xi} = \binom{g}{\gamma},$$

where $x, g \in \mathbb{R}^n$ and $\xi, \gamma \in \mathbb{R}$.

Assume $\tilde{A}$ and A are nonsingular. Then solve the linear systems:

(4.14a) $$Ay = b; \quad Az = g.$$

78 Form the solution of (4.13) as:

(4.14b) $$\xi = \frac{\gamma - c^{T_z}}{d - c^{T_y}}; x = z - \xi y.$$

Note that $d - c^T y = d - c^T A^{-1} b$ is the Schur complement of A in $\tilde{A}$. Hence $d - c^T y \neq 0$ if $\tilde{A}$ nonsingular. Thus if $\tilde{A}$ and A are nonsingular, then the bordering algorithm is valid.

Alternatively we may also write the system (4.13) as:

$$\begin{aligned} Ax + b\xi &= g, \\ c^T x + d\xi &= \gamma. \end{aligned}$$

To solve this, we can proceed by first eliminating ξ, if $d \neq 0$ to get:

$$\xi = \frac{1}{d}(\gamma - c^T x).$$

Then for x we have:

$$(A - \frac{1}{d} bc^T)x = g - \frac{\gamma}{d} b.$$

Note that $(A - \frac{1}{d} bc^T)$ is a rank 1 modification of A. Hence from the 'Sherman-Morrison' formula (see lemma 2.29 in Chapter 2), we can easily determine the inverse of $(A - \frac{1}{d} bc^T)$, once we know the inverse of A. In other words we can easily solve the linear system with the coefficient matrix $(A - \frac{1}{d} bc^T)$. But this procedure required $d \neq 0$ while the bordering algorithm does not. The nonsingularity of A is required by both.

Now we will consider the case when A is singular and $\tilde{A}$ is nonsingular. This occurs at simple points on solution paths (see equation (4.2,ii). **79**
That is we assume:

$$\text{(4.15a)} \qquad \begin{aligned} &\text{(i)} \quad N(A) = \text{ span } \{\Phi\}, \\ &\text{(ii)} \quad b \notin R(A) \\ &\text{(iii)} \quad c^T \notin R(A^T). \end{aligned}$$

These are precisely conditions (4.10) and are equivalent to:

$$\text{(4.15b)} \qquad \psi^T b \neq o \text{ and } c^T \Phi \neq 0,$$

where ϕ and ψ are nontrivial solutions of :

$$A\phi = 0 \quad \text{and} \quad A^T\psi = 0. \tag{4.15c}$$

We write the system (4.13) as :

$$\widetilde{A}\binom{x_o}{\xi_o} = \binom{g}{\gamma},$$

where $x_0, g \in \mathbb{R}^N$ and ξ_0, $U \in \mathbb{R}$. That is:

$$\begin{aligned} Ax_0 + b\xi_0 &= g, \\ c^T x_0 + d\xi_0 &= \gamma, \end{aligned} \tag{4.16}$$

Multiplying the first equation by ψ^T, we get:

$$\xi_0 = \frac{\psi^T g}{\psi^T b}. \tag{4.17a}$$

Hence

$$Ax_0 = g - \frac{\psi^T g}{\psi^T b} b \quad \in R(A). \tag{4.17b}$$

All solutions x_0 of (4.17b) have the form:

$$x_0 = x_p + z\phi,$$

80 where x_p is any particular solution of (4.17b) and z is obtained by substituting the value of x_0 into the second equation of (4.16) to get:

$$z = \frac{\gamma - d\xi_0 - c^T x_p}{c^T \phi}.$$

Hence

$$x_0 = \left[x_p - \frac{c^T x_p}{c^T\phi}\phi\right] + \left(\frac{\gamma - d\xi_0}{c^T\phi}\right)\phi. \tag{4.17c}$$

Hence the unique solution of (4.16) is given by (4.17a,c). To evaluate this solution we need the vectors ϕ, ψ, x_p and the inner products $\psi^T g$, $\psi^T b$, $c^T\phi$ and $c^T x_p$.

The operational count to obtain these vectors is only one inner product more than the count required by the Bordering Algorithm. Thus the solution (x_0, ξ_0) requires only two inner products more. We will show how ϕ and ψ can be obtained with half of a back solve each and hence a total of one back solve.

Left and Right Null Vectors of A:

Assume that A_1 is an $N \times N$ matrix satisfying (4.15a) so that rank $(A_1) = N-1$. The with row and column interchanges determined by some permutation matrices say, P and Q, the transformed matrix,

$$A \equiv PA_1Q, \tag{4.18a}$$

has an LU factorization:

$$A = LU \equiv \begin{bmatrix} \hat{L} & 0 \\ \hat{\ell}^T & 1 \end{bmatrix} \begin{bmatrix} \hat{U} & \hat{u} \\ \hat{0}^T & 0 \end{bmatrix} \tag{4.18b}$$

Here $\hat{L}$ and $\hat{U}$ are lower and upper triangular matrices, respectively, **81**
of order $(N-1)\times(N-1)$ with:

$$\hat{L} = \begin{bmatrix} 1 & & & & \\ \cdots & 1 & 0 & & \\ \cdots & \cdots & 1 & & \\ \cdots & \cdots & \cdots & 1 & \\ \cdots & \cdots & \cdots & \cdots & 1 \end{bmatrix} \qquad \hat{U} = \begin{bmatrix} u_{11} & \cdots & \cdots & \cdots & \cdots \\ & u_{22} & \cdots & \cdots & \cdots \\ & 0 & \cdots & \cdots & \cdots \\ & & & \cdots & \cdots \\ & & & & u_{N-1,N-1} \end{bmatrix}$$

Moreover: $\hat{0}, \hat{u}, \hat{\ell} \in \mathbb{R}^{N-1}$. Of course in actual calculations we do not get the exact zero element in the final diagonal position of U. First we discuss the null vectors and the we will discuss the inexact factorization.

Since L is nonsingular, $A\phi = 0$ if and only if $U\phi = 0$. So with $\hat{\phi} \in \mathbb{R}^{N-1}$ and $\alpha \in \mathbb{R}$, we seek ϕ in the form

$$\phi = \alpha \begin{pmatrix} \hat{\phi} \\ -1 \end{pmatrix}, \quad \alpha \neq 0. \tag{4.19a}$$

It follows because of the nonsingularity of $\hat{U}$ that $\hat{\phi}$ is uniquely determined by

$$\hat{U}\hat{\phi} = \hat{u}.$$

In other words:

$$\hat{\phi} = \hat{U}^{-1}\hat{u}.$$

Since U is in triangular form, we obtain $\hat{\phi}$ and hence ϕ with only **82** one half of a back solve (for a system of order $N - 1$). Similarly, the nonsingularity of $\hat{U}$ implies that $A^T\psi = 0$ if and only if $L^T\psi = \beta\begin{pmatrix}\hat{0}\\ -1\end{pmatrix}$ for $\beta \in \mathbb{R}$. Thus we find that all nontrivial left null vectors are given by:

$$\psi = \begin{pmatrix}\hat{\psi}\\ -1\end{pmatrix}, \quad \beta \neq 0, \tag{4.19b}$$

and $\hat{\psi} \in \mathbb{R}^{N-1}$ is uniquely determined by:

$$\hat{L}^T\hat{\psi} = \ell,$$
$$\hat{\psi} = (\hat{L}^T)^{-1}\hat{\ell}.$$

Again $\hat{\psi}$ and hence ψ are obtained with half of a back solve.

Almost Singular A

We already mentioned that in calculations we do not obtain the exact factorization, but rather an approximation of the form:

$$A = A_\varepsilon = L_\varepsilon U_\varepsilon = \begin{pmatrix}\hat{L} & \hat{o}\\ \hat{\ell}^T & 1\end{pmatrix}\begin{pmatrix}\hat{U} & \hat{u}\\ \hat{0}^T & \varepsilon\end{pmatrix}. \tag{4.20}$$

The quantity ε will be an approximation to zero. If we use full pivoting to determine the permutation matrices P and Q in (4.18a), then under appropriate conditions on A_1 we can bound ε by $C10^{-t}$ for t digit arithmetic where C is a constant. The error analysis of Wilkinson [32] can also be used to estimate the magnitude of ε.

The basic assumptions that we make about the algorithm used to get the form (4.20) and the error growth allowed, are summarized by the requirement that in the singular case (4.15a):

$$\max_{j<N}\left|\frac{\varepsilon}{u_{jj}}\right| \ll 1.$$

83 In actual computations some precise relation must be used if we

are to declare that we are in the singular case. With partial pivoting in columns, one reasonable test is:

$$\min_{j<N}\left|\frac{u_{jj}}{u_{j-1,j-1}}\right| > 10^2\left|\frac{\varepsilon}{u_{N-1,N-1}}\right|.$$

Of course the factor 10^2 may vary from case to case. A better theory is needed here.

Now we use the factorization (4.20) and apply the Bordering Algorithm to solve (4.13). So consider

$$(4.21)\qquad a)\ A_\varepsilon y_\varepsilon = b,\quad (b)\ A_\varepsilon z_\varepsilon = g,$$

with

$$b = \begin{bmatrix}\hat{b}\\ b_N\end{bmatrix},\quad g = \begin{bmatrix}\hat{g}\\ g_N\end{bmatrix}.$$

Now using ϕ and ψ obtained from (4.19a,b) with $\alpha = \beta = 1$, we can easily see that:

$$(4.22)\qquad \begin{aligned} &\text{(a) } y_\varepsilon = \begin{bmatrix}(\hat{L}\,\hat{U})^{-1}\hat{b}\\ 0\end{bmatrix} + \frac{\psi^T b}{\varepsilon}\phi\\ &\text{(b) } z_\varepsilon = \begin{bmatrix}(\hat{L}\,\hat{U})^{-1}\hat{g}\\ 0\end{bmatrix} + \frac{\psi^T g}{\varepsilon}\phi \end{aligned}$$

Now form (as in (4.14b))

$$(4.23)\qquad \begin{aligned} &\text{(a) } \xi_\varepsilon = \frac{\gamma - \hat{c}^T(\hat{L}\,\hat{U})^{-1}\hat{g} - \frac{1}{\varepsilon}(\psi^T g)(c^T\phi)}{d - \hat{c}^T(\hat{L}\,\hat{U})^{-1}\hat{b} - \frac{1}{\varepsilon}(\psi^T b)(c^T\phi)},\\ &\text{(b) } x_\varepsilon = \begin{bmatrix}(\hat{L}\,\hat{U})^{-1}(\hat{g} - \xi_\varepsilon\hat{b})\\ 0\end{bmatrix} + \frac{1}{\varepsilon}[(\psi^T g) - \xi_\varepsilon(\psi^T b)]\phi. \end{aligned}$$

We must compare the solution (4.23) with the exact solution for the singular case (4.17a,c). To do this, identify the particular solution x_p as: **84**

$$x_p \equiv \begin{bmatrix}(\hat{L}\,\hat{U})^{-1}(\hat{g} - \xi_0\hat{b})\\ 0\end{bmatrix}.$$

Now we can expand (4.23) about $\varepsilon = 0$ to obtain the results :

$$\xi_\varepsilon = \xi_0 + 0(\varepsilon),$$
$$x_\varepsilon = x_0 + 0(\varepsilon).$$

In more detail we have :

$$\begin{aligned}\xi_\varepsilon &= \frac{\varepsilon(\gamma - \hat{c}^T(\hat{L}\,\hat{U})^{-1}\hat{g}) - (\psi^T g)(c^T\phi)}{\varepsilon(d - \hat{c}^T(\hat{L}\,\hat{U})^{-1}\hat{b}) - (\psi^T b)(c^T\phi)},\\ &= \frac{\varepsilon(\gamma - \hat{c}^T(\hat{L}\,\hat{U})^{-1}\hat{g})}{\varepsilon(d - \hat{c}^T(\hat{L}\,\hat{U})^{-1}\hat{b}) - (\psi^T b)(c^T\phi)}\\ &\quad - \frac{(\psi^T g)(c^T\phi)}{\varepsilon(d - \hat{c}^T(\hat{L}\,\hat{U})^{-1}\hat{b}) - (\psi^T b)(c^T\phi)},\\ &= \frac{\psi^T g}{\psi^T b} + 0(\varepsilon) = \xi_0 + 0(\xi).\end{aligned}$$

Thus as $\varepsilon \to 0$, the first term of the right hand side of (4.23b) converges to x_p. Also :

$$\begin{aligned}\frac{1}{\varepsilon}[\psi^T g - \xi(\psi^T b)]\phi &= \frac{1}{\varepsilon}\left[\frac{\psi^T g.\varepsilon(d - \hat{c}^T(\hat{L}\hat{U})^{-1}\hat{b}) - \varepsilon(\gamma - \hat{c}^T(\hat{L}\hat{U})^{-1}\hat{g})\psi^T b}{\varepsilon(d - \hat{c}^T(\hat{L}\hat{U})^{-1}\hat{b})(\psi^T b)(c^T\phi)}\right]\phi\\ &= (\psi^T b)\frac{\left[d.\frac{\psi^T g}{\psi^T b} - \gamma + c^T x_p\right]\phi}{\varepsilon\left[d - \hat{c}^T(\hat{L}\hat{U})^{-1}\hat{b}\right] - (\psi^T b)(c^T\phi)}\\ &= \frac{-d\xi_0 + \gamma - c^T x_p}{c^T\phi}\phi + 0(\varepsilon).\end{aligned}$$

Hence

$$x_\varepsilon = x_0 + 0(\varepsilon).$$

85 Note that in the calculation the significant terms $\pm(\psi^T g)(\psi^T b)(c^T\phi)$ cancelled each other. Here ξ_0, x_0 are the exact exact solutions when $\varepsilon = 0$.

Thus we find that the Bordering Algorithm can be used to solve (4.13) whenever $\widetilde{A}$ is nonsingular. If A happens to be singular, then results of some accuracy will be obtained only if a reasonable pivoting strategy is used. Even in this case some accuracy loss must be expected due to the cancellation of significant digits that occurs in forming x as in (4.14b). This cancellation is exactly analogous to the cancellation of the $\dfrac{1}{\varepsilon}$ terms in x_ε of (4.23b). If in the course of the calculations it is recognized that we are at a limit point then the singular A algorithm can be used and a more accurate numerical solution will result as no extra significant digits are lost (see [17] for more details). We will give some numerical examples in the last chapter, in which we have used the above algorithm and they have performed well.

The Tangent Vectors

We will briefly describe how to compute the tangent vectors $(\dot{u}^0, \dot{\lambda}^0)$. They must satisfy :

$$\text{(4.24)} \qquad \begin{aligned} &\text{(a) } G_u^\circ \dot{u}^\circ + G_\lambda^\circ \dot{\lambda}^\circ = 0, \\ &\text{(b) } \|\dot{u}^\circ\|^2 + |\dot{\lambda}^\circ|^2 = 1. \end{aligned}$$

First we consider regular points, where G_u° is nonsingular. We find ϕ_0. from:

$$\text{(4.25)} \qquad G_u^\circ \phi_0 = -G_\lambda^\circ.$$

Then set: 86

$$\text{(4.26)} \qquad \dot{u}^\circ = a\phi_0 \quad \text{and} \quad \dot{\lambda}_0 = a$$

where a is determined from (4.24b) as :

$$\text{(4.27)} \qquad a = \frac{\pm 1}{\sqrt{1 + \|\phi_0\|^2}}.$$

The sign of a is chosen so that the orientation of the path is preserved. More precisely, if $(\dot{u}_{-1}, \dot{\lambda}_{-1})$ is the preceding tangent vector then we require

$$\dot{u}_{-1}^T \dot{u}^0 + \dot{\lambda}_{-1} \dot{\lambda}^0 > 0.$$

Thus we choose the sign of 'a' so that

$$a[\dot{u}_{-1}^T \phi_0 + \dot{\lambda}_{-1}] > 0.$$

Choosing the sign of a is very important in numerical calculations. If we do not choose the sign of a properly, either we will get trapped in the iterations at some point or it will reverse the direction and hence it will compute the same path already computed.

Another important point to recall in actual calculations is that the quantity $||\dot{u}||^2$ is usually meant to approximate some L_2 norm of the continuous formulation of the problem. Thus the net spacing, say h, must be used to form for example :

$$||\dot{u}||^2 = \sum_{j=1}^{J} h^d \dot{u}_j^2.$$

Here $\dot{u}_j$ represent the components of $\dot{u}$ and the underlying contin-
87 uous problem is assumed formulated over a domain in $\mathbb{R}^d$. If this is not done the arclength definition $||\dot{u}||^2 + \dot{\lambda}^2$ is biased too much in the u-subspace and λ is not very significant.

We return to the case of a simple fold at which G_u° is singular. The analysis of the case of almost singular A, (4.20)–(4.22), shows that we get in this case for the solution of (4.25), by setting

$$\phi_0 = \begin{bmatrix} \hat{\phi}_0 \\ \omega_0 \end{bmatrix}, \quad -G_\lambda^\circ = \begin{bmatrix} \hat{g} \\ \gamma \end{bmatrix}; \tag{4.28a}$$

The result :

$$\begin{aligned} \hat{\phi}_0 &= -(\hat{L}\,\hat{U})^{-1}\hat{g} + \left(\frac{\gamma - \hat{\psi}^T \hat{g}}{\varepsilon}\right)\hat{\phi}, \\ \omega_0 &= -\left(\frac{\gamma - \hat{\psi}^T \hat{g}}{\varepsilon}\right) \equiv \frac{\beta}{\varepsilon} \quad \text{(say)} . \end{aligned} \tag{4.28b}$$

Using these results (4.28) in (4.26) and (4.27) we find that we indeed get the tangent to within $0(\varepsilon)$.

Chapter 5

Singular Points and Bifurcation

5.1 Introduction

88

First consider the problem in the uniform formulation, viz .

(5.2) $$G(X) = p.$$

In this formulation we will consider paths which contain singular points and give methods to jump over such points. This includes bifurcation points in the general case.

5.3 Definition

A point $X(s_0)$ on a (smooth) solution path of (5.2) say,

$$\Gamma_{ab} \equiv \left\{ X(s) : X(s) \in \mathbb{R}^{N+1}, \quad G(X)(s)) = p, \quad p \in \mathbb{R}^N, \quad a < s < b \right\}$$

is a simple singular point if $s_0 \in (a, b)$ and Rank $G_x(X(s_0)) = N - 1$.

Note that here s may be any parameter, it need not be arclength. Since G_X is an $N \times (N + 1)$ matrix, at a simple singular point $X(s_0)$, G_X has two independent null vectors, say, Φ_1 and Φ_2 in $\mathbb{R}^{(N+1)}$. Without loss of generality we can require

$$\Phi_i^T \Phi_j = \delta_{ij}; \quad i, j = 1, 2,$$

thus introducing an orthogonal system of co-ordinates for $N\{G_X(X(s_0))\}$. Now $G_X^T(X(s_0))$ is an $(N+1)\times N$ matrix of rank $N-1$, so that,

$$N\left\{G_X^T(X(s_0))\right\} = \text{span}\{\psi\},$$

89 for some nontrivial $\psi \in \mathbb{R}^N$.

Tangents to Γ_{ab}

From (5.2), it follows that:

$$G_X(X(s))\dot{X}(s) = 0.$$

Hence at $s = s_0$, $\dot{X}(s_0)$ has the form :

$$\dot{X}(s_0) = \alpha\phi_1 + \beta\phi_2, \quad \alpha, \beta \in \mathbb{R}.$$

Differentiating again we have :

$$G_X(s)\ddot{X}(s) + G_{XX}(s)\dot{X}(s)\dot{X}(s) = 0.$$

Now multiplying by ψ^T and evaluating at $s = s_0$, the first terms vanishes and we have:

$$\psi^T G_{XX}(s_0)\dot{X}(s_0)\dot{X}(s_0) = 0.$$

Substituting for $\dot{X}(s_0)$, we get :

$$a_{11}\alpha^2 + 2a_{12}\alpha\beta + a_{22}\beta^2 = 0. \tag{5.4a}$$

Here a_{ij}'s are given by:

$$a_{ij} = \psi^T G_{XX}(s_0)\phi_i\phi_j. \tag{5.4b}$$

Since (5.4a) is a quadratic equation, it follows that the roots are governed by the discriminant:

$$\Delta = a_{12}^2 - a_{11}a_{22} \begin{cases} > 0, & \text{two real roots,} \\ = 0, & \text{one real root,} \\ < 0, & \text{no real root.} \end{cases} \tag{5.4c}$$

90 Since Γ_{ab} is a smooth solution path, it has at least one nontrivial tangent at each point on Γ_{ab}. Hence the case $\Delta < 0$ is not possible. So the following lemma holds.

Lemma 5.5. *At a simple singular point $X(s_0)$ on a smooth solution path Γ_{ab} either $\Delta > 0$ or $\Delta = 0$.*

If $\Delta > 0$, then there exist two nontrivial tangents at a singular point. This suggests that bifurcation occurs at that point. Also it gives us an idea for constructing solution paths and switching over from one branch to another.

As in the earlier chapter, we adjoin some scalar normalization

$$N(X, s) = 0,$$

to the equation:

$$G(X) = 0,$$

where $X \in \mathbb{R}^{N+1}$, $s \in \mathbb{R}$ to obtain the augmented system from $\mathbb{R}^{N+2} \to \mathbb{R}^{N+1}$:

$$F(X, s) \equiv \begin{bmatrix} G(X) \\ N(X, s) \end{bmatrix} = 0. \tag{5.6a}$$

Using the normalization

$$N(X, s) = \dot{X}^T(s_0)(X - X(s_0)) - (s - s_0), \tag{5.6b}$$

we get:

$$F_X(X(s)) = \begin{bmatrix} G_X(X(s)) \\ \dot{X}^T(s_0) \end{bmatrix}. \tag{5.6c}$$

F_X is an $(N + 1) \times (N + 1)$ matrix. In the previous chapter we proved 91
that at a simple limit point, the augmented system is nonsingular (using the normalization in (5.6b)). Although G_X has a two dimensional null space, we will show that F_X has only one dimensional null space at a simple singular point.

Lemma 5.7. $\dim N\{F_X(X(s_0)\} = 1$ *at a simple singular point* $X(s_0)$ *on a solution path* Γ_{ab}.

Proof. We use the notation $F_X(X(s_0)) \equiv F_X(s_0)$ and suppose

$$F_X(s_0)\phi = 0.$$

Then the first N equations are:

$$G_X(s_0)\phi = 0,$$

which implies that for some $\alpha_0, \alpha_1 \epsilon \mathbb{R}$:

$$\phi = \alpha_0 \Phi_1 + \alpha_1 \Phi_2.$$

The last equation is :

$$\dot{X}^T(s_0)\phi = 0,$$

which implies by the orthogonality of Φ_1 and Φ_2:

$$\alpha\alpha_0 + \beta\beta_0 = 0.$$

92 That is:

$$\alpha_0 : \beta_0 = -\beta : \alpha.$$

This shows that there is, upto a scalar factor, only one null vector of $F_X(s_0)$. This completes the proof of the lemma. ◻

We recall that:

$$\det F_X(s_0) = \prod_{j=1}^{N+1} n_j(s_0),$$

where $n_j(s_0)$ are eigenvalues of $F_X(s_0)$. Since $F_X(s_0)$ has a one dimensional null space at least one of the eigenvalues is zero. We assume only one is zero. i.e. zero is an algebraically simple eigenvalues. Assume that $n_1(s_0) = 0$ and $n_j(s_0) \neq 0$ for $j = 2, \ldots N+1$. Now consider $\det F_X(X)$ for $\| X - X(s_0) \|$ small and let the corresponding eigenvalues be $n_j(X)$. Suppose

$$\nabla_X n_1(X)|_{X=X(s_0)} \neq 0.$$

Then we can apply the implicit function theorem to show that $n_1(X) = 0$ on a smooth manifold M of dimension N. Thus $\det F_X = 0$ on M. If the solution path is transversal to M, i.e. the tangent at the point of intersection makes an acute angle with the normal to the manifold, then the path crosses M. So the sign of $\det F_X(X(s))$ changes along the path. Then theorem 3.17 allows us to conclude that the point of intersection is a bifurcation point.

Now we compute the angle between the tangent $X(s_0)$ to Γ_{ab} at $s = s_0$ and the normal to the above manifold M at $s = s_0$, viz.$\nabla_X n_1(X(s_0))$. Let $\phi(X(s))$ denote the eigenvector corresponding to $n_1(X(s))$. Then **93**

$$F_X(s))\Phi(X(s)) = n_1(X(s))\Phi(X(s))$$

Differentiating this expression and evaluating at $s = s_0$, where $n_1(X(s_0)) = 0$, yields

$$F_{XX}\dot{X}\Phi + F_X\Phi_X\dot{X} = < \nabla n_1, \dot{X} > \Phi$$

Let

$$N\{F_X^T|_{s=s_0}\} = \text{ span } \{\Psi\}.$$

Then taking innerproduct with Ψ,

$$< \Psi, F_{XX}\dot{X}\Phi > = < \nabla n_1, \dot{X} > < \Psi, \Phi > .$$

To evaluate the required angle $< \nabla n_i, \dot{X} >$, we need to find Φ and Ψ. Lemma 5.7 shows

$$\Phi = \beta\Phi_1 - \alpha\Phi_2$$

Also we have

$$\Psi = (\Psi, 0), \qquad \Psi \in N\{G_X^T(X(s_0))\}$$

and hence using the notations of (5.4a,b,c,),

$$\begin{aligned} < \Psi, F_{XX}\dot{X}\Phi > &= < \Psi, G_{XX}(\alpha\Phi_1 + \beta\Phi_2)(\beta\Phi_1 - \alpha\Phi_2) > \\ &= \alpha\beta a_{11} + (\beta^2 - \alpha^2)a_{12} - \alpha\beta a_{22} \\ &= (\beta^2 - \alpha^2)(a_{11}\frac{\alpha}{\beta} + a_{12}) = (\beta^2 - \alpha^2)\sqrt{\Delta} \end{aligned}$$

Since at $s = s_0$, zero is an algebraically simple eigenvalue of $F_X(X(s_0))$,

$$< \Psi, \Phi > \neq 0.$$

94

Thus we deduce that the required angle is not zero if $\Delta > 0$ and hence the point $s = s_0$ on the solution path is a bifurcation point.

By choosing the basis vectors for $N\{G_X(X(s_0))\}$ suitably, the above computation can be simplified considerably. For example, by taking

$$N\{G_X(X(s_0))\} = \text{ span } \{\Phi_1, \dot{X}\},$$

we get

$$< \Psi, F_{XX}\dot{X}\Phi >= a_{12}$$

This is done in detail in the bifurcation theorem below.

5.8 Bifurcation Theorem

Let $X^0 = X(s_0)$ *be a simple singular point on a smooth solution path*

$$\Gamma_{ab} = \{X = X(s) \in \mathbb{R}^{N+1}, \quad G(X(s)) = p, p \in \mathbb{R}^N, \quad s_a < s < s_b\}.$$

Let $\Delta > 0$ *and 0 be an algebraically simple eigenvalue of* $F_X(X(s_0))$ *(defined as in* (5.4c) *and* (5.6c) *respectively). Then* X^0 *is a bifurcation point on* Γ_{ab}

Proof. Consider the system:

$$F(X, s) \equiv \begin{bmatrix} G(X) \\ N(X, s) \end{bmatrix} = 0$$

Here s is the parameter used to define Γ_{ab}. We will show that F has
95 a bifurcation at $s = s_0$ which in turn will prove that G has a bifurcation at $s = s_0$. Consider the normalization:

$$N(X, s) = \dot{X}(s_0)^T[X - X(s_0)] - t(s)$$

where $t(s)$ is the distance between $X(s_0)$ and the projection of $X(s)$ onto the tangent to Γ_{ab} at $X(s_0)$. See Fig. 5.1.

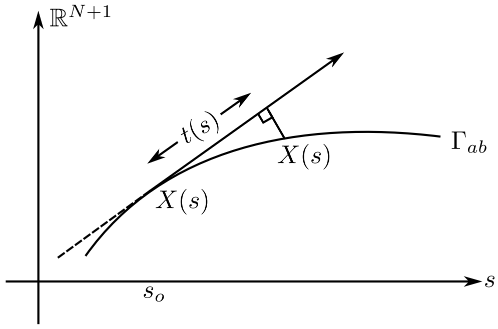


Figure 5.1:

With the indicated normalization, $X(s)$ is a solution of $F(x, s) = 0$. We have :

$$F_X(X(s)) = \begin{bmatrix} G_X(s_0) \\ \dot{X}(s_0)^T \end{bmatrix}$$

At a singular point G_X has a two dimensional null space and the tangent vector, $\dot{X}(s_0)$, is in $N\{G_X(s_0)\}$. Hence $\dot{X}(s_0) = \alpha\Phi_1 + \beta\Phi_2$, for some $\alpha, \beta \in \mathbb{R}$. We choose the basis vectors Φ_1 and Φ_2 of $N\{G_X(s_0)\}$ such that $\alpha = 0, \beta = 1$. We also proved that $N\{G_X(s_0)\}$ has dimension **96**
one and hence by the choice of $X(s_0) = \Phi_2$ we must have:

$$F_X\{X(s_0)\}\Phi_1 = 0$$

Now consider the eigenvalue problem:

$$F_X(X(s))\Phi(s) = n(s)\Phi(s).$$

At $s = s_0$, one of the eigenvalues is zero, say, $n(s_0) = 0$ and so $\Phi(s_0) = \Phi_1$. F_X has a left null vector Ψ^T which is given by $\Psi^T = (\Psi^T, 0)$ where Ψ^T is a left null vector of G_X^0. Now we can easily show that:

$$\dot{n}(s_0) = \frac{\Psi^T F_{XX}^0 \Phi_1 \Phi_2}{\Psi^T \Phi_1} = \frac{a_{12}}{\Psi^T \Phi_1}.$$

Since $\alpha = 0, \beta = 1$ is a root of the equation:

$$a_{11}\alpha^2 + 2a_{12}\alpha\beta + a_{22}\beta^2 = 0$$

we must have:

$$a_{22} = 0$$

Since $\Delta > 0$, this implies that $a_{12} \neq 0$. Hence $\dot{n}(s_0) \neq 0$. Therefore $\det F_X$ changes sign at $s = s_0$. So by theorem 3.17, F and hence G has a bifurcation at $X(s_0)$. □

Note. This theorem gives conditions under which a point on the solution path is a bifurcation point. But we do not obtain two smooth branches of solution. This can be done using the Lyapunov-Schmidt method which we are not going to discuss here (Refer [4], [26]).

97 Now consider the parameter formulation of the problem:

$$G(u, \lambda) = 0 \tag{5.9}$$

5.10 Definition

A point $(u(s_0), \lambda(s_0))$ on a solution path Γ_{ab} is a simple singular point if and only if $\mathrm{Rank}[G_u(s_0), G_\lambda(s_0)] = N - 1$.

At a simple singular point, since Rank $[G_u(s_0), G_\lambda(s_0)] = N - 1$, $N\{G_u(s_0)\}$ has dimension either one or two. If it is one then $G_\lambda(s_0) \in R(G_u)$ and if it is two, then $G_\lambda(s_0) \notin R(G_u)$. Conversely, if either $N\{G_u(s_0)\}$ has dimension one and $G_\lambda(s_0) \in R(G_u)$ or $N(G_u(s_0))$ has dimension two and $G_\lambda(s_0) \notin R(G_u)$, then $(u(s_0), \lambda(s_0))$ is a simple singular point. So we have proven:

Lemma 5.11. *The point $(u(s_0), \lambda(s_0))$ is a simple singular point if and only if either:*

$$\text{(5.11a)} \qquad \begin{array}{rl} \text{(i)} & \dim N\{G_u(s_0)\} = 1, \\ \text{(ii)} & G_\lambda(s_0) \in R(G_u). \end{array}$$

or:

$$\text{(5.11b)} \qquad \begin{array}{rl} \text{(i)} & \dim N\{G_u(s_0)\} = 2, \\ \text{(ii)} & G_\lambda(s_0) \in R(G_u). \end{array}$$

This case (5.11b) is similar to the case we discussed in the uniform formulation, since $N(G_u(s_0))$ has two independent null vectors and we can proceed as above. Now consider the case (5.11a) and let 98

$$N\{G_u(s_0)\} = \text{ span }\{\Phi\},$$

$$N\{G_u^T(s_0)\} = \text{ span }\{\Psi\}.$$

From (5.11a(ii)), the equation:

$$G_u(s_0)\phi_0 = -G_\lambda(s_0),$$

has a solution and it can be made unique by requiring that $\phi^T\phi_0 = 0$. Note that the tangent vector is

$$\dot{X}(s_0) = \begin{bmatrix} \dot{u}(s_0) \\ \dot{\lambda}(s_0) \end{bmatrix}.$$

Then we have for some $\alpha, \beta \in \mathbb{R}$:

$$\dot{\lambda}(s_0) = \beta, \dot{u}(s_0) = \alpha\phi + \beta\phi_0.$$

where:

$$\dot{X}(s_0) = \alpha\Phi_1 + \beta\Phi_2$$

$$\Phi_1 = \binom{\phi}{0},\ \Phi_2 = \binom{\phi_0}{1}.$$

From (5.9) with $(u, \lambda) = (u(s), \lambda(s))$, we get on differentiating twice and setting $s = s_0$:

$$G_u^0\ddot{u}_0 + G_\lambda^0\ddot{\lambda}_0 + (G_{uu}^0\dot{u}_0\dot{u}_0 + 2G_{u\lambda}^0\dot{u}_0\dot{\lambda}_0 + G_{\lambda\lambda}^0\dot{\lambda}_0\dot{\lambda}_0) = 0.$$

Multipling by Ψ^T, we get,

$$a_{11}\alpha^2 + 2a_{12}\alpha\beta + a_{22}\beta^2 = 0, \tag{5.12a}$$

where now:

$$\begin{aligned} a_{11} &= \Psi^T G_{uu}(s_0)\phi\phi \\ a_{12} &= \Psi^T[G_{uu}(s_0)\phi_0 + G_{u\lambda}(s_0)]\phi \\ a_{22} &= \Psi^T[G_{uu}(s_0)\phi_0\phi_0 + 2G_{u\lambda}(s_0)\phi_0 + G_{\lambda\lambda}(s_0)]. \end{aligned} \tag{5.12b}$$

99

Again if $\Delta > 0$, (5.12a) has two real roots and if $\Delta = 0$, it has one real root. It can be shown that $\Delta > 0$, then each root (α^*, β^*) of (5.12a,b) generates a smooth solution are $(u(s), \lambda(s))$ for s and s_0 of the form:

$$\begin{aligned} &u(s) = u_0 + (s - s_0)[\alpha(s)\phi_0 + \beta(s)\phi_1] + (s - s_0)^2 v(s), \\ &\lambda(s) = \lambda_0 + (s - s_0)\alpha(s), \end{aligned}$$

where,

$$\begin{aligned} \Psi^T v(s) &= 0, \\ \alpha(s_0) &= \alpha^*, \\ \beta(s_0) &= \beta^*. \end{aligned}$$

For details see [7]. This result is well known in other forms. See [5], [26].

5.13 Continuation Past Simple Singular Points

Let $\Gamma_{ab} = \{X(s) : X(s) \in \mathbb{R}^{N+1},\ G(X(s)) = 0, s_a < s < s_b\}$ be a smooth path. Assume that at $s = s_0 \in (s_a, s_b)$, the point $X(s_0)$ is an isolated simple singular point, that is rank of $G_X(X(s)) = N$ in the intervals $[s_a, s_0)$ and $(s_0, s_b]$ and the rank is $N - 1$ at $s = s_0$.

100 Let

$$F(X, s) \equiv \begin{bmatrix} G(X) \\ N(X, s) \end{bmatrix} = 0,$$

where,

$$N(X, s) \equiv \dot{X}^T(s_a)[X - X(s_a)] - (s - s_a).$$

We try to construct a solution using the Newton iteration method :

$$\text{(5.14)} \qquad \begin{aligned} &\text{(a)}\ \ A_U(s) \equiv F_X(X_U(s), s), \\ &\text{(b)}\ \ A_U(s)[X_{U+1}(s) - X_U(s)] \equiv -F(X_U(s), s), \\ &\qquad U = 0, 1, 2, \ldots, \end{aligned}$$

with the initial estimate $X_0(s)$ as:

$$X_0(s) = X(s_a) + (s - s_a)\dot{X}(s_a).$$

To assure convergence, we have to show that $X_0(s)$ is in an appropriate domain of convergence. Recall the Newton-Kantorovich theorem 2.23: we get convergence under the assumptions that, for some $s \neq s_0$ in (s_a, s_b):

(a) $F(X(s), s) = 0$,

(b) $\|F_X^{-1}(X, (s), s)\| \leq \beta(s)$,

(c) $\|F_X(X, s) - F_X(Y, s)\| \leq \gamma(s)\|X - Y\|$, for all $X, Y \in B_{\rho(s)}(X(s)), \rho(s) > 0$,

(d) $\rho(s) < \dfrac{2}{3\beta(s)\gamma(s)}$.

Of course as $s \to s_0$, $\|F_X^{-1}(X(s), s)\| \to \infty$. But in the case simple bifurcation point, it can be shown that $\|F_X^{-1}(X(s), s)\| \leq \dfrac{M_0}{|s - s_0|}$ for some $M_0 > 0$, $s \neq s_0$ (See [6]). This shows that there is a full conical neighbourhood, with positive semiangle about the solution are through $X(s_0)$, **101**
and vertex at $X(s_0)$, in which $F_X(Y, s)$ is nonsingular. See Figure 5.2. Note that the tangent $\dot{X}(s_a)$ departs from one cone at the point A and penetrates at B the other cone. We have already seen that for all initial values within this conical neighbourhood, the iterates converge.Hence this allows us to continue our procedure without any trouble at the singular point. The point $X(s_0)$ can be determined by a bisection procedure with $s_1 = s_a$ and $s_U < s_0 < s_{U+1}$, for $U = 1, 2, 3, \ldots$ Each new tangent line through the new solution $X(s_U)$ will have smaller chord lying outside the cone. In the limit the tangent through $X(s_0)$ the bifurcation point, is entirely contained within the cone (locally). The final configuration or a close approximation to it, gives one of the best techniques for computing the bifurcation branch by merely switching the tangent to be used in the normalization. See Figure 5.3. We will discuss later in this chapter how to find the new tangent

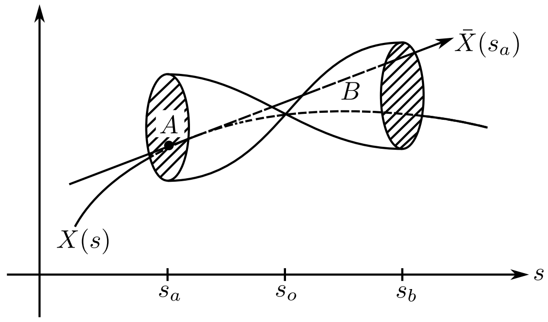


Figure 5.2:

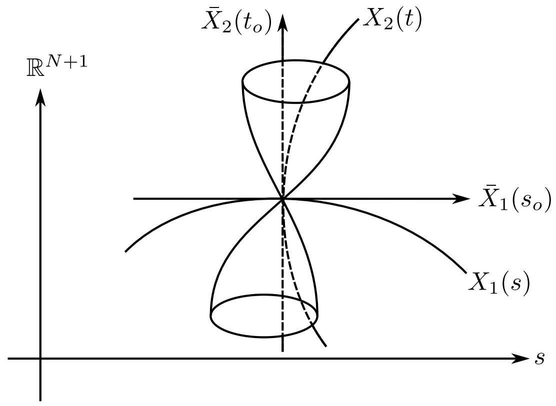


Figure 5.3:

102

5.15 Properties at Folds and Bifurcation Points on Paths

For the parameter formulation

$$G(u, \lambda) = p, \tag{5.16}$$

we consider the eigenvalue problem:

$$(5.17)\qquad \begin{aligned} &\text{(a)}\;\; G_u(s)\phi(s) = n(s)\phi(s),\\ &\text{(b)}\;\; \|\phi(s)\|^2 = 1. \end{aligned}$$

along a solution path of (5.16):

$$\Gamma_{ab} \equiv \{(u(s), \lambda(s)) : G(u(s), \lambda(s)) = p, \quad s_a < s < s_b\}$$

Note that at an algebraically simple eigenvalue, $\phi(s)$ and $n(s)$ are C^∞ functions if G is C^∞. Assume that the problem has a simple limit point **103**
or a simple singular point of type (5.11a) at $s = s_0 \in (s_a, s_b)$. Then we have:

$$N\{G_u^T(s_0)\} = \text{span}\{\psi\},$$
$$N\{G_u(s_0)\} = \text{span}\{\phi\},$$

for some $\psi,\ \phi \in \mathbb{R}^N$. Since $G_u(s_0)$ is singular and the null space is spanned by ϕ, we must have, for some $\{\phi(s), n(s)\}$:

$$n(s_0) = o, \phi(s_0) = \phi$$

Now differentiating (5.17a) twice and multiplying by ψ^T and setting $s = s_0$, we get:

$$(5.18)\qquad \psi^T[G_{uu}(s_0)\dot{u}(s_0)\phi + G_{u\lambda}(s_0)\dot{\lambda}(s_0)\phi] = \dot{n}(s_0)\psi^T\phi.$$

Observe that if the eigenvalue is algebraically simple then $\psi^T\phi \neq 0$. So in this case we can solve for $\dot{n}(s_0)$. We use this in the proof of the following lemma.

Lemma 5.19. *Let $(u(s_0), \lambda(s_0))$ be a simple quadratic fold on Γ_{ab}. Assume that $n(s_0) = 0$ is an algebraically simple eigenvalue of $G_u(s_0)$. Then $\dot{n}(s_0) \neq 0$ and $\det G_u(s_0)$ changes sign at $s = s_0$.*

Proof. At a fold point,

$$\dot{\lambda}(s_0) = 0, \dot{u}(s_0) = \phi.$$

Hence we have: **104**

$$\dot{n}(s_0) = \frac{\psi^T G_{uu}(s_0)\phi\phi}{\psi^T \phi} \neq 0,$$

because $\ddot{\lambda}(s_0) = \frac{\psi^T G_{uu}(s_0)\phi\phi}{\psi^T G_\lambda} \neq 0$ and this implies that $\psi^T G_{uu}(s_0)\phi\phi \neq 0$. Also if $n_j(s)$ are the eigenvalues of $G_u(s)$, then,

$$\det G_u(s) = \prod_{j=1}^{N} n_j(s).$$

Without loss of generality, assume that :

$$n(s_0) = n_1(s_0), n_j(s_0) \neq o \forall j = 2, \ldots N.$$

Since the $n_j(s)$ are continuous, $n_1(s)$ changes sing at $s = s_0$ and all other $n_j(s)$, for $j = 2, \ldots N$ do not change sign in a neighbourhood of s_0. Hence the lemma follows □

Lemma 5.20. *Let* $(u(s_0), \lambda(s_0))$ *be a simple bifurcation points on a smooth path* Γ_{ab} *and* $n(s_0) = 0$ *be an algebraically simple eigenvalue of* $G_u(s_0)$. *Let the discriminant* Δ *of equation (5.12a) be positive. Then* $\dot{n}(s_0) \neq 0$ *and* $\det G_u(s)$ *changes sign at* $s = s_0$ *on one or both the branches through the bifurcation point* $(u(s_0), \lambda(s_0))$ *for which* $\dot{\lambda}(s_0 \neq 0)$.

See figures 5.4 a,b,c. The lemma states that the case shown in fig. 5.4c is not possible, since $\dot{\lambda}(s_0)$ vanishes on both the branches. In fig. 5.4a, $\dot{\lambda}(s_0) \neq 0$ on both the branches, but in fig. 5.4b, $\dot{\lambda}(s_0) \neq 0$ on Γ_+ and $\dot{\lambda}(s_0) = 0$ on Γ_-

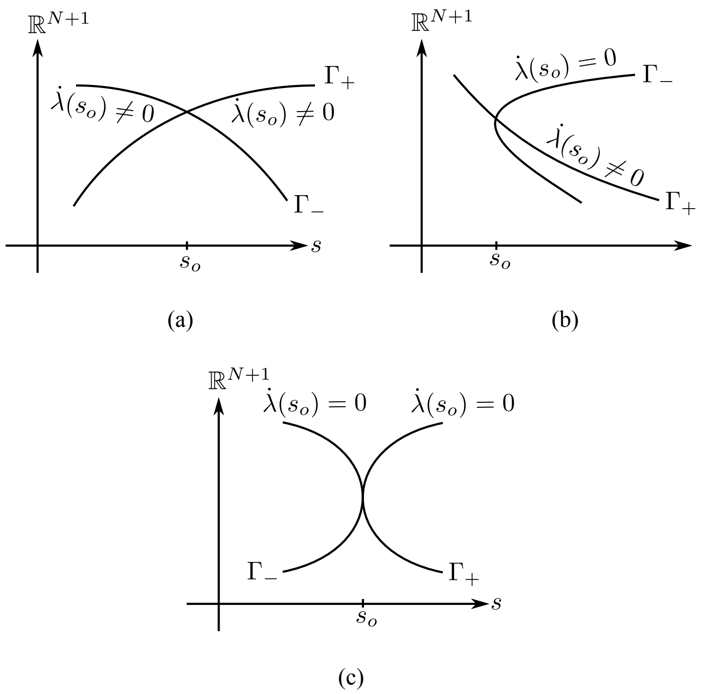


Figure 5.4:

105

Proof. Let (α_+, β_+) and (α_-, β_-) be the roots of the quadratic equation (5.12a). At the bifurcation point, β_+ and β_- gives $\dot\lambda(s_0)$ corresponding to each one of the branches. Therefore if $\dot\lambda(s_0) = 0$ along one of the branches, then one of β_+, β_- say $\beta_+ = 0$. Then α_+ cannot be zero. Hence from equation (5.12a), we have

$$a_{11} = 0$$

Since $\Delta > 0$, $a_{12} \neq 0$. Also both β_+ and β_- cannot vanish together. Thus at least for one branch $\dot\lambda(s_0) \neq 0$. Now for the nonvanishing β,

$$\alpha a_{11} + \beta a_{12} = \beta\sqrt{\Delta} \neq 0.$$

From (5.18), we have **106**

$$\dot{n}(s_0) = \frac{\alpha a_{11} + \beta a_{12}}{\psi^T \phi} \neq 0.$$

Then $\det G_u(s)$ changes sign as in the previous lemma. □

Remark. In the case of simple limit points and simple singular points for which (5.11a) holds, $N\{G_u(s_0)\}$ has dimension one. The only difference between these two types of points is that in the limit point case $G_\lambda \notin R(G_u)$ and in the other case $G_\lambda \in R(G_u)$. Hence if $\psi^T G_\lambda(s_0) = 0$, a bifurcation is effected and if $\psi^T G_\lambda(s_0) \neq 0$, a *fold* is effected.

5.21 Exchange of Stability

The solutions of (5.16) are the steady states of the time dependant problems of the form :

$$\frac{\partial U}{\partial t} = G(U, \lambda) - p. \tag{5.22}$$

Given an arc of solutions $\{u(s), \lambda(s)\}$ of (5.16), it is required to determine the stability of each point as a steady state of (5.22). To use linearized stability theory we seek solutions of (5.22) in the form:

$$\begin{aligned} \lambda &= \lambda(s) \\ U(t, s) &= u(s) + \varepsilon \exp\{tx(s)\}\phi(s) \\ \|\phi(s)\| &= 1. \end{aligned} \tag{5.23}$$

Expanding $G(U(t, s), \lambda(s))$ about $\varepsilon = 0$, we get:

$$\frac{\partial U}{\partial t} = G(u(s), \lambda(s)) + \varepsilon G_u(u(s), \lambda(s)) \exp\{tx(s)\}\phi(s) + 0(\varepsilon^2). \tag{5.24a}$$

107 But from (5.23), we have :

$$\frac{\partial U}{\partial t} = \varepsilon n(s) \exp\{tx(s)\}\phi(s). \tag{5.24b}$$

Equating (5.24a) and (5.24b), since $G(u(s), \lambda(s)) = 0$, it follows that:

$$n(s)\phi(s) = G_u(u(s), \lambda(s))\phi(s) + 0(\varepsilon).$$

Hence we are led to the eigenvalue problem :

$$\text{(5.25)} \qquad \begin{aligned} &\text{(a)}\ \ G_u(u(s), \lambda(s))\phi(s) = n(s)\phi(s),\\ &\text{(b)}\ \ \|\phi(s)\| = 1. \end{aligned}$$

If all eigenvalues $n = n(s)$ of (5.25) have $\mathrm{Re}(n(s)) < 0$ for a given s, we say that $u(s)$ is linearly stable. If at least one eigenvalue has $\mathrm{Re}(n(s)) > 0$, then $u(s)$ is linearly unstable. If all the eigenvalue have $\mathrm{Re}(n(s)) \leq 0$, with at least one equality holding, we say that $u(s)$ is neutrally stable.

Suppose $(u(s_0), \lambda(s_0))$ is a limit point as in lemma 5.19, then $n(s)$ changes sign as it crosses the point $s = s_0$. Hence if a smooth path of solutions has a simple quadratic fold at $s = s_0$ as in lemma 5.19 and solutions for $s > s_0$ (or $s < s_0$) are stable then they are unstable for $s < s_0$ (or $s > s_0$). Hence there is a change of stability at $s = s_0$. Note that here we are not claiming that any branch of solutions is stable or not. We are only proving that if the solution branch is stable at one side, then it is unstable on the other side. We cannot even conclude the converse.

Similarly at a bifurcation point as in lemma 5.20, $\dot{n}(s_0) \neq 0$ there **108**
and $\det G_u(s)$ changes sign at $s = s_0$ on one or both the branches through the bifurcation point for which $\dot{\lambda}(s_0) \neq 0$. Hence here also there may be an exchange of stability on one or both the branches through the bifurcation point. Again here observe that we are not proving any branch of solutions is stable. But we do know that both arcs of one of the branches cannot be stable.

5.26 Switching Branches at Bifurcation Points

Bifurcation points are solutions at which two or more branches of solutions of

$$\text{(5.27)} \qquad G(u, \lambda) - p = 0$$

intersect nontangentially. In this section we consider branch switching only at simple bifurcation points. For more details see [19].

Method I: An obvious way to determine branches bifurcating at (u^0, λ^0) is to determine the two distinct roots of (5.12a) and use them to determine the distinct tangent vector in :

$$\dot{u} = \alpha\phi + \beta\phi_0, \dot{\lambda} = \beta. \tag{5.28}$$

Then we can use our pseudoarclength continuation method to determine the different branches of solutions. If we know one branch of solutions Γ_{ab}, then we know the tangent, $(\dot{u}^0, \dot{\lambda}^0)$, to the curve Γ_{ab} at (u^0, λ^0) and this determines one root of (5.12a). Hence we can determine the other tangent easily. Also note that in finding the values of the coefficients a_{ij} of (5.12a) we need the derivatives G^0_{uu}, $G^0_{u\lambda}$ and $G^0_{\lambda\lambda}$.
109 But we can use the following approximation to these quantities which avoids the need for determining the second derivatives of G.

$$\begin{aligned}
a^{\varepsilon}_{11} &= \frac{1}{\varepsilon}\psi^T[G_u(u^0 + \varepsilon\phi, \lambda^0) - G_u(u^0, \lambda^0)]\phi,\\
&= a_{11} + 0(\varepsilon),\\
a^{\varepsilon}_{12} &= \frac{1}{\varepsilon}\psi^T\{[G_u(u^0 + \varepsilon\phi_0, \lambda^0) - G_u(u^0, \lambda^0)]\phi,\\
&\qquad + [G_\lambda(u^0 + \varepsilon\phi, \lambda^0) - G_\lambda\,(u^0, \lambda^0)]\},\\
&= a_{12} + 0(\varepsilon)\\
a^{\varepsilon}_{22} &= \frac{1}{\varepsilon}\psi^T\{G_u(u^0 + \varepsilon\phi_0, \lambda^0) - G_u(u^0, \lambda^0)]\phi_0\\
&\qquad + 2[G_\lambda(u^0 + \varepsilon\phi_0, \lambda^0) - G_\lambda(u^0, \lambda^0)]\\
&\qquad + [G_\lambda(u^0, \lambda^0 + \varepsilon) - G_\lambda(u^0, \lambda^0)]\},\\
&= a_{22} + 0(\varepsilon).
\end{aligned}$$

Here ϕ, ψ and ϕ_0 are nontrivial solutions of

$$G^0_u\phi = 0; \quad G^{o^T}_u\psi = 0; \quad G^0_u\phi_0 = -G^0_\lambda, \quad \phi^T\phi_0 = 0.$$

Method II: In this method we assume that one branch through the bifurcation point has been determined. Then the tangent $(\dot{u}(s_0), \dot{\lambda}(s_0))$ can also be assumed to be known on this branch. The idea is to seek solutions on some subset parallel to the tangent but displaced from the bifurcation point in a direction normal to the tangent but in a specific plane.

This method avoids the need to evaluate the coefficients a_{ij}, $i, j = 1, 2$. Refore [19].

The solution branch $(u_1(s), \lambda_1(s))$ has a tangent in the direction given by (5.28). An orthogonal to this tangent in the plane spanned by $(\phi, 0)$ and $(\phi_0, 1)$ is given by (5.28) but with α, β replaced by: **110**

$$\hat{\alpha} = \beta(1 + \|\phi_0\|^2), \hat{\beta} = -\alpha\|\phi\|^2.$$

Then we seek solutions in the form :

$$\begin{aligned} u_2 &= u_1(s_0) + \varepsilon(\hat{\beta}\phi_0 + \hat{\alpha}\phi_1) + v, \\ \lambda_2 &= \lambda_1(s_0) + \varepsilon\hat{\beta} + \eta. \end{aligned} \tag{5.29a}$$

These are to satisfy:

$$\begin{aligned} &G(u_2, \lambda_2) = p, \\ &N(u_2, \lambda_2) \equiv (\hat{\beta}\phi_0^T + \hat{\alpha}\phi^T)_v + \hat{\beta}\eta = 0. \end{aligned} \tag{5.29b}$$

We use Newton's method to solve (5.29b) for $v \in \mathbb{R}^N$ and $\eta \in \mathbb{R}$ with the initial estimate $(v_0, \eta_0) = (0, 0)$. Here ε much be taken sufficiently large so that the scheme does not return to $(u_1(s_0), \lambda_1(s_0))$ as the solution.

Method III (Lyapunov-Schmidt): Another way to determine a branch bifurcating from a known branch $(u(s), \lambda(s))$ at $s = s_0$ is to apply a constructive existence theory, as in [22]. We seek the bifurcated branch in the form :

$$\begin{aligned} u &= u_1(\sigma) + \varepsilon(\phi + v), \quad \psi^T v = 0 \\ \lambda &= \lambda_1(\sigma). \end{aligned} \tag{5.30a}$$

Then we have:

$$G_u^0 v - \frac{1}{\varepsilon} G(u_1(\sigma) + \varepsilon(\phi + v), \lambda_1(\sigma)), \psi^T v = 0. \tag{5.30b}$$

111

To ensure that the right hand side is in $R(G_u^0)$, we try pick $\sigma = s$ such that $h(s, \varepsilon, v) = 0$, where,

$$h(s, \varepsilon, v) = \begin{cases} \psi^T[G_u^0 v - \frac{1}{\varepsilon}G(u_1(s) + \varepsilon(\phi + v), \lambda_1(s))], \varepsilon \neq 0, \\ \psi^T[G_u^0 v - G(u_1(s), \lambda_1(s))(\phi + v)]; \varepsilon = 0. \end{cases} \tag{5.30c}$$

It follows that $h(s_0, 0, 0) = 0$ and

$$h_s^0 = h_s(s_0, 0, 0) = -\psi^T [G_{uu}^0 \dot{u}_1(s_0) + G_u^0 \lambda \dot{\lambda}_1(s_0)]\phi$$

Then if $h_s^0 \neq 0$ the implicit function theorem gives the function $s = \sigma(\varepsilon, v)$ and then it can be shown that (5.30b) has a unique solution $v = v(\in)$ for $|\varepsilon|$ sufficiently small, using the contraction mapping theorem.

The main difficulty in applying this method is in solving $h(s, \varepsilon, v) = 0$ for s at each $v = v_U$. If course if λ occurs linearly in the problem and it is used as the parameter s, then this is easy. But in the case, modifications must be introduced. See [3], [8]. This method has also been used for bifurcation from the trivial branch. See [31].

Method IV: Here we use a technique based on a modification of the Crandall and Rabinowitz [5] proof of bifurcation. Thus we seek solutions of the form (5.30a) and define

$$\text{(5.31)} \quad \begin{aligned} &\text{(a)}\, g(v, s, \varepsilon) = \begin{cases} \frac{1}{\varepsilon} G(u_1(s) + \varepsilon(\phi + v), \lambda_1(s)) & \text{if} \quad \varepsilon \neq 0 \\ G_u(u_1(s), \lambda_1(s))(\phi + v) & \text{if} \quad \varepsilon = 0 \end{cases} \\ &\text{(b)}\, N(v, s, \varepsilon) = \psi^T v. \end{aligned}$$

112

Note that

$$g(0, s_0, 0) = 0, \quad N(0, s_0, 0) = 0$$

and

$$A^0 = \left. \frac{\partial(g, N)}{\partial(v, s)} \right|_{0, s_0, 0)} = \begin{bmatrix} G_u^0 & B^0 \\ \psi^T & 0 \end{bmatrix},$$

where,

$$B^0 \equiv \left[G_{uu}^0 \dot{u}_1(s_0) + G_{u\lambda}^0 \dot{\lambda}_1(s_0) \right] \phi.$$

If $\psi^T B^0 \neq 0$, the by the lemma 4.9, A^0 is nonsingular. Now the implicit function theorem shows that :

$$\text{(5.32)} \qquad \begin{aligned} g(v, s, \varepsilon) &= 0, \\ N(v, s, \varepsilon) &= 0. \end{aligned}$$

has a smooth solution $(v(\varepsilon), \rho(\varepsilon))$ for each $|\varepsilon| \leq \varepsilon_0$ and using this solution in (5.31a) yields the bifurcating branch of solutions. In solving (5.32) we never use $\varepsilon = 0$ so that even when applying Newton's method, second derivatives need not be computed.

Method V (Perturbed Bifurcation) : Observe that at a bifurcation point (u^0, λ^0), p is not a regular value for G. Since the set of all regular values is dense, the idea of this method is to perturb p, say $p + \tau q$, $q \in \mathbb{R}^N$, $||q|| = 1$, $\tau \neq 0$, so that $p + \tau q$ is a regular value for G. Consider the two smooth branches of solutions through the bifurcation point 113
(u^0, λ^0). If we delete a small neighbourhood of (u^0, λ^0), we obtain 4 different branches of solutions, say Γ_1, Γ_2, Γ_3, Γ_4. Then consider the perturbed problem :

$$(5.33) \qquad G(u, \lambda) = p + \tau q, \quad q \in \mathbb{R}^N, \quad ||q|| = 1, \quad \tau \neq 0.$$

Assuming $p + \tau q$ is a regular value, this has no bifurcation and Γ_i, $i = 1, 2, 3, 4$ will simply be perturbed and will yield smooth nearby branches of solutions of (5.33). These branches can be connected in 3 ways. See Figs. 5.5 a,b,c.

Case (i) : Here if we are starting from a solution on Γ_1, we will get a perturbed solution on Γ'_1 and then we can continue along Γ'_1 to Γ'_4 which is the perturbed branch of Γ_4. Similarly from Γ_2, we will obtain Γ'_2 and Γ'_3 [Fig. 5.5a].

Case (ii) : In a similar manner we can handle this case also (see Fig. 5.5b).

Case (iii) : In the cases (i) and (ii) we can determine the other branches without any difficulty. Our claim is that the case (iii) doesn't happen. To see this we have to study further about fold following which we are going to describe in the next section. There we will consider the problem (5.33) as a two parameter, (λ, τ) problem and will give an algorithm to obtain a fold .

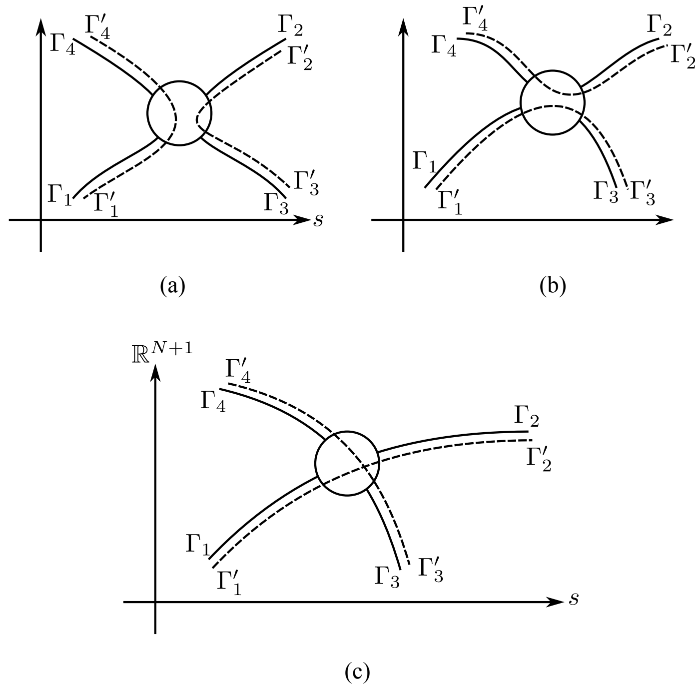


Figure 5.5:

114

5.27 Multi Parameter Problems (Fold Following)

115

We recall the two examples described in the introductory chapter on population dynamics. We examined the steady state solutions of:

$$\frac{d\xi}{dt} = -\xi^2 + \lambda\xi + \tau,$$
$$\frac{d\xi}{dt} = -\xi^3 + \lambda\xi + \tau,$$

that is the solution of:

$$(1) \quad -\xi^2 + \lambda\xi + \tau = 0,$$
$$(2) \quad -\xi^3 + \lambda\xi + \tau = 0,$$

Example 1. $-\xi^2 + \lambda\xi + \tau = 0$.

We note that in the (λ, τ) plane there is a curve across which the number of solutions change. This curve is a fold. See the solution surface sketched in fig. 1.4.

Example 2. $-\xi^3 + \lambda\xi + \tau = 0$,

In Fig. 1.8 we show a curve in the (λ, τ) plane which has a cusp at the origin; as (λ, τ) crosses this curve the number of solutions changes. Look at the solution surface in Fig. 1.12. This curve is a fold on the solution surface.

Hence determining a fold is very important. Other interesting phenomena may occur along the folds. The solutions lie either to the left of the fold curve or to the right of the fold curve. In the latter case, (Fig. 5.6a) the fold point is known as a hyperbolic point and in the former case (Fig. 5.6b) it is called an elliptic point. Suppose the fold point is at (u_0, λ_0, τ_0). In the elliptic case if $\tau > \tau_0$ we have no solution and on the other hand if $\tau < \tau_0$, we obtain a closed loop of solutions. A closed loop of solutions is known as an 'isola'. In the second case as τ changes from τ_0 we get different branches of solutions. For a unified theory of perturbed bifurcation and isola formation see [21]. **116**

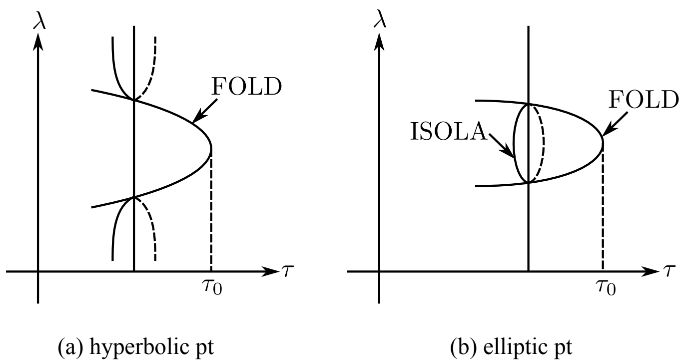


Figure 5.6:

Now let $\tau = \tau_0$ and assume that there is a simple quadratic fold with respect to λ at (u_0, λ_0). We have $G(u_0, \lambda_0, \tau_0) = p$ and

$$\text{(5.35)} \qquad \begin{aligned} &\text{(a)}\ \ N(G_u^0) = \operatorname{span}\{\phi_0\}, \quad \text{for some } \phi_0 \neq 0, \\ &\text{(b)}\ \ G_\lambda^0 \notin R(G_u^0) = \{v \in \mathbb{R}^N : \psi_0^T v = 0\}. \end{aligned}$$

Here ψ_0 is a nontrivial solution of:

$$\text{(5.35)} \qquad \text{(c) } G_u^{0T} \psi_0 = 0,$$

117 and at a simple quadratic fold:

$$\text{(5.35)} \qquad \text{(d) } a = \psi_0^T G_{uu}^0 \phi_0 \phi_0 \neq 0.$$

Now consider the extended system:

$$F_1(u, \psi.\lambda, \tau) \equiv \begin{bmatrix} G(u, \lambda, \tau) - p \\ \psi^T G_u(u, \lambda, \tau) \\ \psi^T G_\lambda^0 - 1 \end{bmatrix} = 0.$$

Here $F_1 : \mathbb{R}^{2N+2} \to \mathbb{R}^{2N+1}$. This system can be written as:

$$\text{(5.36)} \qquad \text{(a) } F(U, \tau) = 0,$$

where

$$\text{(5.36)} \qquad \text{(b) } U = \begin{bmatrix} u \\ \psi \\ \lambda \end{bmatrix}$$

and $F \equiv F_1$. Note that $(u_0, \psi_0, \lambda_0, \tau_0)$ is a solution of this system. We use another formulation:

$$\text{(5.37a)} \qquad F_2(u, \psi, \lambda, \tau) \equiv \begin{bmatrix} F(u, \lambda, \tau) - p \\ G_u(u, \lambda.\tau)\phi \\ \ell^T \phi - 1 \end{bmatrix} = 0,$$

where ℓ is such that

$$\text{(5.37b)} \qquad \ell^T \phi_0 = 1.$$

118 This can also be written in the form (5.36) with $F \equiv F_2$. In this letter case, we have the following theorem

Theorem 5.38. *At a simple quadratic fold* $(u_0.\lambda_0, \tau_0)$*:*

$$F_{\dot{U}}^0 = \frac{\partial F}{\partial(u, \Phi, \lambda)}\bigg|_{(u_0,\Phi_0,\lambda_0)} = \begin{bmatrix} G_u^0 & 0 & G_\lambda^0 \\ G_{uu}^0\Phi_0 & G_u^0 & G_{u\lambda}^0\Phi_0 \\ 0 & \ell^T & 0 \end{bmatrix}$$

is nonsingular.

Proof. Assume $F_U^0\Phi = 0$, for some $\Phi = \begin{bmatrix} p \\ q \\ r \end{bmatrix} \in \mathbb{R}^{2N+1}$

That is

$$\text{(5.39)} \qquad \begin{aligned} &\text{(a)}\ \ G_U^0 p + rG_\lambda^0 = 0, \\ &\text{(b)}\ \ G_{uu}^0\Phi_0 p + G_u^0 q + G_{u\lambda}^0\Phi_0 r = 0, \\ &\text{(c)}\ \ \ell_q^T = 0. \end{aligned}$$

Multiplying (5.39a) by ψ_0^T and using (5.35 b,c), we get $r = 0$. Hence $G_u^0 p = 0$. This shows that $p = \alpha\Phi_0$, for some $\alpha \in \mathbb{R}$. So (5.39b) implies that

$$\alpha\psi_0^T G_{uu}^0 \Phi_0\Phi_0 = \alpha a = 0.$$

So that by (5.35d), α and hence $p = 0$. This implies $G_u^0 q = 0$. Hence $q = \beta\Phi_0$ for some $\beta \in \mathbb{R}$. But by (5.39c) we get $\beta\ell^T\Phi_0 = 0$. Hence by (5.37b), β and $q = 0$ □

Now we can apply the implicit function theorem to obtain

$$u = u(\tau)$$
$$\Phi = \Phi(\tau)$$
$$\lambda = \lambda(\tau)$$

for $|\tau - \delta| < \delta$, for some $\delta > 0$. The arc $(u(\tau), \lambda(\tau), \tau) \in \mathbb{R}^{N+2}$, $\tau \in$ **119**
$(\tau_0 - \delta, \tau_0 + \delta)$ is part of the fold curve. We can also use some other parametrization, s, to to obtain the solution in the form:

$$u = u(s)$$
$$\Phi = \Phi(s)$$

$$\lambda = \lambda(s)$$
$$\tau = \tau(s).$$

Now we will show how to use the Bordering Algorithm to find the Newton iterates applied to solve (5.37a). The linear system has the form:

$$LX = \begin{bmatrix} \underset{\sim}{g} \\ \gamma \end{bmatrix} \tag{5.40}$$

Here

$$L \equiv F_U \equiv \begin{bmatrix} \tilde{A} & \underset{\sim}{b} \\ \underset{\sim}{c}^T & 0 \end{bmatrix}, \quad X \in \mathbb{R}^{2N+2}, \underset{\sim}{g} \in \mathbb{R}^{2N}, \gamma \in \mathbb{R},$$

where

$$\tilde{A} \equiv \begin{bmatrix} A & 0 \\ B & A \end{bmatrix} \text{ with } A = G_u^0,\ B = G_{uu}^0 \Phi_0.$$

First solve:

$$\tilde{A}\underset{\sim}{v} = \underset{\sim}{b}, \underset{\sim}{b}, \underset{\sim}{v} \in \mathbb{R}^{2N},$$
$$\tilde{A}\underset{\sim}{\omega} = \underset{\sim}{g}, \underset{\sim}{g}, \underset{\sim}{\omega} \in \mathbb{R}^{2N}.$$

120 Then form:

$$\xi = \frac{\gamma - \underset{\sim}{c}^T \underset{\sim}{\omega}}{\underset{\sim}{c}^T \underset{\sim}{v}},$$

and

$$\underset{\sim}{x} = \underset{\sim}{\omega} - \xi \underset{\sim}{v}.$$

Note that the system:

$$\tilde{A}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \equiv \begin{bmatrix} A & 0 \\ B & A \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix},$$

can easily be solved as follows:

First solve

$$Ay_1 = \gamma_1,$$

and then

$$Ay_2 = \gamma_2 - By_1.$$

Hence it follows that the system (5.40) can be solved by solving four systems the same coefficient matrix A. So we need do only one LU decomposition of A and then four backsolves. See the Ref. [33] and also [15].

5.41 Paths of Periodic Solutions and Hopf Bifurcation

121

In this section we will discuss the periodic solutions of the system of ODE depending on a parameter:

$$\frac{dy}{dt} = f(y, \lambda). \tag{5.42}$$

First we will briefly describe Poincare's method for containing periodic solution branches and then we discuss Hopf bifurcation. We do not give a detailed description of these things. A good reference for Hopf bifurcation is [25].

We seek periodic solutions of (5.42). Suppose for $\lambda = \lambda^0$ that $\underset{\sim o}{y}(t)$ satisfies (5.42) and $\underset{\sim o}{y}(t+T^0) = \underset{\sim o}{y}(t)$, for some $T^0 > 0$ and for all t. Then $\underset{\sim o}{y}(t)$ is a periodic solution (5.42). Since this system is autonomous, for any σ, the translation of $\underset{\sim o}{y}(t)$. viz.

$$\underset{\sim}{y}(t) = \underset{\sim o}{y}(t + \sigma),$$

is also a periodic solution. We seek periodic solution when λ is pertubed from λ^0

Substituting $\underset{\sim o}{y}(t)$ in (5.42) and differentiating we get:

$$\frac{d}{dt}\underset{\sim o}{\dot{y}}(t) = [f_y(\underset{\sim o}{y}(t), \lambda^0)]\underset{\sim o}{\dot{y}}(t).$$

Let

$$A_0(t) = [f_y(\underset{\sim o}{y}(t), \lambda^0)].$$

Then we can write

$$[\frac{d}{dt} - A_0(t)]\underset{\sim 0}{\dot{y}}(t) = 0.$$

122

Hence the linearised operator $\dfrac{d}{dt} - A_0(t)$ has a nontrivial periodic solutions $\underset{\sim 0}{\dot{y}}(t)$. In the case of nonautonomous system this is not in general true. Then we have to assume that $f(t, y, \lambda)$ is periodic, with period $T(\lambda)$. Then in general the linearized problem does not have nontrivial periodic solution and the continuation in λ can be done yielding solutions with period $T(\lambda)$. In our case (5.42), f has all possible periods in t.

Consider the periodic solution $\underset{\sim 0}{y}(t)$ of (5.42). At $t = t_0 = 0$, $\underset{\sim 0}{y}(0)$ is some arbitrary point in $\mathbb{R}^N$ and let $\pi \subset \mathbb{R}^N$ be the plane which is perpendicular to the tangent, $\underset{\sim 0}{\dot{y}}(0)$ at $y_0(0)$. That is:

$$\pi = \{\underset{\sim}{\xi} \in \mathbb{R}^N : \underset{\sim o}{\dot{y}}(0)^T(\underset{\sim}{\xi} - \underset{\sim o}{y}(0)) = 0\}$$

Consider any point $\underset{\sim}{\xi}\varepsilon\pi$ in some small neighbourhood of $\underset{\sim o}{y}(0)$ and look at the solution curve of (5.42) passing through this point $\underset{\sim}{\xi}$. Suppose the curve intersects this plane π after sometime, then this point of intersection is the image of $\underset{\sim}{\xi}$ under the poincare map. For the existence of Poincare map see [25]. From the definition it is clear that the periodic solution of (5.42) corresponds to the fixed points of the Poincare map.

Consider the initial value problem:

(5.43)
$$\begin{aligned} &\text{(a)}\ \ \frac{dy}{dt} = \underset{,}{\underline{f(y,\lambda)}} \\ &\text{(b)}\ \ \underset{\sim}{y}(0) = \underset{\sim}{\xi} \in \mathbb{R}^N. \end{aligned}$$

123 Let $\underset{\sim}{y} = \underset{\sim}{y}(t, \underset{\sim}{\xi}, \lambda)$ be the solution of (5.43), for $(\underset{\sim}{\xi}, \lambda) \in B_\rho(\underset{\sim o}{y}(0), \lambda^0)$

for some $\rho > 0$. The problem thus reduced to: find $\underset{\sim}{\xi}$ and T such that :

$$
\text{(5.44)} \qquad \begin{aligned}
&\text{(a)} \quad \underset{\sim}{y}(T, \underset{\sim}{\xi}, \lambda) - \underset{\sim}{\xi} = 0,\\
&\text{(b)} \quad \dot{\underset{\sim o}{y}}(0)^T[\underset{\sim}{\xi} - \underset{\sim o}{y}(0)] = 0.
\end{aligned}
$$

These are $(N+1)$ equations in $N+1$ unknowns. Observe that at $\lambda = \lambda^0$, we have the solution $\underset{\sim}{\xi}^0 = \underset{\sim o}{y}(0)$, $T = T^0$. We expect solutions for T near T^0. We can apply the implicit function theorem if the Jacobian of (5.44) with respect to $(\underset{\sim}{\xi}, T)$ is nonsingular. The Jacobian at $(\underset{\sim}{\xi}^0, T^0, \lambda^0)$ is given by:

$$
\frac{\partial}{\partial(\underset{\sim}{\xi}, T)} \begin{bmatrix} \underset{\sim}{y}(T, \underset{\sim}{\xi}, \lambda) - \underset{\sim}{\xi} \\ \dot{\underset{\sim o}{y}}(0)^T[\underset{\sim}{\xi} - \underset{\sim o}{y}(0)] \end{bmatrix} \equiv \begin{bmatrix} (Y^0 - I) & \dot{\underset{\sim o}{y}}(0) \\ \dot{\underset{\sim o}{y}}(0)^T & 0 \end{bmatrix} \equiv A.
$$

Here $Y^0 = Y(T^0, \underset{\sim}{\xi}^0, \lambda^0)$, where Y satisfies :

$$
\text{(5.45)} \qquad \begin{aligned}
&\text{(a)} \quad \tfrac{dY}{dt} - A_0(t)Y = 0,\\
&\text{(b)} \quad Y(0) = I.
\end{aligned}
$$

All solutions of $\dfrac{dv}{dt} - A_0(t)v = 0$ are the form:

$$
\underset{\sim}{v}(t) = Y(t)\underset{\sim}{\zeta}, \underset{\sim}{\zeta} \in \mathbb{R}^N,
$$

and the solution is periodic if and only if:

$$
[Y(T^0) - I]\underset{\sim}{\zeta} = 0.
$$

Already we know that $\dot{\underset{\sim o}{y}}(t)$ is a nontrivial periodic solution. Hence $Y(T^0)$ has an eigenvalue unity. Now we will assume that 1 is a simple
eigenvalue of $Y(T^0)$. Under this hypothesis, we will prove that: **124**

Lemma 5.21. *A is nonsingular.*

Proof. Suppose $A\begin{pmatrix} \underset{\sim}{\zeta} \\ z \end{pmatrix} = 0$. Then

(5.47a) $$[Y(T^0) - I]\underset{\sim}{\zeta} = -z\underset{\sim o}{\dot{y}}(0).$$

Multiplying throughout by $Y(T^0) - I$, since $(Y(T^0) - I)\underset{\sim o}{\dot{y}}(0) = 0$, we get:

(5.47b) $$[Y(T^0) - I]^2\underset{\sim}{\zeta} = 0.$$

This shows that if $z \neq 0$, then 1 is not a simple eigenvalue. Thus we must have $z = 0$. Now the second equation gives:

$$\underset{\sim o}{\dot{y}}(0)^T\underset{\sim}{\zeta} = 0.$$

But $[Y(T^0) - I]\underset{\sim}{\zeta} = 0$ implies that $\underset{\sim}{\zeta}$ is a multiple of $\underset{\sim o}{\dot{y}}(0)$. i.e. $\underset{\sim o}{\dot{y}}(0)^T\ \underset{\sim o}{\dot{y}}(0) = 0$. Hence we must have $\underset{\sim}{\zeta} = 0$. Thus the proof is complete

We can solve the initial value problem (5.43) numerically The main difficulty in applying such numerical method is that the numerical trajectory will blown up if the continuous orbit is not stable. Even in the stable case there may exist unstable orbits arbitrarily near to stable orbits and these cause trouble. In many cases of course this "shooting" method works fine.

We can write the equation (5.44) in the form:

(5.48) $$G(u, \lambda) = 0,$$

125 where $u = (\underset{\sim}{\xi}, T)$. Here also we can do all the analysis of folds, bifurcations, stability etc. as before. We can also compute paths of solutions of (5.48) and hence paths of periodic solutions of (5.42).

We can formulate the continuous problems as follows, after scaling time by the period, T:

(5.49) $$G(\underset{\sim}{y}(t), \underset{\sim}{\xi}, T, \lambda)\begin{bmatrix} \frac{dy}{dt} - T f(\underset{\sim}{y}, \lambda) \\ \underset{\sim}{y}(1) - \underset{\sim}{\xi} \\ \underset{\sim o}{\dot{y}}(0)^T(\underset{\sim}{\xi} - \underset{\sim o}{y}(0)) \end{bmatrix} = 0.$$

Here $G : B \times \mathbb{R}^N \times \mathbb{R}^2 \to \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}$. Now

$$\frac{\partial G}{\partial(\underset{\sim}{y}, \underset{\sim}{\xi}, T)} \equiv \begin{bmatrix} \left[\frac{d}{dt} - TA(t, \lambda)\right] & 0 & f(\underset{\sim}{y}, \lambda) \\ 0 & Y^0 - I & \underset{\sim T}{y}(1) \\ 0 & \underset{\sim o}{\dot{y}}(0)^T & 0 \end{bmatrix}$$

Here $A(t, \lambda) = f_y(\underset{\sim}{y}(t), \lambda)$. As is lemma it can be shown that

$$\frac{\partial G}{\partial(\underset{\sim}{y}, \underset{\sim}{\xi}, T)}\Bigg|\begin{aligned} \underset{\sim}{y} &= \underset{\sim o}{y}(t) \\ \underset{\sim}{\xi} &= \underset{\sim o}{y}(0) \\ T &= T^0 \end{aligned}$$

is nonsingular if 1 is a simple eigenvalue of Y^0.

We consider finite difference approximations to (5.49). We use a uniform net (but this can easily be changed):

$$t_j = jh, \quad h = \frac{1}{M},$$

where M is the number of partitions of the unit time interval. **126**

$$t_0 \quad t_1 \quad \cdots \quad t_{j-1} \quad t_j \quad \cdots \quad t_M$$

Let u_j be an approximation to $u(t_j)$. We can approximate (5.49) as:

$$G^h(u^h, T, \lambda) \equiv \begin{bmatrix} \vdots \\ u_j - u_{j-1} - Th \quad f\left(\frac{1}{2}(u_j + u_{j-1}), \lambda\right) \\ \vdots \\ u_0 - u_N \\ \dot{\xi} o^T (u_N - \xi^0) \end{bmatrix} = 0.$$

This is a system of $MN + N + 1$ equations in $MN + N + 1$ unknowns. Here again we can apply the implicit function theorem to obtain $(u^h(\lambda), T(\lambda))$ if h is sufficiently small. Here again there may exists folds, bifurcations (Hopf bifurcation) etc. Hopf bifurcation is the bifurcation of periodic solution orbit steady state solutions. Refer [14] and [10] for details.

Look at the steady states of (5.42). i.e. solutions of $f(y, \lambda) = 0$. Let $\Gamma_0 = \{y(\lambda)\}$ be a solution branch. Eigenvalues of the linearized problem
127 about states determine the stability of these orbits as solutions of (5.42). So consider the eigenvalue problem:

$$A(\lambda)\phi(\lambda) = \eta(\lambda)\phi(\lambda)$$

Here $A(\lambda) = f_y(y(\lambda), \lambda)$. Since $A(\lambda)$ is real, the eigenvalues will occur as complex conjugate pairs. Assume $\eta_1(\lambda) = \alpha(\lambda) + i\beta(\lambda)$ and $\eta_2(\lambda) = \alpha(\lambda) - i\beta(\lambda)$. We know that at fold points, some real eigenvalues change sign. Also if the real part of any eigenvalue changes sign, there may be a change of stability. We examine the case of the even number,in fact two complex eigenvalues changing sign of the real part. So assume $\alpha(\lambda_0) = 0$, $\dot{\alpha}(\lambda_0) \neq 0$ and $\beta(\lambda_0) \neq 0$. Then the *Hopf bifurcation theorem* [see [25]] states that a periodic solution branch bifurcates from Γ_0 at $\lambda = \lambda_0$. Let

$$\phi(\lambda) = a(\lambda) + ib(\lambda).$$

Substituting this in (**??**), we get:

$$\begin{aligned} Aa &= \alpha a - \beta b, \\ Ab &= \alpha b + \beta a. \end{aligned} \tag{5.51}$$

At Hopf bifurcation, we have $\alpha(\lambda_0) = 0$. Then look at the system (5.52a) for the unknowns $(y, a, b, \beta, \lambda)$.

$$\begin{bmatrix} f(y, \lambda) \\ f_y(y, \lambda)a + \beta b \\ f_y(y, \lambda)b - \beta a \end{bmatrix} = 0. \tag{5.52a}$$

This is a system of $3N$ equations in $3N + 2$ unknowns. So we add two normalizations to obtain a system of $(3N + 2)$ equations, viz:

$$\begin{bmatrix} f(y,\lambda) \\ f_y(y,\lambda)a + \beta b \\ f_y(y,\lambda)b - \beta a \\ a^T a + b^T b - 1 \\ \ell^T a - 1 \end{bmatrix} = 0. \tag{5.52b}$$

128

The choice of the unit vector ℓ need only satisfy $\ell^T a(\lambda_0) = 1$.

Here also we can apply numerical schemes to directly compute points of Hopf bifurcation. In the next chapter we will give some examples in which we employed various schemes that are discussed in the previous chapter. □

Chapter 6

Numerical Examples

6.1 Introduction

129

We present here in some detail, a few worked-out examples showing how the techniques discussed in the lectures are actually employed. Unfortunately we do not show all the procedures, but it should not be difficult for the interested reader to try out the missing ones. We use path following via **Euler-Newton** continuation in both the natural parameter formulation and in the Pseudoarclength variant. Folds are circumvented by switching from the natural parameter to Pseudoarclength when specific tests tell us to do so. Then we switch back after the fold has been traversed. We jump over bifurcation points easily and indeed must continuously test to see if we must return to locate the folds or bifurcation points. We also locate the singular points accurately and switch branches at the bifurcation points. Multiparameter problems are also treated.

The basic logic of our code is illustrated by a simple flow diagram. The numerical schemes employed are given. Details of the tests used and some indication of how they performed are given below. Several problems have been solved and we show graphs of solution paths for some of these problems. Some actual computer listings of convergence data, steplength changes, accurate location of folds and bifurcation points are included with explanation.

6.2 Problems and Difference Schemes

130

Due to limitations of computing power available for these tests, we consider only O.D.E. problems. They are of the form:

$$Lu \equiv \frac{d^2u}{dx^2} + p(x)\frac{du}{dx} = f(u, x, \lambda, \tau, \text{ in } (0, L) \tag{6.3}$$

subject to two point boundary conditions of the form:

$$\begin{aligned} &\text{(a)}\quad B_0 u \equiv q_0\, u(0) + r_0\, \frac{du}{dx}(0) = s_0, \\ &\text{(b)}\quad B_L u \equiv q_L\, u(L) + r_L\, \frac{du}{dx}(L) = s_L. \end{aligned} \tag{6.4}$$

The nonlinearities were of the semilinear form. No conceptual difficulties occur if they are allowed to be quasilinear, i.e. f is in the form $f\left(u, \frac{du}{dx}, x, \lambda, \tau\right)$. We use uniform grids of spacing

$$h = \frac{L}{N}, \tag{6.5a}$$

to define the netpoints $\{x_j\}_0^N$ as :

$$x_0 = 0, x_{j+1} = x_j + h, j = 0, 1, \dots N - 1. \tag{6.5b}$$

The usual centred difference scheme is employed, sometimes modified if singular coefficients occur $\left(\text{i.e. } : p(x) = \frac{m}{\alpha}\right)$ or improved to get fourth order accuracy if $p(x) = 0$. If either $r_0 \neq 0$ or $r_L \neq 0$, we introduce exterior points at $x_{-1=-h}$ and $x_{N+1} = L + h$ respectively. Denoting the discrete approximation to $u(x_j)$ by u_j and using the finite difference operators:

$$D_+ u_j \equiv \frac{u_{j+1} - u_j}{h}; D_{-u_j} \equiv \frac{u_j - u_{j-1}}{h}; D_0 u_j \equiv \frac{u_{j+1} - u_{j-1}}{h}.$$

131 Our basic difference approximation to (6.3) at x_j is given by:

$$L_h u_j \equiv D_+ D_- u_j + p(x) D_0 u_j = f(u_j, x_j, \lambda, \tau); j = 0, 1, \dots N. \tag{6.6}$$

The boundary conditions (6.4) are approximated as:

$$\text{(6.7)} \qquad \begin{aligned} &\text{(a) } B_0^h u_L \equiv q_0 u_0 + r_0 D_0 u_0 = s_0, \\ &\text{(b) } B_L^h u_L \equiv q_L u_N + r_L D_0 u_N = s_L, \end{aligned}$$

The scheme in (6.6) for $j = 0$ and (6.7a) both involve u_{-1}. This quantity is then eliminated using the two equations to get one relation between u_0 and u_1. A similar treatment using (6.6) at $j = N$ and (6.7b) eliminates u_{N+1} and retains one equation in u_N and u_{N-1}. The system thus reduced consists of $N + 1$ equations in as many unknowns u_j, $j = 0, \dots N$. If $r_0 = 0$ and (or) $r_L = 0$, we do not impose (6.6) for $j = 0$ and (or) $j = N$. The corresponding elimination need not be done and we again get $N + 1$ equations in the same number of unknowns. We could also eliminate u_0 and (or) u_N. In this case we not bother to do that in our description here.

The difference scheme in (6.6), (6.7) has second order accuracy. If the continuous problem (6.3), (6.4) has an isolated solution U_0 (say, for fixed λ, τ) and if $p(x)$ and f are smooth, then it is known that, for sufficiently small h, the discrete problem (6.6), (6.7) has a unique solution $\{u_j\}$ in some sphere in R^{N+1} about $\{u_0(x_j)\}$ and further it satisfies:

$$\|u_0(x_j) - u_j\| = 0(h^2).$$

Similar results also apply to the entire solution branches of (6.3), **132**
(6.4) containing simple fold and bifurcation points. However there may be some degradation of accuracy at these special points; this is discussed more thoroughly in the expanded version of these notes.

If $p(x) \equiv 0$, we can easily get fourth order accurate, three point schemes, using Collatz's device as follows. Taylor expansions give:

$$D_+ D_- u(x_j) = \frac{d^2u}{dx^2}(x_j) + \frac{h^2}{12}\frac{d^4u}{dx^4}(x_j) + 0(h^4).$$

Since $u(x)$ satisfies:

$$\frac{d^2u}{dx^2} = f(u(x), x, \lambda, \tau),$$

we have:

$$D_+D_- f(u(x_j), x_j\lambda, \tau) = \frac{d^4u}{dx^4}(x_j) + 0(h^2),$$

and thus we get:

$$D_+D_-u(x_j) - \frac{h^2}{12}D_+D_- f(u(x_j), x_j, \lambda, \tau) = \frac{d^2u}{dx^2}(x_j) + 0(h^4).$$

Thus in place of (6.6) we use, when $p(x) \equiv 0$:

$$L_h u_j \equiv D_+D_-u_j - \frac{1}{12}(f_{j+1} + 10f_j + f_{j-1}),\ j = 0, 1, \ldots N. \tag{6.8}$$

Here we have used the abbreviations:

$$f_j \equiv f(u_j, x_j, \lambda, \tau).$$

We can employ this device even when we have the mixed boundary conditions (6.4). That is even when $r_0 \neq 0$ and (or) $r_L \neq 0$. This does
133 not seem to have been stressed before, so we indicate the details. Again Taylor expansions give:

$$D_0u(x_j) = \frac{du}{dx}(x_j) + \frac{h^2}{12}\frac{d^3u}{dx^3}(x_j) + 0(h^4).$$

Here the factor $h^2/12$ occurs because we use D_0 to approximate d/dx and $2.3! = 12$. Previously this factor came $4!/2$. In any event now

$$D_0\frac{d^2u}{dx^2}(x_j) = \frac{d^3u(x_j)}{dx^3} + 0(h^3).$$

Using (6.3) with $p(x) \equiv 0$ at $x = x_j$, in the above, we get

$$D_0u(x_j) - \frac{h^2}{12}D_0 f(u(x_j), x_j, \lambda, \tau) = \frac{du}{dx}(x_j) + 0(h^2).$$

With $x_j = x_0$, and $x_j = x_N$, we replace (6.7) by :

$$\begin{aligned} &\text{(a)}\quad \tilde{B}_0^h u_0 \equiv q_0u_0 + r_0D_0(u_0 - \frac{h^2}{12}f_0) = s_0, \\ &\text{(b)}\quad \tilde{B}_L^h u_N \equiv q_Lu_N + q_Lu_N + r_LD_0(u_N - \frac{h^2}{12}f_N) = s_L, \end{aligned} \tag{6.9}$$

Now (6.9a) and (6.8), for $j = 0$, both contain non-zero multiples of $(u_{-1} - \frac{h^2}{12} f_{-1})$ which can be eliminated between these two equations. A similar treatment with (6.9b) and (6.8) for $j = N$ eliminates $(u_{N+1} - \frac{h^2}{12} f_{N+1})$. The result is again $N + 1$ equations in the basic unknowns u_j, $j = 0, 1, \ldots N$ and the tridiagonal structure of the system is retained [see the equation (6.10) below]. Of course if $r_0 = 0$ and (or) $r_L = 0$, we need not bother with the above procedure at $j = 0$ and (or) $j = n$. So in either of those cases, we need not impose (6.8) for $j = 0$ and (or) $j = N$ respectively. Even if the elimination of $(u_{-1} - \frac{h^2}{12} f_{-1})$ etc. are not performed we get enhanced accuracy, but the system structure is not as **134**
elegant.

In the case of singular coefficients, (i.e: $p(x) = \frac{m}{x^\alpha}, 0 < \alpha \leq 1$) we modify the scheme as follows. For $j = 0$ we replace (6.6) by:

$$(m + 1)D_+D_-u_0 = f_0, \text{ if } \alpha = 1,$$

and

$$D_+D_-u_0 = f_0 \text{ if } 0 < \alpha < 1.$$

These follow from $u_x(0) = 0$ and hence :

$$\lim_{x\to 0}(u_{xx} + \frac{m}{x^\alpha}u_x) = \begin{cases} (m+1)u_{xx}(0), & \text{if} \quad \alpha = 1, \\ u_{xx}(0), & \text{if} \quad 0 < \alpha < 1. \end{cases}$$

We get only second order accuracy but the error expansion at $x = 0$ proceeds in power of h^2.

All the difference schemes we have described can now be incorporated into the general form:

$$\begin{aligned} g_j(u_{j-1}, u_j, u_{j+1}, \lambda, \tau) \equiv \alpha_j u_{j-1} + \beta_j u_j + \gamma_j u_{j+1} - \hat{\alpha}_j f_{j-1} \\ - \hat{\beta}_j f_j - \hat{\gamma}_j f_{j+1} - \delta_j = 0, 0 \leq j \leq N. \end{aligned} \tag{6.10a}$$

Of course we always have :

$$\alpha_0 = \hat{\alpha}_0 = \gamma_N = \hat{\gamma}_N = 0, \tag{6.10b}$$

and hence :

(6.10c) $$g_0 \equiv g_0(u_0, u_1, \lambda, \tau); g_N \equiv g_N(u_{N-1}, u_N, \lambda, \tau).$$

The system (6.10) is our finite dimensional problem which we can
135 write in the vector form:

$$G(u, \lambda, \tau) = \begin{bmatrix} g_0 \\ g_1 \\ \vdots \\ g_N \end{bmatrix} = 0.$$

Thus our general finite dimensional problem is from $\mathbb{R}^{N+3}$ to $\mathbb{R}^{N+1}$. Note that it is a nonlinear three term recursion, terminated at $j = 0$ and $j = N$ with two term relations.

Of paramount importance are the derivatives of G with respect to u, λ and τ. The j^{th} row of the Jacobian matrix G_u is given by the $N + 1$ vector:

$$\frac{\partial g_i}{\partial \underset{\sim}{u}} = (0, \ldots 0, a_j, b_j, c_j, 0, \ldots 0), \; j = 0, \ldots N.$$

Thus G_u is a tridiagonal matrix with b_j on the diagonal, a_j and c_j are below and above the diagonal respectively. In standard notations for such matrices we have:

(6.11) $$G_u(u, \lambda, \tau) \equiv [a_j, b_j, c_j]_0^N.$$

Differentiating (6.10a) we get that :

(6.12) $$\begin{aligned} a_j &\equiv \frac{\partial g_j}{\partial u_{j-1}} = \alpha_j - \hat{\alpha}_j (f_u)_{j-1}, \\ b_j &\equiv \frac{\partial g_j}{\partial u_j} = \beta_j - \hat{\beta}_j (f_u)_j, \\ c_j &\equiv \frac{\partial g_j}{\partial u_{j+1}} = \gamma_j - \hat{\gamma}_j (f_u)_{j+1}, 0 \le j \le N. \end{aligned}$$

136 The tridiagonal form of (6.11) becomes clear from (6.10b)

and (6.12). Here we use the obvious notation :

$$(f_u)_j = \frac{\partial f}{\partial u}(u_j, x_j, \lambda, \tau).$$

The λ and τ derivatives are not sparse; we have :

$$G_\lambda = \begin{bmatrix} \vdots \\ g_{\lambda,j} \\ \vdots \end{bmatrix}, \tag{6.13}$$

where,

$$g_{\lambda,j} = -\hat{\alpha}_j (f_\lambda)_{j-1} - \hat{\beta}_j (f_\lambda)_j - \hat{\gamma}_j (f_\lambda)_{j+1}, 0 \le j \le N.$$

A similar set of relations define G_τ.

To complete the detailed specification of any of our problems, we need only to give:

$$L,\ p(x),\ \ f(u, x, \lambda, \tau),\ \ (q_0, r_0, s_0),\ \ (q_L, r_L, s_L). \tag{6.14}$$

When these are given, we examine $p(x)$, r_0 and r_L to find out which of the schemes we have described is to be employed. Then the constants

$$(\alpha_j, \beta_j, \gamma_j, \hat{\alpha}_j, \hat{\beta}_j, \hat{\gamma}_j, \delta_j), \quad 0 \le j \le N \tag{6.15}$$

are determined. Now only the net spacing or integer N of (6.5a) need be
given. We give the coefficients of (6.15) in table I for some of the most **137**
basic cases to be tested.

6.16 Algorithm and Tests

The following are the main steps for computing the solution branch starting from a known initial solution (u^o, λ^o).

Step I. Determination of the unit tangent to the path, viz $\pm(\dot{u}, \dot{\lambda})$

Step II. Choice of :

(i) the tangent direction;

(ii) steplength $\triangle$, along the path;

(iii) continuation method; natural parameter $(n \cdot p)$ or pseudoarclength $(p \cdot a)$.

Step III. Construction of the initial iterate : either for $(n \cdot p)$:

$$u^{(1)} = u^o + \triangle u;$$

or for $(p \cdot a)$:

$$u^{(1)} = u^o + \triangle \dot{u},$$
$$\lambda^{(1)} = \lambda^o + \triangle \dot{\lambda}$$

Step IV. Newton's iteration to find (δu) of $(\delta u, \delta\lambda)$.

Step V. Computation of the relation error;

If error $< \varepsilon$, then one step is completed and then go to Step I to determine the next solution. If not, then update (u) or (u, λ) and go to Step IV.

The steps in the algorithm are explained in more detail in the flow-
138 chart. No we explain step II, where certain tests are used to make the
various choices mentioned.

The unit tangent $(\dot{u}, \dot{\lambda})$ is determined only up to its sign. The choice of the sign is crucial in order to avoid going backwards or getting trapped between two points. If the tangent at the previous solution is $(\dot{u}^o, \dot{\lambda}^o)$ then we choose the sign for $(\dot{u}, \dot{\lambda})$, so as to make the inner product $\langle(\dot{u}^o, \dot{\lambda}^o), (\dot{u}, \dot{\lambda})\rangle$ positive.

The choice of the steplength is more delicate as it depends upon a number of parameters. If the steplength is too large, we may require too many Newton iterations to converge (or indeed divergence can occur). It the steplength is too small we may require too many steps to compute the entire solution path. Thus in either case the computational efficiency is reduced. At each point on the branch, we allow the steplength $\triangle$, to vary with U_F, the number of iterations required to converge at the previous step. We fit a decreasing exponential curve in the $(\triangle, U)$-plane and choose $\triangle$ according to this function. Specifically, we use

$$\triangle_{\mathrm{new}} = \theta(U_F)\triangle_{old},$$

where

$$\theta(U) \equiv 2^{\left(\frac{4-U}{3}\right)}.$$

Naturally other choices for $\theta(U)$ can be used.

The choice of the method depends on whether we are near a singular **139**
point or not. Most often we can pass over simple bifurcation points even with natural parameter continuation. Thus the difficulty arises mainly near the folds. Since $\dot{\lambda} = 0$ at the fold, it seems reasonable to switch over to pseudoarclength continuation when $|\dot{\lambda}|$ is small enough. Once the fold is crossed over, $|\dot{\lambda}|$ increases again and we switch back using the same test.

In pseudoarclength continuation, two different types of bordering algorithm can be used depending on whether the matrix G_u is nearly singular or not. To do this we need a test to determine if G_u is “sufficiently” singular. For that we look at the LU decomposition for G_u and test the diagonal elements, $u_{j,j}$, of U as follows:

$$\underset{1 \le j \le N-1}{Min} \quad \left| \frac{u_{j,j}}{u_{j-1,j-1}} \right| > M \left| \frac{u_{NN}}{u_{N-1,N-1}} \right|,$$

where M is a constant depending on the number of points N. If the test is passed, we can use the bordering algorithm in the singular case. Otherwise the usual bordering algorithm is used. In most computations, however, the singular bordering algorithm is invoked only when great accuracy is sought.

6.17 Numerical Examples

The algorithm was used to compute the solution branches for simple nonlinearities in double precision with an error tolerance of 1.E-8.

Example 1. We consider the following nonlinear O.D.E.

$$\begin{aligned} u_{rr} + \frac{(n-1)u_r}{r} + u^3 + \lambda = 0 \text{ in } (0,1), \\ u'(0) = 0; u(1) = 0, n \geq 2 \end{aligned} \tag{6.18}$$

140 A difference scheme, as described in the earlier section was used with twenty points. The coefficients are as listed in case (5) of table-I, with the additional assumption $s_L = 0$ and $q_L = 1$. Starting from the trivial solution at $\lambda = 0$, the branch was computed beyond two folds. Table II gives the convergence data for the first few steps. Notice that near the turning point, pseudoarclength continuation is opted for but later natural parameter continuation. The sign of the determinant changes indicating the crossing of a singular point, a turning point in this case. Fig. 6.1 plots, for $n = 2$, norm $u = (\sqrt{|u'(1)u(0)|})\,\mathrm{sgn}\,(-u'(1))$ versus λ, showing the two folds.

After a certain number of iterations, the continuation is stopped and the program proceeds to the location of singular points. Starting from the stored value of (u, λ), just before the determinant changed sign, using a bisection method, the fold is located accurately. In the vicinity of the fold, the singular bordering algorithm (refer section, 4.11, for the case of almost singular A) is used and the method converges well. Table III gives the convergence pattern near the first fold for $n = 2$.

Example 2. Here we consider a two parameter problem:

$$\begin{aligned} -u'' = u^2 - \lambda u + \mu \quad \text{in}(0,1) \\ u(0) = u(1) = 0 \end{aligned} \tag{6.19}$$

141 For $\mu = 0$, the simple bifurcation points along the trivial branch $u \equiv 0$ are at $\lambda = \lambda_j$ where λ_j are the eigenvalues of $\{-\frac{d^2}{dx^2}\}$ on $(0,1)$. We use the perturbed bifurcation approach, (Method V in section 5.26 of chapter 5) to get the branch bifurcating from the trivial branch at λ_1. Here we use μ to perturb the problem (refer (5.33) where τ_q was used in a similar role.) In fig. 6.2, we plot norm $u = (\sqrt{u_{\max}})\,\mathrm{sgn}\,(u'(0))$ versus λ. The difference scheme used is an in case 1 of table - I with $N = 20$.

Starting from

$$(u, \lambda, \mu) = (0, 0, 0).$$

continuation in μ is done up to $\mu = \varepsilon$, keeping λ fixed at 0. Then keeping $\mu = \varepsilon$, fixed, a λ continuation is done till the solutions are sufficiently large. These solutions are indicated in Fig. 6.2 by two solid lines: upper one for $\mu = .1$ and lower on for $\mu = -.1$. Now to switch from $\mu = \varepsilon$ back to $\mu = 0$, we do a μ-continuation. Then keeping $\mu = 0$ fixed, continuing in λ, we get the branch bifurcating from $(0, \lambda_1)$. This branch is indicated by the dotted line in Fig. 6.2.

Example 3. The following nonlinear ODE is the one dimensional version of (6.18) (i.e. $n = 1$) :

$$\begin{aligned} &u'' + u^3 + \lambda = 0 \text{ in } (0, 1) \\ &u(0) = u(1) = 0 \end{aligned} \tag{6.20}$$

Starting from the trivial solution $(u, \lambda) = (0, 0)$, the branch is continued till $|\ \lambda\ | \geq 16$. Two singular points are found and they are located accurately at $\lambda \simeq 1.1$ and $\lambda \simeq -8.07$ using a bisection method as in **142**
example-1 Table-IV shows the Patten near the second singular point.

Once a good approximation to the singular point is obtained, the right and left null vectors Φ and ψ of G_u are computed as indicated in Chapter IV. At the first singular point , $< \psi, G_\lambda > \simeq 120$. This shows that is must be a fold point. At the second singular point where $\lambda \simeq -8.10$, we find that $< \psi, G_\lambda > \simeq 1.E - 4$. (See table-IV and also the remark after lemma (5.20) in chapter 5). This is indeed a bifurcation point. The branch switching is done using method II of Chapter 5, taking care to use a large step size. This gives one part of the second branch. Reversing the direction, we get the other part of the second branch, bifurcating from the original one.

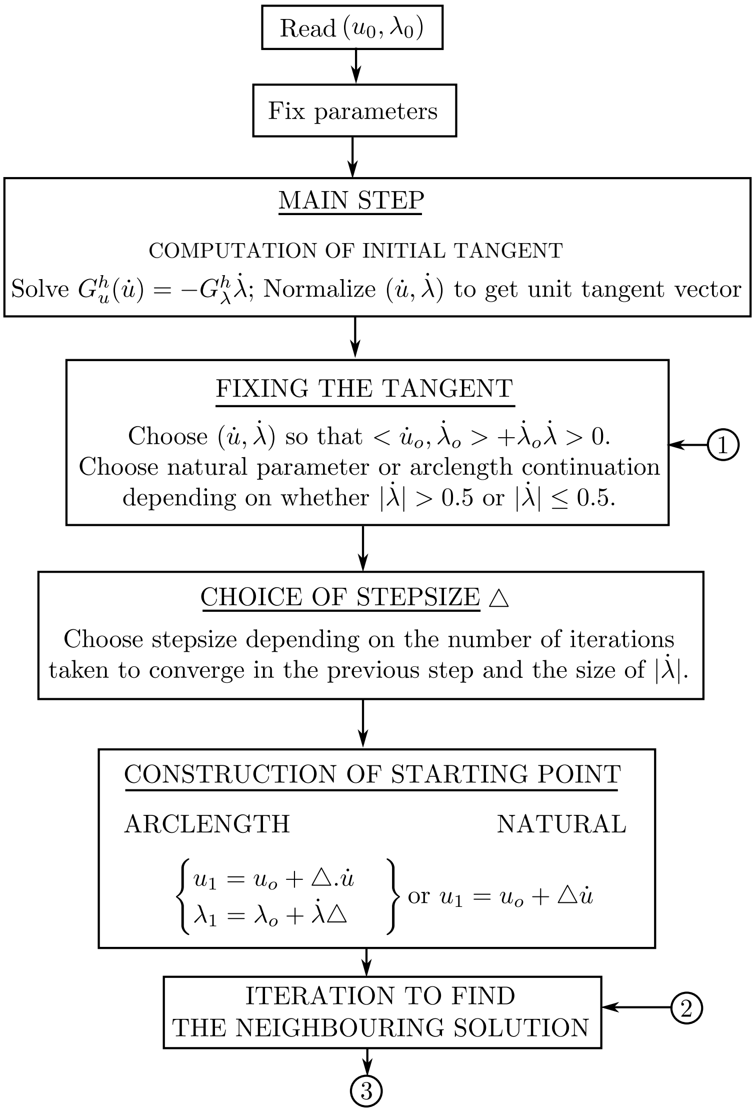
The bifurcating diagram is given in Fig. 6.3. Here norm u actually stands for $(sgnu'(0))u_{\max}$. The branch I represents symmetric solution which are positive till the bifurcating point. After the bifurcation point, the branch develops two nodes near each end. The branch II represents non-symmetric solutions bifurcating from branch I. The solutions in the upper branch have a node near the right end and those in the lower branch near the left end.

143 **Table - 1**

	Case (1)	Case (2)	Case (3)	Case (4)	Case (5)
	$p(x) \equiv 0$	$p(x) \not\equiv 0$	$p(x) \equiv 0$	$p(x) \not\equiv 0$	$p(x) = \frac{m}{x}$
	$r_0 = 0 = r_L$	$r_0 = 0 = r_L$	$r_0 \neq 0, r_L \neq 0$	$r_0 \neq 0, r_L \neq 0$	$r_0 = 1, s_0 = 0$
					$q_o = 0 = r_L$
	4th order accuracy	2nd order accuracy	4th order accuracy	2nd order accuracy	2nd order accuracy
β_0	q_o	q_0	$-\frac{2}{h^2} + \frac{2q_0}{r_0 h}$	$-\frac{2}{h^2} + \frac{2q_0}{r_0 h} - \frac{p_0 q_0}{r_0}$	$-\frac{2}{h^2}(m+1)$
γ_0	0	0	$\frac{2}{h^2}$	$\frac{2}{h^2}$	$\frac{2}{h^2}(m+1)$
$\hat{\beta}_0$	0	0	$5/6$	1	1
$\hat{\gamma}$	0	0	$1/6$	0	0
δ_0	s_0	s_0	$\frac{2s_0}{r_0 h}$	$\frac{s_0}{r_0}(\frac{2}{h} - p_0)$	0
α_j	$1/h^2$	$1/h^2 - p_j/2h$	$1/h^2$	$\frac{1}{h^2} - \frac{p_j}{2h}$	$\frac{1}{h^2} - \frac{m}{2jh^2}$
β_j	$-2/h^2$	$-2/h^2$	$-2/h^2$	$-2/h^2$	$-2/h^2$
γ_j	$1/h^2$	$1/h^2 + p_j/2h$	$1/h^2$	$1/h^2 + p_j/2h$	$1/h^2 + m/2jh^2$
$\hat{\alpha}_j$	$1/12$	0	$1/12$	0	0
$\hat{\beta}_j$	$5/6$	1	$5/6$	1	1
$\hat{\gamma}_j$	$1/12$	0	$1/12$	0	0
α_N	0	0	$2/h^2$	$2/h^2$	$2/h^2$
β_N	q_L	q_L	$-\frac{2}{h^2} - \frac{2q_L}{r_L h}$	$-\frac{2}{h^2} - \frac{2q_L}{r_L h} - \frac{p_N q_L}{r_L}$	$-\frac{2}{h^2} - \frac{q_L}{r_L}(\frac{1}{h} + m)$
$\hat{\alpha}_N$	0	0	$1/6$	0	0
$\hat{\beta}_N$	0	0	$5/6$	1	1
δ_N	s_L	s_L	$-\frac{2s_L}{r_L h}$	$-\frac{s_L}{r_L}(\frac{2}{h} + P_N)$	s_L

Flow Chart

144



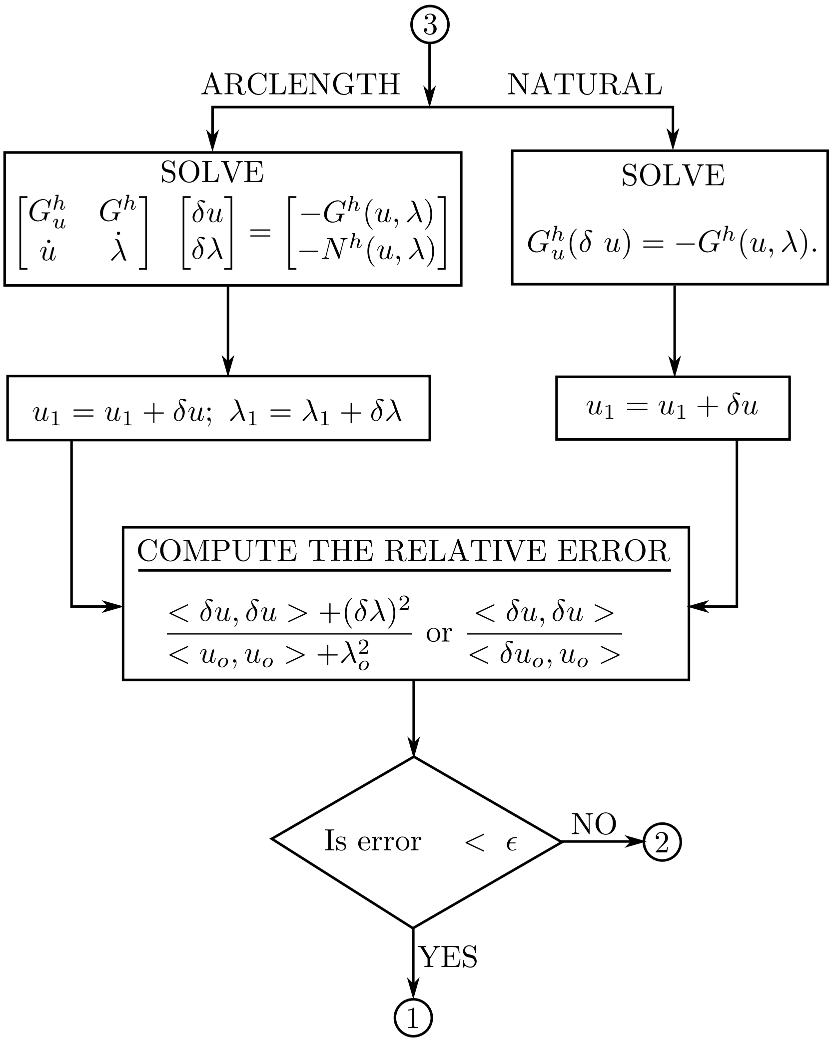

3
ARCLENGTH
NATURAL
SOLVE
$\begin{bmatrix} G_u^h & G^h \\ \dot{u} & \dot{\lambda} \end{bmatrix} \begin{bmatrix} \delta u \\ \delta\lambda \end{bmatrix} = \begin{bmatrix} -G^h(u,\lambda) \\ -N^h(u,\lambda) \end{bmatrix}$
SOLVE
$G_u^h(\delta\ u) = -G^h(u,\lambda).$
$u_1 = u_1 + \delta u;\ \lambda_1 = \lambda_1 + \delta\lambda$
$u_1 = u_1 + \delta u$
COMPUTE THE RELATIVE ERROR
$\frac{< \delta u, \delta u > + (\delta\lambda)^2}{< u_o, u_o > + \lambda_o^2}$ or $\frac{< \delta u, \delta u >}{< \delta u_o, u_o >}$
Is error $< \epsilon$
NO
2
YES
1

TABLE - II **146**

MAIN STEP	METHOD USED	ITERATION	STEPLENGTH Δ	ERROR	λ OBTAINED	SIGN OF THE DETERMINANT
1	(p.a)	1		$.3E-3$		
			0.5		0.23	+1
	nonsingular	2		$.3E-10$		
2	(p.a)	1		$.1E-1$		
	nonsingular	2	0.794	$.2E-5$	0.45	+1
		3		$.5E-13$		
3	(p.a)	1		$.5E-3$		
	nonsingular	2	1.0	$.2E-3$	0.003	−1
		3		$.9E-11$		
4		1		$.5E-1$		
		2		$.5E-2$		
	(n.p)		1.357		−1.35	−1
		3		$.2E-4$		
		4		$.2E-8$		
5		1		$.1E-1$		
		2		$.8E-4$		
	(n.p)		1.357		−2.71	−1
		3		$.1E-7$		
		4		$.5E-15$		

147 **TABLE - III**

MAIN STEP	BORDERING ALGORITHM TYPE	ITERATION	ERROR	λ OBTAINED	LAST PIVOT	SING OF THE DETERMINANT
1	NS	1	.2E-4			
				0.45085	-462.1	-1
	(nonsingular)	2	$.1E-11$			
2	NS	1	$.1E-5$			
				0.45275	66.8	$+1$
		2	$.4E-14$			
3	NS	1	$.9E-9$			
				0.452495	-164.8	-1
		2	$.2E-16$			
4	NS	1	$.4E-6$			
				0.452721	-84.73	-1
		2	$.4E-15$			
5	NS	1	$.8E-6$			
				0.45276	-22.79	-1
		2	$.2E-14$			
6	NS	1	$.1E-5$			
				0.452761	17.51	$+1$
		2	$.2E-14$			
7	S	1	$.9E-6$			
				0.452762	-3.6	-1
	(singular)	2	$.5E-12$			
8	S	1	$.1E-5$			
				0.452762	6.7	$+1$
		2	$.2E-11$			
9	S	1	$.1E-5$			
				0.452762	1.485	$+1$
10	S	1	$.1E-5$			
				0.452762	-1.073	-1
		2	$.4E-13$			

TABLE - IV **148**

MAIN STEP	ITERATION	ERROR	λ OBTAINED	LAST PIVOT	SING OF THE DETERMINANT	$< \Psi, G_\lambda >$
1	1	$.1E-4$				
	2	$.2E-8$	−8.643	−31.9	−1	−2.3
2	1	$.1E-5$				
	2	$.8E-11$	−8.223	−8.4	−1	−.5
3	1	$.8E-7$				
	2	$.5E-13$	−8.014	−3.7	+1	−.26
4	1	$.3E-6$				
	2	$.8E-12$	−8.118	−2.4	−1	−.17
5	1	$.1E-6$				
	2	$.2E-12$	−8.066	−67	+1	−.047
6	1	$.2E-6$				
	2	$.4E-12$	−8.092	−.86	−1	−.061
7	1	$.2E-6$				
	2	$.3E-12$	−8.0792	−.096	−1	−.0068
8	1	$.2E-6$				
	2	$.3E-12$	−8.0727	−.284	+1	−.02
9	1	$.2E-6$				
	2	$.3E-12$	−8.076	−.094	+1	−.0067
10	1	$.2E-6$				
	2	$.3E-12$	−8.07759	−.001	−1	−.00008

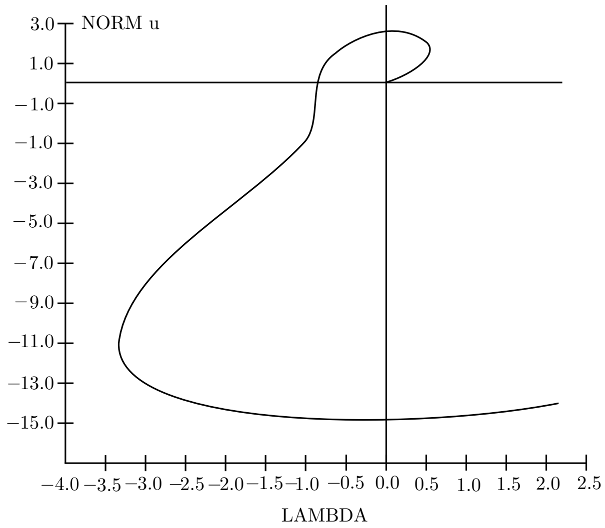


Figure 6.1:

149

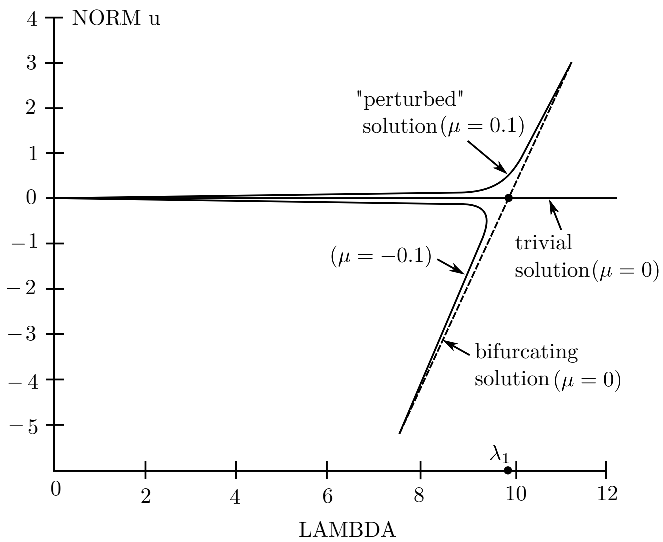


Figure 6.2:

150

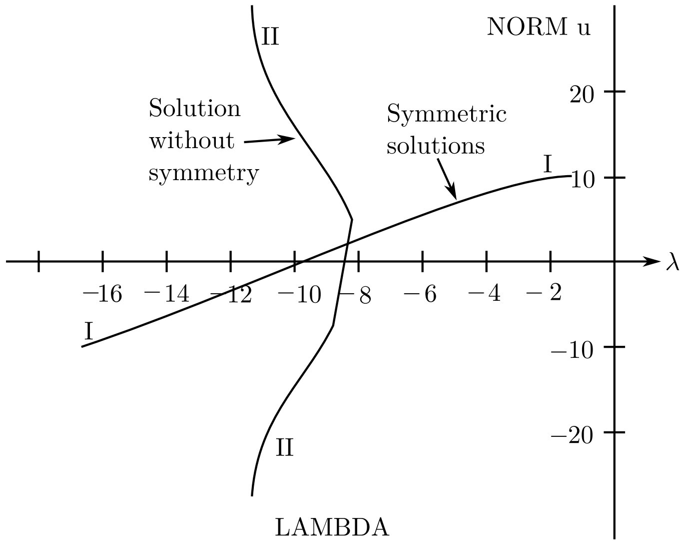

NORM u
II
Solution
without
symmetry
Symmetric
solutions
I
20
10
−16
−14
−12
−10
−8
−6
−4
−2
λ
−10
−20
II
LAMBDA

Figure 6.3:

References

[1] P. ABRAHAM and J. ROBBIN Transversal mapping and flows, Benjamin Press, New York, 1967.

[2] F.H. BRANIN JR., Widely convergent method for finding multiple solutions of simultaneous nonlinear equations, IBM J. Research Develop. 16, 1972, 504 - 522.

[3] Y. M CHEN and P.L. CHRISTIANSEN, Application of a modified Newton's iteration method to construct solutions of eigenvalue problems of nonlinear PDE, SIAM J.Num. Anal. 18, 1970, 335 - 345.

[4] S.N. CHOW and J.K. HALE, Methods of bifurcation theory, Springer Verlag, 1982.

[5] M.G. CRANDALL and P.H. PABINOWITZ, Bifurcation from simple eigenvalues, J. Funct. Anal. 8, 1971, 321 - 340.

[6] D.W DECKER and H. B KELLER, Path following near bifurcation, CPAM, Vol XXXIV, 1981, 149 - 175.

[7] D.W. DECKER and H.B. KELLER, Multiple limit point bifurcation, J1. of Math. Anal. and Appl., Vol. 75, No. 2, June 1980, 417 - 430.

[8] Y.M.J. DEMOULIN and Y.M. CHEN, An interation method for solving nonlinear eigcnvalue problems, SIAM J. Appl. Math., 28, 1975, 588 - 595.

[9] J.E. DENNIS and G.J. MORE, Quasi Newton methods, a motivation and theory, SIAM Review, Vol 19, No. 1, 1977, 46- 89.

[10] E.J. DOEDEL, A. D. JEPSON and H.B. KELLER, Numerical methods for Hopf bifurcation and contiunation of periodic solution paths; in Computing methids in Applied Sciences and Engineering, IV, Ed : R. Glowinski and J.S.Lions, Elsevier Publishers, 1984.

[11] G.F. GAUSS, The struggle for existence, Dover Publications, 1964.

[12] A.H. HOUSEHOLDER, The theory of matrices in Numerical Analysis, Dover Publications, 1975.

[13] G. IOOSS and D. JOSEPH Elementary stability and bifurcation theory, Springer Verlag, 1980.

[14] A.D JEPSON and H.B. KELLER, Steady state and periodic solution paths : their birfurcations and computations ; in *Numerical Methods for Bifuracation Problems, ISNM 70, Ed : Kupper,Mittleman, Weber*, Birkauser Verlag, 1984.

[15] A.D JEPSON and A. SPENCE, Folds in solutions of two parameter systems and their calculations : Part I, SIAM J. Numer. Anal., 22, 1985, 347-368.

[16] L.V.KANTOROVICH, and G.P. AKILOV, Functional Analysis in Normed Spaces, Pergamon Press, 1964.

[17] H.B. KELLER, Practical Procedures in Path following Near Limit Points; in Computing methods in Applied Sciences and Engineering, ed: Glowinski & Lions, North - Holland, 1982.

[18] H.B. KELLER, Constructive methods for Bifurcation and Nonlinear eigenvalue problem; in Computing method in Applied Sciences & Engineering, Ed : Glowinski & Lions, Springer - Varlag, 1977.

[19] H.B. KELLER, Numerical Solutions of Bifurcation and nonlinear eigenvalue problem; in Applications of Bifurcation Theory, Ed : Poul . H. Rabinowitz., Academic Press, 1977, 359-384.

[20] H.B. KELLER, Global Homotopies and Newton methods; in Recent Advances in Numerical Analysis, Ed : C. De Boor and G.H. Golub, Academic press, 1979, 73 -94.

[21] H.B. KELLER, Isolas and perturbed bifurcation theory; in Nonlinear partial differential equations in engineering and applied science, Ed. R.L. Sternberg, A.J Kalinowski and J.S. Papadakis, Marcel Dekker, 1980.

[22] H.B. KELLER and W.F. LANGFORD, Iterations, pertubations and multiplicities for nonlinear bifurcation problem, Arch . Rat. Mech. Anal. 48, 1972, 83 - 108.

[23] H.B. KELLER, and S. ANTMANN Bifurcation Theorey and Nonlinear eigenvalue problem, W.A. Benjamin, 1969.

[24] A.J. LOTKA Elements of Mathematical Biology, Dover Publications

[25] MARSDEN and McCRACKEN, Hopf bifurcation and its applications, Springer Verlag, 1976.

[26] L. NIRENBERG Topics in Nonlinear Functional Analysis, Lecture notes, Courant Inst. MAth Sci. New - York University, 1974.

[27] W.C. RHEINBOLDT, Numerical Analysis of Parametrized nonlinear equations, J. Wiley and Sons, 1986.

[28] D.H. SATTINGER, Topics in Stability and Bifurcation Theory, Lecture Notes in Mathematics, Springer Velag, 1973.

[29] J.T. SCHWARTZ, Nolinear Functional Analysis, Gordon and Breach, 1969.

[30] S. SMALE, A convergent process of price adjustment and global Newton Newton methods J. Math Eco. 3, 1976, 107 - 120.

[31] R.K. WEISS, Bifurcation in difference approximations to two point boundary value problems, Math. of comp. 29, 1975, 746 - 760.

[32] J.H. WILKINSON, The Algebraic Eigenvalue Problem, Oxford Uni. press 1965.

[33] ZHONG - HUA YANG and H.B. KELLER, A direct method for computing order folds, SIAM J. Sci. Stat. Computing, Vol 7, No. 2, April 1986, 351 - 361