COLLOCATION FOR SINGULAR PERTURBATION PROBLEMS I:

FIRST ORDER SYSTEMS WITH CONSTANT COEFFICIENTS

by

U. Ascher* and R. Weiss**

Technical Report 81-2

February 1981

ABSTRACT

The application of collocation methods for the numerical solution of singularly perturbed ordinary differential equations is investigated. Collocation at Gauss, Radau and Lobatto points is considered, for both initial and boundary value problems for first order systems with constant coefficients. Particular attention is paid to symmetric schemes for boundary value problems; these problems may have boundary layers at both interval ends.

Our analysis shows that certain collocation schemes, in particular those based on Gauss or Lobatto points, do perform very well on such problems, provided that a fine mesh with steps proportional to the layers' width is used in the layers only, and a coarse mesh, just fine enough to resolve the solution of the reduced problem, is used in between. Ways to construct appropriate layer meshes are proposed. Of all methods considered, the Lobatto schemes appear to be the most promising class of methods, as they essentially retain their usual superconvergence power for the smooth, reduced solution, whereas Gauss-Legendre schemes do not.

We also investigate the conditioning of the linear systems of equations arizing in the discretization of the boundary value problem. For a row equilibrated version of the discretized system we obtain a pleasantly small bound on the maximum norm condition number, which indicates that these systems can be solved safely by Gaussian elimination with scaled partial pivoting.

---

* Department of Computer Science, University of British Columbia, Vancouver, Canada V6T 1W5. The research of this author was supported in part under NSERC grant A4306.

**Institut für Angewandte und Numerische Mathematik, Technische Universitat Wien, 1040 Wien, Gusshausstrasse 27-29, Austria.

COLLOCATION FOR SINGULAR PERTURBATION PROBLEMS I: FIRST ORDER SYSTEMS WITH CONSTANT COEFFICIENTS

by

U. Ascher and R. Weiss

1. INTRODUCTION

The numerical treatment of singular perturbation problems in ordinary differential equations (ODEs) has received a significant amount of attention in recent years. The analytic solution of such problems usually exhibits thin transition layers, in which the solution varies rapidly. When attempting to approximate such solutions by one of the familiar symmetric difference schemes on a coarse mesh, large oscillations may pollute the numerical solution on the entire interval of integration; see examples in sections 5.3 and 6.3. The efforts to deal with this situation generally fall into two classes which we discuss below.

One approach is to design special purpose methods which would yield accurate solutions, at least away from the transition layers. Thus, large errors, which may be generated at a layer, should decay fast. For very stiff initial value problems (IVPs), methods for which the growth factor $\gamma(\zeta)$, (which is essentially their approximation to e^{ζ}), satisfies $\gamma(\zeta) \to 0$ as $re(\zeta) \to -\infty$ should be used. But for boundary value problems (BVPs) these methods have to be properly upwinded, and this may restrict their applicability for general purpose algorithms. Also, if a

fine resolution of a transition layer is desired then the mesh has to be refined locally in that layer's region.

The second approach, used for BVPs, is to stick with a symmetric scheme but to refine the mesh locally, often using an adaptive algorithm. A number of such algorithms have been proposed and used (see Russell [19] for a survey) and several general purpose codes contain implementations of mesh adapting techniques. These techniques attempt to choose a mesh which equidistributes a certain measure of the error, usually an estimate of the local truncation error or a local estimate of the error itself. These estimates are not valid rigorously unless the maximum mesh size is much smaller than the layers' widths, but this has been ignored in practice, and codes like PASVAR (Lentini-Pereyra [14]) or COLSYS (Ascher, Christiansen, Russell [2]) have solved successfully fairly complex singular perturbation problems (e.g., see Ascher [1]), not backed by theory. On some occasions, however, computations with these codes were not so successful.

Thus, we have set out to extend the theory of convergence and stability of collocation methods to singular perturbation problems with the hope to achieve a better insight to the success or failure of various methods in solving practical problems of this type. The collocation schemes which we discuss include the families of Gauss-Legendre, Gauss-Radau and Gauss-Lobatto points (we refer to them as Gauss, Radau and Lobatto points, respectively), which in turn contain familiar difference schemes like the box scheme, the backward Euler scheme and the trapezoidal scheme as special cases.

We limit our discussion here to boundary value problems for first order systems with constant coefficients which have a bounded inverse. Thus

consider the problem of order n+m, with n equations singularly perturbed,

$$\underset{\sim}{\varepsilon y}' = A_{11}\underset{\sim}{y} + A_{12}\underset{\sim}{z} + \underset{\sim}{f}_1 \qquad 0 \leq t \leq 1 \tag{1.1}$$

$$\underset{\sim}{z}' = A_{21}\underset{\sim}{y} + A_{22}\underset{\sim}{z} + \underset{\sim}{f}_2, \tag{1.2}$$

plus the boundary conditions

$$B_0 \begin{bmatrix} \underset{\sim}{y} \\ \underset{\sim}{z} \end{bmatrix}(0) + B_1 \begin{bmatrix} \underset{\sim}{y} \\ \underset{\sim}{z} \end{bmatrix}(1) = \underset{\sim}{\beta}. \tag{1.3}$$

The matrices A_{11}, A_{12}, A_{21}, A_{22}, B_0 and B_1 are constant and have dimensions nxn, nxm, mxn, mxm, (n+m) x (n+m) and (n+m) x (n+m), respectively.

Our analysis shows that certain collocation methods do perform very well on (1.1) - (1.3), provided that a fine mesh with meshspacings proportional to ε is used in the layers only, and a coarse mesh, just fine enough to resolve the solution of the reduced problem, is used in between. Of all methods considered, the Lobatto schemes appear to be the most promising class of methods for singularly perturbed BVPs.

We also investigate the conditioning of the linear systems of equations arising in the discretization of (1.1) - (1.3). For a row equilibrated version of the system we obtain a pleasantly small bound on the maximum norm condition number, which indicates that these systems can be solved safely by Gaussian elimination with scaled partial pivoting.

An outline of the paper follows. In §2 we have collected the necessary mathematical tools, and in §3 we introduce the collocation methods under investigation and consider their stability properties. In sections 4, 5 and

6 we consider the cases of scalar equations, IVP systems and BVP systems, respectively. For each case we consider four topics: the convergence properties of the numerical methods in the absence of boundary layers, the conditioning of the resulting linear system of equations, and two layer treatments. One layer treatment is designed to obtain an accurate representation of the solution everywhere, while the other is designed to get rid of layer effects away from layers, not worrying about accuracy inside the layer regions. As well, a numerical example is given in each of the last 3 sections, to illustrate the theoretical results.

While our main concern is in BVPs, their analysis requires the prior analysis of IVPs, which in turn requires the analysis of scalar IVPs. Thus in §6 and in §2.3 we decompose the BVP solution into a smooth component and two stiff IVP solution components. We are then able to capitalize on results obtained in §5 for IVP systems. In addition, our result concerning the conditioning of the BVP approximation process is given in theorem 6.2. In §5, our results concerning the convergence properties of the numerical methods for the smooth solution are contained in theorem 5.2. The analysis relies heavily on the feasibility of isolating the slow components and considering each differential equation separately, as shown in §2.2. In §4, the smooth convergence for a scalar equation is considered in theorem 4.1 and the layer treatments are discussed in sections 4.2 and 4.3. These latter sections rely on the stability results of §3.

The extensions of the present results to variable coefficients and to higher order systems will be described in subsequent papers.

2. ANALYTIC BACKGROUND

Throughout this paper we shall assume, for simplicity (and readability), that the functions appearing in the differential equations considered are inifinitely differentiable. Also, unless otherwise indicated, the maximum (or sup) norm is used.

2.1 Scalar equations

Consider the IVP

$$\varepsilon y' = \lambda y + g(t), \qquad 0 \le t \le 1,$$
$$y(0) \text{ given}, \tag{2.1}$$

where $\varepsilon > 0$, $re(\lambda) < 0$ and $g(t) = g(t;\varepsilon)$ is smooth and has an asymptotic power series in ε, $g(t;\varepsilon) = \sum_{\nu=0}^{\infty} \varepsilon^{\nu} g_{\nu}(t)$. The solution of (2.1) is

$$y(t) = y(0)\exp\{\frac{\lambda}{\varepsilon} t\} + \frac{1}{\varepsilon}\int_0^t \exp\{\frac{\lambda}{\varepsilon}(t-s)\}g(s)ds = \tag{2.2}$$

$$= [y(0) + \frac{1}{\lambda} g_0(t)]\exp\{\frac{\lambda}{\varepsilon} t\} - \frac{1}{\lambda} g_0(t) + O(\varepsilon).$$

Hence, for small ε the solution essentially consists of the contributions of a smooth component,

$$y_R(t) = -\frac{1}{\lambda} g_0(t) \tag{2.3}$$

which is the solution to the reduced equation obtained when in (2.1) we set $\varepsilon = 0$, and a fast component, which connects the initial value to $y_R(t)$ via a narrow transition layer of width $O(\varepsilon)$ with a "layer jump" of size $y(0) + \frac{1}{\lambda} g_0(0)$. So this layer is present in the solution unless $y(0) = -\frac{1}{\lambda} g(0)$.

The general solution of (2.1) can be written as

$$(2.4) \quad y(t) = c \exp\{\frac{\lambda}{\varepsilon} t\} + \hat{y}(t), \qquad c = \text{const.},$$

where $\hat{y} = \hat{y}(t;\varepsilon)$ is a particular solution. For any $q > 0$ we can find a $\hat{y}$ such that

$$(2.5) \quad ||\hat{y}^{(\ell)}|| \leq c_q, \quad \ell = 0,1,\ldots,q; \quad 0 < \varepsilon \leq \varepsilon_0; \quad c_q = \text{const.}$$

This particular $\hat{y}$ has a $(q-1)$ term expansion in powers of ε, which is constructed by the standard recursive procedure, see e.g. O'Malley [15]. The principal term in this expansion is $y_R(t)$.

2.2 Systems of equations

Here we consider the system (1.1), (1.2) where we assume that A_{11} is nonsingular. Note that if $A_{21} = 0$ then the differential equations for $\underset{\sim}{z}$ do not have any ε term and so the standard theory applies to $\underset{\sim}{z}$. Thus, our first goal is to transform (1.1), (1.2) into such a form.

Consider the transformation

$$(2.6) \quad \begin{bmatrix} \underset{\sim}{y} \\ \underset{\sim}{z} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \varepsilon L & I \end{bmatrix} \begin{bmatrix} \underset{\sim}{u} \\ \underset{\sim}{v} \end{bmatrix}$$

where L is a constant mxn matrix. Substituting in (1.1), (1.2) we get

$$(2.7)\quad \underset{\sim}{v}' = [A_{21} - LA_{11} - \varepsilon LA_{12}L + \varepsilon A_{22}L]\underset{\sim}{u} + [A_{22} - LA_{12}]\underset{\sim}{v} + [\underset{\sim}{f}_2 - L\underset{\sim}{f}_1].$$

Thus we need to find L so that

$$(2.8)\quad A_{21} - LA_{11} - \varepsilon N(L) = 0; \qquad N(L) = LA_{12}L - A_{22}L$$

Since A_{11} is nonsingular, we can write this as

$$(2.9)\quad L = A_{21}A_{11}^{-1} - \varepsilon N(L)A_{11}^{-1}$$

Now, the question is whether there is a solution L to (2.9). But the answer is obviously affirmative for ε small enough, because then the mapping on the right hand side of (2.9) is contractive in a neighborhood of $A_{21}A_{11}^{-1}$. So we have

Lemma 2.1. If $0 < \varepsilon \leq \varepsilon_0$ with ε_0 small enough, then (2.9) has a unique solution of the form $L = A_{21}A_{11}^{-1} + O(\varepsilon)$; hence the transformation (2.6) is decoupling.

Note that, since the application of the transformation (2.6) and the application of a collocation scheme obviously commute, we do not need to know L; all we need is its existence. We can then apply the numerical method to the coupled system and the analysis to the uncoupled one.

Suppose now that already $A_{21} = 0$ in (1.1) - (1.2). In order to be able to use results obtained for scalar equations for the systems analysis, we need

one more transformation. If

$$A_{11} = EJE^{-1} \tag{2.10}$$

with J the Jordan canonical form of A_{11}, then we will look at

$$\underset{\sim}{w} = E^{-1}\underset{\sim}{y} \tag{2.11}$$

Then, if J is diagonal, for each component of $\underset{\sim}{w}$ we obtain an equation like (2.1), with $\underset{\sim}{z}$ considered as part of the inhomogeneous term g(t), i.e. the analysis can be reduced to analysis for a sequence of scalar problems. Furthermore, this reduction can be achieved even for the defective case by the following device (cf. deHoog-Weiss [12]): Let

$$\hat{J} = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix}, \qquad re(\lambda) < 0 \tag{2.12}$$

be a typical Jordan block and consider the system

$$\varepsilon\underset{\sim}{w}' = \hat{J}\underset{\sim}{w} + \underset{\sim}{g}(t) \tag{2.13}$$

which has the general solution

$$\underset{\sim}{w}(t) = \exp\{\frac{\hat{J}}{\varepsilon}\, t\}\underset{\sim}{w}_0 + \frac{1}{\varepsilon}\int_0^t \exp\{\frac{\hat{J}(t-s)}{\varepsilon}\}\, \underset{\sim}{g}(s)ds. \tag{2.14}$$

Since for any holomorphic function ψ,

$$(2.15)\qquad \psi(\hat{J}) = \frac{1}{2\pi i}\oint_C \psi(\mu)(\mu I - \hat{J})^{-1}d\mu$$

where C is some circle centered at $\mu = \lambda$, it follows that

$$(2.16)\qquad \underset{\sim}{w}(t) = \frac{1}{2\pi i}\oint_C \left[\exp\{\frac{\mu t}{\varepsilon}\}(\mu I-\hat{J})^{-1}\underset{\sim}{w}_0 + \frac{1}{\varepsilon}\int_0^t \exp\{\frac{\mu}{\varepsilon}(t-s)\}(\mu I-\hat{J})\underset{\sim}{g}(s)ds\right]d\mu =$$

$$= \frac{1}{2\pi i}\oint_C \underset{\sim}{w}_\mu(t)d\mu$$

where

$$(2.17)\qquad \begin{aligned} \varepsilon\underset{\sim}{w}_\mu' &= \mu\underset{\sim}{w}_\mu + (\mu I - \hat{J})^{-1}\underset{\sim}{g}(t) \qquad 0 \le t \le 1 \\ \underset{\sim}{w}_\mu(0) &= (\mu I - \hat{J})^{-1}\underset{\sim}{w}_0 \end{aligned}$$

Now note that for each μ, (2.17) is a diagonal system and that we can choose C so that $re(\mu) < 0$.

2.3 Boundary value problems

Consider again the system (1.1) - (1.2) with A_{11} nonsingular and $A_{21} = 0$. For the IVP to be well defined we need that all fast components decay, i.e. that the eigenvalues of A_{11} have negative real parts. The solution then consists of a smooth term plus a possible layer at $t = 0$.

In general, write

$$(2.18)\qquad J = \begin{pmatrix} J_- & 0 \\ 0 & J_+ \end{pmatrix}$$

where J is defined in (2.10), J_- is its n_- x n_- submatrix corresponding to the eigenvalues λ with $re(\lambda) < 0$ and J_+ is its n_+ x n_+ submatrix corresponding to the eigenvalues with $re(\lambda) > 0$. Now, for $\underset{\sim}{w} = E^{-1}\underset{\sim}{y}$ we can write (1.1) as

$$(2.19) \qquad \varepsilon \underset{\sim}{w}' = J\underset{\sim}{w} + B_{12}\underset{\sim}{z} + \underset{\sim}{g}_1$$

with $B_{12} = E^{-1}A_{12}$, $\underset{\sim}{g}_1 = E^{-1}\underset{\sim}{f}_1$. Partitioning $\underset{\sim}{w} = \begin{pmatrix} \underset{\sim}{w}_- \\ \underset{\sim}{w}_+ \end{pmatrix}$ in an obvious way, we consider the following 3 BVPs, where H stands for the system consisting of the homogeneous equations for (2.19), (1.2), and where the matrix boundary conditions are taken one column at a time:

I. $H(\underset{\sim}{w},\underset{\sim}{z}) = 0, \quad \underset{\sim}{w}_-(0) = W_-(\varepsilon), \quad \underset{\sim}{w}_+(1) = W_+(\varepsilon), \quad \underset{\sim}{z}(0) = I$

with the matrices W_-, W_+ chosen so that the solution matrix W_I be smooth.

II. $H(\underset{\sim}{w},\underset{\sim}{z}) = 0, \quad \underset{\sim}{w}_-(0) = I, \quad \underset{\sim}{w}_+(1) = 0, \quad \underset{\sim}{z}(0) = 0$

with the solution matrix W_{II}.

III. $H(\underset{\sim}{w},\underset{\sim}{z}) = 0, \quad \underset{\sim}{w}_-(0) = 0, \quad \underset{\sim}{w}_+(1) = I, \quad \underset{\sim}{z}(0) = 0$

with the solution matrix W_{III}.

Then the general solution of the homogeneous system (2.19), (1.2) can be written as

$$(2.20) \qquad \begin{pmatrix} \underset{\sim}{w}_H \\ \underset{\sim}{z}_H \end{pmatrix} = W_I\underset{\sim}{\eta}_I + W_{II}\underset{\sim}{\eta}_{II} + W_{III}\underset{\sim}{\eta}_{III}$$

where $\underset{\sim}{\eta}_I \in R^m$, $\underset{\sim}{\eta}_{II} \in R^{n_-}$, $\underset{\sim}{\eta}_{III} \in R^{n_+}$ are arbitrary vectors, and the general solution of the inhomogeneous problem is

$$(2.21)\qquad \begin{bmatrix} w \\ z \end{bmatrix} = \begin{bmatrix} w_H \\ z_H \end{bmatrix} + \begin{bmatrix} w_p \\ z_p \end{bmatrix}$$

with $\begin{bmatrix} w_p \\ z_p \end{bmatrix}$ a particular solution which we can choose to be a smooth one (see §2.1).

Substituting (2.21) into the boundary conditions obtained from (1.3) by the transformation (2.11), we obtain a linear system for $\eta = (\eta_I, \eta_{II}, \eta_{III})^T$

$$(2.22)\qquad A(\varepsilon)\eta = \hat{\beta}$$

with $\hat{\beta}$ an appropriate right hand side containing the contribution of the particular solution and

$$(2.23)\qquad A(\varepsilon) = A_0 + \varepsilon A_1 + \ldots$$

The matrix A_0 is nonsingular if and only if the problem is well posed uniformly in ε (which we assume).

To conclude, then, the solution of the BVP consists of a smooth part with possible boundary layers at both ends. In §6 we show that the approximate collocation solution also splits in a similar fashion and thus the results for the IVPs II and III, which are given in §5, become applicable to the BVP.

3. COLLOCATION AND IMPLICIT RUNGE-KUTTA METHODS

3.1 The numerical method

The numerical schemes considered in this paper all belong to the class of collocation methods using continuous piecewise polynomials. Consider the mesh

$$0 = t_1 < t_2 < \ldots < t_N < t_{N+1} = 1$$

(3.1)

$$h_i = t_{i+1} - t_i,\ h = \max_{1\le i\le N} h_i,\ \underline{h} = \min_{1\le i\le N} h_i.$$

On this mesh our k-stage collocation method is completely defined as a function of a fixed set of points

(3.2) $$0 \le \rho_1 < \rho_2 < \ldots < \rho_k \le 1$$

by requiring that the approximate solution $(\underset{\sim}{y}^h, \underset{\sim}{z}^h)$ component-wise be in $C[0,1]$ and reduce to a polynomial of degree at most k on each subinterval (t_i, t_{i+1}), that it satisfy the boundary (or initial) conditions (1.3), and that it satisfy the differential equations (1.1), (1.2) at the collocation points $t_{ij} = t_i + h_i\rho_j$, $i = 1,\ldots,N$, $j = 1,\ldots,k$ (cf. de Boor-Swartz [4], Russell [18], Weiss [22]).

We require of all the methods considered here that they be A-stable, i.e. that their growth function $\gamma(\zeta)$ satisfy

(3.3) $$|\gamma(\zeta)| \le 1, \quad re(\zeta) < 0.$$

Three classes of collocation methods are then considered.

I. $\rho_1 > 0$, $\rho_k < 1$ and the points are placed symmetrically about ½. In particular the Gauss schemes belong to this class, the box scheme being its simplest member. The growth function of these schemes is given by

$$\gamma(\zeta) = \frac{\sum_{j=0}^{k} \gamma_j \zeta^j}{\sum_{j=0}^{k} \gamma_j (-\zeta)^j} \quad ; \quad \gamma_j = \frac{(2k-j)!k!}{(2k)!j!(k-j)!} \tag{3.4}$$

(see Saff-Varga [20]). Thus

$$\gamma(\zeta) \to \begin{cases} +1 & k \text{ even} \\ -1 & k \text{ odd} \end{cases} \qquad re(\zeta) \to -\infty \tag{3.5}$$

II. $\rho_1 > 0$, $\rho_k = 1$. In particular, the Radau schemes (and thus the backward Euler scheme) belong to this class. The growth function satisfies

$$\gamma(\zeta) \to 0 \qquad re(\zeta) \to -\infty \tag{3.6}$$

and these methods exhibit stiff decay (Varah [21], Prothero-Robinson [16]). The application of such methods to singularly perturbed BVPs is examined in Ringhofer [17].

III. $\rho_1 = 0$, $\rho_k = 1$ and the points are placed symmetrically about ½. In particular, the Lobatto schemes (and thus the trapezoidal scheme) belong

to this class and their growth function is given by (3.4) with k replaced by k-1. Consequently (3.5) holds with the even-odd roles interchanged.

3.2 Implicit Runge-Kutta (RK)

The equivalence of the above collocation schemes to certain RK methods is well known (Butcher [6], Axelsson [3], Wright [23], Weiss [22]), but we repeat it here explicitly for the sake of completeness. For simplicity consider a scalar nonlinear differential equation

$$u' = g(t;u). \tag{3.7}$$

The polynomial piece $u^h(t)$ is defined for $t_i \le t \le t_{i+1}$ by

$$u^h(t_i) = u_i, \quad (u^h)'(t_{ij}) = g(t_{ij};u^h(t_{ij})), \quad j = 1,\ldots,k, \tag{3.8}$$

where u_i is obtained from the previous subinterval. Let

$$F_j = g(t_{ij};u^h(t_{ij})), \quad j = 1,\ldots,k \tag{3.9}$$

and express u^h in terms of interpolation to the values $u_i, F_1, \ldots, F_k$:

$$u^h(t) = u_i + h_i \sum_{j=1}^{k} F_j \phi_j\left(\frac{t-t_i}{h_i}\right) \tag{3.10}$$

where $\phi_j(x)$, $j = 1,\ldots,k$ are polynomials of degree at most k on [0,1],

determined by interpolation conditions

$$\phi_j(0) = 0; \quad \phi_j'(\rho_\ell) = \delta_{j\ell}, \quad \ell = 1,\ldots,k. \tag{3.11}$$

Now let $\hat{A}$ be the kxk matrix and $\hat{\underset{\sim}{b}}$ be the k-vector defined by

$$\hat{b}_j = \phi_j(1); \quad \hat{a}_{j\ell} = \phi_\ell(\rho_j). \tag{3.12}$$

Then we get the following equivalent RK method

$$\begin{aligned} u_{i+1} &= u_i + h_i \sum_{j=1}^{k} \hat{b}_j F_j \\ F_j &= g(t_{ij}; u_i + h_i \sum_{\ell=1}^{k} \hat{a}_{j\ell} F_\ell), \qquad j = 1,\ldots,k. \end{aligned} \qquad 1 \le i \le N \tag{3.13}$$

It should be noted that not every RK scheme is equivalent to a collocation scheme. For instance, by (3.12) if $\rho_k = 1$ then the last row of $\hat{A}$ must be identical to $\hat{\underset{\sim}{b}}^T$ for the scheme to be a collocation one. But we feel that, among the most accurate schemes (Gauss, Radau and Lobatto points) the more important RK schemes are, in fact, the collocation ones.

The following lemma is now very easy to verify.

Lemma 3.1. The matrix $\hat{A}$ of (3.12) is nonsingular if and only if $\rho_1 > 0$. If $\rho_1 = 0$ then we can write

$$\hat{A} = \left[\begin{array}{c|c} 0 & 0 \\ \hline \bar{\underset{\sim}{a}} & \bar{A} \end{array}\right] \tag{3.14}$$

with $\tilde{a}$ a (k-1) vector and $\bar{A}$ a (k-1) x (k-1) nonsingular matrix.

3.3 Properties of the growth function $\gamma(\zeta)$

For the scalar equation

$$\varepsilon y' = \lambda y \; ; \quad y(0) = 1, \tag{3.15}$$

a RK method reads

$$y_{i+1} = \gamma(\frac{\lambda h_i}{\varepsilon}) y_i ; \quad y_1 = 1. \tag{3.16}$$

Comparing the solutions of the differential and difference equations, the following result is obtained.

Lemma 3.2. If the RK method is of order p then

$$\gamma(\zeta) = e^{\zeta} + c_{\gamma}(\zeta)^{p+1} + O(|\zeta|^{p+2}), \quad \zeta \in C \tag{3.17}$$

with c_{γ} a computable constant. (In particular for Gauss or Lobatto points, c_{γ} is easily obtained by comparing (3.17) with (3.4)). Thus

$$\gamma^{(j)}(0) = 1, \qquad j = 0,1,\ldots,p. \tag{3.18}$$

Recall that, in particular p = 2k for Gauss, p = 2k-1 for Radau and p = 2k-2 for Lobatto points. Since $|\gamma(\frac{\lambda h_i}{\varepsilon})|$ is the rate of decay in the

difference scheme (3.16), it is of interest to find where $|\gamma(\zeta)|$ attains its minimum for a given fixed $\arg(\zeta)$ and what the minimum value is. For the Radau schemes the answer is given by (3.6), but for the Gauss (and hence Lobatto) schemes we need to know where $|\gamma(\zeta)|$ is small, because appropriate step sizes may damp errors more effectively than others.

In figures 1, 2, 3, 4 and 5, contour plots of $|\gamma(\zeta)|$ in the upper left quadrant of the complex plane for levels 0.1 (0.1) 1.0 are given for Gauss points with k = 1,2,3,4 and 5, respectively. For a Lobatto scheme with k points, figure (k-1) applies. In addition, for each k a curve of solutions for $\min\{|\gamma(\zeta)|;\ \arg(\zeta)\ \text{fixed}\}$ is plotted. Thus, each point on this curve is obtained by minimizing the corresponding one dimensional function $|\gamma(\zeta)|$ along the ray from the origin passing through that point. So, in particular, this curve is othogonal to the contour lines where it crosses them. The actual values for ζ and $|\gamma(\zeta)|$ thus obtained for the negative real axis are given in table 3.1.

From the plots we see that, as k increases, the optimal damping steps increase and, more importantly, for a fixed ray near the imaginary axis, the value of $|\gamma(\zeta)|$ decreases significantly. This points out an advantage that higher order methods may have, as exploited in further sections. (In particular, see example (5.18)).

Table 3.1: Optimal damping steps and rates for Gauss points on the negative real axis

k	ζ	$\lvert\gamma(\zeta)\rvert$
1	-2.	0
2	-3.46	.0718
3	-4.64	0
4	-6.10	.00508
5	-7.29	0

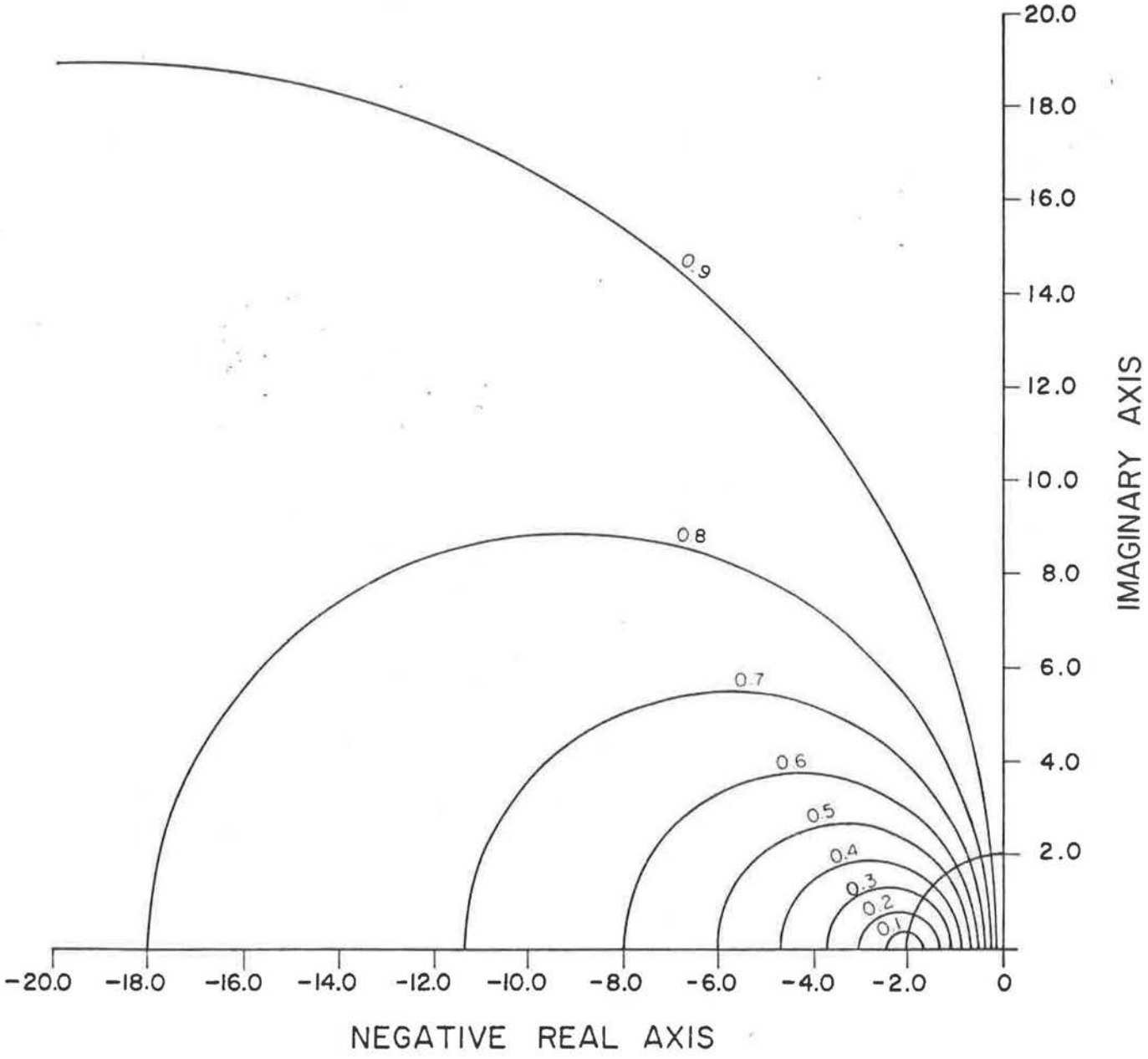

Figure 1: Contour levels of $|\gamma(\zeta)|$ and curves of minimal values: k=1.

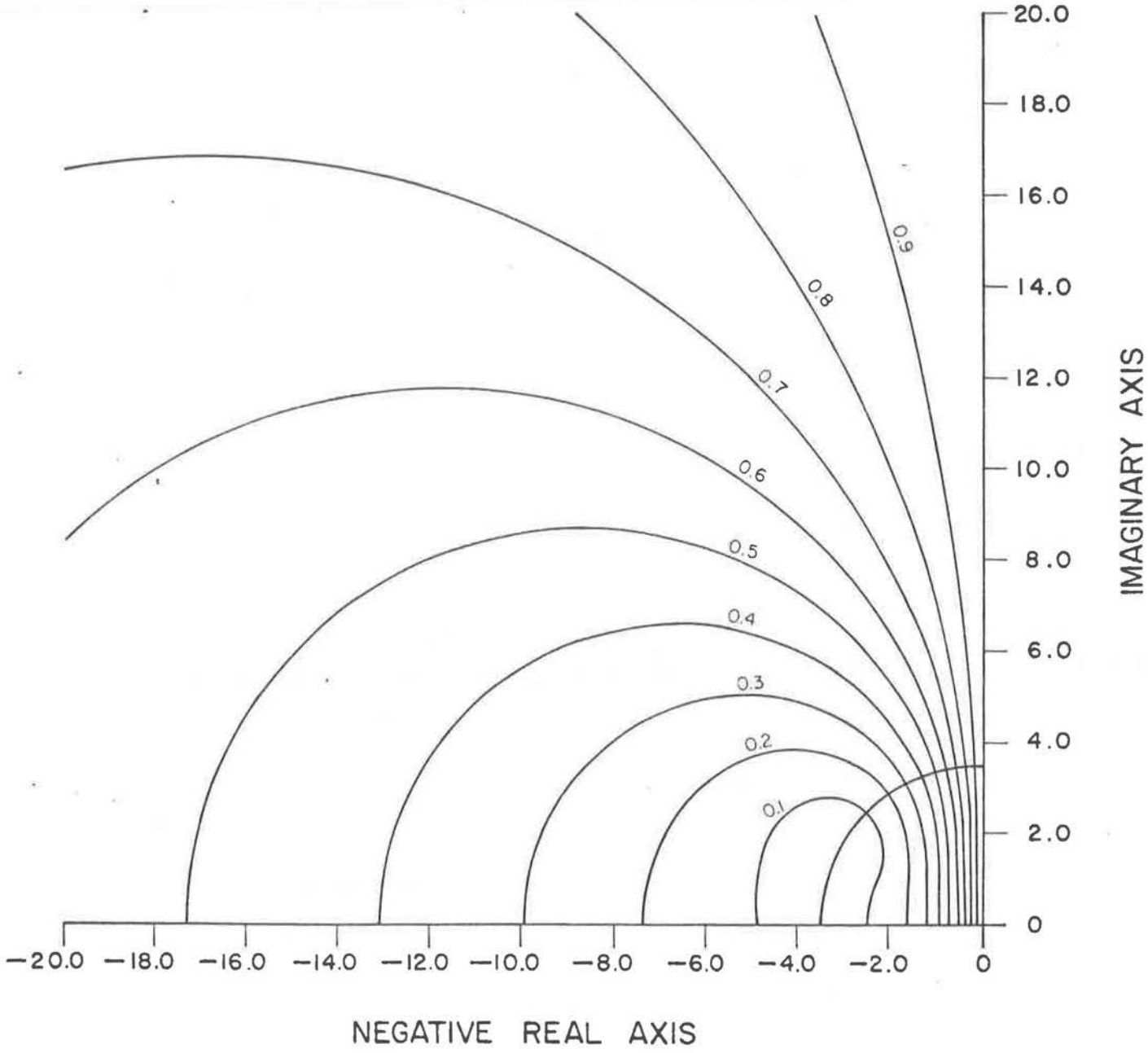

Figure 2: Contour levels of $|\gamma(\zeta)|$ and curves of minimal values: k=2.

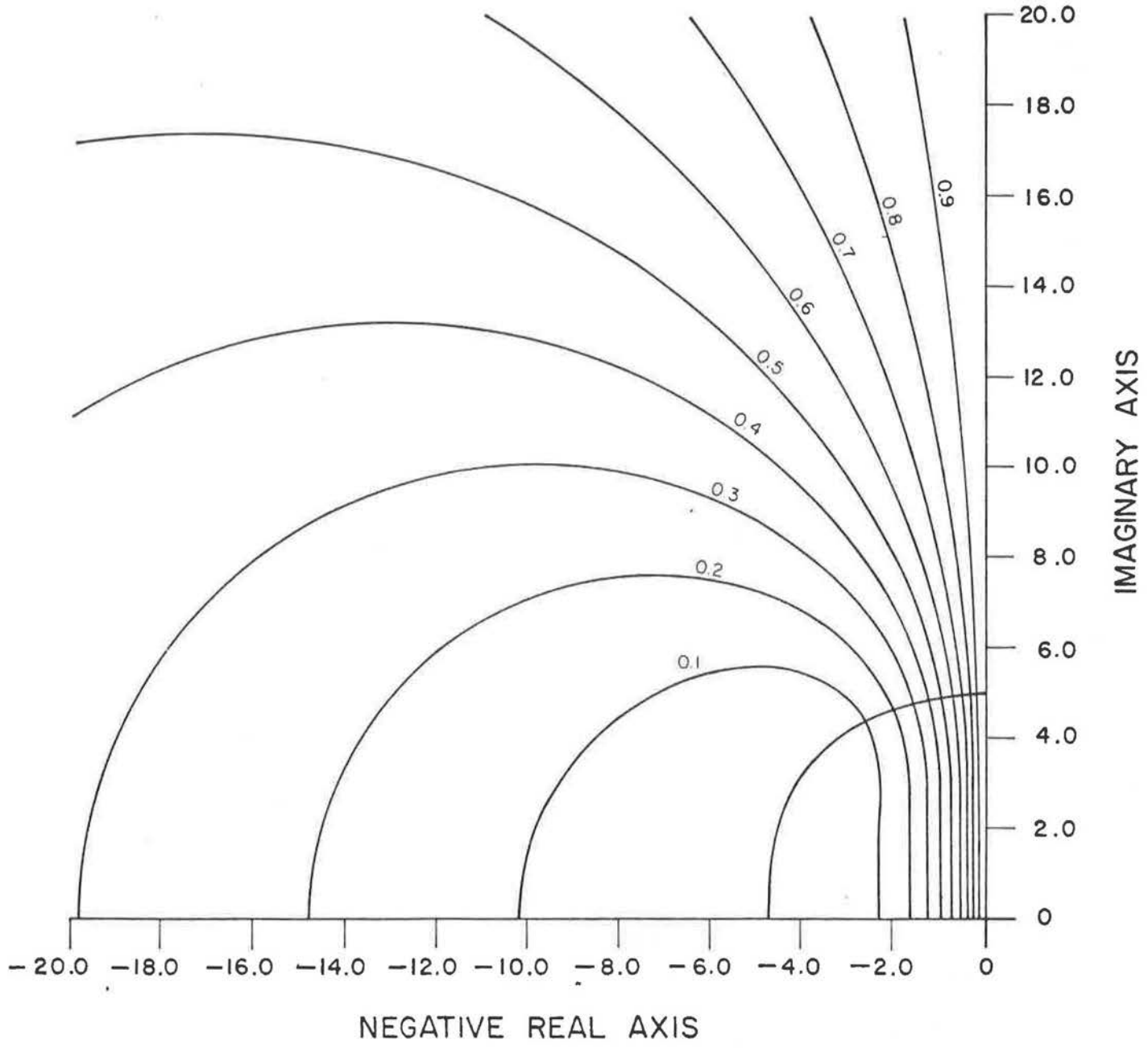

Figure 3: Contour levels of $|\gamma(\zeta)|$ and curves of minimal values: k=3.

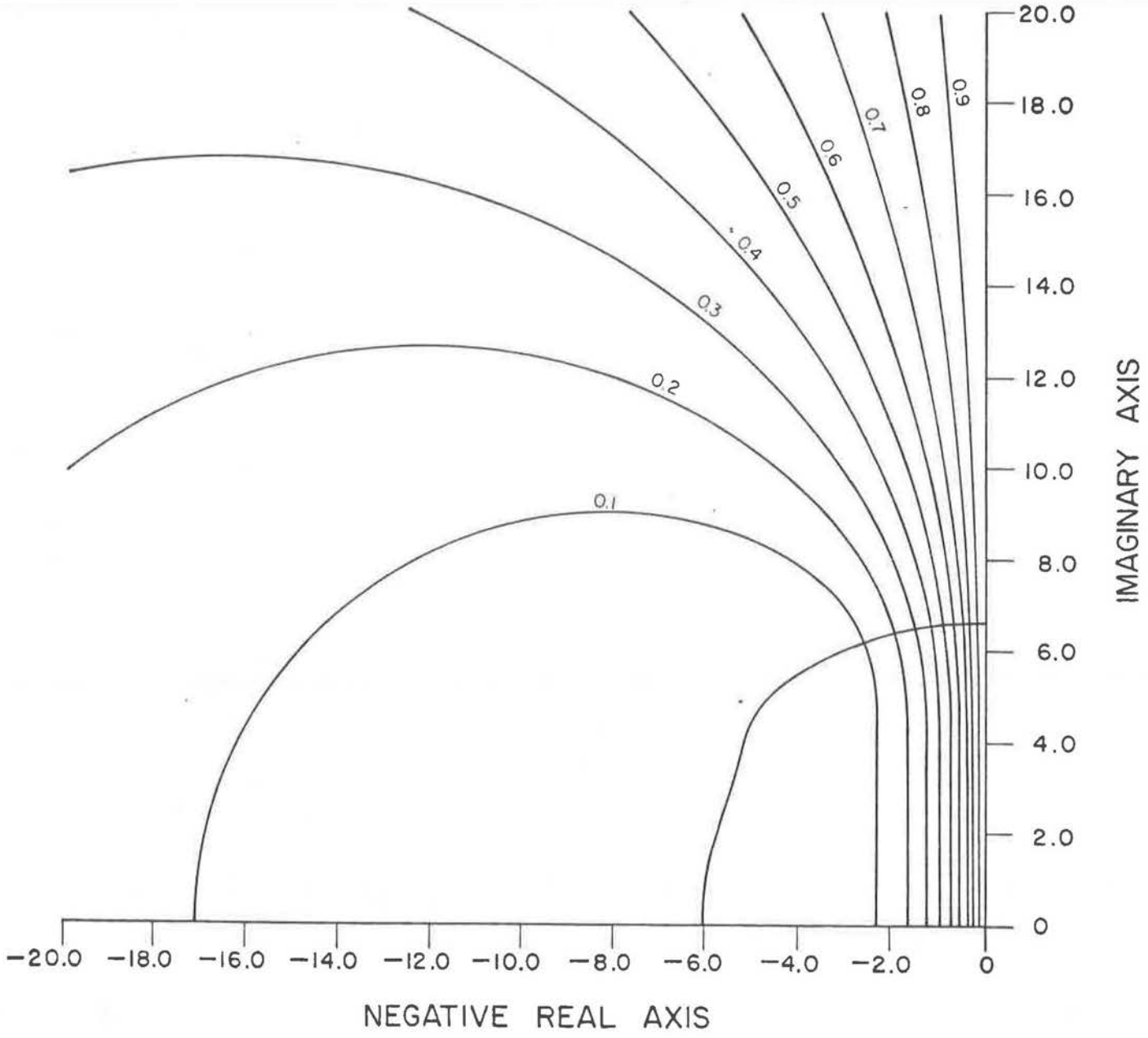

Figure 4: Contour levels of $|\gamma(\zeta)|$ and curves of minimal values: k=4.

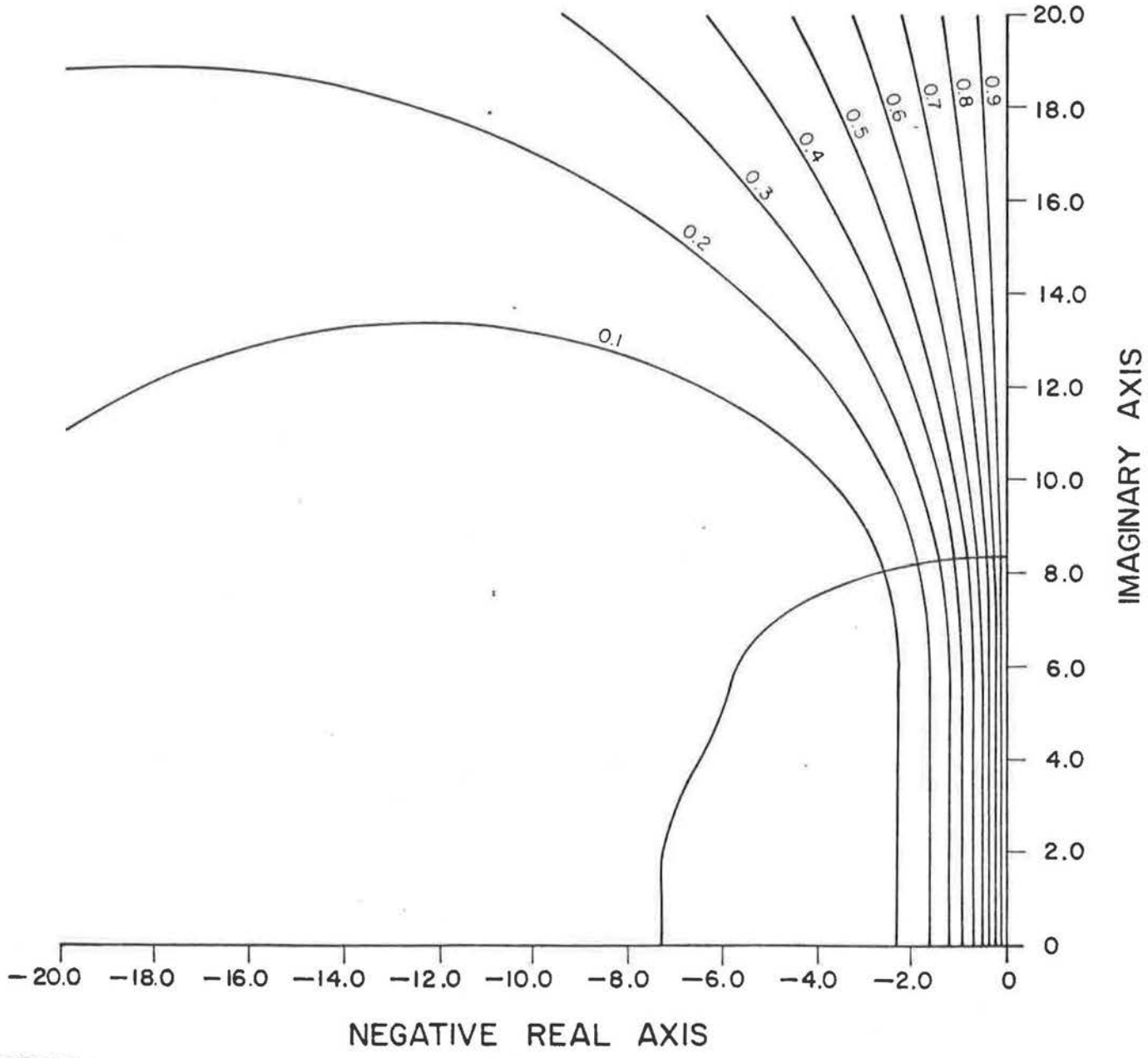

Figure 5: Contour levels of $|\gamma(\zeta)|$ and curves of minimal values: k=5.

4. A SCALAR EQUATION

In this section we consider the initial value problem

$$\varepsilon y' = \lambda y + g(t); \quad y(0) \text{ given}, \tag{4.1}$$

with λ a complex scalar of order unity, $re(\lambda) < 0$, and $g(t)$ a sufficiently smooth inhomogeneous term. First, consider the case where no transition layer is present.

4.1 The smooth solution case

Theorem 4.1. Let $y(t)$ be a smooth solution to (4.1). Denote $e_i = y_i - y(t_i)$, $1 \le i \le N$, and assume that $\frac{\varepsilon}{|\lambda|\underline{h}} < ||\hat{A}^{-1}||^{-1}$ for $\rho_1 > 0$, $\frac{\varepsilon}{|\lambda|\underline{h}} < ||\bar{A}^{-1}||^{-1}$ for $\rho_1 = 0$. Then the following error estimates hold for $1 \le i \le N$.

I. For methods of class I

$$e_i = O(h^k) \tag{4.2}$$

while if k is odd and the mesh is "locally almost uniform", i.e. either

$$h_{j+1} = h_j(1 + O(h_j)) \quad \underline{\text{or}} \quad h_{j-1} = h_j(1 + O(h_j)), \quad j < i \tag{4.3}$$

then, if $\varepsilon << |\lambda|\underline{h}$, for Gauss points

(4.4) $e_i = O(h^{k+1})$.

II. For methods of class II, if $\varepsilon << |\lambda|\underline{h}$,

(4.5) $e_i = \varepsilon O(h^k)$.

III. For methods of class III,

(4.6) $e_i = \varepsilon O(h^{k-1})$

while if k is even and the mesh is "locally almost uniform" then, if $\varepsilon << |\lambda|\underline{h}$, for Lobatto points

(4.7) $e_i = \varepsilon O(h^k)$.

Remarks

(i) The condition of "local almost uniformity" may look very restrictive at a first glance, but it is not. For instance, any mesh which is obtained by halving each subinterval of an arbitrary initial mesh satisfies (4.3).

(ii) Obviously in (4.2) - (4.7) h can be replaced by $\max_{1\le j\le i} h_j$.

Proof. Denote $y_{ij} = y^h(t_{ij})$, $y_i = y^h(t_i)$, $g_{ij} = g(t_{ij})$. The scheme (3.13) applied to (4.1) can be written for $1 \le i \le N$ as

$$(4.8)\qquad \varepsilon\frac{y_{ij}-y_i}{h_i} = \sum_{\ell=1}^{k} \hat{a}_{j\ell}(\lambda y_{i\ell} + g_{i\ell}) \qquad \begin{aligned} &1 \le j \le k \quad \text{if} \quad \rho_1 > 0, \\ &2 \le j \le k \quad \text{if} \quad \rho_1 = 0, \end{aligned}$$

$$(4.9)\qquad \varepsilon\frac{y_{i+1}-y_i}{h_i} = \sum_{\ell=1}^{k} \hat{b}_{\ell}(\lambda y_{i\ell} + g_{i\ell}).$$

For the exact solution, by quadrature,

$$(4.10)\qquad \varepsilon\frac{y(t_{ij})-y(t_i)}{h_i} = \sum_{\ell=1}^{k} \hat{a}_{j\ell}(\lambda y(t_{i\ell}) + g_{i\ell}) + \varepsilon r_{ij}; \quad r_{ij} = O(h_i^k)$$

and, for a method of class I, we also need

$$(4.11)\qquad \varepsilon\frac{y(t_{i+1})-y(t_i)}{h_i} = \sum_{\ell=1}^{k} \hat{b}_{\ell}(\lambda y(t_{i\ell}) + g_{i\ell}) + \varepsilon O(h_i^k).$$

Let $e_{ij} = y(t_{ij}) - y_{ij}$, $e_i = y(t_i) - y_i$. Then we get

$$(4.12)\qquad \varepsilon\frac{e_{ij}-e_i}{h_i} = \sum_{\ell=1}^{k} \hat{a}_{j\ell}\lambda e_{i\ell} + \varepsilon r_{ij}$$

and for a method of class I,

$$(4.13)\qquad \varepsilon\frac{e_{i+1}-e_i}{h_i} = \sum_{\ell=1}^{k} \hat{b}_{\ell}\lambda e_{i\ell} + \varepsilon O(h_i^k).$$

Assume first that $\rho_1 > 0$. Then we write (4.12) in matrix form as

$$(4.14)\qquad \underset{\sim}{e}_i = \frac{\varepsilon}{\lambda h_i}\left(\frac{\varepsilon}{\lambda h_i} I - \hat{A}\right)^{-1} \underset{\sim}{1} e_i + \frac{\varepsilon}{\lambda}\left(\frac{\varepsilon}{\lambda h_i} I - \hat{A}\right)^{-1} \underset{\sim}{r}_i$$

where $\underset{\sim}{e}_i = (e_{i1},\ldots,e_{ik})^T$, $\underset{\sim}{1} = (1,1,\ldots,1)^T$ (a vector of length k) and $\underset{\sim}{r}_i = (r_{i1},\ldots,r_{ik})^T$. Now, since $\varepsilon < \dfrac{|\lambda| h_i}{||\hat{A}^{-1}||}$ we can write

$$(4.15)\qquad (\frac{\varepsilon}{\lambda h_i} I - \hat{A})^{-1} = \frac{\varepsilon}{\lambda h_i} \hat{C}\hat{A}^{-1} - \hat{A}^{-1}$$

with $\hat{C}$ an appropriate nonsingular matrix which is bounded independently of ε.

For a method of <u>class I</u> we get by (4.13), (4.14), (4.15)

$$e_{i+1} = e_i + \frac{\lambda h_i}{\varepsilon} \underset{\sim}{\hat{b}}^T \underset{\sim}{e}_i + \varepsilon O(h_i^k) = e_i + \underset{\sim}{b}^T (\frac{\varepsilon}{\lambda h_i} I - \hat{A})^{-1} \underset{\sim}{1}\, e_i +$$

$$+ \frac{\lambda h_i}{\varepsilon} \underset{\sim}{b}^T (\frac{\varepsilon}{\lambda h_i} \hat{C}\hat{A}^{-1} - \hat{A}^{-1}) \frac{\varepsilon}{\lambda} \underset{\sim}{r}_i + \varepsilon O(h_i^k)$$

$$(4.16)\qquad e_{i+1} = \gamma(\zeta_i) e_i - h_i \underset{\sim}{b}^T \hat{A}^{-1} \underset{\sim}{r}_i + \varepsilon O(h_i^k)$$

where $\gamma(\zeta_i)$ is the growth factor

$$(4.17)\qquad \gamma(\zeta_i) = 1 - \underset{\sim}{\hat{b}}^T (\hat{A} - \zeta_i^{-1} I)^{-1} \underset{\sim}{1}; \qquad \zeta_i = \frac{\lambda h_i}{\varepsilon}\,.$$

Equation (4.16) is a difference equation with $e_1 = 0$. The solution is

$$(4.18)\qquad e_{i+1} = \sum_{j=0}^{i-1} \Big(\prod_{\ell=i-j+1}^{i} \gamma(\zeta_\ell) \Big) h_{i-j} [-\underset{\sim}{b}^T \hat{A}^{-1} \underset{\sim}{r}_{i-j} + \varepsilon O(h_{i-j}^{k-1})]\,.$$

Since $|\gamma(\zeta)| \le 1$ for $re(\zeta) \le 0$, we clearly obtain the result (4.2). Furthermore, if k is odd then as $re(\zeta) \to -\infty$, for Gauss points $(\gamma(\zeta))^j \to (-1)^j$ (cf. (3.5)) and so, provided that (4.3) holds, cancellations to a first order in h occur in

the sum (4.18), since (4.3) allows us to arrange this sum in pairs of terms which are equal in magnitude to a first order in h and alternating in sign. The result (4.4) follows.

For methods of class II we follow the same proof development above until we get to (4.14), where we note that $e_{i+1} = e_{ik}$ and so we simply consider the last element in the vector identity (4.14). Thus we do not lose the ε factor multiplying $\underset{\sim}{r}_i$! Moreover, since as $re(\zeta) \to -\infty$, $|\gamma(\zeta)| \sim |\zeta^{-1}|$, in the sum corresponding to (4.18) all terms are negligible except for the first, so (4.5) is obtained. If $\varepsilon \sim |\lambda|h$ then we only get (4.2) by this analysis.

Now assume that $\rho_1 = 0$ and consider methods of class III. We have $e_i = e_{i1}$ and so the expression (4.12) is written for $2 \le j \le k$ only and in place of (4.14) we obtain

$$(4.19) \quad \bar{\underset{\sim}{e}}_i = (\frac{\varepsilon}{\lambda h_i} I - \bar{A})^{-1}(\frac{\varepsilon}{\lambda h_i} \bar{\underset{\sim}{1}} + \bar{\underset{\sim}{a}})e_i + (\frac{\varepsilon}{\lambda h_i} \bar{C}\bar{A}^{-1} - \bar{A}^{-1})\frac{\varepsilon}{\lambda} \bar{\underset{\sim}{r}}_i$$

where $\bar{\underset{\sim}{e}}_i = (e_{i2},\ldots,e_{ik})^T$, $\bar{\underset{\sim}{1}} = (1,1,\ldots,1)^T$ (of length k-1), $\bar{A}$ and $\bar{\underset{\sim}{a}}$ are given by (3.14), $\bar{\underset{\sim}{r}}_i = (r_{i2},\ldots,r_{ik})^T$ and $\bar{C}$ is defined analogously to (4.15). Further, like in methods for class II, $\rho_k = 1$, and so $e_{ik} = e_{i+1}$. Thus we look at the last row of (4.19):

$$(4.20) \quad e_{i+1} = \gamma(\zeta_i)e_i + \frac{\varepsilon}{\lambda}[(\frac{\varepsilon}{\lambda h_i} \bar{C}\bar{A}^{-1} - \bar{A}^{-1})\bar{\underset{\sim}{r}}_i]_{k-1}.$$

The solution of this difference equation can again be expressed similarly to (4.18) and we see that, while we have not lost the ε factor just like in the class II case, we do lose a power of h in the summation just like in the class I case. The result (4.6) follows. Once again, if $\gamma(\zeta) \to -1$ and the

mesh is locally almost uniform then we gain a power of h back, obtaining (4.7). This occurs for Lobatto points if $\varepsilon << |\lambda|\underline{h}$ and if k is even (Recall that $\gamma(\zeta)$ for k Lobatto points is identical to $\gamma(\zeta)$ for k-1 Gauss points).

QED

Remarks

(i) Since we could nowhere in the proof utilize the usual superconvergence properties of the Gauss, Radau or Lobatto points (note in particular (4.16)), these results are sharp for these methods.

(ii) The results (4.2), (4.5) and (4.6) hold not only at the nodes t_i but at the collocation points t_{ij} as well. The result (4.2) holds for all methods considered in case that $\varepsilon \sim |\lambda|h$, as the usual analysis shows.

4.2 Layer accuracy

The above results (especially (4.5) - (4.7)) may, at a first glance, look too good to be true. Indeed they are in a sense: note that we have only treated one equation with a smooth solution. Unless y(0) is specified in a particular way, the solution of (4.1) usually contains a thin transition layer in which it changes fast (like $\exp\{\frac{\lambda}{\varepsilon}t\}$) from the specified y(0) to the solution of the reduced equation, $y_R(t) = -\frac{1}{\lambda}g(t)$. In this layer we have to take small steps in h if a method from class I or III is used, since such a method would not possess the property of stiff decay, and with any method if an adequate representation of the solution in the layer is desired.

It is sufficient here to consider the homogeneous problem,

(4.21) $\varepsilon y' = \lambda y, \quad y(0) = 1 \qquad (\hat{\lambda} := -re(\lambda) > 0),$

whose solution is, of course, $y(t) = \exp\{\frac{\lambda}{\varepsilon}t\}$. Recalling lemma 3.2 we write (3.17) as

(4.22) $\gamma(\frac{\lambda h}{\varepsilon}) = \exp\{\frac{\lambda h}{\varepsilon}\}[1 + c_\gamma(\frac{\lambda h}{\varepsilon})^{p+1} + O((\frac{|\lambda| h}{\varepsilon})^{p+2})],$

which is valid with $|\lambda| h \leq K\varepsilon$, K a constant, for a method of order p. Now, the solution to the difference equation

(4.23) $y_{i+1} = \gamma(\frac{\lambda h_i}{\varepsilon})y_i \ , \qquad y_1 = 1$

is

$$y_{i+1} = \prod_{j=1}^{i} \gamma(\frac{\lambda h_j}{\varepsilon}) = \prod_{j=1}^{i} \exp\{\frac{\lambda h_j}{\varepsilon}\}[1 + c_\gamma(\frac{\lambda h_j}{\varepsilon})^{p+1} + O((\frac{|\lambda| h_j}{\varepsilon})^{p+2})]$$

(4.24) $y_{i+1} \approx \exp\{\frac{\lambda t_{i+1}}{\varepsilon}\}[1 + c_\gamma \sum_{j=1}^{i} (\frac{\lambda h_j}{\varepsilon})^{p+1}]$

So the absolute error is

(4.25) $|e_{i+1}| \approx |\exp\{\frac{\lambda t_{i+1}}{\varepsilon}\} c_\gamma \sum_{j=1}^{i} (\frac{\lambda h_j}{\varepsilon})^{p+1}| \leq \exp\{\frac{-\hat{\lambda} t_{i+1}}{\varepsilon}\}|c_\gamma| \sum_{j=1}^{i} (\frac{|\lambda| h_j}{\varepsilon})^{p+1}$

Suppose that an error tolerance δ should be satisfied on $[0, T\varepsilon]$, $T \geq \frac{1}{\hat{\lambda}}$; i.e. we require

$$(4.26)\quad |e_{i+1}| \le \delta \qquad 1 \le i \le \mu, \qquad \sum_{j=1}^{\mu} h_j = T\varepsilon.$$

By (4.25),

$$(4.27)\quad |c_\gamma|(\frac{|\lambda|}{\varepsilon})^{p+1} \sum_{j=1}^{i} h_j^{p+1} \le \delta\exp\{\frac{\hat{\lambda}t_{i+1}}{\varepsilon}\}.$$

One choice of a mesh in the layer is uniform, $h_j = h_L$, $j = 1,\ldots,\mu$. Then we get

$$(4.28)\quad |c_\gamma|(\frac{|\lambda|}{\varepsilon})^{p+1} h_L^p\, t_{i+1} \le \delta\exp\{\frac{\hat{\lambda}}{\varepsilon}t_{i+1}\}.$$

Looking at the point where the curves wt and $\delta\exp\{\frac{\hat{\lambda}t}{\varepsilon}\}$ osculate we find $t = \frac{\varepsilon}{\hat{\lambda}}$, $w = \frac{\hat{\lambda}}{\varepsilon}\delta e$, and so the best choice of a uniform stepsize in the layer is

$$(4.29)\quad h_L = \frac{\varepsilon}{|\lambda|}[\frac{\hat{\lambda}}{|\lambda|}\frac{e}{|c_\gamma|}]^{1/p}\,\delta^{1/p}.$$

Thus, h_L is proportional to ε and to $\delta^{1/p}$! This points out an advantage that higher order methods may have. For instance, a Gauss method with $k = 4$ may need roughly 20 steps of size comparable to ε to maintain overall accuracy of 10^{-8} inside the layer. On $(T\varepsilon,1]$ we now have a smooth (reduced) solution and the results of theorem 4.1 apply. Note also that μ is independent of ε!

If uniform accuracy inside the layer is sought then the choice of a uniform mesh there may not be wise. Rather, an error equidistributing mesh seems more appropriate. Moreover, one cannot expect in general an accuracy

of better than $e^{-\hat{\lambda}T}$ away from the layer with a Gauss or Lobatto scheme, because this error is not damped. Thus, e.g. for accuracy of 10^{-8} one needs to take $T \approx \frac{20}{\hat{\lambda}}$, and this gives a rather large region to traverse with the uniform stepsize (4.29). A possible alternative is the variable stepsize given in the following theorem:

Theorem 4.2. The mesh defined by

$$(4.30) \quad h_i := h_{i-1} \exp\{\frac{1}{p} \frac{\hat{\lambda}}{\varepsilon} h_{i-1}\} = h_1 \exp\{\frac{1}{p} \frac{\hat{\lambda}}{\varepsilon} t_i\} \qquad i = 2,\ldots,\mu$$

with

$$(4.31) \quad h_1 := \frac{\varepsilon}{|\lambda|} \left[\frac{\hat{\lambda}}{|\lambda||c_\gamma|}\right]^{1/p} \delta^{1/p}$$

yields

$$(4.32) \quad |e_{i+1}| \lesssim \delta \qquad i = 1,\ldots,\mu$$

provided that $|\lambda|h \leq k\varepsilon$, k a constant.†

The proof involves a straightforward induction and is omitted. Note that (4.31) is essentially the same as the uniform layer step size (4.29) and that $h_i > h_{i-1}$, $i > 1$. A numerical example illustrating the difference between the two strategies (4.29) and (4.30), (4.31) is given in §4.5. With both strategies, μ is independent of ε and is a monotonically decreasing function of p.

---

† We are indebuted to Dr. P. Markowich for helping us in refining this result.

Remark

For use in the subsequent analysis the results of theorem 4.1 and of this section must be combined appropriately, since a mesh with stepsizes proportional to ε is used on $[0,T\varepsilon]$ and a coarse mesh is used thereafter. In fact, if A_{11} of (1.1) has eigenvalues with positive and negative real parts then there are three intervals which need to be considered separately, since meshes with stepsizes proportional to ε must be used on $[0,T^{(0)}\varepsilon]$ and $[1 - T^{(1)}\varepsilon,1]$, with $T^{(0)}$, $T^{(1)}$ independent of ε, while a coarse mesh is used on $[T^{(0)}\varepsilon,1-T^{(1)}\varepsilon]$. But again, it is straightforward to combine the results appropriately

4.3 Layer-damping mesh

Our purpose here is to choose a mesh in the transition layer in order to get rid, as soon as possible, of its effects on the solution away from the layer, using a method of class I or III. We are not interested in an accurate solution in the layer; rather, we want to climb on the smooth solution curve with as few steps as possible.

Here we need step sizes to damp the layer errors as much as possible. But these are provided in figures 1 to 5! In particular, for a given k-Gauss or (k+1)-Lobatto method and a given $\lambda = \alpha + i\beta$, let $r = \beta/\alpha$ and pick $\zeta = \xi+ir\xi$ from the appropriate minimum value curve in figure k. Then the step size which would damp perturbations most effectively is

$$h_D = \varepsilon\, \xi/\alpha \tag{4.33}$$

(note that α and ξ are both negative). If λ is real and k is odd (k even for Lobatto) then one step would do it, since $\gamma_r := \min|\gamma(\zeta)| = 0$! (cf. table 3.1). In general this is not the case, but N_L repeated applications with the same step h_D of (4.33) produce a damping factor of $\gamma_r^{N_L}$ and so N_L is easily determined to meet a prescribed tolerance. Note again the advantage of higher order methods for eigenvalues with significant imaginary parts.

4.4 Conditioning and scaling

In both sections 4.2 and 4.3 we have worked with very nonuniform meshes: in the layer $h_L = O(\varepsilon)$ and outside $h_j >> \varepsilon$. This introduces the question of conditioning of the resulting matrix problem, in view of the BVPs that we intend to solve. Here we show that with an appropriate scaling, no problem is caused by such highly nonuniform meshes.

Rewriting (4.1) as

$$(4.34) \quad Ly \equiv \varepsilon y' - \lambda y = g(t); \qquad y(0) \text{ given},$$

we note that the differential problem (4.34) is well scaled, and so the discretized form (4.8), (4.9) is considered, since it preserves this scaling. We write (4.8), (4.9) together with $y_1 = y(0)$ as

$$(4.35) \quad L^h \underset{\sim}{y}^h = \underset{\sim}{g}^h.$$

Now, if in the layer we have $h_L = \varepsilon/c_L$ (for simplicity suppose that h_L is uniform in the layer) and outside the layer we have stepsizes h_j, $\varepsilon << \underline{h} \le h_j \le h$, then obviously

(4.36) $\quad ||L^h|| \leq \max\{c_1 c_L,\ c_2||\hat{A}||\}$

c_1, c_2 constants. Thus we are left to bound $(L^h)^{-1}$. For this we look at the solution of (4.35) for an arbitrary $\underset{\sim}{g}^h$.

Proceeding as in the proof of theorem 4.1 we obtain, for $\rho_1 > 0$,

$$(4.37) \quad y_{i+1} = \gamma(\frac{\lambda h_i}{\varepsilon})y_i + \frac{h_i}{\varepsilon}\underset{\sim}{b}^T[I + (\frac{\varepsilon}{\lambda h_i} I - \hat{A})^{-1}\hat{A}]\underset{\sim}{g}_i$$

with $\underset{\sim}{g}_i$ an appropriate piece of $\underset{\sim}{g}^h$. Consider the layer region $[0,T\varepsilon]$, $T\varepsilon = N_L h_L$. From (4.37) we get

$$|y_{i+1}| \leq |y_i| + \frac{T}{N_L} c_3||\underset{\sim}{g}_i||$$

c_3 a constant. Thus

$$(4.38) \quad |y_i| \leq c_3 T||\underset{\sim}{g}^h|| \qquad\qquad 1 \leq i \leq N_L + 1.$$

Now, for the outer region $(T\varepsilon,1]$, use (4.15) to obtain

$$(4.39) \quad y_{i+1} = \gamma(\frac{\lambda h_i}{\varepsilon})y_i + \frac{1}{\lambda}\underset{\sim}{\hat{b}}^T\hat{C}\underset{\sim}{g}_i \qquad\qquad i \geq N_L + 1.$$

Thus, with $\zeta_j = \frac{\lambda h_j}{\varepsilon}$,

$$(4.40) \quad y_{i+1} = (\prod_{\ell=N_L+1}^{i}\gamma(\zeta_\ell))y_{N_L+1} + \frac{1}{\lambda}\sum_{j=0}^{i-N_L-1}(\prod_{\ell=i-j+1}^{i}\gamma(\zeta_\ell))\underset{\sim}{b}^T\hat{C}\underset{\sim}{g}_{i-j}.$$

So, again using the fact that $|\gamma(\zeta)| \leq 1$, we get

$$(4.41) \quad |y_{i+1}| \leq c_3 T ||\underset{\sim}{g}^h|| + (i-N_L)c_4 ||\underset{\sim}{g}^h||$$

c_4 a constant. With N being the total number of subintervals we finally get that

$$(4.42) \quad ||(L^h)^{-1}|| \leq c_3 T + c_4(N - N_L).$$

We sum all this up in a theorem.

Theorem 4.3. The maximum norm condition number of the matrix L^h can be bounded by

$$(4.43) \quad \chi(L^h) \leq [c_3 T + c_4(N-N_L)]\max\{c_1 c_L, c_2||\hat{A}||\}.$$

Thus, if the fine mesh is over the layer only ($\underline{h} >> \varepsilon$) and the number of mesh points in the layer is not large (hence c_L and T are not large) then the condition number of L^h is bounded independently of local or global mesh ratios and is linearly proportional to the number of subintervals N.

4.5 Example

To illustrate some of the above results we have made several runs with the simple example

(4.44) $\varepsilon y' = -(y-e^t) + \varepsilon e^t$

whose reduced solution is, of course, $y_R(t) = e^t$.

First, choose $y(0) = 1$, obtaining the smooth solution $y(t) = e^t$. Results for $\varepsilon = 10^{-4}$ are contained in table 4.1. Uniform meshes of N subintervals are used and the error $E = |y(1) - y^h(1)|$ is listed, along with the measured rates of convergence with respect to h. The notation a. - b $\equiv$ a x 10^{-b} is used throughout the paper.

These results confirm theorem 4.1 with respect to h. Results were obtained also for $\varepsilon = 10^{-8}$, which confirm the linear dependence on ε in (4.5), (4.6) and (4.7).

Next, we take $y(0) = 0$, obtaining a transition layer: $y(t) = e^t - e^{-t/\varepsilon} + O(\varepsilon)$. When repeating the experiments of table 4.1, good results are obtained, as expected, only for Radau points. For the Gauss and Lobatto points, the layer errors are not damped and should be dealt with.

In the following treatments of the layer we restrict ourselves to the Lobatto points because, once the layer effect is damped, accurate results are obtained at $t = 1$ for a very small ε, without having to worry about h (cf. (4.6) vs. (4.2)!). First, consider the meshes obtained from (4.30), (4.31) for $\varepsilon = 1. - 8$, $\delta = 1. - 8$, $T = 20$. In table 4.2 we list, under "N" the mesh size $\mu+1$ obtained by repeatedly applying (4.30) until $t_{i+1} \geq T\varepsilon$; under "NU" the mesh size obtained by using a uniform step of width h_1 from (4.31) inside the layer; under "EL" the maximum error in $[0,T\varepsilon]$; and under "E" as before the error at $t = 1$. The meshes constructed by (4.30) do not have additional points in $(T\varepsilon,1)$.

Table 4.1: $\varepsilon(y'-e^t) = -(y-e^t)$; $y = e^t$, $\varepsilon = 10^{-4} \equiv 1.-4$

Method	k	N	E	rate
Gauss	1	2	.52-1	
		4	.13-1	2.0
		8	.34-2	2.0
		16	.89-3	2.0
	2	2	.12-1	
		4	.30-2	2.0
		8	.72-3	2.0
		16	.16-3	2.1
	3	2	.11-3	
		4	.71-5	3.9
		8	.47-6	3.9
		16	.35-7	3.7
	4	2	.13-4	
		4	.78-6	4.0
		8	.45-7	4.1
		16	.21-8	4.4
	5	2	.73-7	
		4	.12-8	5.9
		8	.21-10	5.8
		16	.41-12	5.7
Lobatto	2	2	.35-5	
		4	.89-6	2.0
		8	.23-6	2.0
		16	.59-7	1.9
	3	2	.89-6	
		4	.22-6	2.0
		8	.54-7	2.0
		16	.12-7	2.1

Method	k	N	E	rate
Radau	1	2	.58-4	
		4	.31-4	0.9
		8	.16-4	0.9
		16	.83-5	1.0
	2	2	.62-5	
		4	.17-5	1.9
		8	.45-6	1.9
		16	.11-6	2.0
	3	2	.34-6	
		4	.47-7	2.8
		8	.63-8	2.9
		16	.80-9	3.0
	4	2	.13-7	
		4	.90-9	3.8
		8	.59-10	3.9
		16	.37-11	4.0
	5	2	.37-9	
		4	.13-10	4.8
		8	.42-12	4.9
		16	.13-13	5.1
Lobatto	4	2	.87-8	
		4	.57-9	3.9
		8	.38-10	3.9
		16	.28-11	3.8
	5	2	.10-8	
		4	.65-10	4.0
		8	.37-11	4.1
		16	.17-12	4.5

Table 4.2: $\varepsilon(y'-e^t) = -(y-e^t)$; $y = e^t - e^{-t/\varepsilon} + O(\varepsilon)$; $\varepsilon = 1. - 8$; Lobatto points

k	N	NU	EL	E
2	5780	57737	.99-8	.33-8
3	80	387	.93-8	.41-8
4	21	64	.76-8	.27-9
5	11	24	.56-8	.10-8

Note that δ is a fairly tight bound on EL; also note the superiority of higher order methods for this δ and the usefulness of the layer mesh (4.30), (4.31) compared to the uniform layer mesh.

Next we note that for this trivial example we can get $\gamma(\zeta) = 0$ with the trapezoidal rule (k = 2, Lobatto) by taking a step $h_1 = 2\varepsilon$, according to (4.33). Indeed, with the mesh $\{0, 2\varepsilon, 1\}$ we get E = .35-9 ! In the next section we give another example, more realistic if less striking, of a layer-damping mesh.

5. INITIAL VALUE SYSTEMS

Here we consider the system

$$\underset{\sim}{\varepsilon y}' = A_{11}\underset{\sim}{y} + A_{12}\underset{\sim}{z} + \underset{\sim}{f}_1 \tag{5.1}$$

$$\underset{\sim}{z}' = A_{21}\underset{\sim}{y} + A_{22}\underset{\sim}{z} + \underset{\sim}{f}_2 \tag{5.2}$$

with $\underset{\sim}{y}(0)$, $\underset{\sim}{z}(0)$ given. As in the previous section, we consider the smooth case first. But before we proceed it should be noted that if there are no slow components $\underset{\sim}{z}$ then the results of the previous section are generalized directly for (5.1). It is the presence of different time scales which makes the difference here.

5.1 The smooth solution case

In view of previous considerations, we can assume that $A_{21} = 0$ and that A_{11} is already in Jordan canonical form, see §2.2. Now in (5.2) no significant dependence on ε is left and the usual theory applies. We have (see, eg. Russell [18], Weiss [22]).

Theorem 5.1. For the slow components $\underset{\sim}{z}$ the following results hold: with Gauss collocation points, at the mesh points

$$|\underset{\sim}{z}(t_i) - \underset{\sim}{z}^h(t_i)| = O(h^{2k}) \tag{5.3}$$

while at any other point t, $t_i \leq t \leq t_{i+1}$,

$$(5.4) \qquad |\underset{\sim}{z}(t) - \underset{\sim}{z}^h(t)| = O(h_i^{k+1}) + O(h^{2k}).$$

With Radau collocation points(5.3), (5.4) hold with 2k-1 replacing 2k. With Lobatto collocation points (5.3), (5.4) hold with 2k-2 replacing 2k.

Now, the collocation equations for (5.1) can be written as

$$(5.5) \qquad \varepsilon(\underset{\sim}{y}^h)'(t_{ij}) = A_{11}\underset{\sim}{y}^h(t_{ij}) + [A_{12}\underset{\sim}{z}(t_{ij}) + \underset{\sim}{f}_1(t_{ij})] + [A_{12}(\underset{\sim}{z}^h(t_{ij}) - \underset{\sim}{z}(t_{ij}))].$$

Recall from §2.2 that we can consider the equations in (5.5) one at a time. Moreover, due to the linearity we can consider the two inhomogeneous contributions in (5.5) separately. The first one, a component of $A_{12}\underset{\sim}{z}(t_{ij}) + \underset{\sim}{f}_1(t_{ij})$ under the appropriate initial conditions, is readily seen to conform to the conditions of theorem 4.1 . Hence the relevant result out of (4.2) - (4.7) applies. That leaves us to deal with the other contribution, $A_{12}[\underset{\sim}{z}^h(t) - \underset{\sim}{z}(t)]$, under zero initial conditions. Thus we are looking at the equation (4.1) with an inhomogeneous term which satisfies $|g(t)| = O(h_i^{k+1})$, $t_i \leq t \leq t_{i+1}$, and $|g(t_i)| = O(h^{2k})$, $O(h^{2k-1})$ or $O(h^{2k-2})$ for Gauss, Radau or Lobatto points, respectively, and with $y(0) = 0$.

Following the tracks of the proof of theorem 4.1 once more, we obtain for our special g(t), in case that $\rho_1 > 0$,

$$(5.6) \qquad \underset{\sim}{y}_i = (\frac{\varepsilon}{\lambda h_i} I - \hat{A})^{-1} \frac{\varepsilon}{\lambda h_i} \underset{\sim}{1}\, y_i + \frac{1}{\lambda}(\frac{\varepsilon}{\lambda h_i} I - \hat{A})^{-1}\hat{A}\, \underset{\sim}{g}_i$$

with an obvious notation. Assuming $\varepsilon < \dfrac{|\lambda| h_i}{||\hat{A}^{-1}||}$ all i, we get for Gauss points, by (4.15), (4.17),

$$(5.7)\qquad y_{i+1} = \gamma(\zeta_i)y_i + \frac{1}{\lambda}\hat{\underset{\sim}{b}}^T\hat{C}\underset{\sim}{g}_i \qquad\qquad i \geq 1,\ \zeta_i = \frac{\lambda h_i}{\varepsilon}\ .$$

Hence

$$(5.8)\qquad y_{i+1} = \frac{1}{\lambda}\sum_{j=0}^{i-1}\Big(\prod_{\ell=i-j+1}^{i}\gamma(\zeta_\ell)\Big)\hat{\underset{\sim}{b}}^T\hat{C}\underset{\sim}{g}_{i-j}.$$

From here, since $\underset{\sim}{g}_{i-j} = O(h_{i-j}^{k+1})$, we get

$$(5.9)\qquad |\underset{\sim}{y}(t_i) - \underset{\sim}{y}^h(t_i)| = O(h^k).$$

Furthermore, if k is odd and (4.3) holds then, if $\varepsilon << |\lambda|h$, $\underset{\sim}{g}_{i+1-j} - \underset{\sim}{g}_{i-j} = O(h_{i-j}^{k+2})$. Since (4.4) also holds for the other inhomogeneous contribution we have

$$(5.10)\qquad |\underset{\sim}{y}(t_i) - \underset{\sim}{y}^h(t_i)| = O(h^{k+1}).$$

Now, for Radau points we look at the last row of (5.6). Using (4.15) we have

$$(5.11)\qquad y_{i+1} = \gamma(\zeta_i)y_i - \frac{1}{\lambda}\,g(t_{i+1}) + \frac{\varepsilon}{\lambda h_i}O(h_i^{k+1}).$$

The solution to this difference equation looks similar to (5.8) and, as before, when $\varepsilon << |\lambda|h$ all terms in the sum but for the first are negligible. Combining with (4.5) we obtain

$$(5.12)\qquad |\underset{\sim}{y}(t_i) - \underset{\sim}{y}^h(t_i)| = O(h^{2k-1}) + \varepsilon O(h^k).$$

Finally, for the Lobatto points we have

$$(5.13)\quad \underset{\sim}{\bar{y}}_i = (\frac{\varepsilon}{\lambda h_i} I-\bar{A})^{-1}(\frac{\varepsilon}{\lambda h_i} \bar{1}+\underset{\sim}{\bar{a}})y_i + \frac{1}{\lambda}(\frac{\varepsilon}{\lambda h_i} \bar{C}\bar{A}^{-1}-\bar{A}^{-1})[\bar{A}\underset{\sim}{\bar{g}}_i+\underset{\sim}{\bar{a}}g(t_i)]$$

with notation corresponding to that of (4.19). Now consider the last row of (5.13). Since the last component of $\bar{A}^{-1}\underset{\sim}{\bar{a}}$ is $(-1)^k$ (see the expression for $\gamma(\zeta_i)$ in (5.13) and recall (3.5)),

$$(5.14)\quad y_{i+1} = \gamma(\zeta_i)y_i - \frac{1}{\lambda}(g(t_{i+1}) + (-1)^k g(t_i)) + \frac{1}{\lambda}\frac{\varepsilon}{\lambda h_i} O(h_i^{k+1})$$

So

$$(5.15)\quad y_{i+1} = -\frac{1}{\lambda}\sum_{j=0}^{i-1}\Big(\prod_{\ell=i-j+1}^{i}\gamma(\zeta_\ell)\Big)[g(t_{i-j+1}) + (-1)^k g(t_{i-j}) - \frac{\varepsilon}{\lambda h_{i-j}} O(h_{i-j}^{k+1})]$$

Now, $g(t_{i-j+1})$ and $g(t_{i-j})$ are $O(h^{2k-2})$. In the summation, we lose a power of h, but we gain it back as follows: For k odd, in which case $\gamma(\zeta) \to 1$ as $re(\zeta) \to -\infty$, we get cancellation in the limit sum $\sum_{j=0}^{i-1}(g(t_{i-j+1}) - g(t_{i-j}))$. For k even, in which case $\prod_{\ell=i-j+1}^{i}\gamma(\zeta_\ell) \to (-1)^j$ as $re(\zeta) \to -\infty$, we get cancellation in the limit sum $\sum_{j=0}^{i-1}(-1)^j(g(t_{i-j+1}) + g(t_{i-j}))$. In the latter case, if (4.3) holds, we also get cancellation to a first power in h in the last term of (5.15). Combining this with (4.7), we conclude the proof to the following theorem.

Theorem 5.2. Let $(\underset{\sim}{y}(t),\underset{\sim}{z}(t))$ be a smooth solution to the initial value problem for (5.1), (5.2) and assume that, for $\rho_1 > 0$, $\varepsilon < \frac{\underline{\lambda}\,\underline{h}}{||\hat{A}^{-1}||}$, and for $\rho_1 = 0$, $\varepsilon < \frac{\underline{\lambda}\,\underline{h}}{||\bar{A}^{-1}||}$, where $\underline{\lambda} = \min\{|\lambda|;\ \lambda \text{ is an eigenvalue of } A_{11}\}$. Then

I. For Gauss points (5.9) holds and, if k is odd and (4.3) holds then if $\varepsilon << \underline{\lambda}\underline{h}$, (5.10) holds.

II. For Radau points, if $\varepsilon << \underline{\lambda}\underline{h}$ then (5.12) holds.

III. For Lobatto points,

$$(5.16) \quad |\underset{\sim}{y}(t_i) - \underset{\sim}{y}^h(t_i)| = O(h^{2k-2}) + \varepsilon O(h^{k-1})$$

while, if k is even and (4.3) holds then, if $\varepsilon << \underline{\lambda}\underline{h}$,

$$(5.17) \quad |\underset{\sim}{y}(t_i) - \underset{\sim}{y}^h(t_i)| = O(h^{2k-2}) + \varepsilon O(h^{k}).$$

Thus, for $k \geq 4$ and ε very small compared to h, Lobatto points have a higher order of superconvergence than Gauss points for the same k !

5.2 Matrix condition and initial layers

Unless the initial conditions are very special, a transition layer appears in the solution which connects the initial values to the smooth solution curves. Handling these layers is done as in the previous section: for methods of class II nothing needs to be done if only the accuracy away from the layer is of concern. For the other methods, however, a number of mesh points with stepsize $h_L = O(\varepsilon)$ need to be placed to ensure at least that the layer errors are well damped. For the special choices of such stepsizes, we need to find the eigenvalues of A_{11} (and this time not just for analysis). Then for (4.29) or (4.30), (4.31), the eigenvalue with largest magnitude can be used, while for a layer-damping mesh, a sequence of stepsizes obtained by (4.33) for each eigenvalue in turn is used, repeatedly if necessary.

Next, we need to extend theorem 4.3 for the system case. The problem here is in the components $\underset{\sim}{z}$ which are approximated with stepsizes of order ε in the layer (of $\underset{\sim}{y}$) while not being multiplied by ε, which balances L^h for $\underset{\sim}{y}$ in (4.8), (4.9). We delay the treatment of this problem to the next section, where it is dealt with in the more general context of BVPs.

5.3 Example

Consider the problem

$$
\begin{aligned}
\varepsilon y_1' &= -2y_1 - 2y_2 + 4z + f_1(t;\varepsilon) \\
\varepsilon y_2' &= 2y_1 - 2y_2 + f_2(t;\varepsilon) \\
\varepsilon y_3' &= 2y_1 - y_2 - y_3 + z + f_3(t;\varepsilon) \\
z' &= y_1 - y_3 + f_4(t;\varepsilon)
\end{aligned} \tag{5.18}
$$

with the $f_i(t)$ defined so that the smooth solution is given by

$$
y_1(t) = \sin t, \quad y_2(t) = e^t, \quad y_3(t) = 2e^{-t}, \quad z(t) = \frac{1}{t+1}, \tag{5.19}
$$

for all $\varepsilon > 0$.

Some sample runs for the smooth curve with uniform meshes are given in table 5.1. We take $\varepsilon = 10^{-8}$ in order that the dependence of the accuracy on h be clearly demonstrated. For each k and N we list under "E" and "rate" the four errors at t = 1 and their rates of convergence with respect to h for the four solution components. The results of theorems 5.1 and 5.2 are confirmed.

Table 5.1: Problem (5.18); $y_1 = \sin t$, $y_2 = e^t$, $y_3 = 2e^{-t}$, $y_4 = \frac{1}{t+1}$; $\varepsilon = 1.-8$

Gauss points				Radau points				Lobatto points			
k	N	E	rate	k	N	E	rate	k	N	E	rate
2	4	.15-2		2	4	.50-4		3	4	.40-5	
		.60-2				.50-4				.40-5	
		.38-2				.10-3				.80-5	
		.53-5				.50-4				.40-5	
	8	.39-3	2.0		8	.68-5	2.9		8	.24-6	4.1
		.15-2	2.0			.68-5	2.9			.24-6	4.1
		.96-3	2.0			.14-4	2.9			.48-6	4.1
		.34-6	4.0			.68-5	2.9			.24-6	4.1
3	4	.83-4		3	4	.36-6		4	4	.25-7	
		.86-4				.36-6				.25-7	
		.15-3				.71-6				.51-7	
		.76-8				.36-6				.25-7	
	8	.58-5	3.8		8	.11-7	5.0		8	.42-9	5.9
		.61-5	3.8			.11-7	5.0			.42-9	5.9
		.11-4	3.8			.23-7	5.0			.83-9	5.9
		.12-9	6.0			.11-7	5.0			.42-9	5.9

Next, we consider the effect of layers by running the problem (5.18) with the initial values $\underset{\sim}{y}(0) = (1,2,3)^T$, $z(0) = 1$, with $\varepsilon = 10^{-8}$. Three types of meshes are used:

$\mu 1$ = a uniform mesh of 8 subintervals,

$\mu 2$ = $\mu 1$ plus the mesh generated by (4.30), (4.31) with $\delta = 1.-6$, $T = 15$, $\hat{\lambda} = 2.$

$\mu 3$ = $\mu 1$ plus the 3 mesh points 1.8ε, 3.6ε, 8.24ε. This is the layer-damping mesh from fig. 3 ($k = 4$ for Lobatto).

Results are accumulated in table 5.2. Here, under "method" we list the type of collocation scheme and the type of mesh used. For the error at $t = 1$, the smooth solution was used as the exact one; thus the smallest errors listed under "E" are polluted. The results with the mesh $\mu 1$ are as expected: The Radau schemes perform rather well whereas the other ones do not, because they do not damp the layer errors (of magnitude 1).

The experiments for the meshes of types $\mu 2$ and $\mu 3$ are recorded for the Lobatto points only; the Gauss points give qualitatively the same results. With the meshes of type $\mu 2$ the errors at $t = 1$ for $k = 3$ and especially for $k = 2$ are largely due to the approximation of the smooth curve. Note the accuracy obtained witn the mesh $\mu 3$ and $k = 4$. It is better than that for $k = 5$ since the mesh is tailored for $k = 4$.

Table 5.2: Problem (5.18), $\underset{\sim}{y}(0) = (1,2,3)^T$; $\varepsilon = 1.-8$

Method	k	N	E	Method	k	N	E
Gauss,μ1	3	8	1.0	Lobatto,μ2	2	590	.12-2
			1.0				.12-2
			1.0				.23-2
			.28-8				.12-2
Radau,μ1	1	8	.23-1	Lobatto,μ2	3	35	.70-6
			.23-1				.29-6
			.47-1				.53-6
			.23-1				.24-6
Radau,μ1	2	8	.68-5	Lobatto,μ2	4	19	.42-8
			.68-5				.35-7
			.14-4				.33-7
			.68-5				.23-8
Radau,μ1	3	8	.10-7	Lobatto,μ2	5	15	.75-7
			.10-7				.14-7
			.21-7				.16-7
			.10-7				.18-8
Radau,μ1	4	8	.23-8	Lobatto,μ3	3	11	.24-1
			.23-8				.59-2
			.45-8				.59-2
			.23-8				.24-6
Radau,μ1	5	8	.93-9	Lobatto,μ3	4	11	.87-5
			.93-9				.85-6
			.19-9				.85-6
			.93-9				.23-8
Lobatto,μ1	4	8	1.0	Lobatto,μ3	5	11	.98-4
			1.0				.64-4
			1.0				.64-4
			.27-8				.18-8

6. BOUNDARY VALUE PROBLEMS

In this section we finally consider the BVP (1.1) - (1.3),

$$\varepsilon \underset{\sim}{y}' = A_{11}\underset{\sim}{y} + A_{12}\underset{\sim}{z} + \underset{\sim}{f}_1 \tag{6.1}$$

$$\underset{\sim}{z}' = A_{21}\underset{\sim}{y} + A_{22}\underset{\sim}{z} + \underset{\sim}{f}_2 \tag{6.2}$$

$$B_0\begin{pmatrix}\underset{\sim}{y}\\ \underset{\sim}{z}\end{pmatrix}(0) + B_1\begin{pmatrix}\underset{\sim}{y}\\ \underset{\sim}{z}\end{pmatrix}(1) = \underset{\sim}{\beta} \tag{6.3}$$

assuming that the differential operator has a bounded inverse under (6.3), independent of ε. We also assume, as previously, that all the eigenvalues of A_{11} have nonzero real parts. Some of these real parts may be positive and some negative; so we can expect boundary layers at both ends of the interval and a smooth solution in between.

6.1 Convergence and layers

Consider the following two operators: splitting the problem (6.1) - (6.3) as described in §2.3 and finding an approximate solution by one of the collocation methods considered here. Clearly, the two commute; in fact, we can write the general solution $(\underset{\sim}{w}^h, \underset{\sim}{z}^h)$ of the collocation scheme applied to (6.1) - (6.2) as

$$\begin{pmatrix}\underset{\sim}{w}^h\\ \underset{\sim}{z}^h\end{pmatrix} = W^h_{I}\underset{\sim}{\eta}^h_{I} + W^h_{II}\underset{\sim}{\eta}^h_{II} + W^h_{III}\underset{\sim}{\eta}^h_{III} + \begin{pmatrix}\underset{\sim}{w}^h_p\\ \underset{\sim}{z}^h_p\end{pmatrix}; \qquad \underset{\sim}{w}^h = E^{-1}\underset{\sim}{y}^h. \tag{6.4}$$

where the superscript h denotes the collocation approximation to the corresponding quantity in (2.20), (2.21). The coefficients $\underset{\sim}{\eta}^h_I, \underset{\sim}{\eta}^h_{II}, \underset{\sim}{\eta}^h_{III}$ of the linear combination in (6.4) are obtained upon substituting this into the boundary conditions (6.3), yielding

$$(6.5) \qquad A^h(\varepsilon)\underset{\sim}{\eta}^h = \underset{\sim}{\hat{\beta}}^h; \qquad \underset{\sim}{\eta}^h = (\underset{\sim}{\eta}^h_I, \underset{\sim}{\eta}^h_{II}, \underset{\sim}{\eta}^h_{III})^T.$$

Now W^h_I and $(\underset{\sim}{w}^h_p, \underset{\sim}{z}^h_p)$ are approximations to smooth solutions, and theorems 5.1 and 5.2 apply to them as long as the numerical method is A-stable in both directions of integration. This is the case for methods I and III, since these collocation points are symmetric about 1/2. For W_{II} and W_{III} we have to deal additionally with stable initial value problems, one in t and the other in 1 - t, with transition layers at 0 and at 1, respectively. If sufficiently fine meshes are used in the layers then, by the results of §4.2, W_{II} and W_{III} can be approximated as accurately as desired. Hence $A^h(\varepsilon)$ approximates $A(\varepsilon)$ and $\underset{\sim}{\hat{\beta}}^h$ approximates $\underset{\sim}{\hat{\beta}}$ of (2.22), when h and δ (the boundary layer accuracy) are small. It follows that $(A^h(\varepsilon))^{-1}$ is uniformly bounded for h and δ sufficiently small, whence

$$\underset{\sim}{\eta}^h - \underset{\sim}{\eta} = (A^h(\varepsilon))^{-1}[(A(\varepsilon) - A^h(\varepsilon))\underset{\sim}{\eta} + (\underset{\sim}{\hat{\beta}}^h - \underset{\sim}{\hat{\beta}})]$$

Thus we have established the following convergence result:

Theorem 6.1. For collocation methods of classes I and III, including Gauss and Lobatto schemes but not Radau schemes, the results of §5 hold for the BVP as well, with the layer treatment applied at the interval ends for J_- and J_+, respectively.

Here we see the advantage of the symmetric schemes. With methods of class II, collocation in the "wrong" direction, i.e. using the points $\sigma_j = 1 - \rho_j$ for a very stiff IVP, is disastrously unstable. Thus, to use these methods the transformation (2.10), (2.11) has to be carried out explicitly (at each mesh point for the variable coefficient case). Then an appropriately upwinded collocation scheme of class II can be applied to the transformed system, as was suggested in Kreiss-Nichols [13], Ringhofer [17].

For the symmetric schemes we may want to find the eigenvalues of A_{11} (even in the variable coefficient case) just at $t = 0$ and at $t = 1$ in order to define layer-damping meshes at the two ends. The higher order Lobatto schemes are particularly recommended.

6.2 Conditioning and scaling

Here we derive a bound for the maximum norm condition number of a row scaled version of the collocation equations for (6.1) - (6.3). For brevity we only consider Lobatto points, but all the results are easily extended to any A-stable, symmetric scheme.

We present the analysis for a mesh with uniform steps of size $O(\varepsilon)$ in the layers and a coarse mesh in between. Specifically, we use:

on $[0,T^{(0)}\varepsilon]$: $h_i = h^{(0)} = \varepsilon/c_L^{(0)}$, $i = 1,\ldots,N^{(0)}$; $N^{(0)} = T^{(0)}\, c_L^{(0)}$,

on $[T^{(0)}\varepsilon, 1-T^{(1)}\varepsilon]$: $h_i > \varepsilon||\bar{A}^{-1}||/\underline{\lambda}$, $i = N^{(0)}+1,\ldots,N-N^{(1)}$,

on $[1-T^{(1)}\varepsilon, 1]$: $h_i = h^{(1)} = \varepsilon/c_L^{(1)}$, $i = N-N^{(1)}+1,\ldots,N$; $N^{(1)} = T^{(1)}\, c_L^{(1)}$,

see fig. 6.

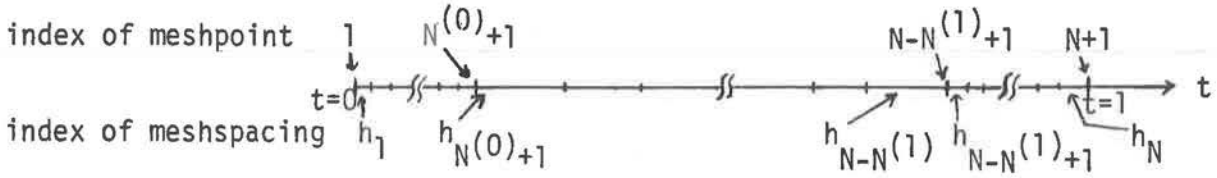

Figure 6: The mesh

We assume, as in §4.4, that $\varepsilon << \underline{h} \le h_i \le h$, $N^{(0)} + 1 \le i \le N - N^{(1)}$ and let $N_c = N - N^{(1)} - N^{(0)}$.

The essentially row equilibrated form of the collocation system we consider is

$$(6.6)\qquad \varepsilon \frac{\underset{\sim}{y}_{ij} - \underset{\sim}{y}_i}{h_i} - \sum_{\ell=1}^{k} \hat{a}_{j\ell}(A_{11}\underset{\sim}{y}_{i\ell} + A_{12}\underset{\sim}{z}_{i\ell}) = \underset{\sim}{p}_{ij} \qquad j = 2,\ldots,k;\quad i = 1,\ldots,N$$

$$(6.7)\qquad \underset{\sim}{z}_{ij} - \underset{\sim}{z}_i - h_i \sum_{\ell=1}^{k} \hat{a}_{j\ell}(A_{21}\underset{\sim}{y}_{i\ell} + A_{22}\underset{\sim}{z}_{i\ell}) = \underset{\sim}{q}_{ij}$$

$$(6.8)\qquad B_0 \begin{pmatrix} \underset{\sim}{y}_1 \\ \underset{\sim}{z}_1 \end{pmatrix} + B_1 \begin{pmatrix} \underset{\sim}{y}_{N+1} \\ \underset{\sim}{z}_{N+1} \end{pmatrix} = \underset{\sim}{\beta}.$$

We write this system in compact form as

$$(6.9)\qquad L^h \begin{pmatrix} y^h \\ z^h \end{pmatrix} = \begin{bmatrix} L_{11}^h & L_{12}^h \\ L_{21}^h & L_{22}^h \\ & B^h \end{bmatrix} \begin{pmatrix} y^h \\ z^h \end{pmatrix} = \begin{pmatrix} p^h \\ q^h \\ \beta^h \end{pmatrix}$$

Our aim is to provide a sharp bound on

(6.10) $\chi(L^h) = ||L^h||\ ||(L^h)^{-1}||.$

Theorem 6.2. The condition number of (6.6) - (6.8) satisfies

(6.11) $\chi(L^h) \le c(1 + \max(C_L^{(0)}, C_L^{(1)}))\{N(T^{(0)}+T^{(1)}) + N(\varepsilon \underline{h}^{-1} N_c + 1 + h^{k+1} N)\}$

where the constant c is independent of ε and the mesh.

Remarks

(i) In the usual case $\varepsilon << \underline{h}$ and $C_L^{(0)}$, $C_L^{(1)}$, $T^{(0)}$, $T^{(1)}$ are moderate constants. Then the condition number is O(N), which is as good as can be expected.

(ii) In practice an explicit equilibration like (6.6) - (6.8) is unnecessary if Gauss elimination with scaled partial pivoting is used. This is the technique used in the package SOLVEBLOK (de Boor-Weiss [5]), which is used in the examples here as well as in COLSYS [2]. We conclude that the discretized collocation equations are safely solved in this way.

Proof. For the analysis we can assume that $A_{21} = 0$, so that $L_{21}^h = 0$ in (6.9), and that A_{11} is diagonal, as given in (2.18). It is clear that

(6.12) $||L^h|| \le c_1(1 + \max(C_L^{(0)}, C_L^{(1)}))$ $\qquad c_1 = \text{const},$

so we have left to consider $(L^h)^{-1}$. First we treat the slow components z^h. We derive a useful representation for the general solution of $L_{22}^h z^h = q^h$. Define

(6.13) $Z_s^h = \{Z_{ij}^{(s)},\ j = 1,\ldots,k;\ i = 1,\ldots,N;\ Z_{ij}^{(s)} \in R^{m\times m}\}$

by

$$(6.14)\quad L_{22}^{h} Z_s^h = 0, \quad Z_{s,1}^{(s)} = I$$

i.e. Z_s^h is the collocation approximation to the fundamental solution matrix $Z_s(t)$ defined by

$$(6.15)\quad \frac{d}{dt} Z_s(t) = A_{22} Z_s(t), \quad Z_s(t_s) = I.$$

From Russell [18],

$$(6.16)\quad ||Z_s(t_{ij}) - Z_{ij}^{(s)}|| \leq c_2 h^{k+1} \qquad c_2 = \text{const.}$$

Since $\underset{\sim}{z}_i = \underset{\sim}{z}_{i1} = \underset{\sim}{z}_{i-1,k}$ for Lobatto points, we have from (6.7) (with $A_{21} = 0$)

$$(6.17)\quad \underset{\sim}{z}_{ij} = \underset{\sim}{z}_{i-1,k} + h_i B_{ij} \underset{\sim}{z}_{i-1,k} + \underset{\sim}{q}_{ij} + h_i C_{ij} \hat{\underset{\sim}{q}}_i; \quad \hat{\underset{\sim}{q}}_i = (\underset{\sim}{q}_{i2}, \ldots, \underset{\sim}{q}_{ik})^T,$$

where the matrices B_{ij}, C_{ij} satisfy $||B_{ij}||$, $||C_{ij}|| \leq c_3 = \text{const.}$ In particular we have for $j = k$

$$\underset{\sim}{z}_{ik} = \underset{\sim}{z}_{i-1,k} + h_i B_{ik} \underset{\sim}{z}_{i-1,k} + \underset{\sim}{q}_{ik} + h_i C_{ik} \hat{\underset{\sim}{q}}_i.$$

Hence, by superposition, the general solution of (6.7) can be written as

$$(6.18)\quad \underset{\sim}{z}_{ij} = Z_{ij}^{(1)} \underset{\sim}{\eta}_I + \sum_{s=1}^{i-1} Z_{ij}^{(s+1)} (\underset{\sim}{q}_{sk} + h_s C_{sk} \hat{\underset{\sim}{q}}_s) + \underset{\sim}{q}_{ij} + h_i C_{ij} \hat{\underset{\sim}{q}}_i =$$

$$= \underset{\sim}{z}_{ij}^{hom} + \underset{\sim}{z}_{ij}^{p}; \qquad \underset{\sim}{\eta}_I \in R^m.$$

For later use we split the particular solution $\underset{\sim}{z}^p_{ij}$ as

$$(6.19) \qquad \underset{\sim}{z}^p_{ij} = \underset{\sim}{u}_{ij} + \underset{\sim}{v}_{ij}$$

where

$$(6.20) \qquad \underset{\sim}{u}_{ij} = \sum_{s=1}^{i-1} Z_{ij}^{(s+1)} \underset{\sim}{q}_{sk}$$

and $\underset{\sim}{v}_{ij}$ is the rest. Clearly

$$(6.21) \qquad ||v^h|| \leq c_4 ||q^h||, \qquad c_4 = \text{const.}$$

$$(6.22) \qquad ||z_p^h|| \leq c_5 N ||q^h||, \qquad c_5 = \text{const.}$$

We now turn to the fast components. We write the general solution of the system

$$(6.23) \qquad L_{11}^h y^h + L_{12}^h z^h = p^h$$

as

$$(6.24) \qquad y^h = Y^h \begin{pmatrix} \underset{\sim}{\eta}_{II} \\ \underset{\sim}{\eta}_{III} \end{pmatrix} + y_p^h,$$

where Y^h is the appropriate matrix solution of $L_{11}^h Y^h = 0$ and

$$(6.25) \qquad L_{11}^h y_p^h + L_{12}^h z_p^h = p^h,$$

with the initial or terminal value of each component of y_p^h specified as zero, depending on the sign of the real part of the relevant eigenvalue of A_{11} (which we assumed to be diagonal). Since (6.25) consists of n scalar equations for y_p^h, it suffices to consider one such equation

$$(6.26) \quad \varepsilon \frac{y_{ij}-y_i}{h_i} - \sum_{\ell=1}^{k} \hat{a}_{j\ell}\lambda y_{i\ell} = - \sum_{\ell=1}^{k} \hat{a}_{j\ell} g_{i\ell} + p_{ij}$$

where g_{ij} is the relevant component of $A_{12}z_{ij}^p$, p_{ij} is the relevant component of p_{ij}, and where we omit the index p on the y_{ij}'s. We assume that in (6.26) we have $re(\lambda) < 0$, so that the initial condition $y_1 = 0$ is appropriate. We now have to consider separately the three intervals $[0,T^{(0)}\varepsilon]$, $[T^{(0)}\varepsilon,1-T^{(1)}\varepsilon]$, $[1-T^{(1)}\varepsilon,1]$.

I. The interval $[0,T^{(0)}\varepsilon]$: From (6.26) it follows immediately that

$$|y_{i+1}| \leq |y_i| + \frac{h^{(0)}}{\varepsilon} c_6(||g^h|| + ||p^h||), \; i = 1,\ldots,N^{(0)}; \quad c_6 = \text{const.},$$

whence

$$(6.27) \quad |y_{ij}| \leq c_7 T^{(0)}(||z^h|| + ||p^h||) \qquad j = 1,\ldots,k, \; i = 1,\ldots,N^{(0)}; \; c_7 = \text{const.}$$

II. The interval $[T^{(0)}\varepsilon,1-T^{(1)}\varepsilon]$: We consider the two parts of the right hand side in (6.26) separately, but denote the solution by y_{ij} in each case.

Part 1: $-\sum_{\ell=1}^{k} \hat{a}_{j\ell} g_{i\ell}$. As in the proof of theorem 5.2 we obtain

$$(6.28)\quad \bar{\underset{\sim}{y}}_i = (\tfrac{\varepsilon}{\lambda h_i} I-\bar{A})^{-1}(\tfrac{\varepsilon}{\lambda h_i} \bar{\underset{\sim}{1}}+\bar{\underset{\sim}{a}})y_i + \tfrac{1}{\lambda}(\tfrac{\varepsilon}{\lambda h_i} \bar{C}\bar{A}^{-1}-\bar{A}^{-1})(\bar{A}\bar{\underset{\sim}{g}}_i+\bar{\underset{\sim}{a}}g_{i1})$$

The last component of this equation is

$$y_{i+1} = \gamma(\zeta_i)y_i - \tfrac{1}{\lambda}(g_{i+1,1}+(-1)^k g_{i,1}) + \tfrac{1}{\lambda}[(\tfrac{\varepsilon}{\lambda h_i} \bar{C}\bar{A}^{-1})(\bar{A}\bar{\underset{\sim}{g}}_i+\bar{\underset{\sim}{a}}g_{i1})]_k$$

The solution to this recursion is

$$(6.29)\quad y_i = \prod_{\ell=N^{(0)}+1}^{i-1} \gamma(\zeta_\ell) y_{N^{(0)}+1} - \tfrac{1}{\lambda} \sum_{j=N^{(0)}+1}^{i-1} [\prod_{\ell=j}^{i-2} \gamma(\zeta_{\ell+1})](g_{j+1,1}+(-1)^k g_{j,1}) +$$

$$+ \tfrac{\varepsilon}{\lambda^2} \sum_{j=N^{(0)}+1}^{i-1} [\prod_{\ell=j}^{i-2} \gamma(\zeta_{\ell+1})] h_i^{-1} \bar{C}\bar{A}^{-1} [\bar{A}\bar{\underset{\sim}{g}}_j+\bar{\underset{\sim}{a}}g_{j1}]_k, \quad i = N^{(0)}+1,\ldots,N-N^{(1)}.$$

Corresponding to the three terms of the right hand side of this equation we write $y_i = y_i^{(1)} + y_i^{(2)} + y_i^{(3)}$. Clearly

$$(6.30)\quad \begin{aligned} &|y_i^{(1)}| \le |y_{N^{(0)}+1}| && i = N^{(0)}+1,\ldots,N-N^{(1)} \\ &|y_i^{(3)}| \le c_8 \varepsilon \underline{h}^{-1} N_c ||g^h|| && c_8 = \text{const.} \end{aligned}$$

For $y_i^{(2)}$ we have to distinguish between the cases when k is odd and k is even.

Case 1: k is odd. Then by (6.16), (6.20), (6.21)

$$|g_{j+1,1}-g_{j,1}| \le c_9 \sum_{s=1}^{N}(h_s+h^{k+1})||q^h|| \le c_9(1+Nh^{k+1})||q^h|| \qquad c_9 = \text{const.}$$

whence

$$(6.31)\qquad |y_i^{(2)}| \le c_{10}N_c(1+Nh^{k+1})||q^h||, \qquad i = N^{(0)}+1,\dots,N-N^{(1)},\ c_{10} = \text{const.}$$

Case 2: k is even. We write

$$\frac{1}{\lambda}\sum_{j=N^{(0)}+1}^{i-1}\Big[\prod_{\ell=j}^{i-2}\gamma(\zeta_{\ell+1})\Big](g_{j+1,1}+g_{j,1}) = \frac{1}{\lambda}\sum_{j=N^{(0)}+2}^{i-1}\Big[\prod_{\ell=j}^{i-2}\gamma(\zeta_{\ell+1})\Big](\gamma(\zeta_j)+1)g_{j1} +$$

$$+\frac{1}{\lambda}\Big[\prod_{\ell=N^{(0)}+1}^{i-2}\gamma(\zeta_{\ell+1})\Big]g_{N^{(0)}+1,1} + \frac{1}{\lambda}g_{i,1}$$

Since for large $|\zeta|$,

$$\gamma(\zeta) \approx -1 + \text{const}\cdot\zeta^{-1}$$

we obtain

$$|y_i^{(2)}| \le c_{11}(\varepsilon\underline{h}^{-1}N_c+1)||g^h||, \quad i = N^{(0)}+1,\dots,N-N^{(1)}, \quad c_{11} = \text{const.},$$

whence, using (6.22),

$$(6.32)\qquad |y_i^{(2)}| \le c_{12}N(\varepsilon\underline{h}^{-1}N_c+1)||q^h||, \quad i = N^{(0)}+1,\dots,N-N^{(1)}, \quad c_{12} = \text{const.}$$

This concludes case 2 and hence part 1.

Part 2: p_{ij}. Here we consider the contribution of p_{ij} in (6.26). This is just the case considered in §4.4 and as in (4.41) we obtain

$$(6.33)\quad |y_{ij}| \le c_{13}N_c||p^h||, \quad i = N^{(0)}+1,\dots,N-N^{(1)}; \quad c_{13} = \text{const.}$$

Now, combining the estimates (6.30), (6.31), (6.32), (6.33) and using (6.22) and (6.28) we finally obtain for the intermediate interval

$$(6.34)\quad |y_{ij}| \le c_{14}\{|y_{N_0+1}|N(\varepsilon\underline{h}^{-1}N_c+1+Nh^{k+1})||q^h||+N_c||p^h||\}, \quad j = 1,\dots,k,$$
$$i = N^{(0)}+1,\dots,N-N^{(1)}, \qquad c_{14} = \text{const.}$$

III. The interval $[1-T^{(1)}\varepsilon,1]$: This is easy. As in the first interval we obtain

$$(6.35)\quad |y_{ij}| \le c_{15}(|y_{N-N^{(1)}+1}| + T^{(1)}(||z^h|| + ||p^h||)), \quad j = 1,\dots,k,$$
$$i = N-N^{(1)}+1,\dots,N; \quad c_{15} = \text{const.}$$

Thus, we have concluded our estimates for y_p^h of (6.24): combining (6.22), (6.27), (6.34) and (6.35) we obtain

$$(6.36)\quad ||y_p^h|| \le c_{16}\{(T^{(0)}+T^{(1)})(N||q^h||+||p^h||) + N(\varepsilon\underline{h}^{-1}N_c+1+Nh^{k+1})||q^h|| +$$
$$+ N_c||p^h||\} \qquad c_{16} = \text{const.}$$

We have then dealt with the particular solution. The general solution of (6.6), (6.7) is

$$(6.37)\qquad \begin{bmatrix} y^h \\ z^h \end{bmatrix} = \begin{bmatrix} 0 \\ Z^h \end{bmatrix} \underset{\sim}{\eta}_I + \begin{bmatrix} Y^h \\ 0 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\eta}_{II} \\ \underset{\sim}{\eta}_{III} \end{bmatrix} + \begin{bmatrix} y_p^h \\ z_p^h \end{bmatrix}$$

and a comparison with (6.4) yields

$$\begin{bmatrix} 0 \\ Z^h \end{bmatrix} = W_I^h \ , \quad \begin{bmatrix} Y^h \\ 0 \end{bmatrix} = (W_{II}^h, W_{III}^h)$$

The concrete values of $\underset{\sim}{\eta}_I$, $\underset{\sim}{\eta}_{II}$, $\underset{\sim}{\eta}_{III}$ are determined by substituting (6.37) into (6.8), which leads to the linear system

$$A^h(\varepsilon)\underset{\sim}{\eta} = \underset{\sim}{\beta} - B_0 \begin{bmatrix} y_1^p \\ z_1^p \end{bmatrix} - B_1 \begin{bmatrix} y_{N+1}^p \\ z_{N+1}^p \end{bmatrix} \ ; \qquad \underset{\sim}{\eta} = (\underset{\sim}{\eta}_I, \underset{\sim}{\eta}_{II}, \underset{\sim}{\eta}_{III})^T .$$

Since $||(A^h(\varepsilon))^{-1}|| \leq c_{17}$, $c_{17} = \text{const}$, this implies that

$$(6.38)\qquad ||\underset{\sim}{\eta}|| \leq c_{18}\left(||\underset{\sim}{\beta}|| + \left|\left|\begin{matrix} y_p^h \\ z_p^h \end{matrix}\right|\right|\right) , \qquad c_{18} = \text{const.}$$

and the result (6.11) finally follows from (6.38), (6.37), (6.22), (6.36) and (6.12). This completes the proof of Theorem 6.2.

QED

6.3 Example

The example presented here is from Flaherty - O'Malley [8]. Consider

$$(6.39)\qquad \mu y'' + \varepsilon y' - y = 0$$

$$(6.40)\qquad y(0) = 1, \quad y(1) = \tfrac{1}{2},$$

with ε and μ small parameters. The behaviour of the asymptotic solution depends on whether $\frac{\mu}{\varepsilon^2}$ tends to 0,1 or ∞ as $\varepsilon \to 0$, but in all of these cases boundary layers at both ends are present, connecting the boundary values to the reduced solution $y_R \equiv 0$. Here we consider the case $\mu = \varepsilon^2$.

Rewriting (6.39) as a first order system as in Kreiss-Nichols [13], we get

$$\varepsilon y_1' = -y_1 + y_2 \tag{6.41}$$

$$\varepsilon y_2' = y_1 \tag{6.42}$$

subject to (6.40). Thus there are no slow components present, and the eigenvalues of A_{11} are $\lambda_- = -\frac{1}{2} - \frac{\sqrt{5}}{2} < 0$ and $\lambda_+ = -\frac{1}{2} + \frac{\sqrt{5}}{2} > 0$.

First, we perform some calculations for $\varepsilon = 1.\text{-}8$ on a uniform mesh of 8 subintervals. As expected, using the Radau points produces large instabilities, both for this mesh and for all other meshes tried! For the Gauss and Lobatto points the resulting nodal values of y_1 are on the straight line joining the boundary values (cf. Hemker [11, p. 86]) while those of y_2 are 4 or 5 orders of magnitude larger, indicating large oscillations of the collocation solution in between the mesh points.

Next, to the uniform mesh above we add the mesh points 2.87-8 ($\approx \frac{4.64\varepsilon}{\lambda_-}$) and .999999925 ($\approx 1 - \frac{4.64\varepsilon}{\lambda_+}$), obtained using table 3.1. For the Gauss points with k = 2,3 and 4 the errors at t = .5 are (.14, .13-1), (.91-4, .12-3) and (.15-1, .20-2), respectively. The same errors are obtained for the Lobatto points for k = 3,4 and 5, respectively. Our point concerning layer-damping meshes is clearly demonstrated.

REFERENCES

1. U. Ascher, "Solving boundary-value problems with a spline-collocation code", J. Comp. Phys. 34 (1980), 401-413.

2. U. Ascher, J. Christiansen and R.D. Russell, "A collocation solver for mixed order systems of boundary value problems", Math. Comp. 33 (1979), 659-679.

3. O. Axelsson, "A class of A-stable methods", BIT 9 (1969), 185-199.

4. C. de Boor and B. Swartz, "Collocation at Gaussian points", SIAM J. Numer. Anal. 10 (1973), 582-606.

5. C. de Boor and R. Weiss, "SOLVEBLOK: A package for solving almost block diagonal linear systems", ACM Trans. Math. Software 6 (1980), 80-87.

6. J.C. Butcher, "Implicit Runge-Kutta processes", Math. Comp. 18 (1964), 50-64.

7. K. Burrage and J.C. Butcher, "Stability criteria for implicit Runge-Kutta methods", SIAM J. Numer. Anal. 16 (1979), 46-57.

8. J.E. Flaherty and R.E. O'Malley, Jr., "The numerical solution of boundary value problems for stiff differential equations", Math. Comp. 31 (1977), 66-93.

9. (_____), "Analytical and numerical methods for nonlinear singular singularly-perturbed initial value problems", SIAM J. Appl. Math 38 (1980), 225-248.

10. (_____), "On the numerical integration of two-point boundary value problems for stiff systems of ordinary differential equations", Proc. BAIL I, Dublin (1980), ed. J. Miller, 93-102.

11. P.W. Hemker, "A numerical study of stiff two-point boundary value problems", Math. Centrum, Amsterdam (1977).

12. F. de Hoog and R. Weiss, "On the boundary value problem for systems of ordinary differential equations with a singularity of the second kind", SIAM J. Math. Anal. 11 (1980), 41-60.

13. H.O. Kreiss and N. Nichols, Numerical methods for singular perturbation problems, Uppsala University, Department of Computer Science Report No. 57, 1975.

14. M. Lentini and V. Pereyra, "An adaptive finite difference solver for nonlinear two point boundary problems with mild boundary layers", SIAM J. Numer. Anal. 14 (1977), 91-111.

15. R.E. O'Malley, "Introduction to Singular Perturbations", (1974), Academic Press, N.Y.

16. A. Prothero and A. Robinson, "On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations", Math. Comp. 28 (1974), 145-162.

17. C. Ringhofer, A class of collocation schemes for singularly perturbed boundary value problem, Thesis, Technische Universitat Wien, 1981.

18. R.D. Russell, "Collocation for systems of boundary value problems", Numer. Math. 23 (1974), 119-133.

19. (_____), "Mesh selection methods", in Lecture Notes in Computer Science 76, Springer-Verlag (1979), 228-242.

20. E.B. Saff and R.S. Varga, "On the zeros and poles of Pade' approximants to e^z", Numer. Math. 25 (1975), 1-14.

21. J.M. Varah, "Stiff stability considerations for implicit Runge-Kutta methods", Tech. Rep. 80-1, Dept. Computer Science, University of British Columbia, Vancouver, Canada.

22. R. Weiss, "The application of implicit Runge-Kutta and collocation methods to boundary-value problems", Math. Comp. 28 (1974), 449-464.

23. K. Wright, "Some relationships between implicit Runge-Kutta, collocation and Lanczos τ methods, and their stability properties", BIT 10 (1970), 217-227.