MATHEMATICAL CENTRE TRACTS 80

P.W. HEMKER

A NUMERICAL STUDY OF STIFF TWO-POINT BOUNDARY PROBLEMS

MATHEMATISCH CENTRUM AMSTERDAM 1977

AMS(MOS) subject classification scheme (1970): 65L10

ISBN 90 6196 146 7

CONTENTS

ACKNOWLEDGEMENT

I thank the Mathematical Centre for the opportunity to publish this monograph in their series Mathematical Centre Tracts and all those at the Mathematical Centre who have contributed to its technical realization.

NOTATIONS

$\mathbb{R}$ denotes the field of real numbers.

$I = (a,b), \quad a,b \in \mathbb{R}, \ a < b.$

$\Pi = \{x_i \mid a = x_0 < x_1 < \dots < x_N = b\}.$

$I_i = (x_{i-1}, x_i); \quad h_i = x_i - x_{i-1}.$

$L^p(I)$, $p = 1,2,\infty$, the Banach space of Lebesgue p-integrable functions on I;

norm: $\|\cdot\|_{0,p}$; $\|\cdot\|_0 = \|\cdot\|_{0,2}$; $\|\cdot\|_\infty = \|\cdot\|_{0,\infty}$.

$C^k(I)$, $k = 0,1,2,\dots,\infty$, the set of real-valued, k times continuously differentiable funtions on I.

$C_0^\infty(I)$, the set of $C^\infty(I)$ functions with compact support in I.

$C(I) = C^0(I)$.

$C^{-1}(I) = \{u \mid \exists \Pi \ u_{\text{restr.}I_i} \in C^0(I_i), \ i = 1,2,\dots,N\}$,

the set of piecewise continuous functions.

$\|\cdot\|_{\pi,p}$, $p = 1,2,\infty$, pointwise norms (see section 2.2).

$H^k(I)$, $H_0^k(I)$, $k = 0,1,2,\dots$, Sobolev spaces;

norm: $\|\cdot\|_k$ (see section 3.1).

$H^{k,\pi}[a,b]$, $k = 0,1,2,\dots$, piecewise Sobolev spaces;

norm: $\|\cdot\|_{k,\pi}$ (see section 3.2);

innerproduct: $(\cdot,\cdot)_{0,\pi} = \sum_{i=1}^{N} (\cdot,\cdot)_{L^2(I_i)}$.

$M^{m,k}(\Pi)$, $M_0^{m,k}(\Pi)$, $k = 0,1,2,\dots$, $m = -1,0,1,2,\dots$,

piecewise polynomial spaces (see section 3.1).

$N^{m,k}(\Pi,\alpha_\pi)$, $N_0^{m,k}(\Pi,\alpha_\pi)$, $k = 0,1,2,\dots,m = 0,1,2,\dots$,

exponentially fitted spaces (see section 3.4).

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \text{ Kronecker's delta.}$$

$D \cdot = \frac{d\cdot}{dx}$ denotes differentiation with respect to x.

C and $K_1, K_2, \dots$ denote generic constants; that means that they are constants of which the value may be different on each appearance.

CHAPTER I

ANALYTIC PROPERTIES OF LINEAR, SINGULARLY PERTURBED TWO-POINT BOUNDARY-VALUE PROBLEMS

This chapter gives an exposition of some essential results in the theory of singularly perturbed two-point boundary-value problems. In view of the many investigations that have been carried out in this field, it is not a survey of the literature on the subject. The main aim of this chapter is to show the fundamental results in the singular perturbation theory that underlie our numerical investigations in the next chapters.

1.1. INTRODUCTION

In the first three chapters of this monograph we consider a second order, linear, singularly perturbed two-point boundary-value porblem.

One standard form of such a problem is

$$(1.1.1a)\qquad \begin{cases} \varepsilon y''(x) + f(x)y'(x) + g(x)y(x) = s(x), \\ x \in (a,b), \quad \varepsilon > 0, \end{cases}$$

$$(1.1.1b)\qquad y(a) = \alpha, \quad y(b) = \beta.$$

We assume f,g and s to be sufficiently smooth functions on [a,b]. In particular we are interested in the solution of these problems for small values of ε. The most striking feature of the differential equation is that its order is lower for $\varepsilon = 0$ than for $\varepsilon \neq 0$. For this lower order equation one of the two boundary conditions is superfluous. Indeed, for small values of ε, it turns out that small regions arise in [a,b], in which the connection with the boundary conditions is made. This causes the solution to have a multiscale character, i.e. the solution is described by slowly and rapidly varying parts. This multiscale character is a characteristic feature of the functions that describe the solutions of singular perturbation problems. It also means that attempts to seek a solution in the form of an ascending series in powers of ε will fail, unlike the case of regular perturbation problems.

The multiscale character, where one scale prevails over the other in each region, should be distinguished carefully from the "multiple time-scale" as used in the two-variable expansion method, where a solution may depend both on a slow and a fast independent variable in the same region, e.g. $y(t) = t \sin(t/\varepsilon)$; cf. COLE [1968].

As we want to solve problems of the type (1.1.1) for small values of ε, we are interested in the asymptotic behaviour for $\varepsilon \to 0$. In a number of cases, a treatment of this behaviour can be given by the theory of matched asymptotic expansions (cf. e.g. ECKHAUS [1973], COLE [1968]). In other cases, however, the Wentzel-Kramer-Brillouin-(WKB-) method seems to be a more expedient tool (cf. e.g. SIROVICH [1971], WASOW [1965]).

If the coefficient g is negative, the maximum principle can be used in order to derive a number of extremely useful a priori bounds. This principle can also be applied in nonlinear problems (cf. DORR, PARTER & SHAMPINE [1973]).

In section 1.2 we will give a summary of results obtained in the investigations of the asymptotic behaviour of (1.1.1) by e.g. PEARSON [1968a], ACKERBERG & O'MALLEY [1970], KREISS & PARTER [1974] and ABRAHAMSSON [1975]. In section 1.3 some examples are given in which the most striking features of singualrly perturbed two-point boundary value problems are demonstrated.

Although we are able to analyse the behaviour for a certain number of special cases, it is rather difficult to compute the solution for more general functions f,g, and s. For this reason algorithms for its numerical approximation are developed in chapters 2 and 3. In chapter 4 these algorithms are applied to nonlinear problems.

In the remaining part of this section we collect some preliminary results.

An integrating factor

In many cases it is convenient to write equation (1.1.1.a) in a slightly different form; it is obtained by multiplying the equation by an integrating factor p(x):

$$(1.1.2) \qquad \varepsilon\,(p\,y')' + p\,g\,y = p\,s,$$

where

$$(1.1.3) \qquad p(x) = \exp\left\{ \int^{x} \frac{f(t)}{\varepsilon}\,dt \right\}.$$

The solution as a stationary point of a quadratic functional

Let ϕ be a function in $C^1[a,b]$ with fixed endpoints $\phi(a) = \alpha$, $\phi(b) = \beta$, and consider the functional

$$(1.1.4)\qquad E[\phi] = \int_a^b p(x)\{\varepsilon(\phi'(x))^2 - g(x)\phi(x)^2 + 2s(x)\phi(x)\}dx.$$

By the classical Euler-Lagrange theory, it can be shown that the solution of problem (1.1.1) is a stationary point for $E[\phi]$. If $g(x) < 0$, $E[\phi]$ is a convex functional which assumes its minimum for $\phi = y$, the solution of (1.1.1). In particular, the functional in this case is a starting point for theoretical considerations; e.g. it can be used to justify the Ritz-Galerkin-method for obtaining approximations to $y(x)$.

The energy norm

It is well known that, for $g \leq 0$, an "energy"-norm can be defined on $C^1[a,b]$ by

$$(1.1.7)\qquad \|\phi\|_E^2 = \int_a^b p(x)\{\varepsilon(\phi'(x))^2 - g(x)\phi(x)^2\}dx.$$

Here the special role of the integrating factor p is clearly demonstrated: it can be considered as a weighting-function.

Transition points

By application of the *Liouville transformation* to the dependent variable y:

$$(1.1.8)\qquad z(x) = y(x)\exp\int_a^x \frac{f(t)}{2\varepsilon}dt,$$

the differential equation (1.1.1.a) can be cast into the form

$$(1.1.9)\qquad 2\varepsilon z'' + q z = r ,$$

where

$$(1.1.10)\qquad q(x) = 2g(x) - f'(x) - \frac{f^2(x)}{2\varepsilon},$$

$$(1.1.11)\qquad r(x) = 2s(x)\exp\int_a^x \frac{f(t)}{2\varepsilon}dt.$$

One observes that y is an oscillating function when $2g - f' > f^2/2\varepsilon$ over a large enough interval. Asymptotic solutions to equation (1.1.9) for small values of ε are only valid within a certain sector of the complex plane. They are not valid for small values of $q(x)$. In particular they are not valid if the solution is changed from a periodic into an exponential function, i.e. in passing through a zero of q. Such a point, where the character of the solution changes, is called a *transition point* or *"classical turning point"*.

Turning points

Zeroes of the function f in equation (1.1.1.a) are also called turning points. These turning points do not entirely coincide with "classical turning points". The relation will be made clear with the aid of the following three examples.

The example of a single (or first order) classical turning point is given by the equation

$$\varepsilon y'' - xy = 0.$$

The transition point is $x = 0$. By the local coordinate $\xi = x\varepsilon^{-1/3}$ this equation is converted into Airy's equation

$$\frac{d^2y}{d\xi^2} - \xi y = 0.$$

The solution, Airy's function $Ai(\xi)$ or $Bi(\xi)$, is oscillating for $\xi < 0$ and non-oscillating for $\xi > 0$.

An example of a double (or second order) transition point is given by the equation

$$\varepsilon^2 y'' + (1-x^2)y = 0$$

at $x = \pm 1$.

By the change of independent variable $\xi = x\sqrt{2/\varepsilon}$ it becomes

$$\frac{d^2y}{d\xi^2} - (\frac{1}{4}\xi^2 - \frac{1}{2\varepsilon})y = 0.$$

The solutions of this equation are the *Weber* - or *parabolic cylinder* *functions*

$$D_{\frac{1}{2\varepsilon}-\frac{1}{2}}(\xi) \quad \text{and} \quad D_{\frac{1}{2\varepsilon}-\frac{1}{2}}(-\xi),$$

which do not oscillate for $|x| > 1$.

Another example of a double transition point is given by the equation

$$(1.1.12) \qquad \varepsilon y'' + xy' + cy = 0.$$

Applying the Liouville transformation we get

$$z'' - \{(\frac{x}{2\varepsilon})^2 + (\frac{1-2c}{2\varepsilon})\}z = 0.$$

If $c > \frac{1}{2}$, this equation has *two* turning points in the classical sense, viz. for $x = \pm 2\sqrt{\varepsilon}\sqrt{c-\frac{1}{2}}$, which both approach $x = 0$ for $\varepsilon \to 0$. In the other sense it has *one* turning point for $x = 0$ since the coefficient of y' in (1.1.12) has one zero.

We will use the word *turning point* (without further indication) only for a zero of the coefficient of y'.

The non-homogeneous equation

If we investigate the asymptotic behaviour of the problem (1.1.1) for $\varepsilon \to 0$, the right-hand-side term $s(x)$ frequently is unimportant in the sense that the equation is easily reduced to its homogeneous form. If there exists a solution $v_1 \in C^2[a,b]$ of the reduced equation

$$(1.1.13) \qquad fv_1' + gv_1 = s,$$

then $u_1 = y - v_1$ satisfies

$$\varepsilon u_1'' + fu_1' + gu_1 = -\varepsilon v_1''.$$

The process can be iterated and the influence of s on the solution y of (1.1.1) can be expressed in a power series in ε. Truncating the process at the n-th stage, the non-homogeneous term is $O(\varepsilon^n)$ which can usually be discarded, leaving the homogeneous equation

$$\varepsilon u_n'' + f u_n' + g u_n = 0.$$

On subintervals $(c,d) \subset [a,b]$ which do not contain zeroes of f,

$$(1.1.14)\qquad y_1(x) = C \exp\{-\int^x g(t)/f(t)dt\}$$

is the general solution of equation (1.1.13) with $s \equiv 0$, and

$$(1.1.15)\qquad v_1(x) = y_1(x)\cdot \left[\int^x \frac{s(t)}{y_1(t)f(t)}dt + C\right]$$

is the solution of the full equation (1.1.13).

1.2. EXPOSITION OF ASYMPTOTIC PROPERTIES

To obtain an insight into the asymptotic properties of the solution of equation (1.1.1.a) for $\varepsilon \to 0$ we first study the homogeneous equation

$$(1.2.1)\qquad \varepsilon y'' + fy' + gy = 0.$$

We are especially interested in the question under what conditions a solution of the problem (1.2.1)-(1.1.1.b) satisfies approximately the *reduced equation*

$$(1.2.2)\qquad fy' + gy = 0.$$

This certainly will be the case in those parts of [a,b] where $y''(x;\varepsilon)$ is uniformly bounded in ε and hence it is important to know where these parts are (if they exist).

We do not intend to study the problem in all generality but we shall consider a number of characteristic cases. Since f(x) is the coefficient in the leading term of the reduced equation, it is natural to consider the following three cases:

A. f is positive (or negative) definite on the whole interval [a,b] , i.e. there are no turning points;

B. f has a simple zero in (a,b), i.e. there is one turning point;

C. f is identical to zero on [a,b].

A. No turning points

First we focus on the case where f is positive or negative definite on [a,b]. According to the WKB-technique we formally solve equation (1.2.1) by writing

$$(1.2.3)\qquad \begin{aligned} y(x) &= \exp\{\frac{1}{\varepsilon}\int^{x} \beta(t)dt\}, \\ \beta(t) &= \sum_{n=0} p_n(t)\varepsilon^n. \end{aligned}$$

This leads - to first order - to two approximate solutions

$$(1.2.4)\qquad y_1 = C \exp\{-\int_a^x \frac{g(t)}{f(t)}\, dt\},$$

$$(1.2.5)\qquad y_2 = C\, f(x)^{-1} \exp\{-\frac{1}{\varepsilon}\int_a^x f(t)dt + \int_a^x \frac{g(t)}{f(t)}\}dt\ .$$

The solution of eq. (1.2.1) which satisfies the boundary condition (1.1.1.b) can be written

$$(1.2.6)\qquad y(x) = C_1 y_1(x) + C_2 y_2(x) + \mathcal{O}(\varepsilon).$$

For $f > 0$, $C_2 y_2$ is exponentially small outside a small region of $\mathcal{O}(\varepsilon)$ near $x = a$. (The region where $C_2 y_2$ is not exponentially small is called a *boundary layer*.) The coefficient C_1 is determined by $C_1 y_1(b) = \beta$ and C_2 by $C_2 y_2(a) = \alpha - C_1 y_1(a)$.

For $f < 0$, $C_2 y_2$ is exponentially small outside a boundary layer near the other end $x = b$. We see that, away from the boundary layer, the solution is approximately described by $C_1 y_1$. This function satisfies the reduced equation (1.2.2) and the boundary condition at the non-boundary-layer end.

B. One turning point

For a single zero of f, we can take $a < 0 < b$ and $f(0) = 0$ without loss of generality. The WKB-analysis shows that

$$(1.2.7)\qquad y_1 = x^{\ell} \exp\{-\int_0^x (\frac{g(t)}{f(t)} + \frac{\ell}{t}) dt\}$$

$$(1.2.8)\qquad y_2 = x^{-\ell-1}\{\frac{x}{f(x)}\}\exp\{-\frac{1}{\varepsilon}\int_0^x f(t)dt + \int_0^x \left(\frac{g(t)}{f(t)} + \frac{\ell}{t}\right)dt\}$$

where

$$(1.2.9) \qquad \ell = -g(0)/f'(0).$$

Since the singularity in the integrand is subtracted, the integrals do exist.

For arbitrary values of C_L and C_R, C_Ly_1 and C_Ry_1 satisfy the homogeneous reduced equation (1.2.2) on [a,0) and (0,b] respectively . For $\ell < 0$, the form (1.2.7) shows that there are no smooth solutions of (1.2.2) on [a,b] except $y \equiv 0$. If $\ell \geq 0$, for any solution $y_1 \in C^k[a,b]$, $k > \ell$, it is seen that $C_L = C_R$ by the smoothness condition. Moreover, if $\ell \neq 0,1,2,...$ then $C_L = C_R = 0$, i.e. the homogeneous equation only admits the trivial equation.

These facts establish the uniqueness of a solution $y \in C^k[a,b]$; $k > \ell$, of the inhomogeneous reduced equation (1.1.13), if $\ell \neq 0,1,2,...$. This solution exists and can be written in the form

$$(1.2.10) \qquad \begin{aligned} y = -\phi(x)&\left\{\frac{\psi(x)}{\ell} + \frac{x}{\ell}\,\frac{\psi'(x)}{(\ell-1)} + \ldots + \frac{x^n\psi^{(n)}(x)}{\ell(\ell-1)\ldots(\ell-n)}\right\} \\ &+ \frac{x^{\ell}\phi(x)}{\ell(\ell-1)\ldots(\ell-n)}\int_0^x t^{n-\ell}\psi^{(n+1)}(t)dt, \\ &\qquad n > \ell - 1, \\ &\qquad n \geq -1, \end{aligned}$$

where

$$\phi(x) = \exp\left\{-\int_0^x \frac{g(t)}{f(t)} + \frac{\ell}{t}\,dt\right\},$$

$$\psi(t) = \frac{s(t)t}{\phi(t)f(t)},$$

(cf. ABRAHAMSSON [1975], lemma 3.2).

If $\ell = 0,1,2,...$ nontrivial solutions y_1 of the homogeneous equation (1.2.2) on [a,b] are possible; e.g. $y_1 = C\,x$ is a solution of $xy' - y = 0$.

On $[a,-\delta]$ and $[\delta,b]$, $\delta > 0$, the WKB-solution of the homogeneous equation can be written as

$$(1.2.11) \qquad y(x) \sim C_1y_1(x) + C_2y_2(x).$$

Near the turning point both solutions, y_1 and y_2 are not in general valid. Hence it is expected that the coefficients C_1 and C_2 differ on either side of the turning point.

For the description of the qualitative behaviour of the solution of equation (1.2.1) we have to distinguish between $f'(0) > 0$ and $f'(0) < 0$.

Case I: $f'(0) > 0$

Similar to the above remarks about y_1; $y_2 \in C^k[a,b]$, $k > -\ell-1$, implies either $y_2 \equiv 0$ or $\ell = -1,-2,-3,\ldots$.

If $\ell = -1,-2,-3,\ldots$ the solutions of (1.2.1) may explode exponentially over the whole interval $[a,b]$; e.g.

$$y = y_2 = \exp((1-x^2)/2\varepsilon)$$

is a solution for the problem

$$\varepsilon y'' + xy' + y = 0,$$
$$y(-1) = y(1) = 1.$$

For $\ell \neq -1,-2,\ldots$, any nontrivial y_2 is not a smooth solution on $[a,b]$ and we consider the WKB-approximation on the intervals $[a,-\delta]$ and $[\delta,b]$, $\delta > 0$, separately. Since $-\frac{1}{\varepsilon}\int_0^x f(t)dt \le 0$ on $[a,b]$, the influence of y_2 is exponentially small outside a region near $x = 0$.
By analogy to the results obtained without a turning point we see that the approximate solution of (1.2.1)-(1.1.1.b) is described by

$$y(x) \sim \frac{\alpha}{y_1(a)} y_1(x) \qquad \text{on } [a,-\delta]$$

and by

$$y(x) \sim \frac{\beta}{y_1(b)} y_1(x) \qquad \text{on } [\delta,b]$$

for some $\delta > 0$.
In this case there is a boundary layer neither at $x = a$ nor at $x = b$. This is rigorously stated in the following

THEOREM 1.2.1. (cf. ABRAHAMSSON [1975]) *Let there be one turning point at* $x = 0$, *let* $f'(0) > 0$ *and* $\ell = -g(0)/f'(0) \neq -1,-2,\ldots$ *and let* v *be the* *solution on* $[a,0) \cup (0,b]$ *of the reduced equation* (1.1.13) *with the boundary conditions*

$$v(a) = \alpha \quad \textit{and} \quad v(b) = \beta,$$

then there are constants δ, K *and* ε_0 *independent of* ε, *such that the solution of problem* (1.1.1) *satisfies*

$$(1.2.12) \qquad \max_{0<\delta\le|x|\le 1} |y(x;\varepsilon) - v(x)| \le K\varepsilon$$

for all $0 < \varepsilon \le \varepsilon_0$.

The behaviour in the turning point region strongly depends on the sign of ℓ. If $\ell > 0$, $y(0)$ converges to zero. If $\ell < 0$ both limits $y(+0)$ and $y(-0)$ are unbounded and a complicated behaviour may be expected in the turning point region. If $\ell = 0$ a shock layer is expected (cf. fig. 1.3.3).

Case II: $f'(0) < 0$

In this case $-\frac{1}{\varepsilon}\int_0^x f(t)dt \ge 0$ and therefore the influence of y_2 grows exponentially for increasing values of $|x|$. Thus y_2 can serve as a boundary layer function both near $x = a$ and $x = b$. Outside these possible boundary layers, y_2 is exponentially small.

In order to investigate the contribution from y_1 to the solution of (1.2.1)-(1.1.1.b), a link has to be made between the WKB-approximations for $x < 0$ and $x > 0$. To this end equation (1.2.1) is approximated in the neighbourhood of $x = 0$ by

$$(1.2.13) \qquad \varepsilon y'' + f'(0)\, x\, y' + g(0)y = 0.$$

The solution in the turning point region can now be expressed in terms of parabolic cylinder functions $D_\ell(z)$. Introducing the local coordinate $\xi = x\sqrt{-f'(0)/\varepsilon}$, we approximate the solution of eq. (1.2.1) near $x = 0$ by

$$(1.2.14) \qquad y_3(x) = e^{\xi^2/4}[A\, D_\ell(\xi) + B\, D_\ell(-\xi)].$$

$D_\ell(\xi)$ and $D_\ell(-\xi)$ yield two independent solutions when $\ell \neq 0,1,2,\ldots$, otherwise an independent solution is given by $D_{-\ell-1}(i\xi)$. For $\ell = 0,1,2,\ldots$, we have

$$(1.2.15) \qquad D_\ell(z) = e^{-z^2/4}\, He_\ell(z),$$

where He_ℓ is the Hermite polynomial of degree ℓ (cf. ABRAMOWITZ & STEGUN [1965]).

The asymptotic behaviour for $|z| \to \infty$ of the function $D_\ell(z)$ is given by

$$(1.2.16) \qquad D_\ell(z) \sim e^{-z^2/4}\, z^\ell, \quad \text{if } |\arg(z)| < \tfrac{3}{4}\pi,$$

and

$$(1.2.17) \qquad D_\ell(z) \sim e^{-z^2/4}\, z^\ell - \frac{\sqrt{2\pi}}{\Gamma(-\ell)}\, e^{\ell\pi i}\, e^{z^2/4}\, z^{-\ell-1},$$

if $\frac{1}{4}\pi < \arg(z) < \frac{5}{4}\pi$.
(WHITTAKER & WATSON [1946], p.348.)

Since all exponentially large terms in y_3 must be absent for $\xi \to \pm\infty$, when matching the local solution (1.2.14) with $c_1y_1 + c_2y_2$, we have to choose $A = B = 0$, unless $\ell = 0,1,2,\ldots$. So we have $y(x) \sim 0$ in the turning-point region if $\ell \neq 0,1,2,\ldots$. This is rigorously stated in the following

THEOREM 1.2.2. *If* $f'(0) < 0$, $\ell \neq 0,1,2,\ldots$ *then there exists an* $\varepsilon_0 > 0$ *such that for all* $0 < \varepsilon \le \varepsilon_0$ *there is a unique solution* $y(x;\varepsilon)$ *of* (1.2.1)-(1.1.1.b), *which is uniformly bounded on* $[a,b]$.
Moreover, for any fixed $\delta > 0$,

$$(1.2.18) \qquad \lim_{\varepsilon\to 0}\ \max_{a+\delta<x<b-\delta} |y(x;\varepsilon)| = 0.$$

PROOF. See KREISS & PARTER [1974].

In the cases when $\ell = -g(0)/f'(0) = 0,1,2,\ldots$, a non-trivial turning point solution is possible:

$$(1.2.19) \qquad y_3(x) = \exp(\xi^2/4)\, D_\ell(\xi) = He_\ell(x\sqrt{-f'(0)/\varepsilon}).$$

The appearance of non-zero limit-solutions for $\varepsilon \to 0$, which can occur for discrete values of ℓ, is called the *resonance phenomenon*. The condition $\ell = 0,1,2,\ldots$ is necessary for resonance, however, it is not a sufficient condition. The class of functions f and g for which there are non-zero interior limit-solutions appears to be rather small (cf. KREISS & PARTER [1974]).

In the case of a turning point with $f'(0) < 0$ the effect of the right-hand side $s(x)$ in eq. (1.1.1) is not immediately clear. ABRAHAMSSON [1975] shows that the smoothness of $s(x)$ is a prerequisite for the solution $y(x;\varepsilon)$ to be uniformly bounded in ε. Compiling some of his results we state the following

<u>THEOREM 1.2.3</u>. *Let* y *be the solution of* (1.1.1), *let there be one turning point at* $x = 0$ *with* $f'(0) < 0$ *and let* $\ell = -g(0)/f'(0)$ *then there exist* K, δ *and* $\varepsilon_0 > 0$, *independent of* ε, *such that*
if $\ell < 0$ *then*

$$\max_{a \le x \le b} |y(x;\varepsilon)| \le K \max(|\alpha|, |\beta|, \max_{a \le x \le b} |s(x)|),$$

if $\ell > 0$, $\ell \ne 0,1,2,\ldots$, *then*

$$\max_{a \le x \le b} |y(x;\varepsilon)| \le K \max(|\alpha|, |\beta|, \max_{a \le x \le b} |s(x)|, \max_{|x|<\delta} |s^{(k)}|),$$

where k *is the nonnegative integer such that* $\ell < k < \ell+1$.

If s is not smooth enough, then y is possibly not bounded for $\varepsilon \to 0$. E.g. if y is the solution of

$$\varepsilon y'' - xy' + \ell y = \begin{cases} x^d & x \ge 0 \\ 0 & x \le 0 \end{cases};$$
$$d = 0,1,2,\ldots \;;\; \ell > 0,\; \ell \ne 0,1,2,\ldots ,$$
$$y(a) = \alpha,\; y(b) = \beta,$$

then

$$\max_{a \le x \le b} |y(x;\varepsilon)| = \mathcal{O}(\varepsilon^{\frac{d-\ell}{2}}) + \mathcal{O}(1).$$

If $\ell = 0,1,2,\ldots$ the solution may grow exponentially even for smooth functions s. E.g. let y be the solution of

$$\varepsilon y'' - xy' + \ell y = x^{\ell}, \qquad \ell = 0,1,2,\ldots ,$$
$$y(-1) = y(1) = 0 ,$$

then

$$y(x) = \mathcal{O}(e^{1/\varepsilon}) \qquad \text{for } x \in (a,b).$$

<u>REMARK</u>. When some numerical method for the problem (1.1.1) is used which approximates the right-hand side s by a function $\tilde{s}$ that is not smooth then we may not expect the corresponding approximation of y to be uniformly bounded for $\varepsilon \to 0$. E.g. if equation

$$\varepsilon y'' - xy' + \ell y = s$$

is solved numerically by a method which approximates the data s by $\tilde{s}$, such that $\tilde{s}$ has a discontinuous derivative at $x = 0$, then - if no further approximations are made - the approximate solution $\tilde{y}$ satisfies

$$\varepsilon \tilde{y}'' - x\tilde{y}' + \ell\tilde{y} = \tilde{s}$$

and we may have $\tilde{y} = \mathcal{O}(\varepsilon^{\frac{1-\ell}{2}})$.
Hence we can guarantee that $y(x;\varepsilon)$ is uniformly bounded only if $\ell < 1$. Similarly, if $\tilde{s}$ is discontinuous at $x = 0$ (e.g. if s is approximated by a step-function), then $y = \mathcal{O}(\varepsilon^{\ell/2})$ and $\tilde{y}$ will be uniformly bounded only for $\ell < 0$.

<u>REMARK</u>. For both $f'(0) > 0$ and $f'(0) < 0$ we notice that, by introducing the local coordinate $\xi = x/\sqrt{\varepsilon}$ in the turning point region, we can remove the singular perturbation character of equation (1.2.1). It is then converted into

$$\text{(1.2.20)} \qquad \frac{d^2y}{d\xi^2} + \frac{f(x)}{x}\,\xi\,\frac{dy}{d\xi} + g(x)y = 0.$$

For numerical purposes this implies that in a turning-point region of $\mathcal{O}(\sqrt{\varepsilon})$ no special methods for the problem are needed, provided that the mesh is sufficiently refined.

This approach may solve the problem for linear equations when an appropriate mesh can be generated after an a priori analysis, which locates the turning points and boundary layers. However, in general this will be a laborious process, especially when nonlinear equations are considered.

C. f identical to zero

In the case $f \equiv 0$, standard WKB-analysis of eq. (1.2.1) yields the approximation (cf. e.g. SIROVICH [1971])

$$(1.2.21) \qquad y(x) \sim g(x)^{-\frac{1}{4}} \exp(\pm \int^{x} \sqrt{-g(t)/\varepsilon}\, dt).$$

The character of the solution depends on the sign of g. If $g < 0$ the solution is exponential. There are boundary layers at either end of the interval $[a,b]$. Outside both boundary layers, which extend over a region $\mathcal{O}(\sqrt{\varepsilon})$, the solution is exponentially small.
If $g > 0$ the solution is oscillating with a period $2\pi\sqrt{\varepsilon/g}$. For the latter case it is evident that numerical approximation by means of piecewise approximation of the solution is not feasible for small values of ε.

1.3. EXAMPLES

In this section we collect a number of special cases of problem (1.1.1). They illustrate the exposition given in section 1.2. A number of these examples allow an explicit solution and, hence, are appropriate for use as model problems for numerical methods.
Equations without a turning point.

$$(1.3.1) \qquad \varepsilon y'' - y' = 0; \qquad y = A + B \exp(+x/\varepsilon).$$

$$(1.3.2) \qquad \varepsilon y'' + y' = 0; \qquad y = A + B \exp(-x/\varepsilon).$$

Equations with one turning point, $f'(0) > 0$.

$$(1.3.3) \qquad \varepsilon y'' + xy' = 0; \ \ell = 0; \ y = A + B\,\mathrm{erf}(x/\sqrt{2\varepsilon}).$$

$$(1.3.4) \qquad \varepsilon y'' + xy' + \tfrac{1}{2}y = 0; \ \ell = -\tfrac{1}{2}.$$

$$(1.3.5) \qquad \varepsilon y'' + xy' + y = 0; \ \ell = -1; y = \exp(-x^2/2\varepsilon)\ [A + B\int_0^x \exp(t^2/2\varepsilon)dt].$$

$$(1.3.6) \qquad \varepsilon y'' + xy' - \tfrac{1}{2}y = 0; \ \ell = \tfrac{1}{2}.$$

$$(1.3.7) \qquad \varepsilon y'' + xy' - y = 0; \ \ell = 1; \ y = Ax + B[\exp(-x^2/2\varepsilon) + \frac{x}{\varepsilon}\int_0^x \exp(-t^2/2\varepsilon)dt].$$

$$(1.3.8) \qquad \varepsilon y'' + xy' - 2y = 0; \ \ell = 2; \ y = A[x\varepsilon \exp(-x^2/2\varepsilon) + $$
$$+ (x^2+\varepsilon)(B + \int_0^x \exp(-t^2/2\varepsilon)dt].$$

Equations with one turning point, $f'(0) < 0$.

(1.3.9) $\varepsilon y'' - xy' = 0;\ \ell = 0;\ y = A + B\int_0^x \exp(t^2/2\varepsilon)\,dt.$

(1.3.10) $\varepsilon y'' - xy' + y = 0;\ \ell = 1;\ y = Ax + B[\exp(x^2/2\varepsilon) - \frac{x}{\varepsilon}\int_0^x \exp(t^2/2\varepsilon)\,dt].$

(1.3.11) $\varepsilon y'' - xy' - y = 0;\ \ell = -1;\ y = \exp(x^2/2\varepsilon)[A + B\int_0^x \exp(-t^2/2\varepsilon)\,dt].$

Equations with $f \equiv 0$.

(1.3.12) $\varepsilon y'' - y = 0$; $y = A\exp(-x/\sqrt{\varepsilon}) + B\exp(x/\sqrt{\varepsilon})$.

(1.3.13) $\varepsilon y'' + y = 0$; $y = A\sin(x/\sqrt{\varepsilon}) + B\cos(x/\sqrt{\varepsilon})$.

Equation with a classical turning point.

(1.3.14) $\varepsilon y'' - xy = 0$; $y = A\,\mathrm{Ai}(x\varepsilon^{-1/3}) + B\,\mathrm{Bi}(x\varepsilon^{-1/3})$.

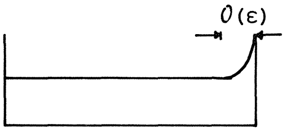


Fig. 1.3.1 $\varepsilon y'' - y' = 0$

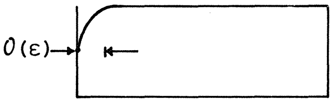


Fig. 1.3.2 $\varepsilon y'' + y' = 0$

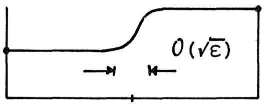


Fig. 1.3.3 $\varepsilon y'' + xy' = 0$
Shock layer

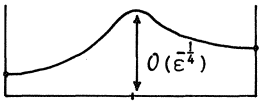


Fig. 1.3.4 $\varepsilon y'' + xy' + 0.5y = 0$

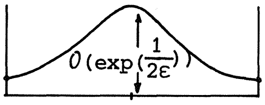


Fig. 1.3.5 $\varepsilon y'' + xy' + y = 0$

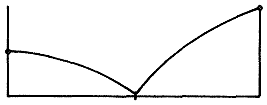


Fig. 1.3.6 $\varepsilon y'' + xy' - 0.5y = 0$
Cusp layer

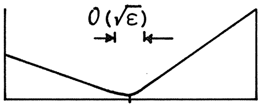


Fig. 1.3.7 $\varepsilon y''+xy'-y=0$
Corner layer

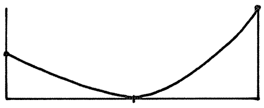


Fig. 1.3.8 $\varepsilon y''+xy'-2y=0$

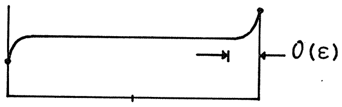


Fig. 1.3.9a
$\varepsilon y''-xy'=0, \ |a| = |b|$

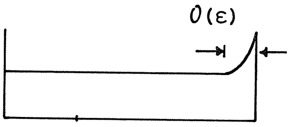


Fig. 1.3.9b
$\varepsilon y''-xy'=0, \ |a| < |b|$

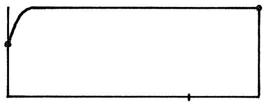


Fig. 1.3.9c
$\varepsilon y''-xy'=0, \quad |a| > |b|$

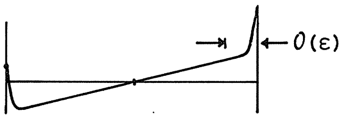


Fig. 1.3.10 $\varepsilon y''-xy'+y=0$
Resonance

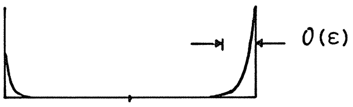


Fig. 1.3.11 $\varepsilon y''-xy'-y=0$

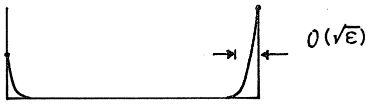


Fig. 1.3.12 $\varepsilon y''-y=0$

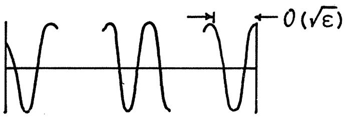


Fig. 1.3.13 $\varepsilon y''+y=0$

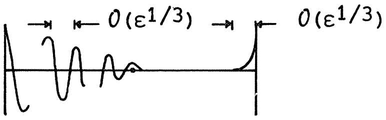


Fig. 1.3.14 $\varepsilon y''-xy=0$

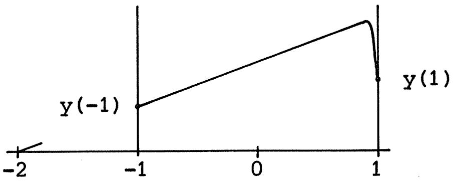


Fig. 1.3.15
$\varepsilon y''-x(2+x)y'+xy = 0$

The general solution of the differential equation

$$(1.3.15)\quad \varepsilon y'' - cxy' + c\ell y = 0$$

with constant coefficients ℓ and c ($c=+1$ or $c=-1$) is described by the parabolic cylinder function $D_\ell(z)$;
if $\ell \neq 0,1,2,\ldots$

$$(1.3.16)\quad y(x) = e^{cx^2/4\varepsilon}\left[A\, D_\ell(x\sqrt{c/\varepsilon}) + B\, D_\ell(-x\sqrt{c/\varepsilon}\right]$$

if $\ell = 0,1,2,\ldots$ $D_\ell(x\sqrt{c/\varepsilon})$ and $D_\ell(-x\sqrt{c/\varepsilon})$ are linearly dependent and an independent solution is given by $D_{-\ell-1}(x\sqrt{-c/\varepsilon})$. In this case we can also write

$$(1.3.17)\quad y(x) = A\, Hh_n(x\sqrt{-c/\varepsilon}) + B\, Hh_n(-x\sqrt{-c/\varepsilon}),$$

where Hh_n is the "probability" function (ABRAMOWITZ & STEGUN [1965])

$$Hh_n(z) = \frac{1}{n!}\int_z^\infty (t-z)^n e^{-t^2/2} dt.$$

When y is subjected to the boundary conditions $y(-1) = \alpha$, $y(1) = \beta$, with the aid of the asymptotic expressions (1.2.16) and (1.2.17), the following asymptotic approximations to y are obtained (O'MALLEY [1970]).
If $c < 0$, $\ell \neq -1,-2,\ldots,$

$$(1.3.18)\quad \begin{aligned} y(x;\varepsilon) &\sim y(\mathrm{sign}(x))|x|^\ell, \\ y(0;\varepsilon) &= O(\varepsilon^{\ell/2}). \end{aligned}$$

If $c < 0$, $\ell = -1,-2,\ldots,$

$$(1.3.19)\quad \begin{aligned} y(x;\varepsilon) &\sim \frac{\beta+(-)^\ell\alpha}{2}\, x^\ell + \frac{\beta-(-)^\ell\alpha}{2}\quad x^{-\ell-1}\; e^{-c(1-x^2)/2\varepsilon}, \\ y(0;\varepsilon) &= O(\exp(-c/2\varepsilon)). \end{aligned}$$

If $c > 0$, $\ell \neq 0,1,2,\ldots,$

$$(1.3.20)\quad \begin{aligned} y(x;\varepsilon) &\sim y(\mathrm{sign}(x))|x|^{-\ell-1} e^{-c(1-x^2)/2\varepsilon}, \\ y(0;\varepsilon) &\text{ is exponentially small.} \end{aligned}$$

If $c > 0$, $\ell = 0,1,2,\ldots$, the asymptotic behaviour is also given by eq. (1.3.19) and $y(0;\varepsilon) = \mathcal{O}(\varepsilon^{\ell/2})$, unless $\alpha = (-1)^{\ell-1}\beta$ or $\ell = 1,3,5,\ldots$.

The class of equations for which the resonance phenomemon can occur is very small. However, the condition that the equation can be reduced to

$$\varepsilon y'' - xy' + ny = 0$$

is overly restrictive. This is demonstrated by the following two examples.

EXAMPLE. The equation (cf. DORR [1970b])

$$\varepsilon y'' - xc(x)y' = 0, \quad c(x) > 0 \text{ on } [a,b], \tag{1.3.21}$$

is easily integrated to obtain

$$y(x) = A + B \int_0^x \exp \int_0^t \frac{\xi c(\xi)}{\varepsilon}\, d\xi dt.$$

Thus, there exists a constant C such that

$$\lim_{\varepsilon\to 0} \max_{a+\delta<x<b-\delta} |y(x;\varepsilon)-C| = 0, \qquad \delta > 0.$$

EXAMPLE. The equation

$$\varepsilon y'' - x(2+x)y' + xy = 0 \quad \text{on } [-1,1] \tag{1.3.22}$$

has a solution (fig. 3.1.15)

$$y(x;\varepsilon) = \sigma(2+x),$$

which is also a solution of the reduced problem. In this case the following statement holds

$$\exists\, \sigma \lim_{\varepsilon\to 0} \max_{-1+\delta<x<1-\delta} |y(x;\varepsilon) - \sigma(2+x)| = 0.$$

According to KREISS & PARTER [1974, thm.2.2] there may exist a boundary layer at the righthand end and σ is determined by $\sigma = y(-1)$.

CHAPTER II

DIFFERENCE METHODS

In this second chapter we treat topics that are basic for the study of the numerical solution of singular perturbation problems. In section 1 we discuss the effect of some analytical transformations and the trouble when standard type discretizations are used. We also briefly consider the application of shooting. In section 2 we discuss methods for representing the numerical approximation of the solution of problem (1.1.1) and we formulate a number of properties that are desirable for methods for solving such problems. In section 3 we give a concise review of the numerical methods that have been used already to solve singular perturbation problems. In section 4 we concentrate on finite difference methods that use exponential fitting and we discuss the features that make these methods interesting for the solution of stiff boundary-value problems.

2.1. INITIAL CONSIDERATIONS

Before we treat new finite difference methods that are specially designed for solving singular perturbation problems, we will show how a number of commonly used discretization methods (forward, backward and central differences) behave when applied with a uniform mesh. This will demonstrate what the problems are and what we should strive for. We also explain why analytical transformations (the integrating factor and the Liouville transformations) are of little use. Finally we show the defects of the shooting technique when applied to problem (1.1.1).

Simple finite difference methods

With the help of a classical example we demonstrate what difficulties may arise when singular perturbation problems are solved by methods that are commonly used. Let us consider the boundary-value problem

$$\begin{aligned} &\varepsilon y'' + y' = 0 \\ (2.1.1)\qquad & \\ &y(0) = 0, \quad y(1) = 1. \end{aligned}$$

In order to compare the solution of this boundary-value problem with the treatment given below we cast it into the form

$$(2.1.2)\qquad y(x) = \frac{1-\nu^{x/h}}{1-\nu^{1/h}}\,,$$

where $h > 0$ and $\nu = \exp(-h/\varepsilon)$. This solution has a boundary layer of thickness $\mathcal{O}(\varepsilon)$ near $x = 0$ and the limit-solution for $\varepsilon \to 0$ is

$$(2.1.3)\qquad \lim_{\varepsilon\to 0} y(x;\varepsilon) = 1 \quad \text{on } [\delta,1],$$

for $\delta > 0$.

We compute the numerical approximation on a set of equally spaced mesh-points $\{x_i\}_{i = 0,\dots,N}$, defined by

$$x_i = ih = i/N, \qquad i = 0,1,\dots,N.$$

On this mesh we seek an approximation y_i to $y(x_i)$ by three distinct difference methods. Successively we use the i) *central difference*, ii) *backward difference* and iii) *forward difference approximation* for representing the first derivative. In particular, we are interested in the approximate solutions for small values of ε, i.e. $\varepsilon \ll h$.

1. Central differences

Here we replace the differential equation (2.1.1) by the difference equation

$$(2.1.4)\qquad \varepsilon\,(y_{i+1}-2y_i+y_{i-1})/h^2 + (y_{i+1}-y_{i-1})/(2h) = 0, \quad i = 1,2,\dots,N-1.$$

With the additional conditions

$$(2.1.5)\qquad y_0 = 0, \qquad y_N = 1,$$

the solution reads

$$(2.1.6)\qquad y_i = \frac{1-\mu_0^i}{1-\mu_0^N}\,, \quad \mu_0 = \frac{2\varepsilon-h}{2\varepsilon+h}\,.$$

We notice that this solution has the same form as (2.1.2), where ν has been replaced by μ_0. Because $|\mu_0-\nu| = \mathcal{O}((\frac{h}{\varepsilon})^3)$ for $\frac{h}{\varepsilon} \to 0$, it is clear that (2.1.6) is a reasonably close approximation to (2.1.2) if $h \ll \varepsilon$. However, the approximation fails completely for $\varepsilon < h$. In particular, for h fixed and $\varepsilon \to 0$ there is no resemblance at all between $y(x_i)$ and y_i, since

$$\text{(2.1.7)}\qquad \begin{aligned} \lim_{\varepsilon\to 0} y_i &= 0 && \text{if i even, N odd; } i = 1,\dots,N-1;\\ &= 1 && \text{if i odd, N odd;}\\ &= i/N && \text{if i even, N even;}\\ &= \infty && \text{if i odd, N even.}\end{aligned}$$

2. Backward differences

Now we replace (2.1.1) by the difference equation

$$\text{(2.1.8)}\qquad \varepsilon\,(y_{i+1}-2y_i+y_{i-1})/h^2 + (y_i-y_{i-1})/h = 0, \quad i = 1,2,\dots,N-1.$$

The solution of the equations (2.1.8) and (2.1.5) is

$$\text{(2.1.9)}\qquad y_i = \frac{1-\mu_1^i}{1-\mu_1^N}, \qquad \mu_1 = \frac{\varepsilon-h}{\varepsilon}.$$

For small values of $\frac{h}{\varepsilon}$ this approximation is less accurate than (2.1.6) since $|\mu_1-\nu| = \mathcal{O}((\frac{h}{\varepsilon})^2)$ for $\frac{h}{\varepsilon} \to 0$. Here again, the approximation completely breaks down for $\varepsilon < h$. For a fixed $h > 0$,

$$\text{(2.1.10)}\qquad \lim_{\varepsilon\to 0} y_i = 0 \quad \text{for } i = 1,2,\dots,N-1.$$

This is not at all an approximation to the limit-solution of the original equation.

3. Forward differences

If the equation (2.1.1) is replaced by

$$\text{(2.1.11)}\qquad \varepsilon(y_{i+1}-2y_i+y_{i-1})/h^2 + (y_{i+1}-y_i)/h = 0, \quad i = 1,2,\dots,N-1,$$

the solution of the difference equation is

(2.1.12) $$y_i = \frac{1-\mu_2^i}{1-\mu_2^N}, \qquad \mu_2 = \frac{\varepsilon}{\varepsilon+h}.$$

Again, for small values of $\frac{h}{\varepsilon}$, y_i approximates $y(i/N)$ and $|\mu_2-\nu| = \mathcal{O}((\frac{h}{\varepsilon})^2)$ for $\frac{h}{\varepsilon} \to 0$. However, in this case, for $\varepsilon \to 0$ the asymptotic behaviour of (2.1.2) is reflected in the approximation, since for fixed h

(2.1.13) $$\lim_{\varepsilon\to 0} y_i = 1 \qquad \text{for } i = 1,2,\ldots,N-1.$$

We conclude that for $h << \varepsilon$ central differences are the most accurate, but forward differences have the property that for $\varepsilon \to 0$ the discrete limit-solution approximates the exact limit-solution. Clearly, this is an important feature if we want to solve the equations with $\varepsilon << h$. Nevertheless, we note that the rate of decay in the boundary-layer is not very well represented since $\exp(-h/\varepsilon) << \frac{\varepsilon}{\varepsilon+h}$ if $\varepsilon << h$.

Another equation

Let us consider the boundary-value problem

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

What happens to this differential equation, in which no first derivative is present, when it is discretized by the common 3-point difference formula? For any $h > 0$, the analytical solution of the boundary-value problem can be written

(2.1.14) $$y(x) = \frac{\nu^{(-2x+1)/2h} + \nu^{(2x-1)/2h}}{\nu^{1/2h} + \nu^{-1/2h}}$$

where

$$\nu = \exp(h/\sqrt{\varepsilon}).$$

The limit solution for $\varepsilon \to 0$ is

$$\lim_{\varepsilon\to 0} y(x;\varepsilon) = 0 \qquad \text{on } [\delta,1-\delta]$$

for $\delta > 0$.

If we replace the differential equation by the difference equation

$$\varepsilon\ (y_{i+1}-2y_i+y_{i-1})/h^2 - y_i = 0, \quad i = 1,2,\ldots,N-1,$$

with the additional conditions

$$y_0 = 1, \quad y_N = 1,$$

the solution reads

$$y_i = \frac{\mu^{-i+N/2} + \mu^{i-N/2}}{\mu^{N/2} + \mu^{-N/2}}, \tag{2.1.15}$$

where

$$\mu = (1 + \frac{h^2}{2\varepsilon}) + \sqrt{(1 + \frac{h^2}{2\varepsilon})^2 - 1}.$$

Again we see that both solutions (2.1.14) and (2.1.15) are of the same form. If $h << \sqrt{\varepsilon}$ a good approximation is obtained:

$$|\mu-\nu| = \mathcal{O}((h/\sqrt{\varepsilon})^3) \text{ for } \frac{h}{\sqrt{\varepsilon}} \to 0.$$

If h is fixed and $\varepsilon \to 0$ then $\mu = \frac{h^2}{\varepsilon} + 2 - \frac{\varepsilon}{h^2} + \ldots$ and

$$\lim_{\varepsilon\to 0} y_i = 0\,, \quad i = 1,2,\ldots,N-1.$$

In other words, for $\varepsilon \to 0$, the limit-solution of the discrete problem is similar to that of the continuous problem; but, again, the rate of decay in the boundary-layers is not accurately represented.

The use of analytic transformations

In the preceding examples we started with the differential equation in its canonical form (1.1.1). But also, if we use other forms such as (1.1.2) or (1.1.9), it turns out that these cannot be of great help in removing the problems related to the smallness of ε. For instance, if we apply the *integrating factor* (1.1.3), equation (2.1.1) is transformed into

(2.1.16) $$(\exp(x/\varepsilon)y')' = 0.$$

Replacing u' by the central difference $(u_{i+1/2}-u_{i-1/2})/h$, we obtain the discrete form of (2.1.16)

(2.1.17) $$\exp(x_{i-1/2}/\varepsilon)y_{i-1} - (\exp(x_{i-1/2}/\varepsilon) + \exp(x_{i+1/2}/\varepsilon))\, y_i + \\ + \exp(x_{i+1/2}/\varepsilon)y_{i+1} = 0, \quad i = 1,2,\dots,N-1.$$

With the boundary conditions (2.1.5), this yields exactly the analytic solution at the mesh-points, i.e.

(2.1.18) $$y_i = \frac{1-\mu^i}{1-\mu^N}, \qquad \mu = \exp(-h/\varepsilon)$$

This seems an excellent discretization method; however (2.1.17) cannot be used in practice since the value $\exp(x_{i\pm 1/2}/\varepsilon)$ will cause overflow, even for values of ε that are not extremely small. Moreover, for equations (1.1.2) with $g \not\equiv 0$ or $s \not\equiv 0$ the discrete equations are

(2.1.19) $$\frac{\varepsilon}{h^2}[\exp(\frac{-hf_i}{2\varepsilon})y_{i-1}-(\exp(\frac{-hf_i}{2\varepsilon})+\exp(\frac{hf_i}{2\varepsilon}))y_i+\exp(\frac{hf_i}{2\varepsilon})y_{i+1}] + g_iy_i = s_i,$$

where $f_i = f(x_i)$, $g_i = g(x_i)$ and $s_i = s(x_i)$. This shows that, for small $|\frac{\varepsilon}{hf_i}|$, the terms g_iy_i and s_i are cancelled by the large term $\varepsilon \exp(|\frac{hf_i}{2\varepsilon}|)$.

Also the *Liouville transformation* (eq. (1.1.8)) is not very useful for computational purposes. The boundary conditions for z and y are related by

(2.1.20) $$\frac{z(b)}{z(a)} = \frac{y(b)}{y(a)} \exp \int_a^b \frac{f(t)}{2\varepsilon}\, dt.$$

This means that the boundary conditions (and equally the right hand side of the equation) are exponentially enlarged by the transformation and hence overflow problems arise. More generally, we can say that by the transformation of the original problem, the (assumed) smooth coefficients f,g and s are replaced by rapidly varying coefficients. This is frequently a disadvantage for numerical purposes.

The shooting method

For the shooting method, a boundary-value problem is rewritten as an initial-value problem (i.v.p.). For the most elementary form of shooting, the homogeneous problem (1.2.1)-(1.1.1.b) is written

$$(2.1.21)\qquad \begin{cases} y' = v, & y(a) = \alpha, \\ v' = -\frac{g}{\varepsilon} y - \frac{f}{\varepsilon} v, & v(a) = p, \end{cases}$$

where the initial value $p = v(a)$ is an unknown parameter. This parameter has to be determined such that the solution (y,v) satisfies the boundary condition at the other end $x = b$. Variants of the shooting method are possible, such as

- starting with the boundary condition at $x = b$ and solving the initial-value problem from b to a;
- starting from both ends and matching the solution at an intermediate point in (a,b);
- introducing a partition $\bigcup_{i=1}^{N} [x_{i-1}, x_i]$ of $[a,b]$, solving the i.v.p. on each subinterval and matching the continuity conditions at each point x_i (*multiple shooting*).

Thus, the method essentially consists of two parts: A. the solution of the initial value problem(s); B. the determination of the unknown parameter(s). (See also K.G. GUDERLEY [1975].)

Both parts introduce numerical trouble when the problem (1.1.1) is solved for small ε.

A. The solution of the initial value problem.

Let us consider problems of type (1.2.1), with $g < 0$. These problems are called stable, because the solution is bounded by the data. For these problems the Jacobian matrix of the i.v.p. (2.1.21) has two eigenvalues, which are approximately $-f/\varepsilon$ and $-g/f$. By switching the direction of the i.v.p. both eigenvalues change sign. In both cases we have to solve an i.v.p. with a positive and a negative eigenvalue. The stable boundary-value problem has been converted into an unstable i.v.p.. Moreover, if the i.v.p. is solved in the direction in which the reduced problem has to be solved, the eigenvalue with larger absolute value is positive, i.e. an exponentially large erroneous component is introduced in (y,v).

B. The determination of the parameters.

The erroneous component of the system makes the equations that have to be solved for the determination of the parameters, very ill-conditioned. We show this by means of example (1.3.3) on the interval [-1,+1].

We apply shooting, starting from -1 and +1 and matching at x = 0. The initial conditions are

$$\begin{aligned} y(-1) &= \alpha\ , & y(+1) &= \beta, \\ v(-1) &= p_1, & v(+1) &= p_2. \end{aligned}$$

We assume that the integration method yields an exact solution (1.3.3) to the i.v.p.'s both from -1 to -0 and from +1 to +0.
Then

$$\begin{pmatrix} y(-0) \\ v(-0) \end{pmatrix} = \begin{pmatrix} 1 & \exp(\frac{1}{2\varepsilon}) \int_0^1 \exp(-t^2/2\varepsilon)\,dt \\ 0 & \exp(\frac{1}{2\varepsilon}) \end{pmatrix} \begin{pmatrix} \alpha \\ p_1 \end{pmatrix}$$

and

$$\begin{pmatrix} y(+0) \\ v(+0) \end{pmatrix} = \begin{pmatrix} 1 & -\exp(\frac{1}{2\varepsilon}) \int_0^1 \exp(-t^2/2\varepsilon)\,dt \\ 0 & \exp(\frac{1}{2\varepsilon}) \end{pmatrix} \begin{pmatrix} \beta \\ p_2 \end{pmatrix}$$

For small values of ε,

$$1 << \exp(\frac{1}{2\varepsilon}) \int_0^1 \exp(-t^2/2\varepsilon)\,dt\ .$$

Therefore, since we require,

$$y(-0) = y(+0), \qquad v(-0) = v(+0),$$

the numerical solution of p_1 and p_2 yields

$$p_1 \doteq 0.0 \doteq p_2$$

and the shooting process does not converge.

The same problems arise when symmetric problems (i.e. with f≡0) are solved. E.g, consider example (1.3.12) on [a,b]. When shooting from a to b is applied, because of the large eigenvalue $\varepsilon^{-1/2}$, a negligible alteration

δ of $y'(a)$ will cause the exponentially large deviation $\delta\sqrt{\varepsilon}\sinh((b-a)/\sqrt{\varepsilon})$ in the computed value of $y(b)$. Multiple shooting over a partition suffers from the same defect: each deviation δ in the guess of $y'(x_i)$ causes a difference $\delta\sqrt{\varepsilon}\sinh((x_{i+1}-x_i)/\sqrt{\varepsilon})$ in the computed value $y(x_{i+1})$. Even this difference will be unmanageably large if $\varepsilon << (x_{i+1}-x_i)^2$. Since the problem is symmetric, reverse shooting does not help. This is the reason why we conclude that (multiple) shooting is an inadequate technique for solving singular perturbation problems of the form (1.1.1).

2.2. REPRESENTATION OF A SOLUTION AND ERROR NORMS FOR AN APPROXIMATION

In this section we discuss what criteria can be applied in order to judge the qualities of a numerical solution of

$$(2.2.1) \qquad \begin{aligned} &Ly \equiv \varepsilon y'' + fy' + gy = s, \\ &y(a) = \alpha, \quad y(b) = \beta, \end{aligned}$$

and we formulate requirements that can be imposed on methods suitable for singular perturbation problems. The choice of criteria for an approximation is a general question which, in fact, forms part of the proper statement of most numerical problems. However, when singular perturbation problems are solved, this deserves our special attention because of their multiscale character.

Representation of a solution

Since the solution of the two-point boundary-value problem (2.2.1) is a function of a real variable, its numerical approximation is given by only a finite number of real numbers. So we are faced with the problem of how we should represent the numerical solution. Generally, this is done in one of the following ways:

1. Given a finite set of *knots* (or *gridpoints*) $\{x_i\}_{i=0}^N$, $a \le x_i \le b$, the value of the solution at each of these points is approximated (*pointwise approximation*). If an approximation is required at other points, it can be obtained by interpolation.
2. Given a set of functions $\{\phi_i\}_{i=1}^N$, defined on $[a,b]$, the solution is approximated by a linear combination of functions ϕ_i (*global approximation*).

In both cases we have to define a suitable measure to quantify the error between the true solution and its numerical approximation.

The norm in which the error is measured may differ from case to case and it should be chosen in accordance with the particular requirements imposed on the approximation. These requirements form part of the proper definition of a numerical problem. For some applications it will be necessary to obtain a solution that is accurate at a number of points specified in advance, for other applications one has to obtain a solution whose error is bounded by a small amount over the whole interval. Also other criteria for a good approximation are possible. All these different criteria lead to the introduction of different norms in which the error between the solution and the approximation can be expressed quantitatively.

Norms for the approximation

We introduce norms both for pointwise and global approximations. Let $\Pi = \{a=x_0<x_1<\ldots<x_N=b\}$ be a given partition of the interval [a,b]. Norms for the *pointwise error* on Π are directly related to vector norms. We define

$$\|y - y_{app}\|_{\pi,1} = \sum_{x_i\in\Pi} |y_i - y(x_i)|, \tag{2.2.2}$$

$$\|y - y_{app}\|_{\pi,2} = \left\{\sum_{x_i\in\Pi} (y_i - y(x_i))^2\right\}^{1/2}, \tag{2.2.3}$$

$$\|y - y_{app}\|_{\pi,\infty} = \max_{x_i\in\Pi} |y_i - y(x_i)|. \tag{2.2.4}$$

Here y is the exact solution and y_i denotes the value of the pointwise approximation y_{app} to the value $y(x_i)$. The pointwise error norms depend crucially on the choice of the knots in Π. This set is not necessarily the set of all points for which an approximate value is available after the computational process; it may be only a subset.

DEFINITION

A numerical approximation y_{app} is called *pointwise exact* on a grid Π, if $\|y - y_{app}\|_{\pi,1} = 0$.

Norms for the *global error* are related to the norms in the function spaces $L^2(a,b)$ and $L^\infty(a,b)$.

$$(2.2.5) \qquad \|y - y_{app}\|_{0,2} = \left\{ \int_a^b (y(z)-y_{app}(z))^2 dz \right\}^{1/2},$$

$$(2.2.6) \qquad \|y - y_{app}\|_{0,\infty} = \max_{a\le x\le b} |y(x) - y_{app}(x)|.$$

Using global approximation, it is also possible to measure the norm of the residual

$$(2.2.7) \qquad \|Ly_{app} - s\|_{0,2} = \left\{ \int_a^b (Ly_{app}-s)^2 dx \right\}^{1/2}.$$

If L is a positive definite, self-adjoint operator, then the energy-norm can also be used

$$(2.2.8) \qquad \|y - y_{app}\|_{E} = \left\{ \int_a^b (y_{app}-y)L(y_{app}-y) dx \right\}^{1/2}.$$

In particular, in the case of a singular perturbation problem, where the solution is a smooth function which locally may change rapidly, the choice of an appropriate error-norm demands care. Here we meet the question whether or not a solution should be accurately represented in *all* regions. The global norm $\|\cdot\|_{0,\infty}$ is appropriate if a good representation of the rapidly varying part is required, and the norm $\|\cdot\|_{0,2}$ if the rapidly varying part of the solution may be roughly represented as long as this does not affect the global behaviour. Nevertheless, to approximate an almost discontinuous solution by a smooth approximation, all global norms require a fine mesh in the neighbourhood of the discontinuity.

For our purposes, here and in the following chapters, we will mainly concentrate on the pointwise error-norm $\|y - y_{app}\|_{\pi,\infty}$, for some arbitrary, but a priori specified, finite set of knots Π. Here we emphasize again that the choice of Π forms part of the proper definition of the numerical problem. By the choice of Π, we decide whether or not we are interested in accurate approximation in particular parts of $[a,b]$. This corresponds to the fact that a large number of gridpoints is required if an accurate approximation in the non-smooth part of the solution is required.

In accordance with this choice of error-norm we represent the computed solution by a sequence of discrete function-values $\{y_i\}$, *corresponding to a given sequence of grid-points* Π.

Desirable features in methods for solving singular perturbation problems

At this point we have to discuss what the desirable features are for a method that is used for singular perturbation problems. In the first place we require that the approximation is - to a certain extent - accurate for an *arbitrary* choice of Π, provided that the global character of the solution can be represented by some interpolation between the gridpoints. In particular we want to obtain a reasonable accuracy with any equidistant mesh which is fine enough to represent the slowly varying parts of the solution. Apart from this, Π may be chosen in such a way that there are parts of the mesh where $x_i - x_{i-1} << \varepsilon$, $x_i - x_{i-1} \approx \varepsilon$, $x_i - x_{i-1} \approx \varepsilon^{1/2}$, $x_i - x_{i-1} >> \varepsilon$ etc.. However, if some local, rapidly varying behaviour (say, between two neighbouring points of Π) is completely missed by the numerical representation on Π, the global numerical solution should be disturbed as little as possible to obtain a small error $\|y - y_{app}\|_{\pi,\infty}$. In general, interpolation fails in the rapidly varying parts. If an accurate representation is wanted in these parts, the set of gridpoints Π should be chosen appropriately.

In order to discuss methods for singular perturbations more rigorously, we formulate some useful properties in the following definitions. We consider the two-point boundary-value problem (2.2.1). We assume that a unique solution y_ε exists for all ε, $0 < \varepsilon \le \varepsilon_0$, and also a *solution* y_0 *of the reduced problem* on $[c,d] \subset [a,b]$ such that

$$\lim_{\varepsilon\to 0} y_\varepsilon(x) = y_0(x)$$

uniformly on $[c,d]$.

For each ε we consider a family of discrete solutions $\{y_{\pi,\varepsilon}\}$ for different partitions $\Pi = \{a=x_0<x_1<x_2<\ldots<x_N=b\}$ of the interval $[a,b]$. To each Π we associate a meshwidth $h = \max_{i=1,\ldots,N} (x_i-x_{i-1})$. Less accurately we sometimes write $y_{h,\varepsilon}$ instead of $y_{\pi,\varepsilon}$.

DEFINITION

A family of discrete functions $\{u_h\}$ has a *limit-function* for $h \to 0$, $U(x)$, defined on an interval $[c,d] \subseteq [a,b]$ and denoted by

$$U = \lim_{h\to 0} u_h,$$

if there exists a continuous function $U(x)$ on $[c,d]$ that satisfies

$$\lim_{i\to\infty} u_{\Pi_i}(x) = U(x)$$

uniformly for all $x \in [c,d] \cap \bigcup_{i=1}^{\infty} \Pi_i$, for any sequence $\{\Pi_i\}_{i=1}^{\infty}$ of partitions of $[a,b]$ with the properties $\lim_{i\to\infty} h_i = 0$ and $i > j \Rightarrow \Pi_i \supset \Pi_j$.

We assume that, for a particular problem and a particular method, a discrete solution $y_{h,\varepsilon}$ exists for all h and ε, $0 < h \leq h_0$, $0 < \varepsilon \leq \varepsilon_0$. We want to show, that for particular methods, the asymptotic behaviour of the finite difference solution for small ε closely approximates that of the continuous solution even without $h \to 0$. To this end we introduce the following definitions.

DEFINITION

A method is *uniformly ε-convergent* of order p (for some class of problems) if there exist constants ε_0 and K, independent of ε, such that

$$\sup_{0<\varepsilon\leq\varepsilon_0} \|y_{h,\varepsilon} - y_\varepsilon\|_{\pi,\infty} \leq Kh^p$$

DEFINITION

A method is *uniformly ε-stable* (for some class of operators L) if there exists a constant K, independent of ε, s, α and β, such that

$$\|y_{h,\varepsilon}\|_{\pi,\infty} \leq K \max(|\alpha|,|\beta|, \max_{a\leq x\leq b}(|s(x)|))$$

for all $0 < \varepsilon \leq \varepsilon_0$, $0 < h \leq h_0$.

DEFINITION

For some class of numerical problems, a method has a *discrete limit solution* $y_{h,0}$ if there exists a discrete function $y_{h,0}$ such that

$$\lim_{\varepsilon\to 0} y_{h,\varepsilon} = y_{h,0}.$$

DEFINITION

For some class of problems a method is called *consistent with the reduced problem* on an interval $[c,d] \subset [a,b]$ if

$$\lim_{h\to 0} y_{h,0} = y_0 \qquad \text{on } [c,d],$$

where y_0 denotes the solution of the reduced problem on $[c,d]$.

As we saw in the examples of section 2.1, it is very useful to be able to fix Π and let $\varepsilon \to 0$ and then to have some assurance that the asymptotic behaviour of the continuous solution is reflected in the discrete solution. If a method is consistent with the solution of the reduced problem, this implies that even in the case where rapidly varying solutions cannot be represented at all on Π, the solution can be accurate in the norm $\|y - y_{app}\|_{\pi,\infty}$ asymptotically for $\varepsilon \to 0$. This includes the possibility of solutions that improve rather than degrade when $\varepsilon \to 0$.

2.3. EXISTING ALGORITHMS FOR THE SOLUTION OF SINGULAR PERTURBATION PROBLEMS

Only a few papers are concerned with the numerical solution of singular perturbation problems, although it seems to be a field of increasing interest. PEARSON [1968a] uses central differences on a non-uniform mesh. He gives the numerical solution of a great variety of singular perturbation problems of the type (1.1.1). IL'IN [1969] introduces the idea of widening the boundary-layer. He uses an equidistant mesh and attains ε-uniform convergence and stability for a limited class of problems. DORR [1970a] uses directional one-sided differences for a particular system of two nonlinear equations and he also gives an extensive discussion for some turning point problems. KREISS [1973,1974] and ABRAHAMSSON et al. [1974] discuss a system of linear equations, without turning points. They use directional one-sided differences or widen the boundary layer. They also prove ε-uniform stability on a uniform net. In this section we treat the essential facts from the above-mentioned papers.

Pearson's algorithm

PEARSON [1968a] gives the first description of a method for the numerical solution of the two-point boundary-value problem (1.1.1). His method is based on the classical 3-point finite difference formula for a non-uniform mesh. Given a partition of the interval $[a,b]$, $a = x_0 < x_1 < \ldots < x_N = b$, this leads to the difference equations

$$(2.3.1)\qquad 2\varepsilon\,\frac{q_j(y_{j+1}-y_j)-p_j(y_j-y_{j-1})}{p_jq_j(p_j+q_j)}+f(x_j)\,\frac{q_j^2(y_{j+1}-y_j)+p_j^2(y_j-y_{j-1})}{p_jq_j(p_j+q_j)}+$$
$$+\,g(x_j)y_j = s(x_j),\qquad j = 1,\ldots,N-1,$$

where $p_j = x_{j+1} - x_j$, $q_j = x_j - x_{j-1}$.

In order that the solution of the difference equation approximates the solution of the differential equation, the mesh should be properly chosen. To this end the mesh spacing is iteratively adjusted such that there is a high density of meshpoints in the regions where y(x) is changing rapidly. Several thousands of meshpoints are used and some simple criterion is chosen for the distribution of the meshpoints, e.g.

$$|y_{j+1} - y_j| \approx 10^{-3} \max_{k,l} |y_k - y_l|.$$

Occasionally, meshpoints had to be added according to a criterion involving steepness in y' rather than steepness in y to obtain accurate solutions. In addition, a mesh smoothing was necessary to ensure that there was no abrupt change in mesh interval size.

Since we saw in section 2.1 that the 3-point scheme may fail completely on a uniform mesh if $\varepsilon << h$, it is quite clear that the adjustment of the mesh, to take into account the effects of small ε, is essential when using this scheme. In order to ensure that the meshpoint set is dense in the right places, the whole process is executed in ε-steps. The process is started with a uniform mesh and with a modest value of ε. The meshpoint set used at the completion of the preceding ε-step forms the initial set for the new step with a smaller ε.

REMARK. This strategy is an application of the *Davidenko-principle*: for modest values of ε, the problem is readily solved. Using the information about this solution, the problem is solved for other values of ε, for which the problem could not be solved before.

Pearson reports that a large number of problems have been solved by his method on a CDC 6600 computer and problems with values of ε as small as 10^{-10} could be solved using single precision arithmetic. The results were found to be accurate to 3 or 4 significant digits. However, the use of the Davidenko principle, together with the iterative adjustment of the mesh in each ε-step, makes the method quite laborious, even for linear problems. Moreover, the large number of meshpoints and the relatively low accuracy gives rise to the question of whether it is possible to make use of the known analytical properties of singular perturbation problems, in order to design an algorithm which is more accurate and less sensitive to the distribution of the meshpoints.

The method of directional differences

DORR [1970a] gives an extensive discussion of the method of *upstream one-sided* (or *directional*) *differences* on a uniform mesh. In fact, he applies this method to special cases of the nonlinear system

$$
\begin{aligned}
& u''(x) = f(x,u,v), \qquad u(0) = u(1) = 0 \ , \\
(2.3.2) \qquad & \varepsilon v''(x) + g(x,u,u')v' - c(x,u,u')v = 0, \\
& c(x,u,u') \geq 0, \qquad v(0) = v_0, \ v(1) = v_1 .
\end{aligned}
$$

Here we shall confine ourselves to the treatment of the two-point boundary-value problem (2.2.1) with $g(x) \leq 0$. The upstream one-sided difference approximates the first order term of eq. (2.2.1) by

$$
(2.3.3.a) \qquad [fy_i']^{\wedge} \overset{\text{def}}{=} \begin{cases} f(x_i)(y_{i+1}-y_i)/h & \text{if } f(x_i) \geq 0, \\ f(x_i)(y_i-y_{i-1})/h & \text{if } f(x_i) < 0. \end{cases}
$$

Hence, the difference equation used reads

$$
(2.3.3.b) \qquad L_h(y_i) = \varepsilon(y_{i+1}-2y_i+y_{i-1})/h^2 + [fy_i']^{\wedge} + g(x_i)y_i = s(x_i).
$$

The primary reason for using these one-sided differences is to ensure that the equations are of positive type and, hence, that there is a unique solution for each set of data and for each $\varepsilon > 0$, $h > 0$. Moreover, for the discrete equation a discrete analogue of the maximum principle (cf. PROTTER & WEINBERGER [1967]) holds. Before we state this in lemma 2.3.1, we first need the following

DEFINITION

A difference operator L_h of the form

$$
(2.3.4) \qquad L_h(y_j) = a_j y_{j-1} + b_j y_j + c_j y_{j+1}
$$

is of *positive type* if

$$(2.3.5)\qquad \begin{array}{l} a_j + b_j + c_j \le 0, \\ a_j > 0,\ c_j > 0 \quad . \end{array}$$

Writing down the difference equations corresponding to equation (2.2.1), using the directional difference (2.3.3), it is immediately seen that it yields a difference operator of positive type, whenever $g \le 0$. The following lemma shows that a discrete maximum principle holds for a difference operator of positive type.

LEMMA 2.3.1. (The discrete maximum principle) *Let* L_h *be a difference operator of positive type and let* $L_h(y_j) \ge 0$ *for* $j = 1,2,\dots,N-1$; *let further* $y_0 = \alpha$, $y_N = \beta$. *If* y_j *assumes a nonnegative maximum value* M *for some* j, $0 < j < N$ *then* $y_j = M$ *for all* j, $j = 0,1,\dots,N$.

REMARK. This lemma is easily verified by a straightforward calculation, and analogously it can be proved that y_h cannot take a non-positive minimum value if $L_h y_h \le 0$.

The following theorem states that meaningful approximations to asymptotic solutions for $\varepsilon \to 0$ of the continuous problem can be obtained, by letting first $\varepsilon \to 0$ and then $h \to 0$ in the discrete problem, if the method of directional differences is used.

THEOREM 2.3.1. *Given a uniform partition of* $[a,b]$ *and a two-point boundary-value problem* (1.1.1) *with* $g < 0$, *then the method of directional differences has a discrete limit-solution for* $\varepsilon \to 0$. *Moreover, the method is consistent with the reduced problem on each closed interval which excludes a turning point.*

REMARK. The theorem still holds if $g \le 0$, provided that $g < 0$ in each turning point region $[x^* - h, x^* + h]$ where $f(x^*) = 0$, $f'(x^*) \le 0$.

PROOF. Let the mesh-function $y_{h,0}$ be determined by

$$(2.3.6)\qquad \begin{cases} f(x_i)(y_{i+1}-y_i)/h + g(x_i) = s(x_i) & \text{if } f(x_i) \ge 0 \\ f(x_i)(y_i-y_{i-1})/h + g(x_i) = s(x_i) & \text{if } f(x_i) < 0 \end{cases}$$
$$y_0 = \alpha,\ y_N = \beta,$$

then $y_{h,0}$ is the (unique) solution of the matrix problem

$$B\, y_{h,0} = c,$$

where $\|B^{-1}\| < \infty$ (Gershgorin).
(Here $\|\cdot\|$ denotes the matrix-norm associated with $\|\cdot\|_{\Pi,2}$.)

The finite difference solutions, $y_{h,\varepsilon}$ are determined by a matrix problem

$$(\varepsilon A+B)y_{h,\varepsilon} = c,$$

where $\|A\| < \infty$. If ε_0 is chosen such that $0 < \varepsilon_0\|A\|\|B^{-1}\| < 1$, then $(\varepsilon A+B)^{-1}$ exists for all $0 < \varepsilon \le \varepsilon_0$ and

$$\|(\varepsilon A+B)^{-1}\| \le \frac{\|B^{-1}\|}{1-\varepsilon\|A\|\|B^{-1}\|}.$$

Hence $y_{h,\varepsilon} = (\varepsilon A+B)^{-1}c$ exists for all $0 < \varepsilon \le \varepsilon_0$, and

$$\|y_{h,\varepsilon}-y_{h,0}\|_{\Pi,2} \le \frac{\varepsilon\|A\|\|B^{-1}\|}{1-\varepsilon\|A\|\|B^{-1}\|}\|y_{h,0}\|_{\Pi,2} \le \frac{\varepsilon\|A\|\|B^{-1}\|^2\|c\|_{\Pi,2}}{1-\varepsilon\|A\|\|B^{-1}\|}.$$

Thus, for all $x_i \in \Pi$, $\lim\limits_{\varepsilon\to 0} y_{h,\varepsilon}(x_i) = y_{h,0}(x_i)$.

On each closed interval which excludes a turning point, the system (2.3.6) integrates the reduced equation

$$fy' + gy = s$$

in the down-stream direction, by the backward Euler method for initial value problems. Hence (cf. HENRICI [1962]), the method is consistent with the reduced problem. □

Il'in's method

Although theorem 2.3.1 reveals the advantages of the directional difference (2.3.3) there are disadvantages too:

i) by the one-sided difference only approximation to first order is attained;

ii) the method is not uniformly ε-convergent on [a,b], even for the simplest case: constant coefficients and $g \equiv 0$.

The latter can be seen by comparing eqs. (2.1.2) and (2.1.12). The difference between the approximation and the exact solution, y_ε, is

$$\|y_{h,\varepsilon} - y_\varepsilon\|_{\pi,\infty} = \max_i \left|\frac{1-\mu_2^i}{1-\mu_2^N} - \frac{1-\nu^i}{1-\nu^N}\right|.$$

As uniform ε-convergence is investigated we can take $h = \varepsilon$, which yields

$$\lim_{h\to 0} \sup_{0<\varepsilon<\varepsilon_0} \|y_{h,\varepsilon} - y_\varepsilon\|_{\pi,\infty} > \lim_{N\to\infty} \max_i \left|\frac{1-2^{-i}}{1-2^{-N}} - \frac{1-e^{-i}}{1-e^{-N}}\right| =$$

$$= \max_i |e^{-i} - 2^{-i}| \geq \frac{1}{2} - \frac{1}{e}.$$

This proves the absence of uniform ε-convergence. Clearly this is caused by the defective representation of the rate of decay from the boundary-layer into the interior. Further, eqs. (2.1.2) and (2.1.12) show that equation (2.1.11) is an approximation to the differential equation

$$(2.3.7) \qquad \varepsilon\left[\frac{h/\varepsilon}{\log(1+h/\varepsilon)}\right]y'' + y' = 0$$

rather than to the original equation (2.1.1). For values h of the same order of magnitude as ε this is approximately

$$(2.3.8) \qquad (\varepsilon + \frac{h}{2})y'' + y' = 0.$$

Equally, if the directional difference method is applied to the equation

$$(2.3.9) \qquad \varepsilon y'' + fy' = 0,$$

the solution of the discrete problem will correspond rather to the equation

$$(2.3.10) \qquad \varepsilon(1+ |\frac{fh}{2\varepsilon}|)y'' + fy' = 0.$$

We see that the boundary layer shows up as "diffused". In order to overcome this effect, which disturbs the representation in the boundary layer, IL'IN [1969] constructs a difference scheme which represents the rate of decay in the boundary layer correctly for the homogeneous case with $g = 0$ and constant coefficients.

He applies this scheme more generally, to the differential equation

$$(2.3.11) \qquad \varepsilon y''(x) + f(x)y'(x) = s(x)$$

and for a uniform net he proposes the following difference scheme

$$(2.3.12) \qquad \gamma_i(y_{i+1}-2y_i+y_{i-1})/h^2 + f(x_i)(y_{i+1}-y_{i-1})/(2h) = s(x_i)$$

where γ_i is selected in accordance with the requirement indicated above. Hence he puts

$$(2.3.13) \qquad \gamma_i = \frac{f(x_i)h}{2} \coth\left(\frac{f(x_i)h}{2\varepsilon}\right).$$

(Note: x coth x is a nice computable function.)
Therefore, the difference operator becomes

$$(2.3.14) \qquad L_h(y_i) = \frac{f_i}{2h}\left[\coth(\frac{f_ih}{2\varepsilon})+1\right]y_{i+1} - \frac{f_i}{h}\coth(\frac{f_ih}{2\varepsilon})y_i + \\ + \frac{f_i}{2h}\left[\coth(\frac{f_ih}{2\varepsilon})-1\right]y_{i-1} = s_i.$$

This difference operator is of positive type. Il'in proves the following errorbound for this operator when applied to (2.3.11).

<u>LEMMA 2.3.2</u>. *The errorbound for the difference method* (2.3.12)-(2.3.13) *when applied to a differential equation* (2.3.11) *with* $\varepsilon > 0$, $f \in C^2[a,b]$, $s \in C^2[a,b]$, $f(x) \neq 0$ *on* $[a,b]$, *is given by*

$$(2.3.15) \qquad \|y_{h,\varepsilon} - y_\varepsilon\|_{\pi,\infty} \le K(\varepsilon)h^2.$$

Moreover, for this class of equations the method is uniformly ε*-convergent of order* 1.

<u>PROOF</u>. see IL'IN [1969].

<u>Some aspects of the work of Abrahamsson, Keller and Kreiss</u>

KREISS [1973] and ABRAHAMSSON et. al. [1974] consider the system of differential equations

$$\begin{aligned} &\varepsilon y'' + Ay' + By = s, \\ &y(a) = \alpha,\ y(b) = \beta, \end{aligned} \tag{2.3.16}$$

where y and s are smooth nth order vector-functions and A and B are smooth n × n matrix functions. The matrix A is symmetrical and block-diagonal, consisting of two blocks, one with all eigenvalues greater than $\eta > 0$, the other with all eigenvalues smaller than $-\eta$. Under these circumstances no problems arise with turning points; the reduced problem is defined by equation (2.3.16) with $\varepsilon = 0$ and the boundary conditions at $x = b$ (resp. $x = a$) for the functions of y that correspond to the positive (resp. negative) definite part of A.

We will confine ourselves to the treatment of the scalar equation (2.2.1) only. For the discretization, a uniform mesh is used and a finite difference scheme is proposed of the form

$$\begin{aligned} &(\varepsilon+h\sigma_i)(y_{i+1}-2y_i+y_{i-1})/h^2 + a_i(y_{i+1}-y_{i-1})/(2h) + b_{1,i}y_{i+1} + \\ &\quad + b_{0,i}y_i + b_{-1,i}y_{i-1} = c_1 s_{i-1/2} + c_0 s_i + c_{-1}s_{i-1/2}. \end{aligned} \tag{2.3.17}$$

Here σ_i is a positive scalar. The coefficients $a_i, b_{j,i}$ and c_j are chosen to give an accurate approximation for the reduced problem. To determine the coefficients we have to distinguish between $f > 0$ and $f < 0$.
If $f < 0$, then

$$\begin{aligned} &\sigma_i = \tfrac{1}{2}\,|f(x_{i-1/2})|,\ a_i = f(x_{i-1/2}), \\ &b_{1,i} = 0,\ b_{0,i} = \tfrac{1}{2}g(x_{i-1/2}),\ b_{-1,i} = \tfrac{1}{2}g(x_{i-1/2}), \\ &c_1 = 0,\ c_0 = 0\quad ,\ c_{-1} = 1. \end{aligned} \tag{2.3.18a}$$

If $f > 0$, then

$$\begin{aligned} &\sigma_i = \tfrac{1}{2}\,|f(x_{i+1/2})|,\ a_i = f(x_{i+1/2}), \\ &b_{1,i} = \tfrac{1}{2}g(x_{i+1/2})\quad ,\ b_{0,i} = \tfrac{1}{2}g(x_{i+1/2}),\ b_{-1,i} = 0, \\ &c_1 = 1\quad ,\ c_0 = 0\quad ;\ c_{-1} = 0. \end{aligned} \tag{2.3.18b}$$

This difference equation can also be written in the form

$$(2.3.19) \qquad L_h(y_i) = \varepsilon(y_{i+1}-2y_i+y_{i-1})/h^2 + [fy']^{=}+[gy]^{\approx} = [s]^{\sim},$$

where

$$[fy_i']^{=} = \begin{cases} f(x_{i+1/2})(y_{i+1}-y_i)/h & \text{if } f(x_i) > 0, \\ f(x_{i-1/2})(y_i-y_{i-1})/h & \text{if } f(x_i) < 0, \end{cases}$$

$$[gy_i]^{\approx} = \begin{cases} g(x_{i+1/2})(y_{i+1}+y_i)/2 & \text{if } f(x_i) > 0, \\ g(x_{i-1/2})(y_i+y_{i-1})/2 & \text{if } f(x_i) < 0, \end{cases}$$

$$[s_i]^{\sim} = \begin{cases} s(x_{i+1/2}) & \text{if } f(x_i) > 0, \\ s(x_{i-1/2}) & \text{if } f(x_i) < 0. \end{cases}$$

We see that this difference approximation can be considered as a refinement of the method of directional differences. Both methods agree as far as the discretization of the 2nd order term is concerned. For the discretization of the terms with y', y and the right-hand side of the equation, midpoint approximations are used.

Provided that the homogeneous reduced problem only has the trivial solution, ABRAHAMSSON et al. [1974] show that the method (2.3.17)-(2.3.18) is uniformly ε-stable. Analogous to the asymptotic expansions as $\varepsilon \to 0$ for the continuous problem, asymptotic expansions in powers of ε, h and ε/h can be given for the discrete problem.

For the reduced problem, the scheme (2.3.17)-(2.3.18) corresponds to the well-known midpoint-rule (cf. KELLER [1974]) and, hence, it gives an approximation which is accurate to second order. However, the refinement also causes the discrete operator to be no longer of positive type for all $g < 0$.

If the matrix A in equation (2.3.16) is not block-diagonal with definite blocks or if the problem is nonlinear, another scheme of type (2.3.17) is proposed, namely

$$(2.3.20) \qquad \begin{aligned} &(\varepsilon+\sigma h)(y_{i+1}-2y_i+y_{i-1})/h^2 + f(x_i)(y_{i+1}-y_{i-1})/(2h) + \\ &\quad + g(x_i)y_i = s(x_i), \end{aligned}$$

where $\sigma > \frac{1}{2}|f|$.

We see that the boundary layer is artificially widened to $\mathcal{O}(h)$. The resemblance between equations (2.3.10) and (2.3.20) is clear. Let us consider this in more detail. By equation (2.1.12) we see that the rate of decay in the boundary layer is not correctly represented if the directional difference approximation (2.3.3) is used. The boundary layer shows up as diffused but oscillations in the numerical solution are suppressed. Consider the scheme (2.3.20) applied to the example

$$(2.3.21)\qquad \varepsilon y'' + fy' = 0, \qquad y(0) = 0,\ y(1) = 1,$$

with constant coefficient f.
The difference equations are

$$(2.3.22)\qquad (\varepsilon+\sigma h)(y_{i+1}-2y_i+y_{i-1})/h^2 + (y_{i+1}-y_{i-1})f/(2h) = 0$$
$$y_0 = 0, \qquad y_N = 1.$$

The solution reads

$$(2.3.23)\qquad y_i = \frac{1-\mu^i}{1-\mu^N}, \quad \text{where } \mu = \frac{\varepsilon+\sigma h - \frac{fh}{2}}{\varepsilon+\sigma h + \frac{fh}{2}}.$$

If $\sigma = 0$ this is equivalent to central differences; if $\sigma = f/2$ to forward differences and if $\sigma = |f/2|$ to directional differences. Oscillations will be absent if $\sigma \geq |\frac{f}{2}| - \frac{\varepsilon}{h}$. To avoid all erroneous oscillatory behaviour, irrespective of the smallness of ε, σ should be chosen such that $\sigma \geq |\frac{f}{2}|$. This is the motivation for scheme (2.3.20). As the correct rate of decay is given by $\mu = \exp(-fh/\varepsilon)$, we see that it is badly represented by (2.3.20), but the numerical boundary-layer is essentially confined to one mesh-width. In contrast with the scheme (2.3.17)-(2.3.18), the accuracy of scheme (2.3.20) is only $\mathcal{O}(h)$.

In order to clarify its relationship with the difference approximations mentioned earlier, scheme (2.3.20) also can be regarded as approximating the first order term fy' of the differential equation (2.2.1) by

$$(2.3.24)\qquad [(f(x_i)+2\sigma)y_{i+1} - 4\sigma y_i + (-f(x_i)+2\sigma)y_{i-1}]/(2h).$$

2.4. EXPONENTIALLY FITTED METHODS

In this section we consider the problem (1.1.1) again and we show that a unifying approach is possible for the methods for the discretization of fy'. Moreover, this new approach enables us to construct a simple method that inherits most of the benefits of the other ones. Refinements, which include the discretization of gy, are also studied. For simplicity, we restrict our investigations to difference schemes on a uniform grid. Extensions and adaptions to non-uniform partitions of [a,b] appear in a natural way when the difference schemes are generated in a more systematic way in chapter 3.

The method of weighted differences

We introduce a new difference approximation to the first order derivative y' in equation (1.1.1)

$$(2.4.1) \qquad y_{\tilde{x}} = ((1+\alpha_i)(y_{i+1}-y_i) + (1-\alpha_i)(y_i-y_{i-1}))/(2h),$$

where α_i is a free parameter, $|\alpha_i| \leq 1$. This approximation is a weighted combination of the forward- and backward- difference approximation. Hence, forward, backward and central differences arise as special cases with α_i fixed, and equal to +1, -1 and 0 respectively. For our purposes we take α_i depending on ε,h and $f(x_i)$. Referring to eq. (2.3.14) we see that Il'in's method is a special case. Also Kreiss' method, eq. (2.3.20), can be cast into the new form by taking $\alpha_i = \frac{2\sigma}{f(x_i)}$. We note that ABRAHAMSSON et al. [1974] also permit $|\alpha_i| > 1$. An advantage of our approach, in particular for Il'in's and Kreiss' method is, that it is clearly seen how the methods behave for $\varepsilon \to 0$.

Having introduced α_i as a free parameter, we can choose it in such a way that a number of requirements are fulfilled. In order to study the properties of the difference quotient (2.4.1), we construct a difference operator L_h, corresponding to the operator L in eq. (2.2.1). We use the difference (2.4.1) and the common 2nd order difference quotient for approximating the 2nd derivative. Thus, we obtain

$$(2.4.2) \qquad \begin{aligned} L_h(y_i) = {} & \left(\frac{\varepsilon}{h^2} + \frac{(1+\alpha_i)f(x_i)}{2h} \right) y_{i+1} + \left(\frac{-2\varepsilon}{h^2} - \frac{2\alpha_i f(x_i)}{2h} + g(x_i) \right) y_i + \\ & + \left(\frac{\varepsilon}{h^2} - \frac{(1-\alpha_i)f(x_i)}{2h} \right) y_{i-1}. \end{aligned}$$

The following lemma is immediate.

LEMMA 2.4.1. *Sufficient and necessary conditions for the difference operator* (2.4.2) *to be of positive type, are*

$$g(x_i) \le 0, \tag{2.4.3}$$

and

$$\frac{-1}{1+\alpha_i} \le \frac{f(x_i)h}{2\varepsilon} \le \frac{1}{1-\alpha_i}, \tag{2.4.4}$$

for $i = 1,2,\ldots,N-1$.

COROLLARY. For an operator L_h of positive type, the values of α_i must be restricted to a subdomain of $[-1,+1]$. This domain depends on the value of $\frac{f(x_i)h}{2\varepsilon}$. In order to yield an operator of positive type, the parameter α_i should satisfy

$$\alpha_i \in [-1,-1+|\frac{2\varepsilon}{f(x_i)h}|] \quad \text{if } f(x_i) < -\varepsilon/h \tag{2.4.5a}$$

$$\alpha_i \in [1-|\frac{2\varepsilon}{f(x_i)h}|, +1] \quad \text{if } f(x_i) > \varepsilon/h. \tag{2.4.5b}$$

The domain of permitted values α_i is indicated in Fig. 2.3.1.

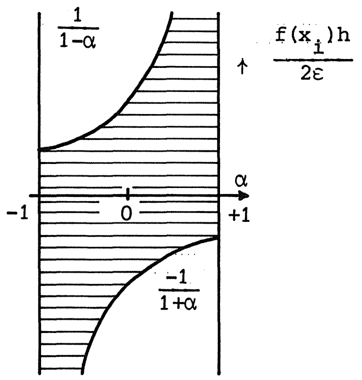


Fig. 2.4.1

The domain of α_i for which L_h, defined by equation (2.4.2) is of positive type.

COROLLARY. It is possible to find continuous functions m: $\mathbb{R} \to [-1,+1]$ such that the operator L_h (eq. (2.4.2)), with

$$(2.4.6) \qquad \alpha_i = m\left(\frac{f(x_i)h}{2\varepsilon}\right),$$

is of positive type, for all $\varepsilon, h > 0$ and all f and g with $g(x) \le 0$.

Exponential fitting of the difference operator (2.4.2)

First we restrict ourselves to the differential equation

$$(2.4.7) \qquad \varepsilon y'' + fy' = 0,$$

with a constant coefficient $f \neq 0$.

For this equation we can construct the parameter α_i in such a way that the rate of decay in the boundary layer is correctly represented.

LEMMA 2.4.2. *With*

$$(2.4.8) \qquad \alpha_i = m\left(\frac{f(x_i)h}{2\varepsilon}\right) = \coth\left(\frac{f(x_i)h}{2\varepsilon}\right) - \frac{2\varepsilon}{f(x_i)h},$$

the difference operator L_h *(eq. (2.4.2)) yields a pointwise exact solution to the two-point boundary-value problem (2.4.7)-(1.1.1.b).*

PROOF. Without loss of generality we restrict ourselves to the boundary conditions (2.1.1). The solution of the difference equation (2.4.2) with these boundary conditions is

$$y_i = \frac{1-\mu^i}{1-\mu^N}, \text{ where } \mu = \frac{2\varepsilon+\alpha_i fh-fh}{2\varepsilon+\alpha_i fh+fh}.$$

The solution of the differential equation is given by eq. (2.1.2) with $\nu = \exp(-fh/\varepsilon)$. Setting μ equal to ν yields (2.4.8). □

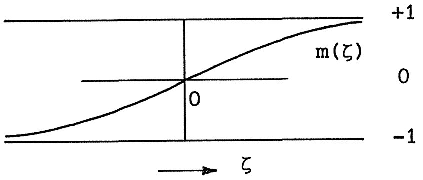


Fig. 2.4.2.

The function m(ζ) defined by eq. (2.4.8).

REMARK. The $m(\zeta)$, defined by equation (2.4.8), is a smooth function $\mathbb{R} \to [-1,+1]$, see fig. 2.4.2. With this $m(\zeta)$ the difference operator L_h (2.4.2) defines a smooth transition from forward to backward differences. For extreme values $\frac{f(x_i)h}{2\varepsilon}$, backward and forward differences are used, just as was the case in the method of directional differences, where the change-over is discontinuous.

THEOREM 2.4.1. *If* f *is positive or negative definite and if* α_i *is defined by* (2.4.8), *then the operator* L_h *in eq.* (2.4.2) *has the following properties:*

i) L_h *is of positive type if* $g(x) \le 0$.

ii) For small ε *(i.e.* $|\varepsilon g| < |f|^2$), *the solution of* $L_h(y_i) = s(x_i)$ *represents the rate of change in the boundary layer with a relative accuracy of* $O(\frac{\varepsilon g}{f^2} \cdot \frac{gh}{f}) + O((\frac{gh}{f})^2)$ *for* $(\frac{gh}{f}) \to 0$.

iii) If $g < 0$, *the difference method described by* $L_h(y_i) = s(x_i)$ *is convergent of order* 2 *and it is uniformly* ε*-convergent of order* 1.

PROOF.

i) A straightforward calculation yields

$$\frac{-1}{1+m(\zeta)} < \zeta < \frac{1}{1-m(\zeta)}.$$

Now lemma 2.4.1 asserts part i) of the present lemma.

ii) In the boundary layer we approximate the homogeneous differential equation (1.2.1) by the differential equation with constant coefficients

$$\varepsilon y'' + fy' + gy = 0.$$

The solution is

$$y(x_i) = C_1 e^{\lambda_1 hi} + C_2 e^{\lambda_2 hi}$$

where λ_1, λ_2 are the roots of

$$\varepsilon\lambda^2 + f\lambda + g = 0.$$

The solution of the difference equation $L_h(y_i) = 0$ is

$$y_i = D_1\mu_1^i + D_2\mu_2^i$$

where μ_1, μ_2 are the roots of

$$(2.4.9) \qquad ((\frac{\varepsilon}{h^2} + \frac{\alpha f}{2h}) + \frac{f}{2h})\mu^2 - [2(\frac{\varepsilon}{h^2} + \frac{\alpha f}{2h}) - g]\mu + ((\frac{\varepsilon}{h^2} + \frac{\alpha f}{2h}) - \frac{f}{2h}) = 0.$$

For a correct representation in the boundary layer , μ_1 should correspond to $\exp(h\lambda_1)$ and μ_2 to $\exp(h\lambda_2)$. A simple calculation shows that, with our particular choice of α, the relation

$$(2.4.10) \qquad \mu_1\mu_2 = \exp(h\lambda_1)\ \exp(h\lambda_2)$$

holds exactly for all values of ε, h, f and g. We seek an asymptotic expression for μ_1 for small values of $|\frac{gh}{f}|$. For convenience, we set $\delta = \frac{\varepsilon}{h} + \frac{\alpha f}{2}$; then we can write eq. (2.4.9), simplifying our notation,

$$(2\delta+f)\mu^2 - (4\delta-2gh)\mu + (2\delta-f) = 0.$$

One root is given by

$$\mu_1 = \frac{2\delta - gh + f\sqrt{1 - \frac{4\delta}{f}(\frac{gh}{f}) + (\frac{gh}{f})^2}}{2\delta+f}$$

$$= 1 - \frac{gh}{f} + O((\frac{gh}{f})^2)$$

$$= \exp(h\lambda_1)[1 + O(\frac{gh}{f}\frac{\varepsilon g}{f^2}) + O((\frac{gh}{f})^2)].$$

Therefore, for small ε such that $|\varepsilon g| < f^2$ and for $|\frac{gh}{f}| \to 0$, the slowly varying component, $\exp(h\lambda_1)$ is represented by μ_1 with a relative error of order $O(\frac{gh}{f} \cdot \frac{\varepsilon g}{f^2}) + O((\frac{gh}{f})^2)$, uniformly for all small ε. It follows from (2.4.10) that the rapidly varying component, $\exp(h\lambda_2)$, is also accurately approximated by μ_2, with the same relative accuracy and uniformly in ε.

iii) Substituting a Taylor series expansion of $y(x_{i+1})$, $y(x_i)$ and $y(x_{i-1})$ for y_{i+1}, y_i resp. y_{i-1} in eq. (2.4.2), we obtain

$$L_h(y_i) = Ly(x_i) + \alpha_i hf(x_i)\frac{y''(x_i)}{2} + h^2 f(x_i)\frac{y'''(x_i)}{6} + \mathcal{O}(h^3).$$

Hence,

$$L_h(y_i) - L_h(y(x_i)) = \mathcal{O}(h)$$

uniformly for $0 < \varepsilon \leq \varepsilon_0$. Since $|f(x)| > 0$ the same technique as was used in theorem 2.3.1 can be used to show that $\|L_h^{-1}\|$ is bounded, uniformly in ε; hence we see also that

$$\|y_i - y(x_i)\| = \mathcal{O}(h),$$

uniformly for $0 < \varepsilon \leq \varepsilon_0$. Moreover, for $|\frac{f(x_i)h}{\varepsilon}| \to 0$, we have

$$\alpha_i = m(\frac{f(x_i)h}{2\varepsilon}) = \frac{f(x_i)h}{6\varepsilon} + \mathcal{O}(|\frac{h}{\varepsilon}|^2).$$

This yields

$$L_h(y_i) = Ly(x_i) + \frac{h^2 f(x_i)}{12\varepsilon}\left[2\varepsilon y'''(x_i) + f(x_i)y''(x_i)\right] + \mathcal{O}(h^3)$$

or

$$L_h(y_i) - L_h(y(x_i)) = \mathcal{O}(h^2) \text{ if } |fh| << \varepsilon.$$

Hence the method is convergent of order 2, if $|fh| << \varepsilon$. □

Asymptotic behaviour for $\varepsilon \to 0$ of the exponentially fitted operator L_h

For $f \neq 0$ and small values of $|\frac{2\varepsilon}{fh}|$, we have

$$\coth(\frac{fh}{2\varepsilon}) = \text{sign}(f) + \mathcal{O}(\exp(-|\frac{fh}{\varepsilon}|)). \tag{2.4.11}$$

If we neglect the exponentially small term, we get $m(z) = \coth(z) - \frac{1}{z} \approx \text{sign}(z) - \frac{1}{z}$ for $|z| \to \infty$. Thus, the difference operator (2.4.2) becomes

$$\begin{aligned} L_h(y_i) \approx {} & \frac{f_i}{2h}(1+\text{sign}(f_i))y_{i+1}+(\frac{-f_i}{h}\ \text{sign}(f_i)+g_i)y_i + \\ & + \frac{f_i}{2h}(\text{sign}(f_i)-1)y_{i-1}. \end{aligned} \tag{2.4.12}$$

We note that the discrete equivalent of the 2nd order term of the differential equation, $\varepsilon y''$, is completely annihilated by the second term of $m(\zeta)$. Hence, if $\exp(-|\frac{fh}{\varepsilon}|) << 1$, our method solves the reduced equation as an initial value problem, from the right to the left if $f > 0$ and from the left to the right if $f < 0$. This is exactly the way the analytical solution behaves for small ε. We note that degeneration to the solution of an initial value problem also happens with the method of directional differences. In that case, the condition reads $|\frac{\varepsilon}{fh}| << 1$ instead of $\exp(-|\frac{fh}{\varepsilon}|) << 1$. As is easily seen from lemma 2.4.2, the latter condition is the more realistic one.

REMARK. In this chapter, the discussion of the exponentially fitted finite difference method (2.4.8) is restricted to uniform partitions of [a,b] only. In chapter 3 it will be generalized to non-uniform partitions (eq. (3.5.12)) and in chapter 4 some numerical results are given. More numerical results, for linear problems, can be found in HEMKER [1974].

A new discretization for g(x)y(x)

Since favourable results have been obtained by introducing a parameter α_i in the difference $y_{\tilde{x}}$ (HEMKER, [1974]), we are in a position to ask the question if it is expedient to introduce parameters in the discretization of the term $g(x)y(x)$ in equation (1.1.1). In the case of constant coefficients, it certainly should be possible to find a discretization of $g(x)y(x)$ which yields a pointwise exact solution for the homogeneous equation. To find this, we consider the discretization

(2.4.13) $$y_{\square} = (\beta_i+\gamma_i)y_{i+1} + (1-2\beta_i)y_i + (\beta_i-\gamma_i)y_{i-1}$$

and we introduce the discrete operator

(2.4.14) $$L_h(y_i) = \varepsilon(y_{i+1}-2y_i+y_{i-1})/h^2 + fy_{\tilde{x}} + gy_{\square}.$$

For this operator L_h we have to determine the parameters α_i, β_i and γ_i. The results are given in the following lemma. Since we have restricted ourselves to non-oscillating solutions, we assume $f^2 - 4\varepsilon g > 0$.

LEMMA 2.4.3. *Suppose we are given the differential equation*

$$(2.4.15)\qquad \varepsilon y'' + fy' + gy = 0,$$

with constant coefficients such that $4\varepsilon g < f^2$.
Let L_h, $y_{\tilde{x}}$ *and* $y_{\square}$ *be defined by* (2.4.14), (2.4.1) *and* (2.4.13) *respectively, and let*

$$(2.4.16a)\qquad \alpha_i = \coth(\frac{fh}{2\varepsilon}) - \frac{2\varepsilon}{fh},$$

$$(2.4.16b)\qquad \beta_i = \gamma_i \coth(\frac{fh}{2\varepsilon}),$$

$$(2.4.16c)\qquad \gamma_i = \frac{-1}{4}\left[\frac{2f}{gh} + \coth(\frac{h\lambda_1}{2}) + \coth(\frac{h\lambda_2}{2})\right].$$

where λ_1, λ_2 *are the roots of* $\varepsilon\lambda^2 + f\lambda + g = 0$.
Then the solution $\{y_i\}$ *of the difference equation* $L_h(y_i) = 0$ *yields a pointwise exact solution to eq.* (2.4.15).

REMARK. The lemma holds for any set of boundary conditions (1.1.1.b). Hence it follows that, if $y_i = y(x_i)$ holds for two distinct points x_i, it holds for all points.

PROOF. In order to deal with the case g = 0 correctly, α_i should be as defined in (2.4.16a); see lemma 2.4.2.
The solution of (2.4.15) reads

$$y(x_j) = C_1 \exp(jh\lambda_1) + C_2 \exp(jh\lambda_2).$$

The solution of the difference equation $L_h(y_i) = 0$ is

$$y_j = D_1\, \mu_1^j + D_2\, \mu_2^j,$$

where μ_1 and μ_2 are the roots of

$$\left[\frac{\varepsilon}{h^2} + \frac{1+\alpha}{2}\frac{f}{h} + (\beta+\gamma)g\right]\mu^2 + \left[\frac{-2\varepsilon}{h^2} - \frac{\alpha f}{h} + (1-2\beta)g\right]\mu +$$
$$+ \left[\frac{\varepsilon}{h^2} - \frac{1-\alpha}{2}\frac{f}{h} + (\beta-\gamma)g\right] = 0.$$

By setting $\mu_1 = \exp(h\lambda_1)$ and $\mu_2 = \exp(h\lambda_2)$ we obtain the expressions (2.4.16b) and (2.4.16c) for β and γ. □

The parameters β_i and γ_i are not well suited for implementation in a realistic algorithm. Nevertheless it is interesting to see how the parameters β and γ behave for small values of ε. Since γ can be expressed as a function of β in a straightforward way we concentrate on β.

We first consider $|\varepsilon g| << f^2$; then

$$\lambda_1 = -f/\varepsilon - \lambda_2$$
$$\lambda_2 = -\frac{g}{f}[1 + \frac{\varepsilon g}{f^2} + \mathcal{O}((\frac{\varepsilon g}{f^2})^2)].$$

Hence

$$(2.4.17) \qquad \beta \approx \frac{\coth(\frac{fh}{2\varepsilon})}{4}\left[\coth(\frac{fh}{2\varepsilon}) + \coth(\frac{gh}{2f}) - \frac{2f}{gh}\right].$$

If, in addition, $\exp(-|\frac{fh}{2\varepsilon}|) << 1$, then

$$(2.4.18) \qquad \beta \approx \frac{1}{4}\left[1 + \mathrm{sign}(f)\left\{\coth(\frac{gh}{2f}) - \frac{2f}{gh}\right\}\right].$$

Note: we already met the function $\coth(z) - z^{-1}$ in equation (2.4.8).

If we consider $|\frac{fh}{2\varepsilon}| << 1$, then

$$(2.4.19) \qquad \beta = \frac{1}{4}\left[\frac{2f}{gh}\coth(\frac{-fh}{2\varepsilon}) + \frac{\cosh(\frac{-fh}{2\varepsilon})}{\sinh(\frac{h\lambda_1}{2})\ \sinh(\frac{h\lambda_2}{2})}\right] \approx$$

$$(2.4.20) \qquad \approx \frac{-\varepsilon}{gh^2} + \frac{1}{4\sinh(\frac{h\lambda_1}{2})\ \sinh(\frac{h\lambda_2}{2})}.$$

In particular, if $f = 0$ then $\gamma = 0$ and

$$(2.4.21) \qquad \beta = \frac{-\varepsilon}{gh^2} - \left[\frac{1}{2\sinh\sqrt{\frac{gh^2}{-4\varepsilon}}}\right]^2.$$

We note that here (as with equation (2.4.12)) the first term in β exactly annihilates the discrete equivalent of the term $\varepsilon y''$. Thus, the difference equation corresponding to $\varepsilon y'' + gy = 0$ reads

$$wy_{i+1} - (4+2w)y_i + wy_{i-1} = 0, \text{ where } w = \operatorname{csch}^2\left(\sqrt{\frac{gh^2}{-4\varepsilon}}\right). \tag{2.4.22}$$

Summarizing, we find for small ε:

i) if $f \neq 0$, $\exp(-|\frac{f^2}{\varepsilon g}|) << 1$ and $\exp(-|\frac{fh}{2\varepsilon}|) << 1$,

$$y_\square = \beta(1+\tanh(\frac{fh}{2\varepsilon}))y_{i+1} + (1-2\beta)y_i + \beta(1-\tanh(\frac{fh}{2\varepsilon}))y_{i-1} \approx \tag{2.4.23}$$

$$\approx \frac{1}{4}(1+z)(1 + \text{sgf})y_{i+1} + \frac{1}{2}(1 - z\ \text{sgf})y_i + \frac{1}{4}(1-z)(1 - \text{sgf})y_{i-1}\ , \tag{2.4.24}$$

where $z = \coth(\frac{gh}{2f}) - \frac{2f}{gh}$ and $\text{sgf} = \text{sign}(f)$.

ii) if $f = 0$, $g \neq 0$ and $\exp(-\sqrt{\frac{gh^2}{-\varepsilon}}) << 1$,

$$y_\square = \beta y_{i+1} + (1-2\beta)y_i + \beta y_{i-1}\ ,$$

where

$$\beta \approx \frac{-\varepsilon}{gh^2} - \frac{1}{4}\exp(-\sqrt{\frac{gh^2}{-\varepsilon}})\ . \tag{2.4.25}$$

If we combine the results obtained in (2.4.12) and (2.4.24), we find the asymptotic behaviour for $\varepsilon \to 0$ of the exponentially fitted operator, that discretizes the differential equation (2.4.15).

If $f \neq 0$, $\exp(-|\frac{fh}{2\varepsilon}|) << 1$ and $\exp(-|\frac{f^2}{\varepsilon g}|) << 1$, then

$$L_h(y_i) \approx (\frac{\text{sgf}+1}{2})(\frac{f}{h} + \frac{g(z+1)}{2})y_{i+1} + \left[\frac{g}{2} - \text{sgf}(\frac{f}{h} + \frac{gz}{2})\right]y_i + +(\frac{\text{sgf}-1}{2})(\frac{f}{h} + \frac{g(z-1)}{2})y_{i-1}. \tag{2.4.26}$$

Here we see, again, that the problem is solved from the left to the right if $f < 0$ and from the right to the left if $f > 0$. Thus, the operator can be regarded as the one-step operator

$$y_{i+1} = -\frac{(-\frac{f}{h} + g\,\frac{1-z}{2})}{(\frac{f}{h} + g\,\frac{1+z}{2})}\,y_i = \exp(\frac{-gh}{f})y_i\ , \tag{2.4.27}$$

for $i = 0,1,\ldots,N-1$ if $f < 0$, or

for $i = 1,2,\ldots,N$ if $f > 0$.

CHAPTER III

GLOBAL METHODS

In contrast to the difference methods treated in chapter 2, global methods yield approximate solutions $y_h(x)$ that are not grid-functions, but functions defined over the whole interval $[a,b]$. Such an approximate solution is selected from a given finite-dimensional subspace of the linear space of all admissible functions. By the proper choice of a basis in this subspace, the global methods can be made to deliver immediately a sequence of discrete function values $\{y_h(x_i)\}$, corresponding to a particular grid Π. Therefore, we can still say that difference schemes are generated by these global methods.

In the first section we describe the general principles of weighted residual methods and we treat the construction of discrete operators. In particular a new, efficient implementation of the Galerkin method is given. In the second section we derive error estimates for weighted residual methods. To this end we introduce the function space $H^{k,\pi}[a,b]$ and we discuss the discrete Green's function. In section 3, we show why standard weighted residual methods fail, when they are applied to singular perturbation problems. We treat: Galerkin's method, Ritz-Galerkin, collocation, least squares and reduction to a system of first order equations. In section 4 we introduce exponentially fitted spaces and we show how they can be used for the construction of weighted residual methods. In section 5 we construct discrete operators by means of exponentially fitted spaces and we also point out the relation to the finite difference methods treated in chapter 2. In the 6th section we describe how exponentially fitted weighted residual methods behave when $\varepsilon \to 0$ and in section 7 we give some numerical results obtained by the new methods.

3.1. INTRODUCTION TO WEIGHTED RESIDUAL METHODS AND THE CONSTRUCTION OF DISCRETE OPERATORS

In this section we discuss global methods of generating difference schemes in a systematic way. A special advantage of global methods is that for non-uniform meshes also the construction of difference schemes follows in a natural way and that the treatment is not essentially more complicated than for uniform meshes.

All methods studied in this chapter provide a way of finding a numerical solution of the form

$$(3.1.1) \qquad y_h(x) = \sum_j a_j\phi_j(x);$$

where $\{\phi_j\}$ is a set of piecewise polynomials.

An extensive literature exists on various methods of this kind. We give an outline of some parts of the theory here, in order to provide the notation and a conceptual framework that will be expanded in the following sections, when exponentially fitted methods are treated. In this section we shall also describe a new, efficient implementation of Galerkin's method.

Generalized solutions

In order to introduce the notation we describe briefly Sobolev spaces and generalized solutions to differential equations. For a comprehensive treatment the reader is referred to YOSIDA [1965].

For any integer $k \geq 0$ we denote by $H^k(a,b)$ the Sobolev space of (classes of) real-valued functions which, together with their destributional derivatives of order $\leq k$, belong to $L^2(a,b)$. These spaces are Hilbert spaces when provided with the innerproduct

$$(u,v)_k = \sum_{\ell=0}^{k} (D^\ell u, D^\ell v),$$

$$(u,v) = (u,v)_0 = \int_a^b u(x)v(x)dx$$

and norm

$$\|u\|_k = \sqrt{(u,u)_k};$$

D denotes the differential operator.
The closure of the set of $C_0^\infty(a,b)$-functions with respect to the norm $\|\cdot\|_k$ is denoted by $H_0^k(a,b)$.

Consider the equation

$$(3.1.2) \qquad Ly \equiv -D(c_0Dy) + c_1Dy + c_2y = s,$$

where $c_0 \in C^1[a,b]$, $c_1,c_2 \in C^0[a,b]$, $c_0 \geq E > 0$, $s \in H^0(a,b)$.
In the classical sense a solution to equation (3.1.2) is a function y, $y \in C^2[a,b]$, such that

$$Ly(x) = s(x) \qquad \text{for all } x \in [a,b].$$

However, it is often convenient to choose y from a larger space S of admissible functions and to define a solution to (3.1.2) as that function $y \in S$ which satisfies the *variational equation*

$$(3.1.3) \qquad \int_a^b \{Ly(x) - s(x)\}\, v(x)\, dx = 0 \qquad \text{for all } v \in V.$$

The *trial space* S, and the *test space* V have to be chosen such that for all $u \in S$, $v \in V$ the integrals

$$\int_a^b Lu(x)\, v(x)\, dx \quad \text{and} \quad \int_a^b s(x)\, v(x)\, dx$$

exist.
The sense in which a solution is obtained is characterized by S and V. E.g. the equation is said to hold *in the strong sense* if $V = H^0(a,b)$ and *in the weak sense* if $S = V = H^1(a,b)$; i.e. after integration by parts.

DEFINITION

The continuous bilinear functional $B\colon H^1(a,b) \times H^1(a,b) \to \mathbb{R}$, defined by

$$(3.1.4) \qquad B(u,v) = (c_0Du,Dv) + (c_1Du,v) + (c_2u,v),$$

is called *the bilinear form associated with* L.

DEFINITION

By $C^{-1}[a,b]$ we denote the subset of functions in $L^2(a,b)$ that are defined and continuous on [a,b], except for a finite number of discontinuities in (a,b).

DEFINITION

Let f(x) be a continuous function on $(x_0-\delta,x_0)$ and on $(x_0,x_0+\delta)$ for some $\delta > 0$, then jmp $f(x_0)$ is defined by

$$\text{jmp } f(x_0) = \lim_{z \downarrow 0} f(x+z) - f(x-z).$$

The following lemma follows immediately by means of integration by parts.

LEMMA 3.1.1. *Let* $u,v \in H^1(a,b)$, *and let* $Du \in C^{-1}[a,b]$ *be continuous except at the set of points* $\Pi = \{x_i \mid a < x_1 < x_2 < \ldots < x_{n-1} < b\}$; *set* $a = x_0$ *and* $b = x_n$, *then*

$$(3.1.5) \qquad B(u,v) = (Lu,v)_{0,\pi} + [c_0 vDu]_\pi,$$

where $(\cdot,\cdot)_{0,\pi}$ *and* $[\cdot]_\pi$ *are defined by*

$$(3.1.6) \qquad \begin{aligned} (u,v)_{0,\pi} &= \sum_{i=1}^{n} (u,v)_{L^2(x_{i-1},x_i)} \qquad \textit{and} \\ [w]_\pi &= w(b) - \sum_{i=1}^{n-1} \text{jmp } w(x_i) - w(a). \end{aligned}$$

COROLLARY. Immediate consequences are

$$(3.1.7) \qquad B(u,v) = (Lu,v)_{0,\pi} \qquad \text{for all } u \in H^2(a,b),\ v \in H_0^1(a,b),$$

and each function that satisfies (3.1.2) also satisfies

$$(3.1.8) \qquad B(y,\phi) = (s,\phi) \qquad \text{for all } \phi \in H_0^1(a,b).$$

REMARK. Since there is no 2nd derivative in equation (3.1.8), this equation can be defined under less restrictive conditions with respect to the function y than equation (3.1.2).

DEFINITION

The *formal adjoint* of the operator L is defined by

$$(3.1.9) \qquad L^T y = -D(c_0 Dy) - D(c_1 y) + c_2 y.$$

By integration by parts, one easily obtains *Green's formula*

$$(3.1.10) \qquad (Lu,v)_{0,\pi} - (u,L^T v)_{0,\pi} = [c_0(uDv-vDu) + c_1 uv]_\pi$$

and in particular, using Sobolev's lemma, the equality

(3.1.11) $\quad (L\phi,\psi) = (\phi,L^T\psi) \quad$ for all $\phi,\psi \in H_0^2(a,b)$.

DEFINITION

The bilinear operator B: $H_0^1(a,b) \times H_0^1(a,b) \to \mathbb{R}$ is called *strictly coercive* if

$$\exists\sigma > 0 \quad \forall v \in H_0^1(a,b) \quad \sigma\|v\|_1^2 \le |B(v,v)|.$$

DEFINITION

Let S and V be two Hilbert-spaces. A bilinear operator B: S × V → ℝ is called *strictly coercive with respect to* S *and* V, if

$$\exists D(S,V) > 0 \quad \forall s \in S \quad \underset{v\neq 0}{\exists v \in V} \quad D(S,V)\|s\|_S\|v\|_V \le |B(s,v)|.$$

DEFINITION

Let S and V be two Hilbert-spaces. A bilinear operator B: S × V → ℝ is called *bounded* if

$$\exists C \in \mathbb{R} \ \forall s \in S,\ v \in V \quad |B(s,v)| \le C\|s\|_S\|v\|_V.$$

Weighted residual methods

The discretization of the differential equation by a weighted residual method is done by starting from the variational equation (3.1.3) and by computing $y_h \in S_h$, such that

$$(Ly_h,v_h) = (s,v_h) \qquad \text{for all } v_h \in V_h$$

(discretization of the strong form), or

$$B(y_h,v_h) = (s,v_h) \qquad \text{for all } v_h \in V_h$$

(discretization of the weak form).

Here S_h and V_h are finite dimensional subspaces of S and V respectively. Thus, corresponding to the different kinds of generalized solutions, we can distinguish between different types of discretization. We will show that discretization of the strong form leads to the collocation method and the weak form to Galerkin-type methods.

In sections 3.4 to 3.6 we shall treat new methods of Galerkin type. For the problem

$$Ly = s \text{ on } [a,b],\ y(a) = \alpha,\ y(b) = \beta,$$

the *classical Galerkin method* is obtained by choosing a basis $\{\phi_i\}_{i=0}^{M}$ in the space $S_h \subset H^1(a,b)$, such that $\{\phi_i\}_{i=1}^{M-1}$ is a basis in $V_h = S_h \cap H_0^1(a,b)$. This leads to an approximate solution $y_h \in S_h$ of the form (3.1.1) and the vector of coefficients (a_j) is determined by the linear system

$$\begin{aligned} &\sum_{j=0}^{M} a_j\, B(\phi_j,\phi_i) = (s,\phi_i), \qquad i = 1,2,\dots,M-1, \\ &\sum_j a_j\, \phi_j(a) \quad = \alpha, \\ &\sum_j a_j\, \phi_j(b) \quad = \beta. \end{aligned} \tag{3.1.12}$$

In general, full polynomial bases $\{\phi_j\}$ on $[a,b]$ lead to dense and ill-conditioned matrices $B(\phi_j,\phi_i)$ and so they are of little use for large M. The practical use of weighted residual methods hinges on the ease with which systems such as (3.1.12) are generated and solved. The revival of global methods is due to the fact that the resulting linear systems are sparse and that the entries are easily calculated. To this end the functions $\{\phi_i\}$ have to be chosen such that they vanish on $[a,b]$, except for a small subinterval. Here piecewise polynomials turn out to be useful tools.

Definition of piecewise polynomial spaces

In order to characterize piecewise polynomial spaces S_h we introduce the following notation. Let $\Pi = \{a = x_0 < x_1 < \dots < x_N = b\}$ be a partition of $[a,b]$ and set $I_i = (x_{i-1},x_i)$ and $h_i = x_i - x_{i-1}$. Let $P_k(E)$ denote the class of all polynomials of degree less than k+1, defined on the set E.

DEFINITION

For $m \le k$ the space of *C^m-piecewise polynomials of degree $\le k$* is defined by

$$M^{m,k}(\Pi) = \{v \in C^m[a,b] \mid v_{\text{restr } I_i} \in P_k(I_i);\ i = 1,2,\dots,N\}. \tag{3.1.13}$$

Similarly,

(3.1.14) $\quad M_0^{m,k}(\Pi) = \{v \in M^{m,k}(\Pi) \mid v(a) = v(b) = 0\}$

denotes the subspace of $M^{m,k}(\Pi)$ of all functions that satisfy homogeneous boundary conditions.

REMARK. By this definition, the space of discontinuous piecewise polynomials of degree k on Π is denoted by $M^{-1,k}(\Pi)$.

Important sub-families of piecewise polynomials are

i) the *Lagrange spaces*: $M^{0,k}(\Pi)$;
ii) the *Hermite spaces* : $M^{m,2m+1}(\Pi)$;
iii) the space of *spline functions*: $M^{m,m+1}(\Pi)$.

The space of piecewise linear functions, $M^{0,1}(\Pi)$, belongs to all three subfamilies.

In contrast to the spaces of spline functions $M^{m,m+1}(\Pi)$, $m > 0$, both Lagrange and Hermite spaces have bases $\{\phi_j\}$ such that the support of each ϕ_j contains at most two neighbouring intervals I_i. This is an expedient feature for computational purposes, since it leads to discrete operators that have a narrow band-matrix structure. To use this property we introduce natural bases.

Natural bases for $M^{0,k}(\Pi)$ and $M^{m,2m+1}(\Pi)$

In Lagrange and Hermite spaces we introduce natural bases; these bases consist of functions that have minimal support on [a,b].

i) The natural basis for a Lagrange space.

Let there be given a set $\{0 = \xi_0^* < \xi_1^* < \ldots < \xi_k^* = 1\}$. As a natural basis $\{\phi_j\}_{j=0,1,\ldots,Nk}$ in $M^{0,k}(\Pi)$ the Nk+1 functions ϕ_j are chosen, such that

(3.1.15)
$$\begin{aligned} &\phi_j \in M^{0,k}(\Pi), \\ &\phi_j(x_i + \xi_\ell^* h_{i+1}) = \delta_{j,\ell+ki}, \end{aligned}$$

for all (i,ℓ), $i = 0,1,\ldots,N-1$ and $\ell = 0,1,\ldots,k$.

ii) The natural basis for a Hermite space.

In $M^{m,2m+1}(\Pi)$ the (N+1)(m+1) natural basis-functions $\{\phi_j\}_{j=0,1,\ldots,m+N(m+1)}$ are chosen such that

$$(3.1.16)\qquad \begin{aligned} &\phi_j \in M^{m,2m+1}(\Pi) \\ &D^{\ell}\phi_j(x_i) = \delta_{j,\ell+i(m+1)} \end{aligned}$$

for all (i,ℓ), $i = 0,1,2,\ldots,N$ and $\ell = 0,1,\ldots,m$.

An additional advantage of the choices (3.1.15) and (3.1.16) is, that a proper selection of the coefficients (a_j) in expression (3.1.1) yields directly the pointwise approximation to y on the grid Π.

Discretization of the differential equation

Having at our disposal bases $\{\phi_j\}$ in S_h and $\{\psi_i\}$ in V_h, we can write the discrete equivalents to the weak and the strong form of the equations respectively as

$$(3.1.17)\qquad \sum_{j=0}^{M} a_j\, B(\phi_j,\psi_i) = (s,\psi_i) \quad \forall\psi_i \in V_h \quad \begin{aligned} &V_h \subset H^1(a,b), \\ i = 1,2,\ldots,M-1, \qquad &S_h \subset H^1(a,b), \end{aligned}$$

and

$$(3.1.18)\qquad \sum_{j=0}^{M} a_j\, (L\phi_j,\psi_i) = (s,\psi_i) \quad \forall\psi_i \in V_h \quad \begin{aligned} &V_h \subset H^0(a,b), \\ i = 1,2,\ldots,M-1, \qquad &S_h \subset H^2(a,b). \end{aligned}$$

In addition to either set of equations, the boundary conditions are given by

$$(3.1.19)\qquad \begin{aligned} &\sum_{j=0}^{M} a_j\, \phi_j(a) = \alpha, \\ &\sum_{j=0}^{M} a_j\, \phi_j(b) = \beta. \end{aligned}$$

THE DISCRETIZATION OF THE WEAK FORM

In this subsection we show how the system (3.1.17) can be described explicitly by means of the discrete equations over a single interval only. For brevity, we introduce the notation

$$b_{ij}(x) = c_0(x)\phi_j'(x)\psi_i'(x) + c_1(x)\phi_j'(x)\psi_i(x) + c_2(x)\phi_j(x)\psi_i(x).$$

The (i,j)-th entry of the discrete operator is

$$B(\phi_j,\psi_i) = \int_a^b b_{ij}(x)\,dx = \sum_{\ell=1}^{N} \int_{x_{\ell-1}}^{x_\ell} b_{ij}(x)\,dx$$

(3.1.20)

$$= \sum_{\ell=1}^{N} \int_0^1 b_{ij}(x_{\ell-1}+h_\ell\xi)\,d(\xi h_\ell) = \sum_{\ell=1}^{N} B_\ell(\phi_j,\psi_i),$$

where $B_\ell(\phi_j,\psi_i)$ denotes the contribution to $B(\phi_j,\psi_i)$ from the interval I_ℓ.

By their definition, all the natural basis functions, ϕ_j and ψ_i, can easily be put into the form

(3.1.21)
$$\begin{aligned} &\phi_{(\ell-1)z+j}\,(x_{\ell-1}+h_\ell\xi) = \Phi_j(\xi) \\ &\psi_{(\ell-1)z+i}\,(x_{\ell-1}+h_\ell\xi) = \Psi_i(\xi) \qquad 0 \le \xi \le 1 \\ &\qquad\qquad\qquad\qquad\qquad\qquad\quad i,j = 0,1,\dots,k \end{aligned}$$

(Where $z = k$ for the Lagrange, and $z = \frac{k+1}{2}$ for the Hermite spaces.) To rescale all functions to local coordinates on I_ℓ, we introduce a local notation for the coefficients $c_i(x)$ on I_ℓ:

(3.1.22)
$$\begin{aligned} C_{0,\ell}(\xi) &= c_0(x_{\ell-1}+\xi h_\ell)/h_\ell, \\ C_{1,\ell}(\xi) &= c_1(x_{\ell-1}+\xi h_\ell) \quad , \\ C_{2,\ell}(\xi) &= c_2(x_{\ell-1}+\xi h_\ell)\ h_\ell, \\ C_{3,\ell}(\xi) &= s\ (x_{\ell-1}+\xi h_\ell)\ h_\ell. \end{aligned}$$

If no confusion is possible we shall omit the index ℓ.
By (3.1.22) a single term $B_\ell(\phi_j,\psi_i)$ from equation (3.1.20) is brought into the form

(3.1.23)
$$\begin{aligned} B_\ell(\phi_{(\ell-1)z+j},\psi_{(\ell-1)z+i}) = \int_0^1 & C_{0,\ell}(\xi)\Phi_j'(\xi)\Psi_i'(\xi) + \\ & C_{1,\ell}(\xi)\Phi_j'(\xi)\Psi_i(\xi) + C_{2,\ell}(\xi)\Phi_j(\xi)\Psi_i(\xi)\ d\xi \overset{\text{def}}{=} \mathcal{B}(\Phi_j,\Psi_i). \end{aligned}$$

Thus, the discrete operator is composed of the N square matrices of order k+1, one for each interval I_ℓ, $\ell = 1,2,\dots,N$,

(3.1.24)
$$(B_\ell(\phi_{(\ell-1)z+j},\psi_{(\ell-1)z+i})), \qquad i,j = 0,1,\dots,k.$$

Analogously, the discretization of the right-hand side of the equation is characterized by a (k+1)-vector

$$(3.1.25)\qquad (s,\psi_{(\ell-1)z+1})_{L^2(I_\ell)} = \int_0^1 c_{3,\ell}(\xi)\Psi_i(\xi)d\xi \overset{\text{def}}{\equiv} S(\Psi_i), \quad i = 0,1,\dots,k.$$

Evaluation of the entries of the discrete operator

The entries of the discrete operator and right-hand side are all integrals and so their evaluation forms an essential part of the method. The evaluation should be efficient, but also accurate enough to guarantee that the order of accuracy, that can be obtained by the discretization, is indeed achieved. We treat two methods:

(1) evaluation by a quadrature rule, and
(2) evaluation by an interpolation rule.

Although the first method is the more efficient when it is properly applied, we shall treat both methods because we need a combination of both when exponentially fitted methods are considered.

1. Evaluation by a quadrature rule.

Let a t-th degree quadrature rule be characterized by a set of nodal points $0 \le \xi_0 < \xi_1 < \dots < \xi_L \le 1$ and a set of positive weights $\{w_i\}_{i\,=\,0,\dots,L}$ such that

$$(3.1.26)\qquad \int_0^1 p(x)dx = \sum_{i=0}^{L} p(\xi_i)w_i,$$

for all polynomials p(x) of degree $\le t$.

The entries $B(\Phi_j,\Psi_i)$ and $S(\Psi_i)$ of the discrete equation are approximated by

$$(3.1.27)\qquad \begin{aligned} B^*(\Phi_j,\Psi_i) = \sum_{k=0}^{L} \Big\{ & c_0(\xi_k)\ (w_k\Phi_j'\Psi_i')(\xi_k) + c_1(\xi_k)\ (w_k\Phi_j'\Psi_i)(\xi_k) + \\ & + c_2(\xi_k)\ (w_k\Phi_j\Psi_i)(\xi_k)\Big\}, \end{aligned}$$

$$(3.1.28)\qquad S^*(\Psi_i) = \sum_{k=0}^{L} c_3(\xi_k)\ (w_k\Psi_i)(\xi_k).$$

2. Evaluation by an interpolation rule.

Let a set of L+1 Lagrange interpolation polynomials $\{X_k\}_{k=0,1,\dots,L}$ of degree L be based on the nodal points $0 \le \xi_0 < \xi_1 < \dots < \xi_L \le 1$. The coefficient functions $c_i(\xi)$, $i = 0,1,2,3$, are replaced by their Lagrange interpolants and the resulting integrands in (3.1.17) are integrated exactly. Thus, the entries $\mathcal{B}(\Phi_j,\Psi_i)$ and $\mathcal{S}(\Psi_i)$ of the discrete equation are approximated by

$$\begin{aligned}\mathcal{B}^*(\Phi_j,\Psi_i) = \sum_{k=0}^{L}\Big\{c_0(\xi_k)\int_0^1 X_k\Phi_j'\Psi_i'd\xi + c_1(\xi_k)\int_0^1 X_k\Phi_j'\Psi_i d\xi + \\ + c_2(\xi_k)\int_0^1 X_k\Phi_j\Psi_i d\xi\Big\},\end{aligned} \tag{3.1.29}$$

$$\mathcal{S}^*(\Psi_i) = \sum_{k=0}^{L} c_3(\xi_k)\int_0^1 X_k\Psi_i d\xi. \tag{3.1.30}$$

REMARK. For each particular method, $(w_k\Phi_j'\Psi_i')(\xi_k)$ etc. in (3.1.27-28) or $\int_0^1 X_k\Phi_j'\Psi_i'd\xi$ etc. in (3.1.29-30) are simple real coefficients that can be computed beforehand.

An efficient implementation of Galerkin's method

We can use the freedom in the choice of a set of base-points $\{\xi_j^*\}_{j=0}^{k}$, (see eq. (3.1.15)) to minimize the amount of computational work. To this end we chose $\{\xi_i^*\}$ in agreement with the quadrature rule (3.1.26). Such an (L+1)-point quadrature rule is characterized by a set of nodal points $0 \le \xi_0 < \xi_1 < \ldots < \xi_L \le 1$, whereas the set $\{\xi_j^*\}$ contains k+1 distinct values $\xi_0^* < \xi_1^* < \ldots < \xi_k^*$ with the additional property $\xi_0^* = 0$, $\xi_k^* = 1$. The corresponding quadrature rule with $L = k$, $\xi_0 = 0$, $\xi_k = 1$ and optimal accuracy is the Lobatto k+1-point rule, which is accurate of degree $t = 2k - 1$ (cf. DAVIS & RABINOWITZ [1967]). If we set $\xi_i^* = \xi_i$, $0 \le i \le k$, an efficient evaluation of (3.1.27) and (3.1.28) is possible, viz.

$$\begin{aligned}\mathcal{B}^*(\Phi_j,\Phi_i) = \sum_{p=0}^{k}\{c_0(\xi_p)\ (w_p\Phi_j'\Phi_i')(\xi_p)\} + c_1(\xi_i)\ w_i\Phi_j'(\xi_i) + \\ + c_2(\xi_i)\ w_i\delta_{ij}\ ,\end{aligned} \tag{3.1.31}$$

$$\mathcal{S}^*(\Phi_i) = c_3(\xi_i)\ w_i\ . \tag{3.1.32}$$

Since $w_0 = w_k$, each i-th row can be divided by w_i. Thus, the amount of computational work is reduced considerably. This is even more true if $c_0(\xi)$ is a constant function.

An operations count for the equation

$$y'' + c_1y' + c_2y = c_3$$

shows that the construction of the discrete system using a (k+1)-point Lobatto method needs

k evaluations of c_1, c_2 and c_3
$2(k+1)^2 + k + 1$ multiplications
$(k+1)^2 + 2$ additions

for each interval $[x_{\ell-1}, x_\ell]$.
These numbers can be compared with those given in RUSSELL [1975] for finite differences and collocation methods.

In the following section we shall show that (k+1)-point Lobatto quadrature is sufficiently accurate to guarantee the optimal error bounds for the discretization with piecewise k-th degree polynomials; that is, the global error is $\mathcal{O}(h^{k+1})$ and the pointwise error on Π is $\mathcal{O}(h^{2k})$.

The advantage of the Galerkin method over the collocation method is that for the Galerkin method the continuity conditions for $y_h(x)$ are less severe and that a symmetric operator L leads to a symmetric discrete operator. An additional advantage of the efficient implementation (3.1.31-32) is that the term $c_2(x)y$ in the continuous operator L contributes to the entries of the discrete operator on the main diagonal only. In particular, this property is useful when problems with non-linear terms in y are considered.

EXAMPLE. To illustrate the Lobatto quadrature method, we give the contribution to the discrete equation from a single interval I_ℓ of length h, for k = 2 and for the equation

$$-y'' + fy' + gy = s$$

with constant coefficients:

$$(\mathcal{B}^*(\Phi_j, \Phi_i)) = \frac{1}{6h}\begin{pmatrix} 14 & -16 & 2 \\ -16 & 32 & -16 \\ 2 & -16 & 14 \end{pmatrix} + \frac{f}{6}\begin{pmatrix} -3 & 4 & -1 \\ -4 & 0 & 4 \\ 1 & -4 & 3 \end{pmatrix} + \frac{gh}{6}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$S^*(\Phi_i) = \frac{sh}{6}\begin{pmatrix}1\\4\\1\end{pmatrix}.$$

<u>The structure of the weak discrete operator</u>

The use of natural basis functions in $M^{0,k}(\Pi)$, $k = 1,2,\ldots$, or $M^{(k-1)/2,k}(\Pi)$, $k = 1,3,5,\ldots$, yields square matrices of order k+1 for the discretization of the operator B on each interval I_ℓ of the partition Π. The operator over the whole interval [a,b] is composed of N of these matrices (3.1.24).

In the case of $M^{0,k}(\Pi)$, we have $z = k$ and the discrete operator consists of N elementary matrices with one entry overlap on the main diagonal for each pair of neighboring intervals. The overlap element $B(\phi_{\ell z},\psi_{\ell z})$ is the sum of two overlapping elements:

(3.1.33) $$B(\phi_{\ell z},\psi_{\ell z}) = B_\ell(\phi_{(\ell-1)z+k},\psi_{(\ell-1)z+k}) + B_{\ell+1}(\phi_{\ell z+0},\psi_{\ell z+0}).$$

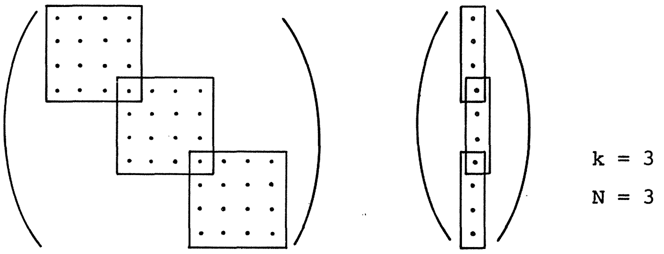


Fig. 3.1.1
The structure of a discrete operator
for Galerkin's method with $S_h = M^{0,k}(\Pi)$.

The particular structure of this discrete operator can be used to reduce the matrix to tridiagonal form during its construction. In this process, called *static condensation*, the intermediate unknown variables are eliminated and only the variables corresponding to $y(x_i)$, $x_i \in \Pi$, are computed by solving the resulting tridiagonal system.

In the case of $M^{m,2m+1}(\Pi)$, $m = 0,1,\ldots$, we have $z = (k+1)/2 = m+1$ and the discrete operator consists of N square (k+1)-th order matrices with an overlap of a square (m+1)-th order matrix for each two neighbouring intervals. The overlap elements $B(\phi_{\ell z+j},\psi_{\ell z+i})$, i,j, = 0,1,...,m, are the sum of the entries of two overlapping matrices

$$(3.1.34)\qquad B(\phi_{\ell z+j},\psi_{\ell z+i}) = B_\ell(\phi_{\ell z+j},\psi_{\ell z+i}) + B_{\ell+1}(\phi_{\ell z+j},\psi_{\ell z+i}), \quad i,j = 0,1,\ldots,m.$$

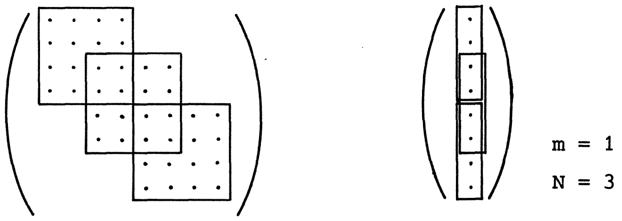


Fig. 3.1.2
The structure of a discrete operator
for Galerkin's method with $S_h = M^{m,2m+1}(\Pi)$.

THE DISCRETIZATION OF THE STRONG FORM

As we did for the weak form of the differential equation, we now go through the same process of constructing discrete operators for the strong form (3.1.18). Here, it turns out that a proper choice of the quadrature rule leads to collocation methods; i.e. we obtain methods that satisfy the original differential equation exactly at a number of specified points.

Let us consider eqs. (3.1.18)-(3.1.19). By application of a quadrature rule

$$(p,\psi_i) \approx \sum_k w_k\, p(\xi_k)\; \psi_i(\xi_k) ,$$

such that the matrix $(w_k\psi_i(\xi_k))$ is square and nonsingular, equation (3.1.18) is equivalent to

$$(3.1.35)\qquad \sum_{j=1}^{M-1} a_j\, L\phi_j(\xi_k) = s(\xi_k) , \quad \xi_i \in [a,b], \quad i = 1,2,\ldots,M-1 , \quad \phi_j \in C^1[a,b].$$

Here, $L\phi_j(\xi_k)$ should exist and hence necessary conditions with respect to continuity of $M^{m,k}(\Pi)$ are $m \geq 1$ and $k \geq 2$. The order of accuracy of these collocation methods is determined by the choice of $M^{m,k}(\Pi)$ and by the degree of accuracy of the quadrature rule. (see: RUSSELL & SHAMPINE [1972], DE BOOR & SCHWARTZ [1973]). We see that in discretizations of the strong form no evaluation of integrals is required, but the continuity conditions on the numerical solution $y_h(x)$ are stronger.

The structure of the strong discrete operator

Collocation by means of functions from $M^{m,k}(\Pi)$ yields k-m degrees of freedom on each subinterval of Π. Since, on each subinterval, an element of $M^{m,k}(\Pi)$ is determined by k+1 coefficients, the elementary matrix for each interval is of order k+1, but it has only k-m nonzero rows. The overlap between two matrices on the main diagonal of the discrete operator is (1+m)/2 entries. Thus, the matrix $L\phi_j(\xi_i)$ is a combination of disjoint rectangular submatrices.

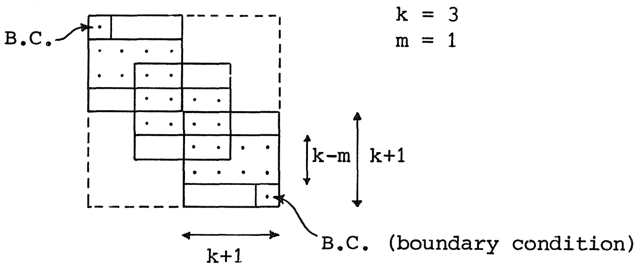


Fig. 3.1.3
The structure of a discrete operator
for a collocation method.

Hence, in the case of collocation over the space $M^{m,k}(\Pi)$, the discrete operator is composed of N characteristic $(k-m)\times(k+1)$ matrices that are the same for each interval, except for the values of the coefficients $c_0,\ldots,c_3$.

Examples of discretization

As an illustration we give four simple difference schemes for equation (1.1.1). The schemes all use $S_h = M^{0,1}(\Pi)$, $V_h = M_0^{0,1}(\Pi)$, i.e. Galerkin's method with piecewise linear functions. Only the way in which the integrals (3.1.23) and (3.1.25) are approximated is different. We apply successively the midpoint and trapezoidal rule for quadrature (3.1.27-28) and piecewise constant and piecewise linear functions for interpolation (3.1.29-30).

The schemes are described by their summand matrix $\mathcal{B}^*(\Phi_j,\Phi_i)$ and vector $\mathcal{S}^*(\Phi_i)$. We consider a characteristic interval $[x_\ell,x_{\ell+1}]$ and we set

$$\begin{aligned} &h = x_{\ell+1} - x_\ell, \quad \xi = (x-x_\ell)/h, \quad x_m = x_\ell + h/2; \\ &f_p = f(x_p), \quad g_p = g(x_p), \quad s_p = s(x_p), \quad p = \ell,m,\ell+1. \end{aligned} \tag{3.1.36}$$

The matrix $\mathcal{B}^*(\Phi_j,\Phi_i)$ and the vector $\mathcal{S}^*(\Phi_i)$ are respectively

i) by the midpoint quadrature rule:

$$\begin{pmatrix} -\frac{\varepsilon}{h} - \frac{1}{2}f_m + \frac{h}{4}g_m & \frac{\varepsilon}{h} + \frac{1}{2}f_m + \frac{h}{4}g_m \\ \frac{\varepsilon}{h} - \frac{1}{2}f_m + \frac{h}{4}g_m & -\frac{\varepsilon}{h} + \frac{1}{2}f_m + \frac{h}{4}g_m \end{pmatrix} \text{ and } \begin{pmatrix} \frac{h}{2}s_m \\ \frac{h}{2}s_m \end{pmatrix}; \tag{3.1.37}$$

ii) by piecewise constant interpolation:

$$\begin{pmatrix} -\frac{\varepsilon}{h} - \frac{1}{2}f_m + \frac{h}{3}g_m & \frac{\varepsilon}{h} + \frac{1}{2}f_m + \frac{h}{6}g_m \\ \frac{\varepsilon}{h} - \frac{1}{2}f_m + \frac{h}{6}g_m & -\frac{\varepsilon}{h} + \frac{1}{2}f_m + \frac{h}{3}g_m \end{pmatrix} \text{ and } \begin{pmatrix} \frac{h}{2}s_m \\ \frac{h}{2}s_m \end{pmatrix}; \tag{3.1.38}$$

iii) by the trapezoidal quadrature rule:

$$\begin{pmatrix} -\frac{\varepsilon}{h} - \frac{1}{2}f_\ell + \frac{h}{2}g_\ell & \frac{\varepsilon}{h} + \frac{1}{2}f_{\ell+1} \\ \frac{\varepsilon}{h} - \frac{1}{2}f_\ell & -\frac{\varepsilon}{h} + \frac{1}{2}f_{\ell+1} + \frac{h}{2}g_{\ell+1} \end{pmatrix} \text{ and } \begin{pmatrix} \frac{h}{2}s_\ell \\ \frac{h}{2}s_{\ell+1} \end{pmatrix}; \tag{3.1.39}$$

iv) by piecewise linear interpolation:

$$\begin{pmatrix} -\frac{\varepsilon}{h} - \frac{1}{6}(2f_\ell+f_{\ell+1}) + \frac{h}{12}(3g_\ell+g_{\ell+1}) & \frac{\varepsilon}{h} + \frac{1}{6}(2f_\ell+ f_{\ell+1}) + \frac{h}{12}(g_\ell+g_{\ell+1}) \\ +\frac{\varepsilon}{h} - \frac{1}{6}(f_\ell+2f_{\ell+1}) + \frac{h}{12}(g_\ell+g_{\ell+1}) & -\frac{\varepsilon}{h} + \frac{1}{6}(f_\ell+2f_{\ell+1}) + \frac{h}{12}(g_\ell+3g_{\ell+1}) \end{pmatrix}$$

$$\text{and} \tag{3.1.40}$$

$$\begin{pmatrix} \frac{h}{6}(2s_\ell+s_{\ell+1}) \\ \frac{h}{6}(s_\ell+2s_{\ell+1}) \end{pmatrix}.$$

We note that (3.1.39) is similar to, but not identical with, the common 3-point discretization for non-equidistant grids, as used e.g. by Pearson (cf. eq. (2.3.1)). For, multiplied by $\frac{2}{h+k}$ and written in the usual notation, (3.1.39) reads

$$(3.1.41) \qquad 2\varepsilon\frac{k(y_{\ell+1}-y_\ell)-h(y_\ell-y_{\ell-1})}{hk(h+k)} + f(x_\ell)\,\frac{y_{\ell+1}-y_{\ell-1}}{h+k} + g(x_\ell) = s(x_\ell)$$

where

$$k = x_\ell - x_{\ell-1}, \qquad h = x_{\ell+1} - x_\ell.$$

3.2. ERROR ESTIMATES

In this section we give lemmas for approximation by piecewise polynomials and treat the error of a weighted residual solution. Some of the results can be carried over to the strong form of the differential equation, but here we confine ourselves to the discretization of the weak form. Thus, we compare the functions y, y_h and y_h^* that satisfy the boundary conditions (1.1.1.b) and one of the variational equations

$$(3.2.1) \qquad y \in S = H^1(a,b), \quad B(y,v) = (s,v) \qquad \forall\, v \in V = H_0^1(a,b);$$

$$(3.2.2) \qquad y_h \in S_h \subset S, \qquad B(y_h,v_h) = (s,v_h) \qquad \forall\, v_h \in V_h \subset V;$$

or

$$(3.2.3) \qquad y_h^* \in S_h \subset S, \qquad B^*(y_h^*,v_h) = (s,v_h)^* \quad \forall\, v_h \in V_h.$$

Here $B^*(\cdot,\cdot)$ and $(\cdot,\cdot)^*$ denote approximations to $B(\cdot,\cdot)$ and $(\cdot,\cdot)$ obtained by quadrature as described in the preceding section.

Approximation in $H^{k,\pi}[a,b]$

First we introduce the linear space of functions that, together with their derivatives up to order k, are Lebesgue integrable over all subintervals of a partition of [a,b]. Thereafter we give lemmas with respect to approximation in these spaces.

DEFINITION

Let $\Pi = \{a = x_0 < x_1 < \ldots < x_N = b\}$ be a partition of $[a,b]$, then we introduce the norm $\|\cdot\|_{k,\pi}$, defined by

$$(3.2.4) \qquad \|v\|^2_{k,\pi} = \sum_{i=1}^{N} \|v\|^2_{H^k(x_{i-1},x_i)},$$

for all functions that have finite norms $\|\cdot\|_{H^k(x_{i-1},x_i)}$, $i = 1,2,\ldots,N$. It is easily verified that $\|\cdot\|_{k,\pi}$ is indeed a norm.

DEFINITION

By $H^{k,\pi}[a,b]$ we denote the linear space of functions, y, that have a finite norm $\|y\|_{k,\pi}$.

If $\Pi_1 \subset \Pi_2$ then $H^{k,\pi_1}[a,b] \subset H^{k,\pi_2}[a,b]$ and if $v \in H^{k,\pi_1}[a,b]$ then $\|v\|_{k,\pi_1} = \|v\|_{k,\pi_2}$.

Since $H^k(a,b) \subset H^{k,\pi}[a,b]$ for all Π, we have

$$(3.2.5) \qquad \|v\|_{j,\pi} = \|v\|_j, \quad j = 0,1,\ldots,k, \quad \text{for all } v \in H^k(a,b).$$

In the following lemmas we consider only *quasi-uniform partitions* of $[a,b]$ with a *meshwidth* h; i.e. we consider partitions Π for which there is a $\lambda > 0$ such that

$$\lambda h \le x_j - x_{j-1} \le h \quad \forall\, x_j \in \Pi,\ j > 0.$$

Often, we use sequences of quasi-uniform partitions $\{\Pi_i\}_{i=0}^{\infty}$ such that $\Pi_{i+1} \supset \Pi_i$ and $h_{i+1} \le h_i$. For a sequence of partitions such that $\Pi_m \supset \Pi_n$ if $m > n$ and $\lim_{i\to\infty} h_i = 0$, we sometimes use the notation

$$\lim_{h\to 0} \|\cdot\|_{k,\pi}$$

instead of

$$\lim_{i\to\infty} \|\cdot\|_{k,\pi_i}.$$

LEMMA 3.2.1. *Let* $\ell = 0$ *or* $\ell = 1$ *and let* $u \in H^{k+1,\pi_0}[a,b] \cap H^{\ell}(a,b)$, $k \ge \ell$, *then for all* $\Pi \supset \Pi_0$ *there exists a* $w_h \in M^{\ell-1,k}(\Pi)$ *such that*

$$(3.2.6)\qquad \|u-w_h\|_{m,\pi} \le K\|D^{k+1}u\|_{0,\pi}\, h^{k+1-m}, \qquad 0 \le m \le k+1,$$

where K *is a constant independent of* u *and* h.

PROOF.

i) $\ell = 0$. Restricted to a subinterval $I_i = [x_{i-1},x_i]$, $x_i \in \Pi$, $i > 0$, we know that $u \in H^{k+1}(x_{i-1},x_i)$ and, therefore, by Sobolev's lemma, $u \in C^k[x_{i-1},x_i]$. We take a k-th degree piecewise polynomial $w_h \in M^{-1,k}(\Pi)$, interpolating u at k+1 points in each subinterval I_i. Now a standard error estimate for the interpolants (cf. e.g. DAVIS [1963] Chapter 3 or CIARLET & RAVIART [1972] p.196 thm. 5) immediately yields (3.2.6).

ii) $\ell = 1$. The same arguments hold for $\ell = 1$, except that $w_h \in M^{0,k}(\Pi)$ should interpolate u at the gridpoints $x_i \in \Pi$ and at k-1 gridpoints inside each interval I_i. □

The following lemma is also frequently used in the computation of error estimates.

LEMMA 3.2.2. *Let* $\ell = 0$ *or* $\ell = 1$ *and let* $h_0 > 0$, *then there exists a* K *such that for all* $v \in M^{\ell-1,k}(\Pi)$, $k \ge \ell$,

$$(3.2.7)\qquad \|v\|_{k,\pi}\, h^{k-\ell} \le K\, \|v\|_{\ell},$$

provided that $h < h_0$; K *depends on* k, ℓ *and* h_0, *but is independent of* v *or* h.

PROOF.

i) $\ell = 0$. We first prove that

$$\int_{-1}^{+1} \left[\left(\frac{d}{d\xi}\right)^j w \right]^2 d\xi \le C \int_{-1}^{+1} w^2\, d\xi, \qquad 0 \le j \le k,$$

for all polynomials w of degree ≤k.

Let $p_i(x) = \sqrt{\frac{2i+1}{2}}\, P_i(x)$, $P_i(x)$ the Legendre polynomial of degree i, then

$$w(\xi) = \sum_{i=0}^{k} a_i p_i(\xi); \quad \int_{-1}^{+1} w^2 \, d\xi = \sum_{i=0}^{k} a_i^2;$$

$$\int_{-1}^{+1} \left[(\frac{d}{d\xi})^j \, w \right]^2 d\xi = \int_{-1}^{+1} (\sum_i a_i (\frac{d}{d\xi})^j \, p_i)^2 \, d\xi$$

$$= \int_{-1}^{+1} (\sum_i a_i \sum_{\ell=0}^{i-j} c_{ij\ell} \, p_\ell(\xi))^2 \, d\xi$$

$$\le \int_{-1}^{+1} \sum_i a_i^2 \sum_{m=0}^{k} (\sum_{\ell=0}^{m-j} c_{mj\ell} \, p_\ell(\xi))^2 \, d\xi$$

$$\le \max_{j=0,1,\dots,k} \left[\sum_{m=0}^{k} \sum_{\ell=0}^{m-j} c_{mj\ell}^2 \right] \int_{-1}^{+1} w^2 \, d\xi.$$

Therefore, there exists a $K > 0$, independent of h and $w \in M^{-1,k}(\Pi)$, such that

$$\int_{x_{i-1}}^{x_i} ((\frac{d}{dx})^j \, w)^2 \, dx \, h^{2k} \le K \int_{x_{i-1}}^{x_i} w^2 \, dx,$$

for $i = 1,2,\dots,N$ and $j = 0,1,\dots,k$.

Summation over $j = 0,1,\dots,k$ and $i = 1,2,\dots,N$ yields

$$\|w\|_{k,\pi}^2 \, h^{2k} \le (k+1) K \, \|w\|_0^2,$$

which proves the lemma for $\ell = 0$.

ii) $\ell = 1$. Following the same lines as in the proof for $\ell = 0$, but substituting $w = Dv$, we obtain

$$\sum_{j=1}^{k} \|D^j v\|_{0,\pi}^2 \, h^{2(k-1)} \le k \, K \, \|Dv\|_0^2.$$

Also

$$\|v\|_{0,\pi}^2 \, h^{2(k-1)} \le k \, K \, \|v\|_0^2,$$

if h is small enough, and hence

$$\|v\|^2_{k,\pi}\ h^{2(k-1)} \le k\ K\ \|v\|^2_1,$$

which proves the lemma for $\ell = 1$. □

Global error estimates for weighted residual methods

A comprehensive literature exists on error estimates for Ritz-Galerkin methods. An extension to more general weighted residual methods based on eq. (3.1.17) is found in BABUŠKA & AZIZ [1972]. We quote two essential therems, the proofs of which can be found in the paper mentioned. The proofs are recommended for reading because of their charming simplicity.

The first theorem is a generalization of the well-known Lax-Milgram theorem.

THEOREM 3.2.1. *Let* S *and* V *be two real Hilbert spaces with scalar product* $(\cdot,\cdot)_S$ *and* $(\cdot,\cdot)_V$ *respectively. Let* $B(u,v)$ *be a bilinear form* $S \times V \to \mathbb{R}$ *such that*

(3.2.8) $\quad \exists\ C_1 > 0 \quad \forall\ u \in S,\ v \in V \quad |B(u,v)| \le C_1\ \|u\|_S\ \|v\|_V,$

(3.2.9) $\quad \exists\ C_2 > 0 \quad \forall\ u \in S \quad \underset{v \neq 0}{\exists\ v \in V} \quad |B(u,v)| \ge C_2\ \|u\|_S\ \|v\|_V,$

(3.2.10) $\quad \forall\ v \in V,\ v \neq 0 \quad \exists\ u \in S \quad |B(u,v)| > 0,$

then

$$\forall\ f \in V' \quad \exists!\ u_0 \in S \quad \forall\ v \in V,\ \ B(u_0,v) = f(v),$$

where V' *denotes the linear space of bounded linear functionals on* V.
($\exists!$ denotes: there exists a unique...)

PROOF. See BABUŠKA & AZIZ [1972] pp.113-115.

The second theorem states that under certain conditions the weighted residual solution to a problem, found in a finite dimensional trial space, is essentially as good as the best possible approximation in that space, except for a certain factor that depends on the norm of the bilinear form B and on $D(S_h,V_h)$, *the coercivity of* $B(\cdot,\cdot)$ *with respect to* S_h *and* V_h (see page 56).

THEOREM 3.2.2. *Let the hypotheses of theorem* 3.2.1 *hold and let* S_h *and* V_h *be linear subspaces of* S *and* V *respectively, such that*

$$(3.2.11) \qquad \exists\; D(S_h,V_h) > 0 \quad \forall\; u \in S_h \quad \exists\; \underset{v \neq 0}{v \in V_h} \quad |B(u,v)| \geq D(S_h,V_h)\, \|u\|_S\, \|v\|_V$$

$$(3.2.12) \qquad \forall\; v \in V_h,\; v \neq 0 \quad \exists\; u \in S_h \quad |B(u,v)| > 0.$$

Let, for a given $f \in V'$, $u_0 \in S$ *denote the unique element such that* $B(u_0,v) = f(v)$, $\forall\; v \in V$; *and let*

$$\delta = \inf_{w \in S_h} \|u_0 - w\|_S.$$

Let $\hat{u}_0 \in S_h$ *denote the unique element such that* $B(\hat{u}_0,v) = f(v) \quad \forall\; v \in V_h$, *then*

$$(3.2.13) \qquad \|u_0 - \hat{u}_0\|_S \leq \left[1 + \frac{C_1}{D(S_h,V_h)}\right]\delta.$$

PROOF. See BABUŠKA & AZIZ [1972] p.187-188.

EXAMPLE 3.2.1. Let us consider the Galerkin method applied to equation (3.1.2) with homogeneous boundary conditions, then $B\colon H_0^1(a,b) \times H_0^1(a,b) \to \mathbb{R}$ is given by (3.1.4) and $S_h = V_h$. Therefore,

$$D(S_h, V_h) = \inf_{\substack{u \in S_h \\ u \neq 0}} \sup_{\substack{v \in V_h \\ v \neq 0}} \frac{|B(u,v)|}{\|u\|_1 \|v\|_1} \geq \inf_{u \in S_h} \frac{|B(u,u)|}{\|u\|_1^2} = \sigma,$$

where σ is the coercivity constant of $B(\cdot,\cdot)$, see page 56, and thus

$$(3.2.14) \qquad \|u_0 - \hat{u}_0\|_1 \leq \left[1 + \frac{C_1}{\sigma}\right] \inf_{s \in S_h} \|u_0 - s\|_1.$$

We derive the following lemma 3.2.4 in order to show how the asymmetry of B, caused by $c_1(x)$, gives rise to the requirement of a fine enough mesh in the Galerkin method. First we need a definition and a lemma to determine the relation between $\|y-y_h\|_0$ and $\|y-y_h\|_1$.

DEFINITION

Let B be a bilinear form $H \times H \to \mathbb{R}$, then *the symmetric part of* B is defined by

$$B_{\mathrm{sym}}(u,v) = \frac{1}{2}B(u,v) + \frac{1}{2}B(v,u).$$

LEMMA 3.2.3. *Let* B *be the bilinear form* (3.1.4), *let* y *be the solution of* (3.2.1) *and* y_h *the solution of* (3.2.2). *Furthermore, let* V_h *be such that for any* $\phi \in H^2(a,b) \cap H_0^1(a,b)$

$$\inf_{v \in V_h} \|\phi-v\|_0 + \varepsilon\|\phi'-v'\|_0 \leq M(h)\ \|L^T\phi\|_0, \tag{3.2.15}$$

where

$$\lim_{h \to 0} M(h) = 0,$$

then

$$\|y-y_h\|_0 \leq \|y-y_h\|_1\ M(h)\ \max(\frac{\|c_0\|_\infty}{\varepsilon},\ \|c_1\|_\infty+\|c_2\|_\infty). \tag{3.2.16}$$

PROOF. Set $\zeta = y - y_h$ and let ϕ denote the solution of

$$L^T\phi = \zeta,$$

then ϕ satisfies

$$B(v,\phi) = (v,\zeta) \qquad \text{for all } v \in H_0^1(a,b).$$

For all $v_h \in V_h$ $B(\zeta,v_h) = 0$ and, hence,

$$\begin{aligned}
\|\zeta\|_0^2 &= B(\zeta,\phi) = B(\zeta,\phi-v_h) \\
&\leq |(c_0\zeta',\phi'-v_h')| + |(c_1\zeta',\phi-v_h)| + |(c_2\zeta,\phi-v_h)| \\
&\leq \|c_0\|_\infty\|\zeta'\|_0\|\phi'-v_h'\|_0 + \|c_1\|_\infty\|\zeta'\|_0\|\phi-v_h\|_0 + \|c_2\|_\infty\|\zeta\|_0\|\phi-v_h\|_0 \\
&\leq \|\zeta\|_1\ \{\frac{\|c_0\|_\infty}{\varepsilon}\ \varepsilon\|\phi'-v_h'\|_0 + (\|c_1\|_\infty+\|c_2\|_\infty)\|\phi-v_h\|_0\} \\
&\leq \|\zeta\|_1\ \max(\frac{\|c_0\|_\infty}{\varepsilon},\|c_1\|_\infty+\|c_2\|_\infty)\cdot\ \{\varepsilon\|\phi'-v_h'\|_0+\|\phi-v_h\|_0\}
\end{aligned}$$

for all $v_h \in V_h$.

Hence,

$$\|\zeta\|_0^2 \le \|\zeta\|_1 \max(\frac{\|c_0\|_\infty}{\varepsilon}, \|c_1\|_\infty + \|c_2\|_\infty) M(h) \|\zeta\|_0 ,$$

which proves the lemma. □

<u>COROLLARY</u>. It is known by the regularity of the solution that there is a constant K, depending on L, such that $\|\phi\|_2 \le K\|L^T\phi\|_0$; now, by lemma 3.2.1 it is clear that, if $V_h \supset M^{0,1}(\Pi)$,

$$\inf_{v\in V_h} \|\phi-v\|_0 + \varepsilon\|\phi'-v'\|_0 \le (1+\varepsilon) \inf_{v\in V_h} \|\phi-v\|_1 \le K_1 h \|\phi\|_2 \le$$

$$\le K_2 h \|L^T\phi\| .$$

So we obtain, by lemma 3.2.3,

$$\|y-y_h\|_0 \le Ch \|y-y_h\|_1 ,$$

where C depends on the operator L, but is independent of y and h.

<u>LEMMA 3.2.4</u>. *Let the conditions of lemma* 3.2.3 *be satisfied and let* S_h *and* V_h *be finite dimensional subspaces of* $H_0^1(a,b)$ *such that*

$$(3.2.17) \qquad \inf_{\substack{s\in H_0^1(a,b),s\neq 0 \\ B(s,w)=0\ \forall w\in V_h}} \ \sup_{\substack{v\in H_0^1(a,b),v\neq 0 \\ B(v,u)=0\ \forall u\in S_h}} \frac{|B_{sym}(s,v)|}{\|s\|_1 \ \|v\|_1} \overset{\text{def}}{=} D^*(S_h,V_h) > 0$$

then, if h *is small enough,*

$$(3.2.18) \qquad \|y-y_h\|_1 \le \frac{K \inf_{s\in S_h} \|y-s\|_1}{D^*(S_h,V_h) - C_4C_5M(h)} ,$$

where

$$C_4 = \max(\frac{\|c_0\|_\infty}{\varepsilon}, \|c_1\|_\infty + \|c_2\|_\infty) ,$$

$$C_5 = \|c_1\|_\infty + \frac{1}{2}\|c_1'\|_\infty .$$

PROOF. Set $\zeta = y-y_h$, then there exists a $\psi \in V$ such that $\psi \neq 0$, $B(\psi,s_h) = 0$ for all $s_h \in S_h$, and

$$\begin{aligned}
D^*(S_h,V_h)\ \|\zeta\|_1\ \|\psi\|_1 &\le |B_{sym}(\zeta,\psi)| \le \\
&\le \frac{1}{2}|B(\zeta,\psi) - B(\psi,\zeta)| + |B(\psi,\zeta)| \\
&\le C_5\ \|\zeta\|_0\ \|\psi\|_1 \qquad + |B(\psi,\zeta-s_h)| \\
&\le C_5\ \|\zeta\|_0\ \|\psi\|_1 \qquad + |B(\psi,y-s_h)| \\
&\le C_5C_4M(h)\ \|\zeta\|_1\ \|\psi\|_1 + C_1\ \|\psi\|_1\ \|y-s_h\|_1 \\
D^*(S_h,V_h)\ \|\zeta\|_1 &\le C_4C_5M(h)\ \|\zeta\|_1 + C_1 \inf_{s_h \in S_h} \|y-s_h\|_1 .
\end{aligned}$$

Thus, if $C_4C_5M(h) < D^*(S_h,V_h)$, we have

$$\|\zeta\|_1 \le \frac{C_1 \inf\limits_{s \in S} \|y-s\|_1}{D^*(S_h,V_h) - C_4C_5M(h)} . \qquad \square$$

REMARK. We see that, for a symmetric problem (i.e. $c_1(x) \equiv 0$), we have $C_5 = 0$ and the requirement $C_4C_5M(h) < D^*(S_h,V_h)$ is automatically satisfied.

EXAMPLE 3.2.2. Let us again consider the Galerkin method applied to equation (3.1.2) with homogeneous boundary conditions, then $S = V = H_0^1(a,b)$ and $S_h = V_h$.

Therefore,

$$\begin{aligned}
D^*(S_h,V_h) &= \inf_{\substack{s \in H_0^1(a,b),\, s \neq 0 \\ B(s,w)=0\ \ \forall w \in S_h}} \ \sup_{\substack{v \in H_0^1(a,b),\, v \neq 0 \\ B(v,u)=0\ \ \forall u \in S_h}} \frac{|B_{sym}(s,v)|}{\|s\|_1\ \|v\|_1} \\
(3.2.19) \qquad &\ge \inf_{\substack{s \in H_0^1(a,b),\, s \neq 0 \\ B(s,v)=0\ \ \forall v \in S_h}} \frac{|B_{sym}(s,s)|}{\|s\|_1^2} \\
&\ge \inf_{s \neq 0} \frac{|B_{sym}(s,s)|}{\|s\|_1^2} = \sigma^* ;
\end{aligned}$$

σ^* is the coercivity constant of the symmetric part of the operator B and, hence, is independent of $c_1(x)$. Large values of $|c_1(x)|$ are represented in C_4, C_5 and $M(h)$ and, in applying the estimate, must be compensated by small values of h.

Green's function and the discrete Green's function

Green's function $G(x,\xi)$ with respect to operator L (eq. (3.1.2)) and homogeneous boundary conditions on the interval $[a,b]$, is the function defined on the closed square $a \le x,\xi \le b$ by

$$(3.2.20)\qquad \begin{array}{ll} 1. & G(x,\cdot) \in H_0^1(a,b) \cap C^2((a,x) \cup (x,b)), \\ 2. & L^T G(x,\cdot) \equiv 0 \text{ on } (a,x) \cup (x,b), \\ 3. & \underset{\xi=x}{\text{jmp}} \frac{\partial}{\partial\xi} G(x,\xi) = + \frac{1}{c_0(x)} . \end{array}$$

The following two properties of Green's function are classical (cf. e.g. YOSIDA [1960]).

i) The solution of the two-point boundary-value problem

$$(3.2.21)\qquad Ly = s \quad \text{on } [a,b],\ s \in L^2(a,b),$$

with homogeneous boundary conditions is given by

$$(3.2.22)\qquad y(x) = - \int_a^b G(x,\xi)\, s(\xi)\, d\xi.$$

ii) Green's function can be constructed from two fixed solutions ϕ_1 and ϕ_2 of $L^T\phi = 0$. Let ϕ_1, ϕ_2 be defined on $[a,b]$ by

$$L^T\phi_1 = 0,\ \phi_1(a) = 0,\ \phi_1'(a) = 1,$$
$$L^T\phi_2 = 0,\ \phi_2(b) = 0,\ \phi_2'(b) = 1,$$

then

$$(3.2.23)\qquad G(x,\xi) = \frac{\underline{\text{if}}\ \xi<x\ \underline{\text{then}}\ \phi_1(\xi)\phi_2(x)\ \underline{\text{else}}\ \phi_1(x)\phi_2(\xi)}{c_0(x)\ (\phi_1(x)\phi_2'(x) - \phi_1'(x)\phi_2(x))} .$$

Note that the denominator $z(x) = c_0(x)\ (\phi_1(x)\phi_2'(x) - \phi_1'(x)\phi_2(x))$ satisfies the differential equation

$$c_0 z' + c_1 z = 0.$$

Therefore, either $z \equiv 0$ or $|z| > 0$ on $[a,b]$. If $z \equiv 0$ then ϕ_1 and ϕ_2 are linearly dependent, the homogeneous problem $L^T\phi = 0$, $\phi(a) = \phi(b) = 0$, has a nontrivial solution and Green's function is not defined. Otherwise $L\phi = s$ has a unique solution given by

$$\phi(x) = -(G(x,\bullet),s).$$

LEMMA 3.2.5. *Let there exist a unique solution* y_h *to the problem* (3.2.2), *then there exists a discrete Green's function* $G_h(x,\xi)$ *relative to the operator* B *and to the spaces* S_h *and* V_h, *such that* y_h *is given by*

$$y_h(x) = -\int_a^b G_h(x,\xi)\; s(\xi)\; d\xi. \tag{3.2.24}$$

PROOF. Let $\{\phi_j\}$ and $\{\psi_i\}$ be bases in S_h and V_h respectively, then $y_h = \sum_j a_j\phi_j$ is determined by

$$\sum_j a_j\; B(\phi_j,\psi_i) = (s,\psi_i).$$

Since the matrix $B(\phi_j,\psi_i)$ is non-singular, it has a unique inverse, the entries of which are denoted by $B^{-1}_{i,j}$. It follows that

$$\begin{aligned} y_h(x) &= \sum_j \phi_j(x) \sum_i B^{-1}_{i,j}\; (s,\psi_i) \\ &= \int_a^b s(t) \sum_{ij} \phi_j(x)\; B^{-1}_{i,j}\; \psi_i(t)\; dt. \end{aligned}$$

Thus, we obtain the form (3.2.24), with

$$G_h(x,\xi) = -\sum_{ij} \phi_j(x) B^{-1}_{i,j}\; \psi_i(\xi). \qquad \square$$

Pointwise error estimates

In theorem 3.2.2 it was shown that interpolating properties of the space S_h carry over to the global error of a weighted residual approximation; in this subsection we show that properties of V_h can produce additional pointwise accuracy. This phenomenon, called *superconvergence*, has been studied by DOUGLAS & DUPONT [1974] for Galerkin methods.

THEOREM 3.2.3. *Let* $B(\bullet,\bullet)$ *be the bilinear form associated with* L *(eq.* (3.1.2)*) and let* $y \in S = H^1(a,b)$ *be the unique solution to the varia-*

tional problem (3.2.1). *Let* $V_h \subset V = H_0^1(a,b)$ *be such that the conditions of theorem* 3.2.2 *are satisfied and let* $y_h \in S_h \subset S$ *be the solution of the corresponding discrete variational problem* (3.2.2), *then a pointwise error-bound is given by*

$$(3.2.25) \qquad |(y-y_h)(x)| \le K\|y-y_h\|_1 \inf_{v\in V_h} \|G(x,\cdot)-v\|_1,$$

where $G(x,\xi)$ *is Green's function.*

<u>PROOF</u>. Set $\zeta = y - y_h$. Now $\zeta \in H_0^1(a,b)$ and by Sobolev's lemma, $\zeta \in C^0[a,b]$. Moreover,

$$B(\zeta,v) = 0 \qquad \text{for all } v \in V_h.$$

We know that $G_\xi(x,\cdot)$ has a discontinuity at $\xi = x$; therefore, by (3.1.5) and (3.1.10)

$$\begin{aligned}\zeta(x) &= -[c_0\zeta G_\xi(x,\cdot) + c_1\zeta G(x,\cdot)]_\pi \\ &= -B(\zeta,G(x,\cdot))+(\zeta,L^TG)_{0,\pi} \\ &= -B(\zeta,G(x,\cdot)) = -B(\zeta,G(x,\cdot)-v)\end{aligned}$$

for all $v \in V_h$.
Therefore,

$$|\zeta(x)| = |B(\zeta,G(x,\cdot)-v)| \le K \|\zeta\|_1 \|G(x,\cdot) - v\|_1$$

for all $v \in V_h$. □

<u>COROLLARY</u>. Applying lemma 3.2.1 and the estimate (3.2.25) we obtain, if $y \in H^{k+1,\pi}[a,b]$, the pointwise error estimate

$$(3.2.26) \qquad \|y - y_h\|_{\pi,\infty} = \mathcal{O}(h^{k+p}) \qquad \text{for } h \to 0$$

if $S_h \supset M^{0,k}(\Pi)$, $V_h \supset M^{0,p}(\Pi)$ and $G(x_i,\cdot) \in H^{p+1,\pi}[a,b]$ for all $x_i \in \Pi$. Application of the corollary of lemma 3.2.3 yields the global estimate in the L^2-norm

$$(3.2.27) \qquad \|y - y_h\|_0 = \mathcal{O}(h^{k+1}).$$

THEOREM 3.2.4. *Let* $B(\cdot,\cdot)$ *be the bilinear form associated with* L *(eq.* (3.1.2)*) and let* $y \in S = H^1(a,b)$ *be the unique solution to the variational problem* (3.2.1)*, where* $V = H_0^1(a,b)$. *Let the conditions of theorem* 3.2.2 *be satisfied and let* $y_h \in S_h \subset S$ *be the solution of the corresponding discrete variational problem* (3.2.2).
Let $\psi_i \in V_h$ *be such that*

$$\psi_i \in H_0^1(a,b) \cap C^1((a,x_i) \cup (x_i,b))$$
$$\text{jmp}\ \psi_i'(x_i) \neq 0,$$

then

$$(y-y_h)(x_i) = \frac{(y-y_h,L^T\psi_i)_{0,\pi}}{c_0(x_i)\ \text{jmp}\ \psi_i'(x_i)} .$$

PROOF. Set $\zeta = y - y_h$ then $\zeta \in H_0^1(a,b)$ and $B(\zeta,v) = 0$ for all $v \in V_h$. Hence, $\zeta \in C^0[a,b]$ and by (3.1.5) and (3.1.10)

$$\begin{aligned} 0 = B(\zeta,\psi_i) &= (\zeta,L^T\psi_i)_{0,\pi} + [c_0\zeta\psi_i' + c_1\zeta\psi_i]_\pi \\ &= (\zeta,L^T\psi_i)_{0,\pi} - c_0(x_i)\ \zeta(x_i)\ \text{jmp}\ \psi_i'(x_i). \end{aligned}$$

Consequently,

$$\zeta(x_i) = \frac{(\zeta,L^T\psi_i)_{0,\pi}}{c_0(x_i)\ \text{jmp}\ \psi_i'(x_i)} . \qquad \square$$

COROLLARY. The above expression for $y(x_i) - y_h(x_i)$ leads immediately to the following pointwise error bound for the discretization (3.2.2),

$$(3.2.28) \qquad |y(x_i)-y_h(x_i)| \le \frac{\|y-y_h\|_0}{|c_0(x_i)|} \inf_{\psi_i} \frac{\|L^T\psi_i\|_{0,\pi}}{|\text{jmp}\ \psi_i'(x_i)|} .$$

REMARK. The estimate (3.2.28) can also be derived for the solution obtained by discretization of the strong form.

REMARK. If there exists a non-trivial $\psi_i \in V_h$ that satisfies $L^T\psi_i = 0$ on $(a,x_i) \cup (x_i,b)$ then $y(x_i) = y_h(x_i)$. Since each ψ_i that satisfies this condition is a scalar multiple of $G(x_i,\cdot)$, this conclusion could also be derived from theorem 3.2.3.

Quadrature and error estimates

In the following theorem we prove that, when k-th degree piecewise polynomials are used for S_h and V_h, a (2k)-th order quadrature rule is sufficiently accurate to guarantee the same order of accuracy for y_h^* as for y_h. Thus the theorem gives a justification for the use of (k+1)-point Lobatto quadrature as described in (3.1.31)-(3.1.32).

THEOREM 3.2.5. *Let* Π *be a quasiuniform partition. Let* $y \in S = H^1(a,b)$, *the solution of equation* (3.2.1) *with* $s \in H^{2k,\pi}[a,b]$, *be approximated by* $y_h^* \in S_h = M^{0,k}(\Pi)$, *which is determined by* (3.2.3) *where* $V_h = M_0^{0,k}(\Pi)$ *and let the operator* B *be such that the hypotheses of theorem* 3.2.2 *hold. Let* $B^*(\cdot,\cdot)$ *and* $(\cdot,\cdot)^*$ *be computed by a* (2k)*-th order quadrature rule, then the error estimates*

$$\|y - y_h^*\|_1 = O(h^k) \tag{3.2.29}$$

and

$$\|y - y_h^*\|_{\pi,\infty} = O(h^{2k}) \tag{3.2.30}$$

hold if h *is sufficiently small.*

PROOF. For all $v \in V_h$

$$\begin{aligned} |B(y_h - y_h^*,v)| &\le |B(y_h,v) - B^*(y_h^*,v)| + |B(y_h^*,v) - B^*(y_h^*,v)| \\ &\le |(s,v) - (s,v)^*| + |B(y_h^*,v) + B^*(y_h^*,v)| \\ &\le C\ \|s\|_{2k,\pi}\ \|v\|_{k,\pi}\ h^{2k} + C\ \|y_h^*\|_{k,\pi}\ \|v\|_{k,\pi}\ h^{2k}. \end{aligned} \tag{3.2.31}$$

B is such that there exist a $D(S_h,V_h) > 0$ and a $v \in V_h$ such that

$$D(S_h,V_h)\ \|y_h - y_h^*\|_1\ \|v\|_1 < |B(y_h - y_h^*,v)|.$$

Hence, by lemma 3.2.2, if h is small enough,

$$D(S_h,V_h)\ \|y_h - y_h^*\|_1\ \|v\|_1 \le |B(y_h-y_h^*,v)| \le$$

$$\le C\ \|s\|_{2k,\pi}\ \|v\|_1\ h^{k+1} + C\ \|y_h^*\|_{k,\pi}\ \|v\|_1\ h^{k+1};$$

$$D(S_h,V_h)\ \|y_h - y_h^*\|_1 \le C\ \|s\|_{2k,\pi}\ h^{k+1} + C\ \|y_h^*\|_{k,\pi}\ h^{k+1}$$

$$\le C[\|s\|_{2k,\pi}+\|y\|_{k,\pi}]h^{k+1} + C\|y-y_h\|_{k,\pi}\ h^{k+1} + C\|y_h-y_h^*\|_{k,\pi}\ h^{k+1} \tag{3.2.32}$$

$$\le C[\|s\|_{2k,\pi}+\|y\|_{k,\pi}]h^{k+1} + C\|y\|_{k+1,\pi}\ h^{k+2} + C\|y_h-y_h^*\|_1\ h^2.$$

So, if h is small enough,

$$\|y_h - y_h^*\|_1 \le \frac{C[\|s\|_{2k,\pi}+\|y\|_{k,\pi}]h^{k+1}+\mathcal{O}(h^{k+2})}{D(S_h,V_h)-Ch^2}\ . \tag{3.2.33}$$

Combination of this inequality with the results of lemma 3.2.1 and theorem 3.2.2 yields the estimate (3.2.29).

Now let $G(x,\xi)$ be Green's function corresponding to L; let G^i denote $G(x_i,\cdot)$, then for all $v \in M_0^{0,k}(\Pi)$

$$|y_h(x_i) - y_h^*(x_i)| \le |B(y_h - y_h^*,G^i)| \le$$

$$\le |B(y_h-y_h^*,\ G^i-v)| + |B(y_h-y_h^*,v)| \tag{3.2.34}$$

$$\le K\|y_h - y_h^*\|_1\ \|G^i - v\|_1 + C\|s\|_{2k,\pi}\ \|v\|_{k,\pi}\ h^{2k} + C\ \|y_h^*\|_{k,\pi}\ \|v\|_{k,\pi}\ h^{2k}.$$

If h is small enough, v can be selected such that

$$\|G^i - v\|_1 \le C\ \|D^{k+1}\ G^i\|\ h^k \text{ and } \|v\|_{k,\pi} \le 2\ \|G^i\|_k .$$

These inequalities, together with (3.2.34), (3.2.29) and the application of lemma 3.2.2 yield (3.2.30). □

3.3. STANDARD GLOBAL METHODS APPLIED TO SINGULAR PERTURBATION PROBLEMS

At first sight, it might be expected that none of the global methods mentioned this far, will be able to handle singular perturbation problems properly. Indeed, in all discrete operators, the contribution due to the second derivative is insignificant as compared with contributions from the other terms of the differential operator. Still both boundary conditions are imposed with the same strength. Thus, in the actual discrete operator no information remains to determine which boundary condition has to be respected. Nevertheless, it is meaningful to study to what extent the various methods may succeed. To this end we investigate their behaviour for the model problem

$$(3.3.1) \qquad \begin{aligned} &\varepsilon y'' + y' = 0, \\ &y(0) = 0,\ y(1) = 1, \end{aligned}$$

on a uniform mesh.

We consider respectively Galerkin's method, collocation, reduction to a system of first order equations, least squares and the Ritz-Galerkin method.

Galerkin's method

First we consider Galerkin methods. If we take $S_h = M^{0,1}(\Pi)$, $V_h = M_0^{0,1}(\Pi)$, it follows from (3.1.38) that the discrete operator coincides with the one obtained with central differences. This operator was studied thoroughly in section 2.1. Discrete operators obtained by means of higher order Lagrange spaces $M^{0,k}(\Pi)$ will give better error bounds. This is a consequence of (3.2.14) and of the relation

$$M^{0,k}(\Pi) \supset M^{0,\ell}(\Pi) \qquad \text{if } k \geq \ell,$$

whence

$$\inf_{v \in M^{0,k}(\Pi)} \|y - v\| \leq \inf_{v \in M^{0,\ell}(\Pi)} \|y - v\|.$$

Nevertheless, small values of ε still yield bad estimates since, in equation (3.2.14), $\sigma = \mathcal{O}(\varepsilon)$.

Whereas a Galerkin method improves when Lagrange spaces of higher order are used, it may degrade with the use of the higher order Hermite spaces $H^{m,2m+1}(\Pi)$. For the latter type, lower order spaces are not subspaces of the higher order ones. Thus the approximation in the higher order spaces may be worse. In particular this will be the case if an approximand is not sufficiently smooth. As was shown in chapter 1,y(x) is not in general smooth for small ε. Hence the factor $\|y - y_h\|_1$ in the error bounds (3.2.14) and (3.2.25) can be larger for larger m. This is even more likely to occur for the factor $\|G(x_i,\cdot) - v_h\|$ in (3.2.25), since $G(x_i,\cdot)$ has a discontinuous derivative at $x = x_i$. Of course, a larger error bound does not imply that the error will in fact be larger. However, an actual computation for the problem (3.3.1) shows that for $M^{2,5}(\Pi)$ the error is larger than for $M^{1,3}(\Pi)$, if ε/h is small. This is illustrated in the tables 3.3.1 and 3.3.2.

	N = 1/h	4	8	16	32	64
S_h	ε					
$M^{0,1}(\Pi)$	1.0	6.2(-4)	1.6(-4)	3.9(-5)	9.8(-6)	2.5(-6)
	1.0(-2)	2.9	8.7(-1)	5.2(-1)	2.6(-1)	8.7(-2)
	1.0(-4)	3.1(+2)	7.8(+1)	1.9(+1)	4.9	1.5
	1.0(-6)	3.1(+4)	7.8(+3)	2.0(+3)	4.9(+2)	1.2(+2)
$M^{1,3}(\Pi)$	1.0	5.3(-6)	4.0(-7)	2.9(-8)	1.9(-9)	1.2(-10)
	1.0(-2)	2.3(-1)	7.9(-2)	3.9(-2)	1.1(-2)	1.7(-3)
	1.0(-4)	5.2(-1)	5.1(-1)	4.8(-1)	3.7(-1)	1.2(-1)
	1.0(-6)	5.2(-1)	5.2(-1)	5.2(-1)	5.2(-1)	5.2(-1)
$M^{2,5}(\Pi)$	1.0	4.5(-10)	7.9(-12)	8.5(-14)	6.4(-14)	4.8(-13)
	1.0(-2)	1.5(-1)	3.5(-2)	3.8(-3)	1.8(-4)	6.1(-6)
	1.0(-4)	4.6(+1)	1.6(+1)	4.8	1.3	4.2(-1)
	1.0(-6)	4.7(+3)	1.7(+3)	5.2(+2)	1.5(+2)	4.0(+1)
$M^{0,3}(\Pi)$	1.0	2.9(-10)	3.8(-12)	1.1(-12)	3.8(-12)	1.6(-11)
	1.0(-2)	4.1(-1)	1.5(-1)	2.7(-2)	1.8(-3)	5.2(-5)
	1.0(-4)	5.2(+1)	1.3(+1)	3.3	1.1	8.6(-1)
	1.0(-6)	5.2(+3)	1.3(+3)	3.3(+2)	8.1(+1)	2.0(+1)
$M^{0,5}(\Pi)$	1.0	4.2(-10)	2.5(-9)	1.7(-8)	8.3(-8)	3.6(-7)
	1.0(-2)	9.4(-2)	1.0(-2)	2.6(-4)	1.5(-6)	1.6(-9)
	1.0(-4)	2.1(+1)	5.1	1.4	8.3(-1)	6.8(-1)
	1.0(-6)	2.1(+3)	5.2(+2)	1.3(+2)	3.3(+1)	8.2

Table 3.3.1. Pointwise errors $\|y - y_h\|_{\pi,\infty}$ for problem (3.3.1). The Galerkin method (3.1.12) has been used for various spaces $M^{m,k}(\Pi)$ on a uniform mesh Π.

	N = 1/h	3	7	15	31	63
S_h	ε					
$M^{0,1}(\Pi)$	1.0	1.1(-3)	2.1(-4)	4.5(-5)	1.0(-5)	2.5(-6)
	1.0(-2)	8.7(-1)	6.2(-1)	5.4(-1)	2.7(-1)	8.9(-2)
	1.0(-4)	1.0	1.0	9.9(-1)	9.9(-1)	9.6(-1)
	1.0(-6)	1.0	1.0	1.0	1.0	1.0
$M^{1,3}(\Pi)$	1.0	1.5(-5)	6.7(-7)	3.8(-8)	2.2(-9)	1.3(-10)
	1.0(-2)	3.2(-1)	9.3(-2)	4.3(-2)	1.1(-2)	1.7(-3)
	1.0(-4)	5.2(-1)	5.1(-1)	4.9(-1)	3.8(-1)	1.3(-1)
	1.0(-6)	5.2(-1)	5.2(-1)	5.2(-1)	5.2(-1)	5.2(-1)
$M^{2,5}(\Pi)$	1.0	2.4(-9)	1.7(-11)	1.2(-13)	1.2(-13)	5.6(-13)
	1.0(-2)	2.7(-1)	5.0(-2)	4.9(-3)	2.1(-4)	6.6(-6)
	1.0(-4)	7.4(-1)	7.3(-1)	7.0(-1)	6.0(-1)	4.2(-1)
	1.0(-6)	7.5(-1)	7.5(-1)	7.5(-1)	7.5(-1)	7.5(-1)
$M^{0,3}(\Pi)$	1.0	1.6(-9)	9.6(-12)	1.3(-12)	9.9(-13)	4.2(-11)
	1.0(-2)	3.3(-1)	1.9(-1)	3.3(-2)	2.1(-3)	5.7(-5)
	1.0(-4)	9.9(-1)	9.8(-1)	9.6(-1)	8.7(-1)	8.6(-1)
	1.0(-6)	1.0	1.0	1.0	1.0	1.0
$M^{0,5}(\Pi)$	1.0	5.9(-10)	1.6(-9)	1.5(-8)	7.7(-8)	3.5(-7)
	1.0(-2)	1.6(-1)	1.8(-2)	3.9(-4)	2.0(-6)	2.1(-9)
	1.0(-4)	9.8(-1)	9.5(-1)	8.7(-1)	8.2(-1)	6.9(-1)
	1.0(-6)	1.0	1.0	1.0	1.0	1.0

Table 3.3.2 Numerical results as in Table 3.3.1. However, for this table an odd number of subintervals has been used.

Collocation

The same model problem (3.3.1) is used to show how collocation fails for $\varepsilon/h \to 0$. We consider the simplest function space that can be used for collocation: $S_h = M^{1,3}(\Pi)$. On each interval I_i, we take the two collocation points at $x = x_{i-1} + ch_i$ and at $x = x_i - ch_i$, $0 \leq c < \frac{1}{2}$. Then the characteristic rectangular submatrix for the discretization of $Ly = \varepsilon y'' + y'$ is

(3.3.2) $$-\frac{1}{h}\begin{pmatrix} 6c(1-c)\;+ & -(1-c)(1-3c)h+ & -6c(1-c)\;- & c(2-3c)h+ \\ +6(\frac{\varepsilon}{h})(1-2c) & +2(\frac{\varepsilon}{h})(2-3c)h & -6(\frac{\varepsilon}{h})(1-2c) & +2(\frac{\varepsilon}{h})(1-3c)h \\ 6c(1-c)\;- & c(2-3c)h- & -6c(1-c)+ & -(1-c)(1-3c)h- \\ -6(\frac{\varepsilon}{h})(1-2c) & -2(\frac{\varepsilon}{h})(1-3c)h & +6(\frac{\varepsilon}{h})(1-2c) & -2(\frac{\varepsilon}{h})(2-3c)h \end{pmatrix}.$$

With the boundary conditions(3.3.1) the discrete problem, for $\varepsilon/h \to 0$, reduces to

(3.3.3) $$\left(\begin{array}{ccc} 1 \quad 0 \;\cdots\cdots\cdots & & 0 \\ \boxed{\begin{array}{cccc} 1 & Ah_1 & -1 & Bh_1 \\ 1 & Bh_1 & -1 & Ah_1 \end{array}} & & \bigcirc \\ & \ddots & \\ \bigcirc & & \boxed{\begin{array}{cccc} 1 & Ah_N & -1 & Bh_N \\ 1 & Bh_N & -1 & Ah_N \end{array}} \\ 0 \;\cdots\cdots\cdots & & 0 \quad 1 \quad 0 \end{array}\right) \begin{pmatrix} y_0 \\ y_0' \\ y_1 \\ y_1' \\ \vdots \\ y_N \\ y_N' \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

where $A = \frac{3c-1}{6c}$, $B = \frac{3c-2}{6(c-1)}$.
The solution of this linear system is

$$y_i = \sum_{j=1}^{i} h_i = x_i, \qquad y_i' = \frac{1}{A+B}.$$

Thus, $\{y_i\}$ yields a pointwise (but not a global) approximation to a straight line, irrespective of the choice of the mesh Π. Sad to say, $y(x_i) = x_i$ is not at all an approximation to a solution of our boundary-value problem. We may conclude that, in general, the result obtained by collocation - for any grid Π with $\varepsilon << \min\limits_{i=1,\ldots,N}(h_i)$ - is not a good approximation to $y(x)$.

Reduction to a system of two first order equations

The second order equation (1.1.1.a) can also be reduced to a system of two first order equations. Then the problem (1.1.1) is written as

(3.3.4) $$\begin{pmatrix} \varepsilon & 0 \\ f & 1 \end{pmatrix} Y' + \begin{pmatrix} 0 & -1 \\ g & 0 \end{pmatrix} Y = \begin{pmatrix} 0 \\ s \end{pmatrix},$$

$$y(a) = \alpha, \; y(b) = \beta,$$

where Y is the vector $Y = (y,v)^T = (y,\varepsilon y')^T$.
In general, global methods are well suited for the solution of two-point boundary-value problems written in this form and it is known (cf. WEISS [1974]) that collocation schemes for systems of first order equations, based on piecewise polynomials $M^{0,k}(\Pi)$, are equivalent to implicit Runge-Kutta methods based on interpolatory quadrature formulas. Since integration by parts is out of the question for first order systems, collocation is also equivalent to a Galerkin method, for which the quadrature is effected by means of a k-point quadrature rule. As was shown by WEISS [1974] and HULME [1972 a,b], the pointwise error for these methods is $\mathcal{O}(h^{t+1})$, where t denotes the degree of precision of the related quadrature rule. Thus the pointwise error is $\mathcal{O}(h^{2k})$ if k Gaussian base-points are used for collocation (cf. DE BOOR & SCHWARTZ [1974]). The pointwise error is $\mathcal{O}(h^{2k-1})$ if Radau and $\mathcal{O}(h^{2k-2})$ if Lobatto points are chosen as collocation points on each subinterval of the mesh. This theory holds when ε is kept fixed and h → 0; however, if ε << h we may not expect the approximation to be accurate. This is shown by the following argument.

Let us consider the analogue of equation (3.1.18) for the system (3.3.4) of first order equations. We write

$$(3.3.5)\qquad \begin{aligned} y_h &= \sum a_j\phi_j, \\ v_h &= \sum b_j\phi_j \end{aligned}$$

where $\{\phi_j\}$ is a basis in $M^{0,k}(\Pi)$. The Galerkin equations now read

$$(3.3.6)\qquad \begin{aligned} &\int (\varepsilon y_h'-v_h)\phi_i dx = 0 && i = 0,\dots,N, \\ &\int (fy_h'+v_h'+gy_h)\phi_i dx = \int s\phi_i dx, && i = 1,\dots,N-1, \end{aligned}$$

with the boundary conditions

$$y_h(a) = \alpha,\ y_h(b) = \beta.$$

Here the coefficients $\{a_j\}$ and $\{b_j\}$ are to be determined.
The discrete system of equations is

$$(3.3.7)\qquad \sum a_j\left[\frac{\varepsilon}{h}\int \phi_j'\phi_i dx\right] - \sum b_j\left[\int \phi_j\phi_i d(\tfrac{x}{h})\right] = 0, \qquad i = 0,\dots,N;$$

(3.3.8) $$\sum a_j\left[\int f\phi_j'\phi_i dx + h\int g\phi_j\phi_i d(\tfrac{x}{h})\right] + \sum b_j\left[\int \phi_j'\phi_i dx\right] = h\int s\phi_i d(\tfrac{x}{h}),$$
$$i = 1,\dots,N-1.$$

If $\{\phi_j\}$ are such that

$$\begin{array}{lll} \phi_0(a) = 1, & \phi_i(a) = 0, & i = 1,\dots,N; \\ \phi_N(b) = 1, & \phi_i(b) = 0, & i = 0,\dots,N-1; \end{array}$$

then $a_0 = \alpha$, $a_N = \beta$.

Now keep h fixed and let $\varepsilon \to 0$. Since the matrix $\int \phi_j\phi_i dx$ is nonsingular the system (3.3.7)-(3.3.8) becomes

(3.3.9) $$\sum a_j\left[\int f\phi_j'\phi_i dx + h\int g\phi_j\phi_i d(\tfrac{x}{h})\right] = h\int s\phi_i d(\tfrac{x}{h}),$$
$$i = 1,\dots,N-1,$$
$$a_0 = \alpha,\ a_N = \beta.$$

This linear system is exactly the same as the one obtained if the Galerkin discretization is directly applied to the second order differential equation and we let $\varepsilon \to 0$. Therefore, for singular perturbation problems, there is no advantage in setting up the larger system (3.3.7)-(3.3.8).

Least squares

We will briefly show that again no success can be expected if we try to find a numerical solution to our problem by the least squares method, i.e. when we seek a function y_h of the form (3.1.1.) that minimizes the functional

(3.3.10) $$Q[y] = \int_a^b [\varepsilon\, y''(x) + f(x)y'(x) + g(x)y(x) - s(x)]^2\, dx$$

and satisfies the boundary conditions.

Minimization of (3.3.10) yields the linear system $\sum_j A_{ij}\, a_j = S_i$, where

(3.3.11) $$A_{ij} = \int_a^b (\varepsilon\, \phi_j''+f\, \phi_j'+g\phi_j)(\varepsilon\, \phi_i''+f\, \phi_i'+g\, \phi_i)dx$$

and

$$(3.3.12) \qquad s_i = \int_a^b s(\varepsilon\ \phi_i''+f\phi_i'+g\phi_i)dx.$$

For large values of $|\frac{fh}{\varepsilon}|$ or $|\frac{gh^2}{\varepsilon}|$ we approximately minimize

$$(3.3.13) \qquad \int_a^b (f(x)y'(x)+g(x)y(x)-s(x))^2\ dx,$$

where y is subjected to both boundary conditions. The functional (3.3.13) corresponds to the residual of the reduced equation, but as the sign of f/ε plays no role in (3.3.13), essential information is lost. Thus the minimization of (3.3.10) scarcely has any relationship to the original problem. This is illustrated by the problem

$$(3.3.14) \qquad \begin{aligned} &\varepsilon y''(x) + f(x)y'(x) = 0, \\ &y(a) = \alpha,\ y(b) = \beta. \end{aligned}$$

For large $|fh/\varepsilon|$, the function that minimizes $Q[y]$ approximately minimizes the functional

$$\int f^2(x)\ (y'(x))^2\ dx.$$

Hence, y is an approximation to the solution of the boundary-value problem

$$(3.3.15) \qquad \begin{aligned} &fy'' + 2f'y' = 0, \\ &y(a) = \alpha,\ y(b) = \beta, \end{aligned}$$

rather than to the original problem (3.3.14).

The Ritz-Galerkin method

In the positive definite case (i.e. if $g \le 0$), it makes sense to search for a global approximation (3.1.1) which is optimal in the energy norm $\|\cdot\|_E$, see eq. (1.1.7). Such an approximation is obtained by solving the linear system

$$\sum_{j=0}^{M} A_{ij}a_j = b_i, \qquad i = 1,\ldots,M-1,$$

where

$$(3.3.16) \qquad A_{ij} = \int_a^b (-\varepsilon\phi_i'\phi_j'+g\phi_i\phi_j) \exp\left(\int_a^x \frac{f(t)}{\varepsilon} dt\right) dx$$

and

$$(3.3.17) \qquad b_i = \int_a^b s(x)\phi_i(x) \exp\left(\int_a^x \frac{f(t)}{\varepsilon} dt\right) dx.$$

In contrast to the previous methods, in the present method f/ε plays a very important role for small values of ε. In fact, the factor $p(x) = \exp(\int_a^x \frac{f(t)}{\varepsilon} dt)$ lays a heavier weight on the side of the more relevant boundary condition. In the limit, for $\varepsilon/(fh) \to 0$, the boundary condition at the end where the boundary layer occurs, is completely neglected. This directional dependence is a great advantage but, because of the exponential magnitude of $p(x)$, practical problems arise in setting up the linear system. If the entries of the symmetric matrix A are calculated in a straightforward manner, overflow problems arise in the computation of $p(x)$. Even if this is circumvented by introducing row-scaling (which disturbs the symmetry), $p(x)$ remains an unmanageable, rapidly varying function. Indeed, for extreme values of hf/ε, asymptotic expressions can be developed for A_{ij} and b_i, but the approach using the integrating factor $p(x)$ remains cumbersome. Another approach, which shares the benefit of directional preference and which overcomes to a certain extent the inconveniences induced by the exponential function $p(x)$, is the exponentially fitted weighted residual method that will be treated in the following sections.

3.4. EXPONENTIALLY FITTED SPACES AND THEIR USE

In section 3.2 we saw that the pointwise error bound on a mesh Π is related to the capacity of the space V_h to represent solutions of the adjoint equation. In this section we investigate how this knowledge and the freedom in the choice of a space V_h can be exploited to obtain better methods for the solution of singular perturbation problems. First, we have to study the properties of the solutions of the adjoint equation, especially the form of Green's function for this kind of problem. Then we construct a space V_h in which these functions are well approximated. This V_h is used to construct methods in which the requirement of a small enough h/ε ratio is relaxed and in which a certain given order of accuracy is attained.

It will become apparent that piecewise exponentials have to be included in V_h and as a result a sense of directional preference, which is also present in the differential operator, is carried over to the corresponding discrete operator. In the extreme case when $\varepsilon << fh$, our particular choice of V_h will mean that essentially an initial value problem is solved, using the correct boundary condition.

Green's function for a singular perturbation problem

In theorem 3.2.3 it was shown that Green's function plays an important role in the determination of pointwise error bounds. Therefore, in studying the numerical solution of the equation

$$(3.4.1) \qquad Ly \equiv \varepsilon y'' + fy' + gy = s, \qquad 0 < \varepsilon << 1,$$

we require information concerning the properties of its Green's function for $\varepsilon \to 0$. To this end we first consider the case $|f| \geq f_0 > 0$. The asymptotic behaviour of Green's function for $\varepsilon \to 0$ is formulated in the following lemma and its corollaries.

LEMMA 3.4.1. *Let* L *be the differential operator defined on the interval* $[a,b]$ *by* (3.4.1), *where* $f \in C^1[a,b]$, $g \in C^0[a,b]$ *and* $|f(x)| \geq f_0 > 0$. *Let the function* $\psi \in H_0^1(a,b) \cap C^2((a,x) \cup (x,b))$ *be the solution of* $L^T\psi = 0$ *on* (a,x) *and on* (x,b) *and let* $\mathrm{jmp}(\psi'(x)) = -1$. *Then, the asymptotic approximation of* ψ *for* $\varepsilon \to 0$ *is given by*

$$(3.4.2) \qquad \begin{aligned} \psi(\xi) &= k_1\, \psi_R(\xi) + k_2\, \psi_{BL}(\xi) \qquad \textit{for } \xi \in [a,x], \\ \psi(\xi) &= k_3\, \psi_R(\xi) + k_4\, \psi_{BL}(\xi) \qquad \textit{for } \xi \in [x,b], \end{aligned}$$

where

$$(3.4.3) \qquad \begin{aligned} \psi_R(\xi) &= \exp \int_a^\xi (g-f')/f\, dt\,, \\ \psi_{BL}(\xi) &= \exp[\tfrac{1}{\varepsilon} \int_a^\xi f dt - \int_a^\xi g/f\, dt]. \end{aligned}$$

The constants k_1, k_2, k_3, k_4 *are determined by*

$$(3.4.4) \qquad \psi(a) = \psi(b) = 0, \textit{ and}$$

$$(3.4.5)\qquad \psi(x) = \frac{-\varepsilon}{F(x)}\,\frac{(E(a)-E(x))(E(x)-E(b))}{E(x)(E(a)-E(b))},$$

where

$$(3.4.6)\qquad F(x) = f(x) + \varepsilon\,\frac{f'(x)}{f(x)} - 2\varepsilon\,\frac{g(x)}{f(x)},$$

$$(3.4.7)\qquad E(x) = \exp[\frac{1}{\varepsilon}\int_a^x F(x)dt].$$

PROOF. Application of the WKB-technique to the differential equation

$$\varepsilon y'' - (fy)' + gy = 0$$

yields, to first order in ε, the two approximate general solutions ψ_R and ψ_{BL}. Hence, the solution $\psi(\xi)$ is given by (3.4.2) and k_1, k_2, k_3, k_4 are determined by the boundary conditions at $\xi = a$, $\xi = b$ and $\xi = x$. From these conditions $\psi(x)$ is determined

$$\psi(x) = k_1\psi_R(x) + k_2\psi_{BL}(x) =$$

$$= \frac{-\varepsilon f}{(f^2+\varepsilon f'-2\varepsilon g)}\,\frac{[\psi_R(a)\psi_{BL}(x)-\psi_{BL}(a)\psi_R(x)][\psi_R(b)\psi_{BL}(x)-\psi_{BL}(b)\psi_R(x)]}{\psi_R(a)\psi_{BL}(b)-\psi_{BL}(a)\psi_R(b)}\,.$$

Introducing F by eq. (3.4.6) and E by $E(\xi) = \psi_{BL}(\xi)/\psi_R(\xi)$, a simple computation yields (3.4.5) and (3.4.7). □

COROLLARY. The function $\psi(\xi)$ has two boundary layers:

if $f < 0$ then at $\xi = a$ and at $\xi = x + 0$, or,

if $f > 0$ then at $\xi = b$ and at $\xi = x - 0$.

This boundary layer behaviour is described by $\psi_{BL}(\xi)$. Outside the boundary layer regions, we obtain the limit-behaviour of $\psi(\xi)$ for $\varepsilon \to 0$ by neglecting the exponentially small terms:

$$(3.4.8.a)\qquad \begin{aligned}\psi(\xi) &\approx 0 && (\xi<x)\\ &\approx \psi(x)\,\exp\int_x^\xi \frac{g-f'}{f}\,dt && (\xi\geq x) \quad \text{if } f > 0,\end{aligned}$$

or

$$(3.4.8.b) \qquad \psi(\xi) \approx \psi(x) \exp \int_x^{\xi} \frac{g-f'}{f}\,dt \qquad (\xi \leq x)$$

$$\approx 0 \qquad\qquad (\xi > x) \qquad \text{if } f < 0.$$

<u>COROLLARY</u>. An asymptotic approximation for $\varepsilon \to 0$, of the Green's function corresponding to the operator L, eq. (3.4.1), is given by

$$(3.4.9) \qquad G(x,\xi) = \frac{1}{+\varepsilon}\,\psi(\xi).$$

<u>EXAMPLE 3.4.1</u>. Green's function corresponding to equation

$$(1.3.1) \qquad L\,y = \varepsilon y'' - y' = 0 \qquad \text{on } [0,1]$$

reads

$$(3.4.10)\ G(x,\xi) = \frac{\underline{\text{if}}\ \xi<x\ \underline{\text{then}}\ (1-e^{-\xi/\varepsilon})(1-e^{(1-x)/\varepsilon})\ \underline{\text{else}}\ (1-e^{-x/\varepsilon})(1-e^{(1-\xi)/\varepsilon})}{-e^{-x/\varepsilon}(e^{+1/\varepsilon}-1)}\ .$$

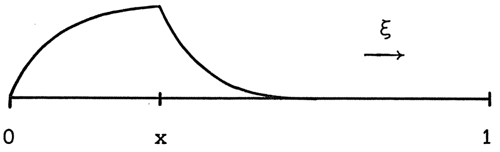


Fig. 3.4.1
Green's function for equation (1.3.1) on [0,1].

<u>Exponentially fitted spaces</u>

In view of the error bound given in theorem 3.2.3 it is expedient to have at one's disposal a space of test functions V_h, in which the functions $G(x_i,\cdot)$, $i = 1,2,\ldots,N$, can be closely approximated. From lemma 3.4.1 and its corollaries we know that, for large values of f/ε, exponential boundary layers appear in $G(x_i,\cdot)$ at $\xi = x_i$. The exponentials cannot be closely approximated in a piecewise polynomial space V_h if fh/ε is large. Hence we introduce function spaces that not only contain piecewise polynomials, but also piecewise exponentials.

DEFINITION

For each subinterval $I_i \subset [a,b]$, let $\alpha_i \in \mathbb{R}$ and let $K(I_i,\alpha_i)$ denote the (one-dimensional) linear space of scalar multiples of the function exp $(x\alpha_i)$, restricted to I_i. Furthermore, let for $k = 1,2,\ldots$,

$$R_k(I_i,\alpha_i) = \text{span}\,(P_{k-1}(I_i),\ K(I_i,\alpha_i)), \tag{3.4.11}$$

$$N^{m,k}(\Pi,\alpha_\pi) = \{v \in C^m[a,b] \mid v_{\text{restr.}I_i} \in R_k(I_i,\alpha_i),\ i = 1,2,\ldots,N\}, \tag{3.4.12}$$

$$N_0^{m,k}(\Pi,\alpha_\pi) = \{v \in N^{m,k}(\Pi,\alpha_\pi) \mid v(a) = v(b) = 0\}. \tag{3.4.13}$$

In eqs. (3.4.12) and (3.4.13), α_π denotes a mapping which gives an $\alpha_i \in \mathbb{R}$ for each interval I_i of the partition Π. The spaces $N^{m,k}(\Pi,\alpha_\pi)$ and $N_0^{m,k}(\Pi,\alpha_\pi)$ are called *exponentially fitted spaces.*

Since a proper test space should be able to represent the discontinuity in the derivative of $G(x_i,\bullet)$ all interesting exponentially fitted spaces have $m = 0$. As we did for the Lagrange spaces $M^{0,k}(\Pi)$, we select basis functions $\{\psi_j\}$ in $N^{0,k}(\Pi,\alpha_\pi)$ such that the support of each ψ_j is contained in at most two neighbouring intervals I_i.

The most obvious way to construct such a set of basis functions in $N^{0,k}(\Pi,\alpha_\pi)$, $k > 1$, is to take a set of natural basis functions in $M^{0,k-1}(\Pi)$, based on a set $\{0 = \xi_0^* < \xi_1^* < \ldots < \xi_{k-1}^* = 1\}$, (cf. eq. (3.1.15)), and to add for each I_i a function $\psi_i^{EF*} \in H_0^1(a,b)$ such that

$$\begin{cases} \psi_i^{EF*}(x) = 0 & x \notin I_i, \\ \psi_i^{EF*}(x) = \exp((x-x_{i-1})\alpha_i) - \phi_i^*(x), & x \in I_i, \end{cases} \tag{3.4.14}$$

where $\phi_i^* \in M^{0,k-1}(\Pi)$ is such that

$$\psi_i^{EF*}(x_{i-1}+\xi_\ell^* h_i) = 0, \qquad \ell = 0,1,\ldots,k-1.$$

For $k = 2,3,4$ these functions ψ_i^{EF*} are illustrated in figure 3.4.2.

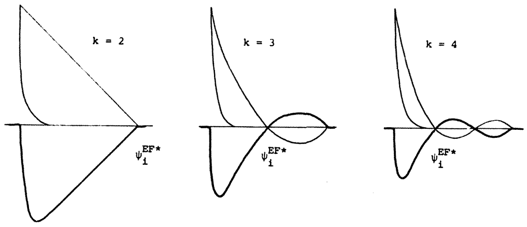


Fig. 3.4.2

Exponentially fitted basis functions ψ_i^{EF*} in $N^{0,k}(\Pi,\alpha_\pi)$; $k > 1$.

This ψ_i^{EF*} is a possible exponential basis function in $N^{0,k}(\Pi,\alpha_\pi)$ for $\alpha_i h_i \neq 0$; however, it vanishes on I_i for $\alpha_i h_i \to 0$. Therefore, it should be normalized e.g. by division by $\max_{x\in[a,b]} |\psi_i^{EF*}(x)|$. When we consider the normalized function as depending on the parameter α_i, it is easy to see that, for continuity reasons, only a unique choice can be made for $\alpha_i = 0$, viz. the k-th degree polynomial which vanishes for $x = x_{i-1} + \xi_\ell^* h_i$, $\ell = 0,1,\ldots,k-1$. For $k > 1$, this suggests the construction of another, more practical, set of basis functions that will be considered in the next subsection. First we consider the case $k = 1$.

If $k = 1$, a function ψ_i^{EF*} cannot be found, but a basis in $N^{0,1}(\Pi,\alpha_\pi)$ is readily constructed by a linear combination of a piecewise constant and a piecewise exponential function, see fig. 3.4.3. Thus, for $k = 1$, a single exponential basis function extends over two intervals and so it has to be described by two exponential coefficients: α_i and α_{i+1}. Introducing the function

$$(3.4.15) \qquad \Psi(\xi,-\alpha) = \frac{e^{-\alpha\xi}-e^{-\alpha}}{1-e^{-\alpha}} ,$$

we describe the basis functions in $N^{0,1}(\Pi,\alpha_\pi)$ by

$$(3.4.16) \qquad \psi_i^{EF}(x) = \begin{cases} \Psi((x-x_i)/h_{i+1},+\alpha_{i+1}h_{i+1}) & \text{if } x \in I_{i+1}, \\ 1-\Psi((x-x_{i-1})/h_i,+\alpha_i h_i \quad) & \text{if } x \in I_i, \end{cases}$$

$$i = 1,2,\ldots,N-1.$$

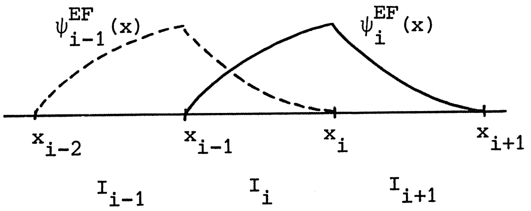


Fig. 3.4.3

Exponentially fitted basis functions in $N^{0,1}(\Pi,\alpha_\pi)$.

This basis $\{\psi_i^{EF}\}_{i=0}^N$ in $N^{0,1}(\Pi,\alpha_\pi)$ can be used for computational purposes and it is easily seen that ψ_i^{EF} reduces to the piecewise linear basis function in $M^{0,1}(\Pi)$, if both $\alpha_i h_i$ and $\alpha_{i+1}h_{i+1}$ vanish. Hence, $N^{0,1}(\Pi,\alpha_\pi)$ reduces to $M^{0,1}(\Pi)$ if $\alpha_i h_i \to 0$ for all i. We denote this by

(3.4.17) $$\lim_{|\alpha_\pi|\to 0} N^{0,1}(\Pi,\alpha_\pi) = M^{0,1}(\Pi),$$

where

(3.4.18) $$|\alpha_\pi| = \max_{i=1,\dots,N} (|\alpha_i h_i|).$$

Analogously, because the exponential basis functions in $N^{0,k}(\Pi,\alpha_\pi)$ degenerate to k-th degree piecewise polynomials if $|\alpha_\pi| \to 0$, we have, also for $k > 1$,

(3.4.19) $$\lim_{|\alpha_\pi|\to 0} N^{0,k}(\Pi,\alpha_\pi) = M^{0,k}(\Pi).$$

We shall not give a more formal description of this property, which can be given by means of the concept of "the aperture of subspaces of a Hilbert-space" as introduced by KRASNOSEL'SKII et. al.[1972], Chap. 4, section 13.5.

Natural bases in $N^{0,k}(\Pi,\alpha_\pi)$

The exponential basis function ψ_i^{EF*} (eq. (3.4.14)) is not convenient for computational purposes, since, for large values of $-\alpha_i h_i$ it is equal to $-\phi_i^*$ in the interior of I_i, except for an exponentially small term, and therefore, it leads to an extremely ill-conditioned basis (cf. VARAH [1974]). For this reason we use a more practical basis in $N^{0,k}(\Pi,\alpha_\pi)$, which will be called the *natural basis*. This basis, related to the natural basis of

$M^{0,k}(\Pi)$ rather than to that of $M^{0,k-1}(\Pi)$, is formed in the following way.

Let the natural basis functions of $M^{0,k}(\Pi)$ be constructed by means of $\{0 = \xi_0^* < \xi_1^* < \ldots < \xi_k^*=1\}$, (cf. eq. (3.1.15)). On each interval I_i, where $\alpha_i h_i = 0$, we use the basis functions of $M^{0,k}(\Pi)$ also for $N^{0,k}(\Pi,\alpha_\pi)$; if $\alpha_i h_i < 0$, the basis functions are defined on I_i by

$$\begin{cases} \psi_{i-1}^{EF}(x) \quad \text{and} \\ \phi_{ij}^*(x) - \phi_{ij}^*(x_{i-1})\psi_{i-1}^{EF}(x_{i-1}) \qquad j = 1,2,\ldots,k, \end{cases}$$

where $\phi_{ij}^* \in M^{-1,k-1}(\Pi)$ is such that

$$\phi_{ij}^*(x_{i-1}+\xi_\ell^* h_i) = \delta_{j\ell}, \qquad j = 1,2,\ldots,k.$$

If $\alpha_i h_i > 0$, then the basis functions on I_i are

$$\begin{cases} \psi_i^{EF}(x) \quad \text{and} \\ \phi_{ij}^*(x) - \phi_{ij}^*(x_i)\psi_i^{EF}(x), \qquad j = 0,1,\ldots,k-1, \end{cases}$$

where $\phi_{ij}^* \in M^{-1,k-1}(\Pi)$ is such that

$$\phi_{ij}^*(x_{i-1}+\xi_\ell^* h_i) = \delta_{j\ell}, \qquad j = 0,1,\ldots,k-1.$$

Thus, given a set of nodal points $\{0 \le \xi_0^* < \xi_1^* < \ldots < \xi_k^* \le 1\}$, the characteristic set of *natural basis functions* $\{\Psi_j\}$ on an interval of length h is described as follows:
if $\alpha < 0$, then

$$\text{(3.4.20.a)} \quad \begin{aligned} \Psi_0(\xi) &= \Psi(\xi,\alpha h), \\ \Psi_j(\xi) &= \Phi_j^*(\xi) - \Phi_j^*(0)\Psi(\xi,\alpha h), \qquad j = 1,2,\ldots,k, \end{aligned}$$

where the (k-1)-th degree polynomial Φ_j^* is defined by

$$\Phi_j^*(\xi_m^*) = \delta_{jm} \qquad j,m = 1,2,\ldots,k;$$

if $\alpha > 0$, then

$$(3.4.20.b)\quad \begin{aligned} \Psi_j(\xi) &= \Phi_j^*(\xi) - \Phi_j^*(1)\Psi(1-\xi,-\alpha h), \qquad j = 0,1,\dots,k-1, \\ \Psi_k(\xi) &= \Psi(1-\xi,-\alpha h) = 1 - \Psi(\xi,\alpha h), \end{aligned}$$

where the (k-1)-th degree polynomial Φ_j^* is defined by

$$\Phi_j^*(\xi_m^*) = \delta_{jm} \qquad j,m = 0,1,\dots,k-1.$$

EXAMPLE 3.4.2. We consider the restriction to $\bar{I}_i = [x_{i-1},x_i]$ of the basis functions in $N^{0,2}(\Pi,\alpha_\pi)$, see fig. 3.4.4. As the set of nodal points $\{\xi_m^*\}$ we use $\{0,\frac{1}{2},1\}$.

First we assume $\alpha_i < 0$. The characteristic basis functions for $M^{-1,1}(\Pi)$ are

$$\begin{aligned} \Phi_1^*(\xi) &= -2\xi + 2, \\ \Phi_2^*(\xi) &= 2\xi - 1. \end{aligned}$$

The basis functions in $N^{0,2}(\Pi,\alpha_\pi)$ on I_{i+1} become

$$\begin{aligned} \psi_{2i}(x) &= \Psi(\xi,\alpha_i h_i), \\ \psi_{2i+1}(x) &= \Phi_1^*(\xi) - 2\,\Psi(\xi,\alpha_i h_i), \\ \psi_{2i+2}(x) &= \Phi_2^*(\xi) + \Psi(\xi,\alpha_i h_i), \end{aligned}$$

where $\xi = (x-x_{i-1})/h$.

If $\alpha_i > 0$, then the characteristic basis functions for $M^{-1,1}(\Pi)$ are

$$\begin{aligned} \Phi_0^*(\xi) &= 2\xi + 1, \\ \Phi_1^*(\xi) &= 2\xi \quad . \end{aligned}$$

The basis functions in $N^{0,2}(\Pi,\alpha_\pi)$ on I_{i+1} become

$$\begin{aligned} \psi_{2i}(x) &= \Phi_0^*(\xi) + \Psi(1-\xi,-\alpha_i h_i), \\ \psi_{2i+1}(x) &= \Phi_1^*(\xi) - 2\,\Psi(1-\xi,-\alpha_i h_i), \\ \psi_{2i+2}(x) &= \Psi(1-\xi,-\alpha_i h_i). \end{aligned}$$

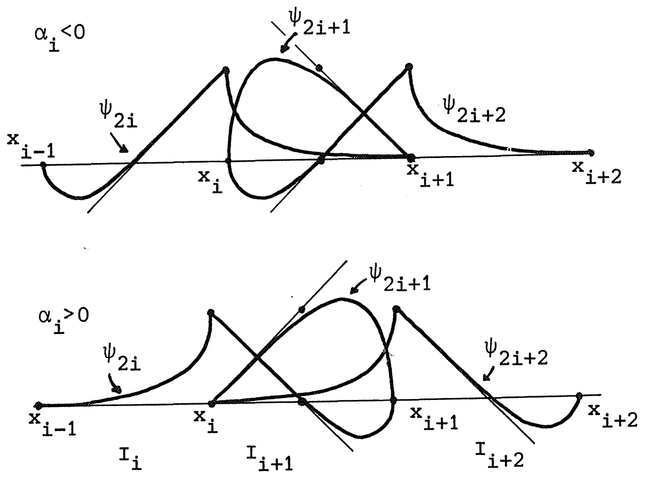


Fig. 3.4.4

Natural basis functions in $N^{0,2}(\Pi,\alpha_\pi)$.

The use of exponentially fitted spaces

Exponentially fitted spaces have been designed to approximate functions that exhibit an exponential behaviour with a large exponential factor (exponential rate) that must be known in advance. Since the exponential rate of the exponential boundary layers that appear in singular perturbation problems can be determined, we can seek a numerical solution y_h in an exponentially fitted trial space S_h and / or we can use an exponentially fitted test space V_h, in which case we fit Green's function.

Exponential fitting of S_h can be applied in two ways:

1. it can be used throughout the whole interval [a,b] (*complete fitting*).

or

2. it can be applied only in a region where a boundary layer is expected, *(partial fitting)*.

In the first case, the disadvantage is that the exponentials can introduce spurious internal boundary layers in the numerical approximation; either the contribution of the exponentially fitted component is negligible or the numerical approximation behaves almost discontinuously, even where the analytical solution behaves smoothly.

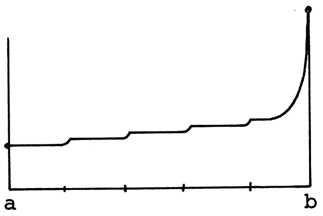


Fig. 3.4.5

A numerical approximation in an exponentially fitted space S_h.

On the other hand, when exponential fitting is applied only in the boundary layers, a priori knowledge about the solution is assumed. This information may be easily available for homogeneous linear problems, but one will meet serious difficulties in non-linear problems. Moreover, even when the differential operator is discretized with the help of a priori knowledge about the solution of the homogeneous equation, the inhomogeneous problem will not fully share in the profit of exponential fitting. This is illustrated by the following argument.

Let the operator L, eq. (3.1.2), be given and let $s \in L^2(a,b)$. Consider the following problem: find an approximation to $y \in H_0^1(a,b)$, the solution of

$$Ly = s \quad \text{on } [a,b].$$

Given a particular choice of a trial space $S_h \subset H_0^1(a,b)$ and test space $V_h \subset H_0^1(a,b)$, the approximation y_h is given by

$$y_h = - \int_a^b G_h(x,\xi) s(\xi) \, d\xi,$$

while the solution y is given by

$$y(x) = - \int_a^b G(x,\xi) s(\xi) \, d\xi.$$

So we obtain

$$(3.4.21) \qquad |(y-y_h)(x)| = \left| \int_a^b \{G(x,t) - \sum_{ij} \phi_j(x) B^{-1}_{i,j} \psi_i(t)\} s(t) \, dt \right|$$
$$\leq \|G(x,\cdot) - \sum_{ij} \phi_j(x) B^{-1}_{i,j} \psi_i(\cdot)\|_0 \, \|s\|_0.$$

To minimize the error independently of s we have to seek S_h and V_h such that the first norm is minimal. It is seen that exponential fitting of the functions $\phi_j \in S_h$ cannot be of help except for particular choices of s, whereas fitting of the $\{\psi_i\}$ in such a way that $G(x,\cdot)$ is closely approximated in V_h, always will result in a small pointwise error at x.

Thus we have obtained an argument in favor of the exponential fitting of V_h instead of the exponential fitting of S_h. In V_h exponential functions can be included that fit the boundary layers of $G(x_i,\cdot)$. As a result small pointwise errors are obtained at the nodal points x_i. Therefore, we shall consider only exponential fitting of V_h, except in the following examples, where exponential fitting of V_h and S_h are compared for two simple problems. The examples show that exponential fitting of V_h is indeed better than exponential fitting of S_h.

<u>EXAMPLE 3.4.3</u>. In this example we show with an inhomogeneous equation that exponential fitting has different effects when it is applied to S_h and to V_h. We consider the problem

$$\varepsilon y'' + y' = -1 \quad \text{on } [0,2], \tag{3.4.22}$$

with homogeneous boundary conditions. The solution is

$$y(x) = 2\,\frac{1-\exp(-x/\varepsilon)}{1-\exp(-2/\varepsilon)} - x\ . \tag{3.4.23}$$

The discretization is executed on the mesh $\Pi = \{0,1,2\}$; thus, S_h is spanned by a single function ϕ and V_h by a single function ψ. The discrete operator and right hand side are

$$B(\phi,\psi) = \int_0^2 (-\varepsilon\phi'\psi'+\phi'\psi)\,dt, \tag{3.4.24}$$

and

$$(s,\psi) = \int_0^2 (-\psi)\,dt, \tag{3.4.25}$$

and the approximate solution at x = 1 is given by

$$y_h(1) = \phi(1)\,(s,\psi)/B(\phi,\psi). \tag{3.4.26}$$

Now we consider complete and partial exponential fitting, both for S_h and for V_h.

A. Complete exponential fitting of S_h.

Here we use $S_h = N^{0,1}(\Pi,\alpha_\pi)$, $V_h = M^{0,1}(\Pi)$. The exponential rate of the boundary layer can directly be derived from the equation (cf. section 1.2), so we take $\alpha_1 = \alpha_2 = -1/\varepsilon$. Hence ϕ and ψ are given by

(3.4.27) $$\phi(x) = \begin{cases} \dfrac{\exp(-x/\varepsilon) - 1}{\exp(-1/\varepsilon) - 1} & \text{if } x \in [0,1], \\ \dfrac{\exp(-x/\varepsilon) - \exp(-2/\varepsilon)}{\exp(-1/\varepsilon) - \exp(-2/\varepsilon)} & \text{if } x \in [1,2], \end{cases}$$

and

(3.4.28) $$\psi(x) = \begin{cases} x & \text{if } x \in [0,1], \\ 2 - x & \text{if } x \in [1,2]. \end{cases}$$

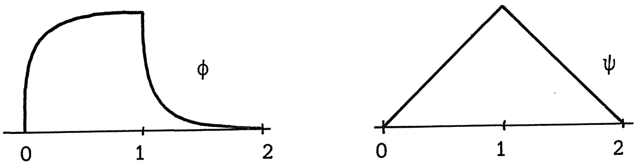


Fig. 3.4.6

The functions ϕ and ψ when complete exponential fitting of S_h is applied.

Evaluation of (3.4.24) and (3.4.25) yields

(3.4.29) $$B(\phi,\psi) = -(1+e^{-1/\varepsilon})/(1-e^{-1/\varepsilon})$$

and

(3.4.30) $$(s,\psi) = -1.$$

Hence, the approximate solution at the point $x = 1$ is

(3.4.31) $$y_h(1) = \frac{(s,\psi)}{B(\phi,\psi)} = \frac{1 - \exp(-1/\varepsilon)}{1 + \exp(-1/\varepsilon)},$$

which is the exact solution.

In example 3.4.4 we will show that this result is due to the particular choice of the right hand side which is a constant. The global approximation to the solution is

$$y_h(x) = \frac{1 - \exp(-1/\varepsilon)}{1 + \exp(-1/\varepsilon)}\ \phi(x)\,.$$

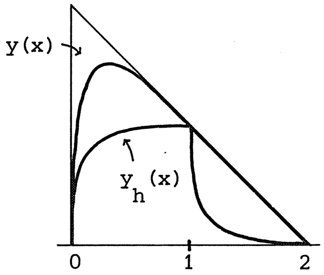


Fig. 3.4.7

The solution y of eq. (3.4.22) and the approximation y_h with complete exponential fitting of S_h.

B. Partial exponential fitting of S_h.

We use again $S_h = N^{0,1}(\Pi,\alpha_\pi)$ and $V_h = M^{0,1}(\Pi)$, but we apply exponential fitting in the boundary-layer region only, i.e. on [0,1]. So, we take

$\alpha_1 = -\frac{1}{\varepsilon}$, $\alpha_2 = 0$;

ψ is still given by (3.4.28), but now ϕ is given by

$$\phi(x) = \begin{cases} \dfrac{\exp(-x/\varepsilon) - 1}{\exp(-1/\varepsilon) - 1} & \text{if } x \in [0,1], \\ 2 - x & \text{if } x \in [1,2]. \end{cases} \qquad (3.4.32)$$

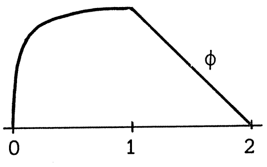


Fig. 3.4.8

The trial function ϕ, when partial exponential fitting is applied to S_h.

Evaluation of $y_h(1)$ now yields

$$y_h(1) = \frac{1}{\frac{1}{2} + \varepsilon + \frac{\exp(-1/\varepsilon)}{1 - \exp(-1/\varepsilon)}} .$$

If $e^{-1/\varepsilon} << 1$, then $y_h(1) \approx \frac{2}{1+2\varepsilon}$ and the global approximation is given by

$$y_h(x) \approx \frac{2}{1+2\varepsilon} \phi(x)$$

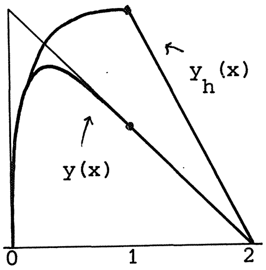


Fig. 3.4.9

The solution y of eq. (3.4.22) and the approximation y_h with partial exponential fitting of S_h.

C. Complete exponential fitting of V_h.

Now we use $S_h = M^{0,1}(\pi)$, $V_h = N^{0,1}(\Pi,\alpha_\pi)$. The exponential rate α corresponds to the exponential rate of $G(x_i,\cdot)$; hence $\alpha_1 = \alpha_2 = 1/\varepsilon$. Now ϕ and ψ are

$$(3.4.33) \qquad \phi(x) = \begin{cases} x & \text{if } x \in [0,1], \\ 2 - x & \text{if } x \in [1,2], \end{cases}$$

$$(3.4.34) \qquad \psi(x) = \begin{cases} \dfrac{\exp(x/\varepsilon) - 1}{\exp(1/\varepsilon) - 1} & \text{if } x \in [0,1], \\ \dfrac{\exp(x/\varepsilon) - \exp(2/\varepsilon)}{\exp(1/\varepsilon) - \exp(2/\varepsilon)} & \text{if } x \in [1,2]. \end{cases}$$

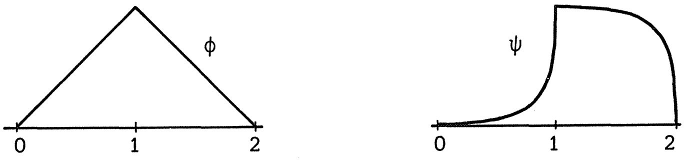


Fig. 3.4.10

The functions ϕ and ψ when complete exponential fitting of V_h is applied.

In this case evaluation of (3.4.26) again yields the pointwise exact solution at x = 1 and the global solution is

$$y_h(x) = \frac{-\phi(x)}{B(\phi,\psi)} = \frac{1 - \exp(-1/\varepsilon)}{1 + \exp(-1/\varepsilon)}\,\phi(x).$$

D. Partial exponential fitting of V_h.

To complete our exposition we find the approximate solution when V_h is partially exponentially fitted. Using $S_h = M^{0,1}(\Pi)$, $V_h = N^{0,1}(\Pi,\alpha_\pi)$, $\alpha_1 = 1/\varepsilon$, $\alpha_2 = 0$, we obtain

$$y_h(1) = \frac{(s,\psi)}{B(\phi,\psi)} = \frac{\dfrac{1}{\exp(1/\varepsilon) - 1} - \dfrac{1}{2} - \varepsilon}{\dfrac{-1}{\exp(1/\varepsilon) - 1} - \dfrac{1}{2} - \varepsilon}.$$

For small values of ε, $y_h(1) \approx 1$ and the global solution is $y_h(x) \approx \phi(x)$; see figure 3.4.11.

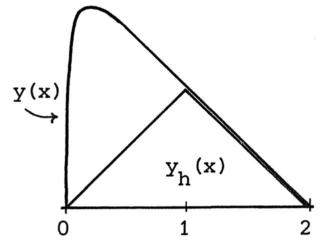


Fig. 3.4.11

The solution y of eq. (3.4.22) and the approximation y_h with complete or partial exponential fitting of V_h.

This example demonstrates that fitting of S_h is inferior to fitting of V_h in the following sense. When S_h is fitted to the behaviour of the solution y, the error caused by the boundary layers also effects the smooth part of the solution, whereas the error due to the boundary layer is restricted to the boundary layer region when V_h is fitted.

EXAMPLE 3.4.4. In this example we show how exponential fitting of S_h has a different effect when we take another right hand side s in the equation of example 3.4.3. We consider the problem

(3.4.35) $\quad \varepsilon y'' + y' = 1 - \varepsilon - x \quad$ on [0,2], with homogeneous boundary conditions.

The solution is

$$y(x) = \frac{1}{2}\, x(2-x).$$

A. Complete exponential fitting of S_h.

The discretization is executed as in example 3.4.3.A, i.e. $S_h = N^{0,1}(\Pi,\alpha_\pi)$, $\alpha_1 = \alpha_2 = -1/\varepsilon$, $V_h = M^{0,1}(\Pi)$; the functions ϕ and ψ are given by (3.4.27) and (3.4.28). Now

$$(s,\psi) = \int_0^2 (1-\varepsilon-x)\psi(x) = -\varepsilon$$

and so

$$y_h(1) = \frac{(s,\psi)}{B(\phi,\psi)} = \varepsilon\,\frac{1 - \exp(-1/\varepsilon)}{1 + \exp(-1/\varepsilon)}.$$

Thus, for small values of ε, $y_h(1)$ is a poor approximation to $y(1)$ and so is the global approximation.

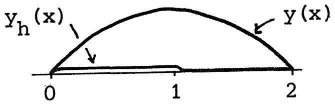


Fig. 3.4.12

The solution of eq. (3.4.35) and the approximation y_h with exponential fitting of S_h.

B. Complete exponential fitting of V_h.

The discretization is done as in example 3.4.3.C, ϕ and ψ are given by (3.4.33) and (3.4.34). Now

$$(s,\psi) = \int_0^2 (1-\varepsilon-x)\psi(x)dx = -\frac{1}{2}\,\frac{1 - \exp(-1/\varepsilon)}{1 + \exp(-1/\varepsilon)}$$

and

$$y_h(1) = \frac{(s,\psi)}{B(\phi,\psi)} = \frac{1}{2} = y(1).$$

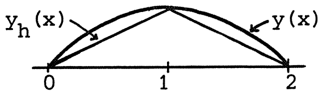


Fig. 3.4.13

The solution of eq. (3.4.35) and the approximation with exponential fitting of V_h.

In view of equation (3.4.21), it is clear that by complete exponential fitting of V_h, $y_h(1) = y(1)$ for any function s since

$$\psi(\cdot) = -B(\phi,\psi)G(1,\cdot)/\phi(1).$$

3.5. EXPONENTIALLY FITTED DISCRETE OPERATORS

In this section we discretize the weak form of the differential equation (1.1.1), using piecewise polynomial trial spaces and exponentially fitted test spaces. Thus we construct difference schemes that are especially designed to solve the singular perturbation problem (1.1.1) in the case of a large $|hf/\varepsilon|$ ratio. The schemes aim at a pointwise accurate approximation on a given mesh Π and good interpolatory properties in the smooth part of the solution.

The choice of S_h and V_h

The following lemma gives an indication of what kind of functions of a limited support should be included in a trial space V_h in order to obtain pointwise accurate approximations on a given mesh Π.

LEMMA 3.5.1. *Let there exist a unique solution to equation* (3.2.1), *let* Π *be a partition of* $[a,b]$ *and let* $G_i(\cdot,\cdot)$ *denote Green's function with respect to the interval* $[x_{i-1},x_{i+1}]$. *Let the functions* $\{G_i(x_i,\cdot)\}_{i=1}^{N-1}$ *form a subset of* $V_h \subset H_0^1(a,b)$ *and let there exist a unique solution* $y_h \in S_h \subset H_0^1(a,b)$ *of the equation*

$$B(y_h,v_h) = (s,v_h) \qquad \textit{for all } v_h \in V_h,$$

then y_h *is a pointwise exact solution on* Π.

PROOF. By theorem 3.2.3

$$|(y-y_h)(x_i)| \le K\|y - y_h\|_1 \inf_{v\in V_h} \|G(x_i,\bullet) - v\|_1.$$

Since $B(\phi_j,\psi_i)$ is nonsingular, y_h is bounded for all $i = 0,1,\dots,N$ and so is $\|y - y_h\|_1$. Let the function u be defined by

$$u = \sum_{j=1}^{N-1} \frac{G(x_i,x_j)}{G_j(x_i,x_j)} G_j(x_j,\bullet)$$

then $u \in V_h$.
On each interval $[x_{m-1},x_m]$, $G(x_i,\bullet)$ - u satisfies the equation

$$L^T(G(x_i,\bullet)-u) = L^T G(x_i,\bullet) - L^T u = 0$$

and the boundary conditions

$$G(x_i,x_j) = u(x_j), \qquad j = m-1,m.$$

Thus we have $u = G(x_i,\bullet)$ on each interval $[x_{m-1},x_m]$.
Hence $\inf_{v\in V_h} \|G(x_i,\bullet) - v\|_1 = 0$ and

$$|(y-y_h)(x_i)| = 0,$$

which proves the lemma. □

In contrast to the functions $G(x_i,\bullet)$, the functions $G_i(x_i,\bullet)$ have a support of only two intervals. This property makes the latter appropriate as basis functions in V_h when discrete operators are constructed. Of course, accurate computation of each $G_i(x_i,\bullet)$ would require the same effort as the solution of the original boundary-value problem, but a space in which they are sufficiently approximated is readily found in most cases. For smooth problems the space $M^{0,k}(\Pi)$ suffices. For singular perturbation problems of the form (1.1.1) with a large ratio $|f/\varepsilon|$ we found in lemma 3.4.1 that a boundary layer appears with a known exponential rate and so $N^{0,k}(\Pi,\alpha_\pi)$ can be used.

As far as the pointwise accuracy of the approximation is concerned, the choice of a trial space S_h should be such that $\|y - y_h\|_1$ can take small values. Further, S_h should be selected with a view to computational convenience and good interpolatory properties (global accuracy). Since the solution of a singular perturbation problem may behave almost discontinuously in some parts of the interval $[a,b]$, we use a space $M^{m,k}(\Pi)$ with the lowest possible number of continuity constraints: $m = 0$; the degree k of the piecewise polynomials depends on the accuracy required.

The choice of the parameters α_i

The parameters α_i represent the exponential rate of the local Green's functions $G_j(x_j,\xi)$ on the interval $I_i = (x_{i-1},x_i)$. The WKB approximation of the fast component of the adjoint equation

$$\varepsilon y'' - (fy)' + gy = 0$$

is

$$\exp \int^{x} \{\frac{f(t)}{\varepsilon} - \frac{g(t)}{f(t)} + \mathcal{O}(\varepsilon)\}dt.$$

Thus, the exponential rate, which depends on x, is given by $f(x)/\varepsilon - g(x)/f(x) + \mathcal{O}(\varepsilon)$. The local Green's functions $G_j(x_j,\xi)$, $j = i-1,i$, have boundary layers at x_{i-1} if $f/\varepsilon < 0$ or at x_i if $f/\varepsilon > 0$. The fast component dominates in the boundary layer and so we take

$$(3.5.1) \qquad \alpha_i = \frac{f(x_j)}{\varepsilon} - \frac{g(x_j)}{f(x_j)},$$

$$j = \begin{cases} i-1 & \text{if } f/\varepsilon \le 0, \\ i & \text{if } f/\varepsilon > 0. \end{cases}$$

If f is not a constant function, then the effective difference $\Delta\alpha$ between α_i and the exponential rate in the boundary layer will be of order $f'(x_j)\Delta/\varepsilon + \mathcal{O}(\varepsilon)$, here Δ is the length of the region where the fast component is significant. This boundary layer extends over an interval of $\mathcal{O}(\varepsilon)$, and so $\Delta\alpha$ is of order $\mathcal{O}(f'(x_i)) + \mathcal{O}(\varepsilon)$.

The WKB method yields an approximation that is asymptotically correct for $\varepsilon/f \to 0$. On the other hand, for small values of f/ε, the first order term fy' does not play the dominant role which is characteristic of nonsymmetric singular perturbation problems. Small values of h correspond to small values of $|\alpha_\pi|$, and as we saw in section 3.4

$$N^{0,k}(\Pi,\alpha_\pi) \to M^{0,k}(\Pi) \text{ as } |\alpha_\pi| \to 0.$$

This means that, with this choice of α_π, exponentially fitted operators reduce to ordinary Galerkin operators for small values of $\alpha_i h_i$. Let h be the meshwidth of the (quasi uniform) grid Π, then $|\alpha_\pi| \to 0$ as $h \to 0$ and the following consequence is immediate:

For a fixed $\varepsilon > 0$ *and* $h \to 0$, *all convergence results for the classical Galerkin method* (3.1.12) *carry over to our exponentially fitted methods.*

The evaluation of the entries of the discrete equation

An important problem that arises is the *efficient* evaluation of the integrals $B(\phi_j,\psi_i)$ and (s,ψ_i). Because of the possibly, rapidly varying components in $\psi_i \in V_h$, a simple quadrature rule cannot be used. We may proceed in two ways. We may use either

(1) an *interpolation rule.*

The coefficients of the differential equation are approximated by Lagrangian approximation, whereupon the quadrature is executed exactly (analog of the interpolation rule in section 3.1); or

(2) a combination of an interpolation and a *quadrature rule.*

The part due to the polynomial components in $N^{0,k}(\Pi,\alpha_\pi)$ is computed by a quadrature rule and only the part involving the exponential component is computed by an interpolation rule.

We illustrate both approaches by showing the discretization of the term c_2y in the differential equation (3.1.2). We use the natural basis (3.4.20) in the space $N^{0,k}(\Pi,\alpha_\pi)$. Without loss of generality we assume $\alpha < 0$. The contribution from c_2y to $B(\Phi_j,\Psi_i)$ is

$$\int c_2\Phi_j\Psi_i d\xi = \int c_2\Phi_j(\Phi_i^*-\lambda_i\Psi(\cdot,z))d\xi \tag{3.5.2}$$

where $z = h\alpha$ and $\lambda_i = \Phi_i^*(0)$.

Let the *Lagrangian interpolation* be based on the nodes $\{\eta_m\}_{m=1}^M$ and let the corresponding polynomials be $\{X\}_{m=1}^M$, then the integral is approximated by

(3.5.3) $$\int_0^1 c_2\Phi_j\Psi_i d\xi \approx \sum_{m=1}^{M} c_2(\eta_m)\ \{ \int X_m\Phi_j\Phi_i^* d\xi - \lambda_i \int X_m\Phi_j\Psi(\cdot,z)d\xi\}.$$

Here $\int_0^1 X_m\Phi_j\Phi_i^* d\xi$ are real constants independent of z and

$$\int_0^1 X_m\Phi_j\Psi(\cdot,z)d\xi$$

depend on z.

Using a *quadrature rule* for the polynomial parts we approximate the integral by

(3.5.4) $$\int_0^1 c_2\Phi_j\Psi_i d\xi \approx \sum_{m=1}^{M} c_2(\eta_m)\ \{w_m\Phi_j(\eta_m)\Phi_i^*(\eta_m) - \lambda_i \int X_m\Phi_j\Psi(\cdot,z)d\xi\}.$$

In (3.5.4) the coefficients depending on z are the same as in (3.5.3) and, in general, the amount of computational work is the same in both cases. However, if $\{\eta_m\}$ and $\{\Phi_j\}$ are properly selected, (3.5.4) can be computed more efficiently.

In both cases we need to evaluate integrals of the form

$$\int_0^1 P(x)\Psi(x,z)dx$$

where P(x) is a polynomial. Introducing the notation

(3.5.5) $$\bar{w}_n(z) = \int_0^1 x^n\Psi(x,-z)dx$$

and

(3.5.6) $$T(z) = \frac{-e^{-z}}{1-e^{-z}} = \frac{1}{2}(1-\coth(z/2)),$$

we can calculate the integral $\bar{w}_n(z)$, $z \neq 0$, by recursion from

(3.5.7) $$\bar{w}_0(z) = T(z) + \frac{1}{z} \qquad ,$$

$$\bar{w}_n(z) = \frac{T(z)}{n+1} + \frac{n}{z}\bar{w}_{n-1}(z).$$

Moreover, we have the relations

(3.5.8) $$\int_0^1 x^n \frac{d}{dx}\Psi(x,-z)\ dx = -n\bar{w}_{n-1}(z)$$

and

$$(3.5.9)\qquad \int_0^1 P(x)\,\frac{d}{dx}\,\Psi(x,-z)\;dx = -z\int P(x)\Psi(x,-z)dx \\ +z\;T(z)\int P(x)dx.$$

We notice also the relation between $\bar{w}_0(z)$ and the function m as defined by (2.4.8), namely

$$(3.5.10)\qquad \bar{w}_0(z) = \frac{1}{2} - \frac{1}{2}\,m(\frac{z}{2}).$$

The use of an interpolation rule

We illustrate discretization with exponential fitting by generating two finite difference schemes which yield piecewise linear approximations to the solution of (1.1.1). In both examples we use the natural basis in $M^{0,1}(\Pi)$ and in $N^{0,1}(\Pi)$ and we compute the integrals using formula (3.5.3).

EXAMPLE 3.5.1. We take M = 1 and $X_1 \equiv 1$. The coefficients of the differential equation are thus approximated by piecewise constants on the grid Π. The evaluation of the matrix

$$B_\ell(\phi_j,\psi_i) = \int_{I_\ell} \{-\varepsilon\phi_j'\psi_i' + f\phi_j'\psi_i + g\phi_j\psi_i\}\,dx$$

and the vector

$$\int_{I_\ell} s\psi_i dx$$

yield respectively

$$(3.5.11)\qquad \begin{pmatrix} -\frac{\varepsilon}{h} - w_0 f + (w_0-w_1)gh & \frac{\varepsilon}{h} + w_0 f + w_1 gh \\ \frac{\varepsilon}{h} - (1-w_0)f + (\frac{1}{2} - w_0 + w_1)gh & -\frac{\varepsilon}{h} + (1-w_0)f + (\frac{1}{2} - w_1)gh \end{pmatrix}$$

and

$$\begin{pmatrix} w_0 sh \\ (1-w_0)sh \end{pmatrix},$$

where $h = x_\ell - x_{\ell-1}$; $w_i = \bar{w}_i(-h\alpha_\ell)$, $i = 0,1$.

EXAMPLE 3.5.2. Now we use (3.5.3) and take $M = 2$, $X_1(\xi) = 1 - \xi$, $X_2(\xi) = \xi$; thus, approximating f, g and s by piecewise linear functions, we obtain

$$B_\ell(\phi_j,\psi_i) = \begin{pmatrix} b_{00} & b_{01} \\ b_{10} & b_{11} \end{pmatrix}, \quad (s,\psi_i) = \begin{pmatrix} d_0 \\ d_1 \end{pmatrix},$$

where

$$b_{00} = \frac{-\varepsilon}{h} + (-w_0+w_1)f_0 + (-w_1)f_1 + g_0h(w_0-2w_1+w_2) + g_1h(w_1-w_2),$$

$$b_{01} = \frac{\varepsilon}{h} + (w_0-w_1)f_0 + w_1f_1 + g_0h(w_1-w_2) + g_1hw_2,$$

$$b_{10} = -\frac{1}{2}f_0 - \frac{1}{2}f_1 + \frac{1}{3}g_0h + \frac{1}{6}g_1h - b_{00},$$

$$b_{11} = +\frac{1}{2}f_0 + \frac{1}{2}f_1 + \frac{1}{6}g_0h + \frac{1}{3}g_1h - b_{01},$$

$$d_0 = (w_0-w_1)s_0h + w_1s_1h,$$

$$d_1 = \frac{1}{2}s_0h + \frac{1}{2}s_1h - d_0.$$

Here, the subscripts 0 and 1 in f_0, f_1, g_0, g_1, s_0, s_1 denote function values of f, g and s at $x = x_{\ell-1}$ and $x = x_\ell$ respectively; $h = x_\ell - x_{\ell-1}$ and $w_j = \bar{w}_j(-h\alpha_\ell)$, $j = 0,1,2$.

Relationship to other difference methods

In example 3.5.1 the discretizations of the terms $\varepsilon y''$ and fy' are identical with those obtained by the method of exponentially fitted differences (2.4.2)-(2.4.8). This follows directly from (3.5.10). Moreover, scheme (3.5.11) suggests the adaptation of the exponentially fitted finite difference scheme for a non-uniform mesh, which can be written as

$$\begin{pmatrix} \frac{-2\varepsilon}{h} - f(1+w) + gh(1+w) & \frac{2\varepsilon}{h} + f(1+w) \\ \frac{2\varepsilon}{h} - f(1-w) & \frac{-2\varepsilon}{h} + f(1-w) + gh(1-w) \end{pmatrix} \begin{pmatrix} sh(1+w) \\ sh(1-w) \end{pmatrix} \tag{3.5.12}$$

where $w = m(\frac{fh}{2\varepsilon})$, m is defined by (2.4.8). This scheme has been implemented in ALGOL 68. The program and some of its results are listed in chapter 4.

For $\alpha_\ell h_\ell \to 0$, the method (3.5.11) reduces to the one described by (3.1.38). In the limit for $\alpha_\ell h_\ell \to \infty$ ($\alpha_\ell h_\ell \to -\infty$), y' is discretized by backward (forward) differences and y by the trapezoidal rule on the backward (forward) interval. If the interpolants of f and g on I_ℓ are taken equal to the midpoint values in this subinterval, then, in the limit for $|\alpha_\ell h_\ell| \to \infty$, the discretization of fy' + gy is the same as given by eq. (2.3.19). Thus, for $|fh/\varepsilon| \to \infty$, eq. (3.5.11) yields exactly the scheme as proposed by ABRAHAMSSON et al. [1974] for linear problems without a turning point.

The use of a quadrature rule

In this subsection we describe the discretization of the differential equation

$$(3.5.13)\qquad (\varepsilon(x)y'(x))' + f(x)y'(x) + g(x)y(x) = s(x),$$

where we allow $\varepsilon(x)$ to be a slowly varying positive definite function of x. We use a quadrature rule and we select this quadrature rule and the functions $\{\Phi_j\}$ and $\{\Psi_i\}$ so as to minimize the amount of computational work.

The description is given for a characteristic interval I_ℓ with $\alpha_\ell < 0$. On this interval we introduce

$$(3.5.14)\qquad \begin{aligned} E(\xi) &= \varepsilon(x)/h, & & \\ F(\xi) &= f(x)\ , & h &= (x_\ell - x_{\ell-1})\ , \\ G(\xi) &= g(x)h\ , & \xi &= (x - x_{\ell-1})/h\ , \\ K(\xi) &= s(x)h\ , & z &= -h\alpha_\ell\ . \end{aligned}$$

We select a (k+1)-point symmetrical quadrature rule, characterized by its nodes $0 = \xi_0 < \xi_1 < \ldots < \xi_k = 1$, $\xi_i + \xi_{k-i} = 1$. The natural basis functions $\{\Phi_j\}_{i=0}^k$ in $S_h = M^{0,k}(\Pi)$ and $\{\Psi_i\}_{i=0}^k$ in $V_h = N^{0,k}(\Pi,\alpha_\pi)$ are chosen in agreement with the nodes of this rule, i.e. the set $\{\xi_i\}$ used in the construction of both Φ_j and Ψ_i is taken to be the same as the set $\{\xi_i^*\}$ of nodal points. The entries of the discrete operator, for $\alpha_\ell < 0$, are

$$(3.5.15)\qquad \mathcal{B}(\Phi_j,\Psi_i) = \begin{cases} \mathcal{B}(\Phi_j,\Psi_0) \\ \mathcal{B}(\Phi_j,\Phi_i^*) - \Phi_i^*(0)\mathcal{B}(\Phi_j,\Psi_0), & i = 1,2,\ldots,k, \end{cases}$$

where

$$(3.5.16)\qquad \begin{aligned} \mathcal{B}(\Phi_j,\Psi_0) &= \mathcal{B}(\Phi_j,\Psi(\bullet,\alpha_\ell h)) = \mathcal{B}(\Phi_j,\Psi(\bullet,-z)) = \\ &= -\ zT(z)\int_0^1 E\Phi_j' d\xi + \int_0^1 (zE+F)\Phi_j'\Psi(\bullet,-z)d\xi \\ &\qquad + \int_0^1 G\Phi_j\Psi(\bullet,-z)d\xi \\ &\approx -\ z\ T(z)\sum_{m=0}^{k}\{E(\xi_m)\ [w_m\Phi_j'(\xi_m)]\} \\ &\quad + \sum_{m=0}^{k}\{(zE+F)(\xi_m)\int_0^1 \Phi_m\Phi_j'\Psi(\bullet,-z)d\xi\} \\ &\quad + \sum_{m=0}^{k}\{G(\xi_m)\int_0^1 \Phi_m\Phi_j\Psi(\bullet,-z)d\xi\} \\ &\overset{\text{def}}{\equiv} \mathcal{B}^*(\Phi_j,\Psi(\bullet,-z)). \end{aligned}$$

$$(3.5.17)\qquad \begin{aligned} \mathcal{B}(\Phi_j,\Phi_i^*) &= \int_0^1 \{-E\Phi_j'\Phi_i^{*\prime} + F\Phi_j'\Phi_i^* + G\Phi_j\Phi_i^*\}d\xi \\ &\approx \sum_{m=0}^{k}\{-\ E(\xi_m)\ [w_m\Phi_j'(\xi_m)\Phi_i^{*\prime}(\xi_m)]\} \\ &\qquad + F(0)\ [w_0\Phi_j'(0)\Phi_i^*(0)] + F(\xi_i)\ [w_i\Phi_j'(\xi_i)] \\ &\qquad + G(0)\ [w_0\delta_{0j}\Phi_i^*(0)\quad] + G(\xi_i)\ [w_i\delta_{ij}\quad] \\ &\overset{\text{def}}{\equiv} \mathcal{B}^*(\Phi_j,\Phi_i^*). \end{aligned}$$

The right hand side of the equation is

$$(3.5.18)\qquad \mathcal{S}(\Psi_i) = \begin{cases} \mathcal{S}(\Psi_0) & ,\quad i = 0, \\ \mathcal{S}(\Phi_i^*) - \Phi_i^*(0)\mathcal{S}(\Psi_0), & \quad i = 1,2,\dots,k, \end{cases}$$

$$(3.5.19)\qquad \mathcal{S}(\Psi_0) = \int_0^1 K\Psi(\cdot,-z)d\xi$$

$$\approx \sum_{m=0}^{k} \{K(\xi_m) \int_0^1 \Phi_m\Psi(\cdot,-z)d\xi\} \overset{\text{def}}{\equiv} \mathcal{S}^*(\Psi(\cdot,-z)).$$

$$(3.5.20)\qquad \mathcal{S}(\Phi_i^*) = \int_0^1 K\Phi_i^* d\xi$$

$$\approx K(0)\ [w_0\Phi_i^*(0)] + K(\xi_i)[w_i] \overset{\text{def}}{\equiv} \mathcal{S}^*(\Phi_i^*)$$

REMARK. The coefficients between square brackets all denote real constants depending only on the choice of the set $\{\xi_i\}$.

If ε is independent of x, then summation over $E(\xi_m)$ can be avoided in (3.5.17) and in (3.5.16) where

$$(3.5.21)\qquad \sum_{m=0}^{k} E(\xi_m)\ [w_m\Phi_j'(\xi_m)] = \begin{cases} -\varepsilon/h & \text{if } j = 0, \\ 0 & \text{if } j = 1,2,\dots,k-1, \\ \varepsilon/h & \text{if } j = k. \end{cases}$$

An algorithm based on formulas (3.5.15)-(3.5.20) has been written in ALGOL 60 and, in order to demonstrate the effect of exponential fitting, numerical results are given in section 3.7.

Further approximation of the α_i-dependent entries

Since α_i is determined in (3.5.1) with a relative accuracy of only $\mathcal{O}(\alpha_i^{-2})$, we can approximate (3.5.16) and (3.5.19) further to

$$\mathcal{B}^*(\Phi_j,\Psi(\cdot,-z)) =$$

$$\{-z^3T(z) \sum_{m=0}^{k} \{E(\xi_m)[w_m\Phi_j'(\xi_m)]\}$$

$$(3.5.22)\qquad + \sum_m (zE+F)(\xi_m)q_{mj}$$

$$+ (\underline{\text{if}}\ j{=}0\ \underline{\text{then}}\ zG(\xi_0) + \sum_m G(\xi_m)p_{m1}\ \underline{\text{else}}\ 0)\}/z^2$$

$$+ \mathcal{O}(z^{-3}) + \mathcal{O}(e^{-z}),$$

$$\mathcal{S}^*(\Psi(\cdot,-z)) = \{zK(\xi_0) + \sum_{m=0}^{k} K(\xi_m)p_{m1}\}/z^2 \qquad (3.5.23)$$
$$+ \mathcal{O}(z^{-3}) + \mathcal{O}(e^{-z}),$$

where $p_{j1} = \Phi_j'(0)$,

$p_{j2} = \Phi_j''(0)$,

$q_{mj} = p_{m1}p_{j1} + \delta_{m0}p_{j2}$.

The algorithm obtained from (3.5.22)-(3.5.23) by truncation of the exponentially small and $\mathcal{O}(z^{-3})$ terms is more efficient than the one given by (3.5.15)-(3.5.20), but it is less accurate for small values of $|\alpha_\ell h_\ell|$. An algorithm that uses a classical Galerkin method for small values of $|\alpha_\ell h_\ell|$ and the formulas (3.5.22)-(3.5.23) for larger values of $|\alpha_\ell h_\ell|$, combines the advantages of both. A program that uses such a combination of both methods has been written in ALGOL 68. It is listed in chapter 4, where some of its results are also reported.

EXAMPLE 3.5.3. If we use a quadrature rule for k = 1, we obtain the exponentially fitted difference scheme

$$\mathcal{B}^*(\Phi_j,\Psi_i) = \begin{pmatrix} b_{00} & b_{01} \\ b_{10}^* & b_{11}^* \end{pmatrix},$$

$$\mathcal{S}^*(\Psi_i) = \begin{pmatrix} d_0 \\ d_1 \end{pmatrix},$$

where

$$b_{10}^* = -\frac{1}{2}f_0 - \frac{1}{2}f_1 + \frac{1}{2}g_0h - b_{00},$$

$$b_{11}^* = \frac{1}{2}f_0 + \frac{1}{2}f_1 + \frac{1}{2}g_1h - b_{01},$$

and where b_{00}, b_{01}, d_0, d_1, f_0, f_1, g_0, g_1 are defined as in example 3.5.2.

3.6. THE ASYMPTOTIC BEHAVIOUR OF EXPONENTIALLY FITTED METHODS

In this section we study the behaviour of the exponentially fitted weighted residual (EFWR) method (3.5.15)-(3.5.20) as $\varepsilon \to 0$. In the preceding section we saw that exponentially fitted global methods lead to linear systems of type (3.1.17), where the operator $B(\phi_j,\psi_i)$ as well as the right hand side (s,ψ_i) can be split into a polynomial and an exponential part:

$$(3.6.1) \qquad \begin{aligned} B(\phi_j,\psi_i) &= B(\phi_j,\phi_i^*) + B(\phi_j,\psi_i^*), \\ (s,\psi_i) &= (s,\phi_i^*) + (s,\psi_i^*). \end{aligned}$$

For the EFWR method this splitting was explicitly given for each interval I_ℓ by eqs. (3.5.15) and (3.5.18). By letting $z \to \infty$ in the exponential parts (3.5.16) and (3.5.19), we see that these parts vanish as $|\alpha_\ell h_\ell| \to \infty$ and, therefore, a one-sided coupling remains in $B(\phi_j,\phi_i^*)$. We now derive a sufficient condition on ε to allow us to neglect the down-stream influence. We then study how exponentially fitted methods degenerate to one-step methods for initial-value problems.

Asymptotic behaviour for small values of ε

We assume that ε is independent of x, $f \le p_0 < 0$ and $|g| < M$. On an interval I_ℓ we consider the exponential part of the discrete operator, $\mathcal{B}(\Phi_i,\Psi_0)$, and of the right-hand side, $\mathcal{S}(\Psi_0)$. Using equations (3.5.16) and (3.5.19) we obtain asymptotic expressions for $\varepsilon \to 0$, namely

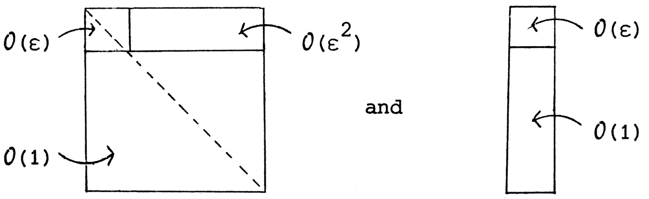
$$(3.6.2) \qquad \begin{aligned} \mathcal{B}(\Phi_j,\Psi_0) = G(0)\Phi_j(0)\frac{1}{z} + [F'(0)\Phi_j'(0) + G'(0)\Phi_j(0)]\frac{1}{z^2} + \\ + \mathcal{O}(z^{-3}) + \mathcal{O}(e^{-z}), \end{aligned}$$

$$(3.6.3) \qquad \mathcal{S}(\psi_0) = K(0)\frac{1}{z} + K'(0)\frac{1}{z}2 + \mathcal{O}(z^{-3}) + \mathcal{O}(e^{-z}),$$

where $z = -F(0)h/\varepsilon + G(0)F(0)$.

Similar expressions are obtained for $\mathcal{B}^*$ and $\mathcal{S}^*$ if $F'(0)$ is replaced by $\sum_m F(\xi_m)\Phi_m'(0)$ and if analogous substitutions are made for $G'(0)$ and $K'(0)$.

Since $\Phi_j(0) = \delta_{0,j}$, the elementary parts, $\mathcal{B}(\Phi_j,\Psi_i)$ and $\mathcal{S}(\Psi_i)$, of the discrete equation have the structure


$\mathcal{O}(\varepsilon)$
$\mathcal{O}(\varepsilon^2)$
$\mathcal{O}(1)$
and
$\mathcal{O}(\varepsilon)$
$\mathcal{O}(1)$

This means that only a one-way coupling remains in the system if the entries of order $\mathcal{O}(\varepsilon^2)$ can be neglected. In that case the method degenerates to a one-step method, which integrates the differential equation from one end to the other, starting with the relevant boundary condition. Under these circumstances we distinguish between two possibilities: whether or not the terms $\mathcal{O}(\varepsilon)$ can be neglected.

EXAMPLE 3.6.1. The EFWR scheme given in example 3.5.3 reads, for $f < 0$ and $\varepsilon \to 0$, written as a power series in z^{-1} (except for exponentially small terms and $\mathcal{O}(\varepsilon^3)$ terms)

$$(\mathcal{B}^*(\Phi_i,\Psi_j)) \approx \left[\begin{pmatrix}0 & 0\\ -\frac{1}{2} & +\frac{1}{2}\end{pmatrix}(f_0+f_1) + \begin{pmatrix}0 & 0\\ \frac{1}{2}g_0h & \frac{1}{2}g_1h\end{pmatrix}\right] +$$

$$+\begin{pmatrix}1 & 0\\ -1 & 0\end{pmatrix}\frac{g_0h}{z} + \left[\begin{pmatrix}-1 & 1\\ 1 & -1\end{pmatrix}(f_1-f_0) + \begin{pmatrix}1 & 0\\ -1 & 0\end{pmatrix}(g_1-g_0)h\right]\frac{1}{z^2},$$

$$(\mathcal{S}^*(\Psi_j)) \approx \begin{pmatrix}0\\ \frac{1}{2}(s_0+s_1)h\end{pmatrix} + \begin{pmatrix}1\\ -1\end{pmatrix}\frac{s_0h}{z} + \begin{pmatrix}1\\ -1\end{pmatrix}\frac{(s_1-s_0)h}{z^2},$$

where $z = -\frac{f_0h}{\varepsilon} + \frac{g_0h}{f_0}$; f_0, f_1, g_0, g_1, s_0 and s_1 denote the function values of f, g and s at $x = x_{\ell-1}$ and $x = x_\ell$ respectively.

EXAMPLE 3.6.2. Another scheme, which is not of the form (3.5.22)-(3.5.23), is given in example 3.5.1. Because the splitting (3.6.1) is still valid, we can give an asymptotic expression similar to (3.6.2) and (3.6.3). When $f < 0$, we take the piecewise constant approximations to f, g and s equal to $f_0 = f(x_{\ell-1})$, $g_0 = g(x_{\ell-1})$ and $s_0 = s(x_{\ell-1})$. We thus obtain, neglecting $\mathcal{O}(\varepsilon^3)$ and $\mathcal{O}(\exp(-1/\varepsilon))$ terms,

$$(\mathcal{B}(\Phi_j,\Psi_i)) \approx \left[\begin{pmatrix} 0 & 0 \\ -1 & -1 \end{pmatrix} f_0 + \begin{pmatrix} 0 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} g_0 h\right] + \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \frac{g_0 h}{z},$$

$$(\mathcal{S}(\Psi_i)) \approx \begin{pmatrix} 0 \\ 1 \end{pmatrix} s_0 h + \begin{pmatrix} 1 \\ -1 \end{pmatrix} \frac{s_0 h}{z},$$

where $z = -\frac{f_0 h}{\varepsilon} + \frac{g_0 h}{f_0}$.

EXAMPLE 3.6.3. If, in example 3.6.2, we take for the exponential rate the cruder approximation $z = -f_0 h/\varepsilon$, then we obtain

$$(\mathcal{B}(\Phi_j,\Psi_i)) \approx \left[\begin{pmatrix} 0 & 0 \\ -1 & -1 \end{pmatrix} f_0 + \begin{pmatrix} 0 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} g_0 h\right] + \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \frac{g_0 h}{z} + $$
$$+ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \frac{g_0 h}{z^2}.$$

$\mathcal{O}(\varepsilon^2)$ terms are neglected, $\mathcal{O}(\varepsilon)$ terms are not

In each step of the integration process, the initial value $y_h(x_{\ell-1})$ is given and the value $y_h(x_\ell)$ is computed from

$$(3.6.4) \quad \begin{cases} a_0 = y_h(x_{\ell-1}) \;, \\ \sum_{j=0}^{k} \mathcal{B}_\varepsilon(\Phi_j,\Psi_i) a_j = \mathcal{S}_\varepsilon(\Psi_i) \quad , \quad i = 1,2,\ldots,k, \\ y_h(x_\ell) = a_k. \end{cases}$$

Here, $\mathcal{B}_\varepsilon(\Phi_j,\Psi_i)$ and $\mathcal{S}_\varepsilon(\Psi_i)$ are equal to $\mathcal{B}(\Phi_j,\Psi_i)$ and $\mathcal{S}(\Psi_i)$, except that the contributions of $\mathcal{O}(\varepsilon)$ that originate from the interval $I_{\ell+1}$ have been added.

In order to study the influence of this contribution, we consider the problem

$$\begin{cases} \varepsilon y'' + fy' + gy = s \quad \text{on } I_\ell, \\ y(x_{\ell-1}) = a_0, \\ -\varepsilon y'(x_\ell) + \gamma\, g(x_\ell) y(x_\ell) = \gamma s(x\ ). \end{cases}$$

The weak form of this problem reads:
find $y \in H^1(x_{\ell-1},x_\ell)$ such that $y(x_{\ell-1}) = a_0$ and

$$B(y,\psi) + (\gamma g y \psi)(x_\ell) - (s,\psi) - (\gamma s \psi)(x_\ell) = 0$$
$$\text{for all } \psi \in \{\phi \mid \phi \in H^1(x_{\ell-1},x_\ell),\ \phi(x_\ell) = 0\}.$$

The discretization of this problem is given by

$$\begin{cases} a_0 = y(x_{\ell-1}) \\ \sum_{j=0}^{k} [\mathcal{B}(\Phi_j,\Psi_i) + \gamma g(x_\ell)\delta_{jk}\delta_{ik}]a_j = \mathcal{S}(\Psi_i) + \gamma s(x_\ell)\delta_{ik}, \\ \qquad i = 1,2,\dots,k, \end{cases}$$

This system is identical with (3.6.4) when $\gamma = -\varepsilon/(f(x_\ell)h)$. It follows that the EFWR method computes $y(x_\ell)$ step by step, for $\ell = 1,2,\dots,N-1$, by discretization of the weak form of the boundary problem

$$\begin{cases} \varepsilon y'' + fy' + gy = s \quad \text{on } [x_{\ell-1},x_\ell], & (3.6.5.a) \\ f(x_\ell)y'(x_\ell) + g(x_\ell)y(x_\ell) = s(x_\ell), & (3.6.5.b) \end{cases}$$

using the starting condition $y(x_0) = \alpha$.

Thus, we have formulated the influence due to the interval $I_{\ell+1}$, in terms of the mixed boundary condition (3.6.5.b) at x_ℓ.

Terms of order $\mathcal{O}(\varepsilon)$ are neglected

If all terms of $\mathcal{O}(\varepsilon)$ are neglected in (3.6.4), the exponential part of the discrete operator is neglected completely. In each step of the integration process the initial value $y_h(x_{\ell-1})$ is given and $y_h(x_\ell)$ is computed by

$$(3.6.6) \quad \begin{cases} a_0 = y_h(x_{\ell-1}), \\ \sum_{j=0}^{k} a_j\, \mathcal{B}(\Phi_j,\Phi_i^*) = \mathcal{S}(\Phi_i^*) \quad , \quad i = 1,\dots,k, \\ y_h(x_\ell) = a_k \quad . \end{cases}$$

Also the terms of $\mathcal{O}(\varepsilon)$ in $\mathcal{B}(\Phi_i,\Phi_j^*)$ vanish and the reduced equation (1.1.13) is solved on each I_ℓ by a weighted residual method with $S_h = M^{0,k}(\Pi)$ and $V_h = M^{-1,k-1}(\Pi)$.

When the quadrature described by eqs. (3.5.17) and (3.5.20) is applied, the discrete equations satisfied by the solution in I_ℓ are

$$(3.6.7)\qquad \begin{cases} a_0 = y_h(x_{\ell-1}) \\ \sum\limits_{j=0}^{k} a_j \sum\limits_{p=0}^{k} w_p\Phi_i^*(\xi_p)\ \{F(\xi_p)\Phi_j'(\xi_p) + G(\xi_p)\delta_{jp}\} = \sum\limits_{p=0}^{k} w_p\Phi_i^*(\xi_p)K(\xi_p). \end{cases}$$

Thus, the value $y_h(x_\ell) = a_k$ is determined by $y_h(x_{\ell-1})$, $F(\xi_p)$, $G(\xi_p)$ and $K(\xi_p)$. This one-step method (3.6.7), to which the exponentially fitted global method (3.5.15)-(3.5.20) reduces for $\frac{\varepsilon}{fh} \to 0$, will be called the *reduced EFWR method.*

The following is an immediate conclusion:

In the limit for $\frac{\varepsilon}{fh} \to 0$, *the exponentially fitted global method* (3.5.15)-(3.5.20) *solves the reduced problem by the reduced EFWR method.*

The accuracy of the reduced EFWR method

Since the matrix $(w_p\Phi_i^*(\xi_p))$ is not square, the one-step method (3.6.7) is *not* equivalent to the collocation method for the reduced equation based on the nodes $\{\xi_i\}_{i=1,\dots,k}$. However, using the following lemma, we show that it collocates at k points that are not known in advance.

LEMMA 3.6.1. *Let be given a set* $\{x_0 < x_1 < \dots < x_k\}$ *and a set* $\{w_p \mid w_p > 0,\ p = 0,1,\dots,k\}$. *Let* f *be a continuous function on* $[x_0,x_k]$ *such that*

$$\sum_{p=0}^{k} w_p\ \Phi(x_p)f(x_p) = 0$$

for any polynomial $\Phi(x)$ *of degree* $<k$, *then* f *has at least* k *distinct zeros on* $[x_0,x_k]$.

PROOF. This lemma is easily verified by a standard technique; see e.g. DAVIS [1963], thms. 10.1.3 and 10.1.4. □

COROLLARY. Let the solution obtained with the reduced EFWR method be denoted by y_h^*, then y_h^* is determined on each interval I_ℓ by

$$\sum_{p=0}^{k} w_p\Phi_i^*(\xi_p)\ \{F(\xi_p)hy_h^{*\prime}(\zeta_p) + G(\xi_p)y_h^*(\zeta_p) - K(\xi_p)\} = 0,$$

$$h = h_\ell,\ \zeta_p = x_{\ell-1} + h\xi_p\ ,$$

for all $\Phi_i^* \in P^{k-1}(I_\ell)$.

Applying the lemma, we conclude that

$$f(x)y_h^{*\prime}(x) + g(x)y_h^*(x) - s(x)$$

has at least k distinct zeros on each $[x_{\ell-1},x_\ell]$; i.e. $y_h^*(x)$ collocates the reduced equation at k (unknown) points in each closed interval I_ℓ.

The relation between implicit Runge-Kutta and collocation methods was established by HULME [1972a, 1972b] and WEISS [1974]. Direct application of theorems 2.2. and 5.1 of WEISS [1974] yields the following result.

If $f,g \in C^{k+1}[a,b]$, $|f| \geq p_0 > 0$, $|g| < M$ *then the reduced EFWR method yields a unique solution if the partition is sufficiently refined and the truncation error of this one-step method is at least of order* h^{k+1}. *This result also holds for quasilinear problems, provided that* g/f *is Lipschitz-continuous with respect to* y.

Better error estimates are derived in the following two theorems. These theorems show that, using EFWR methods, we can obtain accurate approximations to asymptotic solutions for $\varepsilon \to 0$ of the continuous problem, by letting first $\varepsilon \to 0$ and then $h \to 0$. In theorem 3.6.1 we give error bounds for the reduced exponentially fitted method, assuming exact evaluation of the integrals involved. In theorem 3.6.2 we show that a quadrature rule of order $\geq$ 2k is required to realize the bounds given in theorem 3.6.1.

THEOREM 3.6.1. *Let* y_ε *be the solution of problem* (1.1.1) *with* $f,g,s \in C^k[a,b]$, $|f| > 0$. *Let* $y_{h,\varepsilon}$ *be the approximation to* y_ε *obtained by the weighted residual method* (3.1.17), *that is characterized by* $S_h = M^{0,k}(\Pi)$, $V_h = N^{0,k}(\Pi,\alpha_\pi)$, α_π *being determined by eq.* (3.5.1). *Let* R *denote a subinterval of* [a,b], *containing at least the mesh-interval at the boundary-layer end of* [a,b], *then the method is consistent with the reduced problem on* [a,b]\R. *Moreover, we can find constants* C *and* h_0 *such that for all* $h < h_0$

$$\|y_0 - y_{h,0}\|_{H^0(a,b)\backslash R} \leq C\, h^{k+1} \|y_0\|_{H^{k+1}(a,b)\backslash R}, \tag{3.6.8}$$

$$\|y_0 - y_{h,0}\|_{H^1(a,b)\backslash R} \leq C\, h^{k} \|y_0\|_{H^{k+1}(a,b)\backslash R}, \tag{3.6.9}$$

$$|(y_0-y_{h,0})(x_i)| \leq C\, h^{2k} \|y_0\|_{H^{k+1}(a,b)\backslash R}, \tag{3.6.10}$$

where y_0 *and* $y_{h,0}$ *are defined on* $[a,b]\backslash R$ *by* $y_0 = \lim_{\varepsilon\to 0} y_\varepsilon$ *and* $y_{h,0} = \lim_{\varepsilon\to 0} y_{h,\varepsilon}$.

PROOF. Without loss of generality we give the proof for $f < 0$; then y_0 is the solution of the reduced problem

$$L_0 y_0 \equiv f y_0' + g y_0 = s \quad \text{on } [a,b]\backslash R$$
$$y_0(a) = \alpha.$$

In view of eq. (3.5.1) and the definition of $N^{0,k}(\Pi,\alpha_\pi)$, the discrete limit solution $y_{h,0} \in M^{0,k}(\Pi)$ is determined by

$$(L_0 y_{h,0}, v) = (s,v) \qquad \forall v \in V_h = M_0^{-1,k-1}(\Pi),$$
$$y_{h,0}(a) = \alpha.$$

In the remaining part of this proof we shall consider only the interval $[a,b]\backslash R$ and for convenience we denote L_0, y_0 and $y_{h,0}$ by L, y and y_h. Defining $e_h = y_h - y$, we have $e_h(a) = 0$ and

$$(Le_h, v) = 0 \qquad \forall v \in V_h.$$

Since

$$\|Le_h\|_0^2 = |(Le_h, Le_h - v)| \le \|Le_h\|_0 \, \|Le_h - v\|_0$$

for all $v \in V_h$, we have by lemma 3.2.1

$$\|Le_h\|_0 \le \inf_{v\in V_h} \|Le_h - v\|_0 \le C\,h^k \, \|D^k Le_h\|_{0,\pi}$$
$$\le Ch^k \, \|D^k s\|_0 + C\,h^k \, \|D^k(fy_h' + gy_h)\|_{0,\pi},$$

$$\|D^k(fy_h' + gy_h)\|_{0,\pi} \le K_1 \|y_h\|_{k+1,\pi} = K_1 \|y_h\|_{k,\pi}$$
$$\le K_1 \, \|y\|_{k,\pi} + K_1 \|e_h\|_{k,\pi}.$$

By lemma 3.2.1 there also exists a $z_h \in M^{0,k}(\Pi)$ such that

$$h^m \, \|e_h - z_h\|_{m,\pi} \le C\,h^{k+1} \, \|D^{k+1} e_h\|_{0,\pi} = C\,h^{k+1} \, \|D^{k+1} y\|_0,$$
$$0 \le m \le k+1.$$

Hence, using lemma 3.2.2, we obtain

$$\begin{aligned} h^k \|e_h\|_{k,\Pi} &\le h^k \|z_h\|_{k,\Pi} + h^k \|e_h - z_h\|_{k,\Pi} \\ &\le C\,h\,\|z_h\|_{1,\Pi} + C\,h^{k+1}\,\|D^{k+1}y\|_0 \qquad (3.6.11) \\ &\le C\,h\,\|e_h\|_{1,\Pi} + C\,h^{k+1}\,\|D^{k+1}y\|_0 . \end{aligned}$$

Since $f \le f_0 < 0$, the inverse operator L^{-1}: $\mathcal{H}^0(a,b)\backslash R \to \mathcal{H}^1(a,b)\backslash R$ is bounded and

$$\begin{aligned} \sigma\,\|e_h\|_1 \le \|Le_h\|_0 &\le C\,h^k\,\|D^k s\|_0 + C\,h^k\|y\|_k \\ &+ C\,h\,\|e_h\|_1 + C\,h^{k+1}\,\|D^{k+1}y\|_0 . \end{aligned}$$

Hence, if $h < h_0 = \sigma/C$,

$$(3.6.12) \qquad \|e_h\|_1 \le \frac{C\,h^k\|y\|_{k+1}}{\sigma - hC}$$

which proves (3.6.9). Notice that $h_0 = \sigma/C$ depends on f and g, but is independent of ε.

Let $x_\ell \in \Pi$, $x_\ell \ne a$, $x_\ell \ne b$ and let $\phi \in C^1[a,x_\ell]$ be defined by

$$\begin{cases} L^T\phi \equiv -(f\phi)' + g\phi = e_h \\ \phi(x_\ell) = 0 \end{cases}$$

then L^T: $\mathcal{H}^1(a,x_\ell) \to \mathcal{H}^0(a,x_\ell)$ has a bounded inverse,

$$\|e_h\|_0^2 = (L^T\phi, e_h) = (\phi, Le_h) = \inf_{v\in V_h} (\phi - v, Le_h)$$

and

$$\inf_{v\in V_h} \|\phi - v\|_0 \le C\,h\,\|D\phi\|_0 \le C\,h\,\|\phi\|_1 \le C\,h\,\|e_h\|_0 .$$

Hence

$$\|e_h\|_0 \le C\,h\,\|Le_h\|_0 \le C\,h\,\|e_h\|_1 \le \frac{C\,h^{k+1}\|y\|_{k+1}}{\sigma - hC} ,$$

which proves (3.6.8).

Green's function with respect to the operator L on $[a,x_\ell]$ is

$$(3.6.13)\qquad G(x,\xi) = \begin{cases} -f(\xi)^{-1} \exp \displaystyle\int_x^\xi \frac{g(t)}{f(t)}\, dt & \text{if } \xi < x, \\ 0 & \text{if } \xi > x. \end{cases}$$

For $u \in H^1(a,x_\ell)$

$$u(x) = -(G(x,\cdot),Lu) \quad , \quad x \in [a,x_\ell].$$

Hence

$$|e_h(x_i)| = |(Le_h,G(x_i,\cdot))| = \inf_{v \in V_h} |(Le_h,G(x_i,\cdot)-v)|$$
$$\leq \|Le_h\|_0 \inf_{v \in V_h} \|G(x_i,\cdot) - v\|_0.$$

For all $x_i \in \Pi$ the functions $v \in V_h = M^{-1,k-1}(\Pi)$ can represent the discontinuity of $G(x_i,\xi)$ at $\xi = x_i$. Therefore,

$$(3.6.14)\qquad \begin{aligned} |(y-y_h)(x_i)| &\leq \|Le_h\|_0\; C\, h^k\, \|D^k G(x_i,\cdot)\|_{H^0(a,x_i)} \\ &\leq \frac{C\, h^{2k} \|y\|_{k+1} \|D^k G(x_i,\cdot)\|_{H^0(a,x_i)}}{\sigma - hC}, \end{aligned}$$

which proves (3.6.10). □

In order to study the influence of quadrature on the accuracy of the reduced EFWR method and in order to determine its stability, we need the following two lemmas.

LEMMA 3.6.2. *Let the hypotheses of theorem* 3.6.1 *hold and let* $e_h = y_h - y$, *then, if* h *is small enough,*

$$\|e_h\|_{H^{k,\pi}(a,b)\setminus R} \leq C\, \|y\|_{k+1}$$

and

$$\|D^k Le_h\|_{H^{0,\pi}(a,b)\setminus R} \le C\ \|y\|_{k+1}.$$

PROOF. Following the same lines as used for the proof of (3.6.12), we see that

$$h^k\|e_h\|_{k,\pi} \le C\ h^k\|D^{k+1}y\|_{0,\pi} + C\ h\ \|e_h\|_1$$
$$\le C\ h^k\ \|D^{k+1}y\|_0 + \frac{C\ h^{k+1}\|y\|_{k+1}}{\sigma - hC}$$

Hence

$$h^k\ \|e_h\|_{k,\pi} \le C\ h^k\ \|y\|_{k+1}$$

and

$$\|D^k Le_h\|_{0,\pi} \le \|D^k s\| + C\ \|y_h\|_{k+1,\pi}$$
$$\le \|D^k s\| + C\ \|y\|_k + C\ \|e_h\|_{k,\pi}$$
$$\le C\ \|y\|_{k+1}. \quad \square$$

LEMMA 3.6.3. *The truncation error of the one-step Galerkin method* (3.6.6), *where* $S_h = M^{0,k}(\Pi)$ *and* $V_h = M^{-1,k-1}(\Pi)$, *is* $\mathcal{O}(h^{2k+1})$.

PROOF. To determine the truncation error, we consider the error in a single step in the solution of an initial-value problem, assuming correct initial values. Without loss of generality we can consider the first step of the integration process over [a,b]. The same arguments as were used in the proof of theorem 3.6.1 yield, for h small enough,

$$|(y_h - y)(x_1)| = \inf_{v\in V_h} |(Le_h, G(x_1,\cdot)-v)| \le$$
$$\le \inf_{v\in V_h} \|Le_h - v\|_{L^2(x_0,x_1)} \inf_{v\in V_h} \|G(x_1,\cdot)-v\|_{L^2(x_0,x_1)} ;$$

$$\inf_{v\in V_h} \|G(x_1,\cdot)-v\|^2_{L^2(x_0,x_1)} \le \inf_{v\in V_h} \|G(x_1,\cdot)-v\|^2_{L^\infty(x_0,x_1)} (x_1-x_0) \le$$
$$\le C\ \|D^kG(x_1,\cdot)\|^2_{0,\pi}\ h^{2k+1};$$

$$\inf_{v\in V_h} \|Le_h-v\|^2_{L^2(x_0,x_1)} \le C\ \|D^kLe_h\|^2_{L^2(x_0,x_1)}\ h^{2k} \le C\ \|D^kLe_h\|^2_{0,\pi} h^{2k+1}.$$

By lemma 3.6.2 we know that $\|D^kLe_h\|_{0,\pi}$ is bounded, independently of h, so that

$$|(y_h-y)(x_1)| \le C\ h^{2k+1},$$

which was to be proved. □

<u>THEOREM 3.6.2</u>. *Let the conditions of theorem* 3.6.1 *be satisfied and let* $f,g,s \in H^{t+1,\pi}[a,b]$ *with* $t \ge k$. *If a* t-*th degree quadrature rule is used for the evaluation of the integrals* $B^*(\Phi_j,\Phi_i^*)$ *and* $S^*(\Phi_i^*)$, *then the pointwise error* $\|y_0 - y_{h,0}\|_{\pi,\infty}$ *of the reduced EFWR method, characterized by* $S_h = M^{0,k}(\Pi)$ *and* $V_h = M^{-1,k-1}(\Pi)$, *is of order* p; p = min (2k,t+1).

<u>PROOF</u>. We use the same notation as in the proof of theorem 3.6.1, i.e. we consider the reduced operator L ($Ly \equiv fy' + gy$) on $[a,b]\backslash R$, where R denotes the meshinterval containing the boundary layer. Furthermore, $(v,w)^*$ denotes the approximation to (v,w) computed by application of the quadrature rule on each interval I_ℓ and $y_h^* \in S_h$ denotes the solution of

$$(Ly_h^*,v)^* = (s,v)^* \qquad \text{for all } v \in V_h.$$

Set $e_h^* = y_h^* + y_h$, then, for all $v \in V_h$,

$$|(Le_h^*,v)| \le |(Ly_h,v) - (Ly_h^*,v)| \le |(s,v) - (s,v)^*| +$$
$$+ |(Ly_h^*,v)^* - (Ly_h^*,v)| \le (K_1\|s\|_{t+1,\pi}+K_2\|y_h^*\|_{k,\pi})\|v\|_{k-1,\pi}h^{t+1}. \tag{3.6.15}$$

By eqs. (3.6.11) and (3.6.12) it follows that, if h is small enough, $\|e_h\|_{k,\pi} \le Ch\|y\|_{k+1}$ and, therefore,

$$h^k\|y_h^*\|_{k,\pi} \le h^k\|e_h^*\|_{k,\pi} + h^k\|e_h\|_{k,\pi} + h^k\|y\|_k$$
$$\le Ch\|e_h^*\|_1 + Ch^k\|y\|_{k+1}. \tag{3.6.16}$$

By the corollary to lemma 3.6.1, $Ly_h^* - s$ has at least k distinct zeros on each I_ℓ, so that

$$\begin{aligned}\|Ly_h^* - s\|_0 &\le C\|D^k(Ly_h^*-s)\|_{0,\pi}\ h^k\\ &\le C\|D^k s\|_{0,\pi}\ h^k + C\|y_h^*\|_{k,\pi}\ h^k\\ &\le C\|y\|_{k+1}\ h^k + C\|e_h^*\|_1\ h.\end{aligned}$$

By theorem 3.6.1, if h is small enough,

$$\begin{aligned}\|Le_h^*\|_0 = \|Ly_h^* - Ly_h\|_0 &\le \|Ly_h^* - s\|_0 + \|Le_h\|_0\\ &\le C\|e_h^*\|_1\ h + C\|y\|_{k+1}\ h^k\\ &\le \frac{C}{\sigma}\ \|Le_h^*\|_0\ h + C\|y\|_{k+1}\ h^k;\end{aligned}$$

$$(3.6.17)\qquad \sigma\|e_h^*\|_1 \le \|Le_h^*\|_0 \le \frac{\sigma C\|y\|_{k+1}\ h^k}{\sigma - hC} \le C\|y\|_{k+1}\ h^k.$$

Let $G(\cdot,\cdot)$ be defined by (3.6.13), then by lemma 3.2.1 we can select a $v \in V_h$ such that

$$\|G(x_i,\cdot) - v\| \le C\|D^k G(x_i,\cdot)\|_{0,\pi} h^k, \qquad \|v\|_{k-1,\pi} \le C.$$

Therefore, by repeated use of (3.6.15)-(3.6.17), we obtain

$$\begin{aligned}|(y_h^*-y_h)(x_i)| = |(G(x_i,\cdot),Le_h^*)| &\le |(G(x_i,\cdot) - v,Le_h^*)| + |(Le_h^*,v)|\\ &\le \|G(x_i,\cdot) - v\|_0\|Le_h^*\|_0 + (K_1\|s\|_{t+1,\pi}+K_2\|y_h^*\|_{k,\pi})\ h^{t+1}\\ &\le C\ h^{2k}\|y\|_{k+1} + (C\|s\|_{t+1} + C\|y\|_{k+1})\ h^{t+1}.\end{aligned}$$

Combination of this result with inequality (3.6.10) proves the theorem. □

COROLLARY. *If a* (k+1)*-point Lobatto rule is used for the evaluation of* $B^*(\Phi_j,\Psi_i)$ *and* $S^*(\Psi_i)$, *then the order of the resulting reduced EFWR method is* 2k. *If a* (k+1)*-point Newton-Cotes rule is used, the order equals* k + 1 *if* k *is odd and* k + 2 *if* k *is even.*

Stability of the reduced EFWR method as an initial-value problem method

As a consequence of the maximum principle, the boundary-value problem (1.1.1) with $s \equiv 0$ is stable with respect to the boundary conditions if $g \le 0$; i.e. for two solutions y_1, y_2 of equation (1.1.1)

$$|y_1(x) - y_2(x)| \le |y_1(a) - y_2(a)| + |y_1(b) - y_2(b)|$$

for all $x \in [a,b]$.

We will show that this property is preserved at meshpoints by the EFWR method in the limit for $\varepsilon/f \to 0$. Since the EFWR method reduces to a one-step method in the down-stream direction, we have to show that the reduced EFWR method is a stable method for the solution of the reduced problem.

DEFINITION

A method for the solution of an initial value problem is called *A-stable* in the sense of Dahlquist, if $|y_{i+1}| < |y_i|$ when the method is applied with a positive step-length h to any differential equation of the form $y' = \lambda y$, where λ is a complex constant with negative real part.

THEOREM 3.6.3. *If a* (2k-1)*-th degree quadrature rule is applied for the evaluation of* $B^*(\Phi_j,\Psi_i)$ *and* $S^*(\Psi_i)$, *then the reduced EFWR method, characterized by* $S_h = M^{0,k}(\Pi)$ *and* $V_h = M^{-1,k-1}(\Pi)$, *is A-stable.*

PROOF. We apply the method to the equation $y' = \lambda y$ and we consider a single step in the integration process. We set $z = \lambda(x_{i+1}-x_i)$ and $y_i = a_0 = 1$; then $y_{i+1} = a_k$ is determined by

$$\sum_{j=1}^{k} a_j \left[\sum_{p=0}^{k} w_p\Phi_i^*(\xi_p)\ \{\Phi_j'(\xi_p) + z\delta_{jp}\}\right] = - \sum_{p=0}^{k} w_p\Phi_i^*(\xi_p)\ \{\Phi_0'(\xi_p) + z\delta_{0p}\}.$$

Writing down the stability function,

$$F(z) = \frac{a_k}{a_0},$$

by means of Cramer's rule, we observe that F(z) has the form

$$F(z) = \frac{Q_1(z)}{Q_2(z)},$$

where Q_1 and Q_2 are polynomials of degree k. Since we know that the truncation error of the one-step method is of order 2k+1,

$$F(z) = \exp(z) + O(z^{2k+1}) \qquad \text{for } z \to 0.$$

This relation determines Q_1 and Q_2 completely:
$F(z) = P_{k,k}(z)$, the Padé approximant to exp(z) with in fact the denominator and the numerator both of degree k. A-stability follows from the fact that $|P_{kk}(z)| < 1$ for $\mathrm{Re}(z) < 0$; cf. EHLE [1968] and AXELSSON [1969]. □

REMARK. By direct computation of F(z) for all real $z < 0$, it is also readily verified that the reduced EFWR-method is stable if a (k+1)-point Newton-Cotes quadrature rule is used for the evaluation of $\mathcal{B}^*(\Phi_j,\Psi_i)$ and $\mathcal{S}^*(\Psi_i)$, $k = 1,2,\ldots,6$.

REMARK. For each quadrature rule having the properties $\xi_0 = 0$ and $\xi_i + \xi_{k-i} = 1$, $i = 0,1,\ldots,k$,

$$\lim_{z\to-\infty} \left|\frac{y_{i+1}}{y_i}\right| = |\Phi_k^*(\xi_0)| = 1.$$

This means that the reduced EFWR method is weakly stable as $|gh/f| \to \infty$. However, the reduced EFWR method is only applied if $|\varepsilon/(fh)| \to 0$. This means that the condition to suffer from the weak stability is

$$(3.6.18) \qquad \frac{\varepsilon}{h} << |f| << -gh.$$

REMARK. We can also construct methods that are consistent with the reduced equation and strongly A-stable as $\varepsilon \to 0$. To this end we use spaces V_h with incomplete sets of polynomials. These spaces contain, for each subinterval I_ℓ, an exponential basis function and k k-th degree polynomials. However, these methods show no superconvergence of the pointwise error. Such a method for k = 1 is shown in the following example.

EXAMPLE 3.6.4. Let a basis in V_h be defined by (cf. equation (3.4.20))

$$\Psi_0(\xi) = \Psi(\xi,\alpha h) \text{ and } \Psi_1(\xi) = 1 - \xi \qquad \text{if } \alpha \le 0,$$
$$\Psi_0(\xi) = \xi \qquad \text{and } \Psi_1(\xi) = 1 - \Psi(\xi,\alpha h) \text{ if } \alpha > 0;$$

then the discrete equation is described by

$$\mathcal{B}^*(\Phi_j,\Psi_i) = \begin{pmatrix} b_{00} & b_{01} \\ b_{10}^{**} & b_{11}^{**} \end{pmatrix}, \; \mathcal{S}^*(\Psi_i) = \begin{pmatrix} d_0 \\ d_1^{**} \end{pmatrix},$$

where

$$\begin{aligned} b_{10}^{**} &= \frac{\varepsilon}{h} - \frac{1}{2} f_1 - b_{00}, \\ b_{11}^{**} &= -\frac{\varepsilon}{h} + \frac{1}{2} f_1 + \frac{1}{2} g_1 h - b_{01}, \\ d_1^{**} &= \frac{1}{2} s_1 h - d_0; \end{aligned}$$

f_1, g_1, s_1, b_{00}, b_{01}, d_0 being defined as in example 3.5.2.

As $\varepsilon \to 0$, this method reduces to the backward Euler method. This is in contrast to the method in example 3.5.2, which reduces to the trapezoidal rule.

3.7. NUMERICAL RESULTS

In this section we show the effect of exponential fitting of Galerkin-type methods. Some results obtained with exponentially fitted weighted-residual (EFWR) methods are compared with those obtained with the corresponding classical Galerkin (GAL) methods. In this section, linear problems of the form (1.1.1) are solved. In chapter 4, nonlinear problems are treated and results are given which were obtained with the exponentially fitted finite-difference method (3.5.12).

The GAL methods used in this section all have $S_h = M^{0,k}(\Pi)$. The EFWR methods are of type (3.5.15)-(3.5.20), without further approximation of the exponential terms. The methods are somewhat less efficient than those based on (3.5.22)-(3.5.23), but they show more clearly the effect of exponential fitting. Compared to the more efficient ones, the methods used show essentially the same behaviour; they are only slightly more accurate for intermediate values of ε.

The various EFWR and GAL methods used in this section differ only in the number and the choice of the nodal points $\{0 = \xi_0 < \xi_1 < \ldots < \xi_k = 1\}$. These points are chosen in agreement with the (k+1)-point Lobatto or Newton-Cotes quadrature rules. We identify these methods as LOBk and NCk methods respectively. We notice that the LOB1 and NC1 methods are identical, since both are characterized by $\xi_0 = 0$, $\xi_1 = 1$ (trapezoidal rule). Also the LOB2 and NC2 methods are identical (Simpson's rule). For $k > 2$ the LOBk and NCk methods are different.

Five problems have been selected. For each problem and for various values of ε and h, we give the error of approximation and the computed order of convergence. The programs were written in ALGOL 60 and were run on a CDC CYBER 73/28 computer. The machine precision is about 14 decimal digits.

The error of approximation is given by $e = \| y - y_{h_i} \|_{\pi_0,\infty}$, where Π_0 is a fixed, equidistant grid. The computation of the approximate solution y_{h_i} is made on equidistant grids $\Pi_i \supset \Pi_0$. The *order of convergence* r is computed as

$$(3.7.1) \qquad r = \frac{\log(\| y-y_{h_i} \|_{\pi_0,\infty} / \| y-y_{h_{i+1}} \|_{\pi_0,\infty})}{\log(\; h_i \; / \; h_{i+1} \;)} \; ,$$

where h_j denotes the meshwidth of the grid Π_j.

EXAMPLE 3.7.1.

Problem:

$$(3.7.2) \qquad \begin{aligned} &\varepsilon y'' + (2+\cos(\pi x))y' - y = -(1+\varepsilon\pi^2)\cos(\pi x) - (2+\cos(\pi x))\sin(\pi x) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{on } [-1,+1], \\ &y(-1) = y(1) = -1. \end{aligned}$$

Solution: $y(x) = \cos(\pi x)$.

Characteristics: the problem has neither turning points nor boundary layers.

Using five different quadrature rules, the corresponding EFWR and GAL methods were applied to this problem. The results are shown in tables 3.7.1 and 3.7.2. In this case, where the solution is smooth over the whole interval, both methods yield acceptable approximations, the EFWR methods being more accurate.

METHOD	ε	h = 1/4		h = 1/8		h = 1/16
		e	r	e	r	e
LOB1	10^{-1}	7.47(-2)	1.6	2.41(-2)	1.2	1.04(-2)
EFWR	10^{-3}	7.15(-2)	2.0	1.74(-2)	2.0	4.33(-3)
	10^{-5}	7.15(-2)	2.0	1.74(-2)	2.0	4.33(-3)
	10^{-10}	7.15(-2)	2.0	1.74(-2)	2.0	4.33(-3)
LOB2	10^{-1}	1.02(-3)	2.5	1.82(-4)	2.5	3.25(-5)
EFWR	10^{-3}	8.72(-4)	4.0	5.30(-5)	4.0	3.28(-6)
	10^{-5}	8.75(-4)	4.0	5.32(-5)	4.0	3.31(-6)
	10^{-10}	8.75(-4)	4.0	5.32(-5)	4.0	3.31(-6)
LOB3	10^{-1}	5.07(-5)	5.5	1.16(-6)	3.3	1.21(-7)
EFWR	10^{-3}	4.83(-6)	6.0	7.55(-8)	5.5	1.72(-9)
	10^{-5}	4.70(-6)	6.2	6.56(-8)	6.0	1.01(-9)
	10^{-10}	4.70(-6)	6.2	6.56(-8)	6.0	1.01(-9)
LOB4	10^{-1}	1.20(-6)	7.2	8.09(-9)	6.4	9.44(-11)
EFWR	10^{-3}	3.92(-8)	4.7	1.56(-9)	4.3	7.78(-11)
	10^{-5}	1.32(-8)	7.7	6.16(-11)	5.3	1.55(-12)
	10^{-10}	1.32(-8)	7.7	6.16(-11)	5.3	1.54(-12)
NC4	10^{-1}	5.96(-6)	3.1	6.82(-7)	2.9	9.43(-8)
EFWR	10^{-3}	4.48(-7)	5.9	7.45(-9)	5.5	1.68(-10)
	10^{-5}	4.49(-7)	6.0	6.95(-9)	6.0	1.09(-10)
	10^{-10}	4.49(-7)	6.0	6.95(-9)	6.0	1.10(-10)

Table 3.7.1

The error and order of convergence when problem (3.7.2) is solved by EFWR methods. The error $e = \|y - y_h\|_{\pi_0,\infty}$ was measured on the equidistant grid $\Pi_0 = \{-1 = x_0 < x_1 < \ldots < x_8 = +1\}$.

METHOD	ε	h = 1/4		h = 1/8		h = 1/16
		e	r	e	r	e
LOB1	10^{-1}	1.76(-1)	2.1	4.20(-2)	2.0	1.04(-2)
GAL	10^{-3}	2.02(-1)	2.1	4.74(-2)	2.1	1.13(-2)
	10^{-5}	2.03(-1)	2.1	4.80(-2)	2.0	1.18(-2)
	10^{-10}	2.03(-1)	2.1	4.80(-2)	2.0	1.18(-2)
LOB2	10^{-1}	8.98(-3)	4.0	5.44(-4)	4.1	3.25(-5)
GAL	10^{-3}	4.95(-2)	2.1	1.15(-2)	2.5	2.05(-3)
	10^{-5}	5.04(-2)	2.0	1.24(-2)	2.0	3.09(-3)
	10^{-10}	5.04(-2)	2.0	1.24(-2)	2.0	3.10(-3)
LOB3	$1{-}^{-1}$	2.70(-4)	5.3	6.65(-6)	5.8	1.21(-7)
GAL	10^{-3}	1.07(-3)	4.1	5.88(-5)	4.2	3.24(-6)
	10^{-5}	1.10(-3)	4.1	6.53(-5)	4.0	4.02(-6)
	10^{-10}	1.10(-3)	4.1	6.54(-5)	4.0	4.04(-6)
LOB4	10^{-1}	3.45(-6)	7.4	2.11(-8)	7.8	9.44(-11)
GAL	10^{-3}	1.07(-4)	4.4	5.09(-6)	5.2	1.36(-7)
	10^{-5}	1.13(-4)	4.0	6.98(-6)	4.0	4.32(-7)
	10^{-10}	1.13(-4)	4.0	6.99(-6)	4.0	4.36(-7)
NC4	10^{-1}	2.02(-4)	5.3	5.12(-6)	5.8	9.43(-8)
GAL	10^{-3}	2.57(-4)	3.1	2.94(-5)	3.7	2.21(-6)
	10^{-5}	2.31(-4)	4.0	1.41(-5)	4.0	8.61(-7)
	10^{-10}	2.32(-4)	4.0	1.43(-5)	4.0	8.89(-7)

Table 3.7.2

The error and order of convergence when problem (3.7.2) is solved by GAL methods. The error $\|y - y_h\|_{\Pi_0,\infty}$ was measured on the equidistant grid $\Pi_0 = \{-1 = x_0 < x_1 < \ldots < x_8 = +1\}$.

The results in table 3.7.1 show that the orders of convergence determined in theorems 3.2.5 and 3.6.2 are strictly attained. This means that they cannot be improved.

EXAMPLE 3.7.2.

Problem:

$$(3.7.3)\qquad \begin{aligned} &\varepsilon y'' + y' - (1+\varepsilon)y = 0 \quad \text{on } [-1,+1], \\ &y(-1) = 1 + \exp(-2) \quad , \\ &y(+1) = 1 + \exp(-2(1+\varepsilon)/\varepsilon) \quad . \end{aligned}$$

Solution:

$$(3.7.4)\qquad y(x) = e^{x-1} + e^{-(1+\varepsilon)(1+x)/\varepsilon}.$$

Characteristics: the equation has no turning points; the solution has a boundary layer near x = -1.

The results are shown in table 3.7.3. In this case, where a boundary layer is present, the GAL methods fail, whereas the EFWR methods are able to represent the smooth part of the solution with a certain order of accuracy.

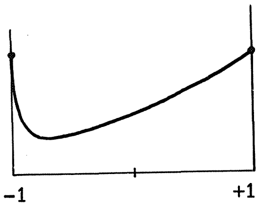


Fig. 3.7.1
The solution of problem (3.7.3).

METHOD	ε	h = 1/4		h = 1/8		h = 1/16
		e	r	e	r	e
LOB1	10^{-1}	1.94(-3)	-3.5	2.18(-2)	2.1	5.13(-3)
EFWR	10^{-3}	1.93(-3)	2.0	4.80(-4)	2.0	1.20(-4)
	10^{-5}	1.93(-3)	2.0	4.80(-4)	2.0	1.20(-4)
	10^{-10}	1.93(-3)	2.0	4.80(-4)	2.0	1.20(-4)
LOB2	10^{-1}	1.41(-4)	-2.4	7.65(-4)	4.1	4.43(-5)
EFWR	10^{-3}	2.00(-6)	4.0	1.25(-7)	4.0	7.80(-9)
	10^{-5}	2.00(-6)	4.0	1.25(-7)	4.0	7.80(-9)
	10^{-10}	2.00(-6)	4.0	1.25(-7)	4.0	7.80(-9)
LOB3	10^{-1}	7.47(-6)	-0.4	1.02(-5)	6.1	1.51(-7)
EFWR	10^{-3}	1.00(-8)	9.5	1.37(-11)	10.1	1.24(-14)
	10^{-5}	8.93(-10)	6.0	1.39(-11)	7.9	5.86(-14)
	10^{-10}	8.93(-10)	6.0	1.39(-11)	11.3	5.33(-15)
LOB1	10^{-1}	1.48(-1)	2.8	2.18(-2)	2.1	5.13(-3)
GAL	10^{-3}	7.98(-1)	0.1	7.25(-1)	0.3	6.01(-1)
	10^{-5}	8.15(-1)	0.0	7.70(-1)	0.0	7.67(-1)
	10^{-10}	8.15(-1)	0.0	7.70(-1)	0.0	7.70(-1)
LOB2	10^{-1}	1.61(-2)	4.4	7.65(-4)	4.1	4.43(-5)
GAL	10^{-3}	8.26(-1)	0.2	7.22(-1)	0.8	4.09(-1)
	10^{-5}	8.61(-1)	0.0	8.60(-1)	0.0	8.55(-1)
	10^{10}	8.62(-1)	0.0	8.61(-1)	0.0	8.61(-1)
LOB3	10^{-1}	7.94(-4)	6.3	1.02(-5)	6.1	1.51(-7)
GAL	10^{-3}	9.07(-1)	0.5	6.25(-1)	1.7	1.98(-1)
	10^{-5}	1.08(+0)	0.3	8.90(-1)	0.0	8.81(-1)
	10^{-10}	1.08(+0)	0.3	8.93(-1)	0.0	8.93(-1)

Table 3.7.3

The error and order of convergence for problem (3.7.3). The error $e = \|y - y_h\|_{\pi_0,\infty}$ was measured on the equidistant mesh $\Pi_0 = \{-1 = x_0 < x_1 < \ldots < x_8 = +1\}$.

EXAMPLE 3.7.3.

Problem (cf. eq. (1.3.11)):

$$\varepsilon y'' - xy' - y = -(1+\varepsilon\pi^2)\cos(\pi x) + \pi x \sin(\pi x) \quad \text{on } [-1,+1], \qquad (3.7.5)$$
$$y(-1) = y(+1) = -1.$$

Solution: $y(x) = \cos(\pi x)$.

Characteristics: the equation has a turning point at $x = 0$; the solution has no rapidly varying behaviour.

The results are shown in table 3.7.4. The GAL methods are able to yield a meaningful approximation, however, the EFWR methods are more accurate. Analogously to example 3.7.1, we see that, as $\varepsilon \to 0$, the order of convergence of the GAL methods reduces to $\mathcal{O}(h^k)$ for k even and $\mathcal{O}(h^{k+1})$ for k odd.

EXAMPLE 3.7.4.

Problem (cf. equation (1.3.3)):

$$\varepsilon y'' + xy' = -\varepsilon\pi^2 \cos(\pi x) - (\pi x)\sin(\pi x) \quad \text{on } [-1,+1], \qquad (3.7.6)$$
$$y(-1) = -2, \quad y(+1) = 0.$$

Solution:

$$y(x) = \cos(\pi x) + \operatorname{erf}(x/\sqrt{2\varepsilon})/\operatorname{erf}(1/\sqrt{2\varepsilon}) \qquad (3.7.7)$$

Characteristics: the solution has a shock layer in the turning-point region near $x = 0$.

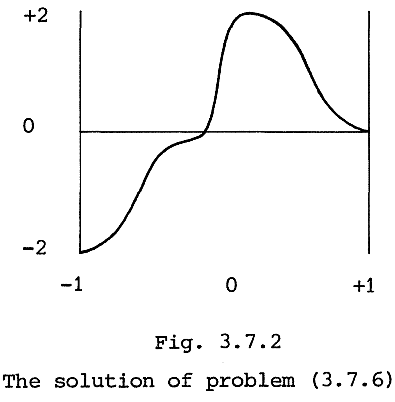


Fig. 3.7.2
The solution of problem (3.7.6).

METHOD	ε	h = 1/4		h = 1/8		h = 1/16
		e	r	e	r	e
LOB1	10^{-1}	9.81(-2)	2.1	2.23(-2)	2.0	5.45(-3)
EFWR	10^{-3}	1.30(-1)	2.0	3.20(-2)	2.0	7.98(-3)
	10^{-5}	1.31(-1)	2.0	3.24(-2)	2.0	8.06(-3)
	10^{-10}	1.31(-1)	2.0	3.24(-2)	2.0	8.06(-3)
LOB2	10^{-1}	1.85(-3)	3.9	1.20(-4)	4.0	7.54(-6)
EFWR	10^{-3}	1.06(-3)	2.8	1.53(-4)	3.8	1.07(-5)
	10^{-5}	5.84(-4)	4.0	3.58(-5)	4.0	2.23(-6)
	10^{-10}	5.84(-4)	4.0	3.58(-5)	4.0	2.23(-6)
LOB3	10^{-1}	1.06(-5)	6.1	1.57(-7)	6.0	2.43(-9)
EFWR	10^{-3}	2.45(-4)	3.7	1.94(-5)	5.4	4.58(-7)
	10^{-5}	4.54(-6)	1.9	1.20(-6)	2.1	2.81(-7)
	10^{-10}	9.74(-7)	6.0	1.49(-8)	6.0	2.32(-10)
LOB1	10^{-1}	9.81(-2)	2.1	2.23(-2)	2.0	5.45(-3)
GAL	10^{-3}	1.87(-1)	4.5	8.12(-3)	1.5	2.90(-3)
	10^{-5}	1.89(-1)	4.8	6.74(-3)	2.0	1.67(-3)
	10^{-10}	1.89(-1)	4.8	6.72(-3)	2.0	1.66(-3)
LOB2	10^{-1}	1.85(-3)	3.9	1.20(-4)	4.0	7.54(-6)
GAL	10^{-3}	2.69(-2)	2.5	4.65(-3)	2.8	6.70(-4)
	10^{-5}	3.08(-2)	2.0	7.63(-3)	2.0	1.88(-3)
	10^{-10}	3.08(-2)	2.0	7.67(-3)	2.0	1.92(-3)
LOB3	10^{-1}	1.06(-5)	6.1	1.57(-7)	6.0	2.43(-9)
GAL	10^{-3}	8.63(-4)	5.3	2.21(-5)	4.3	1.13(-6)
	10^{-5}	9.16(-4)	4.4	4.37(-5)	4.0	2.65(-6)
	10^{-10}	9.17(-4)	4.4	4.42(-5)	4.0	2.76(-6)

Table 3.7.4

The error and order of convergence for problem (3.7.5). The error $e = \|y - y_h\|_{\pi_0,\infty}$ was measured on the equidistant grid $\Pi_0 = \{-1 = x_0 < x_1 < \ldots < x_7 = +1\}$.

METHOD	ε	h = 1/7		h = 1/14		h = 1/28
		e	r	e	r	e
LOB1	10^{-1}	6.32(-2)	2.0	1.57(-2)	2.0	3.91(-3)
EFWR	10^{-3}	7.80(-2)	1.6	2.62(-2)	-2.1	1.12(-1)
	10^{-5}	7.90(-2)	1.5	2.87(-2)	1.6	9.37(-3)
	10^{-10}	7.90(-2)	1.5	2.87(-2)	1.6	9.37(-3)
LOB2	10^{-1}	2.64(-4)	3.9	1.80(-5)	4.0	1.14(-6)
EFWR	10^{-3}	1.18(-3)	3.7	8.93(-5)	-1.7	2.98(-4)
	10^{-5}	1.23(-3)	3.8	8.88(-5)	0.7	5.35(-5)
	10^{-10}	1.23(-3)	3.8	8.95(-5)	3.8	6.36(-6)
LOB3	10^{-1}	2.89(-6)	6.0	4.65(-8)	6.0	7.31(-10)
EFWR	10^{-3}	4.46(-5)	-2.6	2.68(-4)	1.6	8.53(-5)
	10^{-5}	1.09(-4)	4.0	6.74(-6)	-2.6	4.20(-5)
	10^{-10}	1.11(-4)	4.0	7.13(-6)	4.0	4.49(-7)
LOB1	10^{-1}	6.32(-2)	2.0	1.57(-2)	2.0	3.91(-3)
GAL	10^{-3}	8.49(-1)	0.8	4.82(-1)	2.3	9.60(-2)
	10^{-5}	1.05	-5.9	6.50(+1)	2.3	1.36(+1)
	10^{-10}	1.05	-22.6	6.52(+6)	2.2	1.39(+6)
LOB2	10^{-1}	2.64(-4)	3.9	1.80(-5)	4.0	1.14(-6)
GAL	10^{-3}	2.77(-1)	2.9	3.85(-2)	3.6	3.25(-3)
	10^{-5}	5.02(-1)	-0.2	5.58(-1)	-0.0	5.70(-1)
	10^{-10}	5.05(-1)	-0.2	5.70(-1)	-0.1	6.23(-1)
LOB3	10^{-1}	2.89(-6)	6.0	4.65(-8)	6.0	7.31(-10)
GAL	10^{-3}	5.83(-2)	4.4	2.77(-3)	5.0	8.95(-5)
	10^{-5}	9.80(-1)	-2.9	7.45	2.4	1.39
	10^{-10}	1.00	-19.6	7.90(+5)	2.2	1.75(+5)

Table 3.7.5

Problem (3.7.6). The error in the shock-layer region, $e = \|y - y_h\|_{\pi,\infty}$, was measured over the whole equidistant grid of respectively 14,28 and 56 subintervals.

METHOD	ε	h = 1/7		h = 1/14		h = 1/28
		e	r	e	r	e
LOB1	10^{-1}	6.32(-2)	2.0	1.56(-2)	2.0	3.88(-3)
EFWR	10^{-3}	4.00(-2)	2.0	9.93(-3)	2.0	2.48(-3)
	10^{-5}	4.00(-2)	2.0	9.94(-3)	2.0	2.48(-3)
	10^{-10}	4.00(-2)	2.0	9.94(-3)	2.0	2.48(-3)
LOB2	10^{-1}	2.64(-4)	3.9	1.74(-5)	4.0	1.10(-6)
EFWR	10^{-3}	1.18(-3)	4.7	4.43(-5)	4.7	1.68(-6)
	10^{-5}	3.94(-4)	4.0	2.48(-5)	4.0	1.55(-6)
	10^{-10}	3.94(-4)	4.0	2.48(-5)	4.0	1.55(-6)
LOB3	10^{-1}	2.76(-6)	6.0	4.40(-8)	6.0	6.92(-10)
EFWR	10^{-3}	4.30(-5)	3.1	4.89(-6)	4.4	2.28(-7)
	10^{-5}	1.72(-6)	6.0	2.73(-8)	5.9	4.51(-10)
	10^{-10}	1.50(-6)	5.8	2.73(-8)	5.9	4.47(-10)
LOB1	10^{-1}	6.32(-2)	2.0	1.56(-2)	2.0	3.88(-3)
GAL	10^{-3}	2.16(-1)	0.0	2.04(-1)	5.1	6.03(-3)
	10^{-5}	6.71(-2)	-2.9	5.01(-1)	0.3	4.08(-1)
	10^{-10}	6.53(-2)	-2.9	5.05(-1)	0.2	4.25(-1)
LOB2	10^{-1}	2.64(-4)	3.9	1.74(-5)	4.0	1.10(-6)
GAL	10^{-3}	2.77(-1)	4.2	1.52(-2)	6.2	2.10(-4)
	10^{-5}	5.02(-1)	0.3	4.13(-1)	0.4	3.15(-1)
	10^{-10}	5.05(-1)	0.2	4.25(-1)	0.2	3.68(-1)
LOB3	10^{-1}	2.76(-6)	6.0	4.40(-8)	6.0	6.92(-10)
GAL	10^{-3}	5.83(-2)	5.4	1.42(-3)	7.7	7.04(-6)
	10^{-5}	1.14(-2)	-4.9	3.35(-1)	0.6	2.15(-1)
	10^{-10}	1.10(-4)	-11.7	3.61(-1)	0.2	3.20(-1)

Table 3.7.6

Problem (3.7.6). The error, outside the shock-layer region, $e = \|y - y_h\|_{\pi_0,\infty}$, was measured on the equidistant grid $\Pi_0 = \{-1 = x_0 < x_1 < \dots < x_7 = 1\}$.

Because of the almost discontinuous character of the solution in the turning point region, the solution is badly approximated by any global approximation on a coarse mesh. Also the pointwise approximation at the gridpoints near the turning point is not very accurate when the EFWR methods are used for this problem, but the EFWR methods are not sensitive to these errors in the down-stream direction.

If we measure the error over all points of the grid Π_i (i.e. the same grid as was used for the construction of the difference scheme), then the error $\|y - y_{h_i}\|_{\pi_i,\infty}$ shows the pointwise behaviour of the approximate solution in the shock-layer region (table 3.7.5). If we take $\Pi_0 = \{i/7 \mid i = -7,-5,\dots,+7\}$ then the grid points in the shock layer are not included when the error $\|y - y_h\|_{\pi_0,\infty}$ is measured (table 3.7.6).

EXAMPLE 3.7.5.

Problem (cf. equation (1.3.7)):

$$\text{(3.7.8)} \qquad \varepsilon y'' + xy' - y = -(1+\varepsilon\pi^2)\cos(\pi x) - (\pi x)\sin(\pi x) \quad \text{on } [-1,+1],$$
$$y(-1) = -1, \quad y(+1) = +1.$$

Solution:

$$\text{(3.7.9)} \qquad y(x) = \cos(\pi x) + x + \frac{x\,\mathrm{erf}(x/\sqrt{2\varepsilon}) + \sqrt{2\varepsilon/\pi}\,\exp(-x^2/2\varepsilon)}{\mathrm{erf}(1/\sqrt{2\varepsilon}) + \sqrt{2\varepsilon/\pi}\,\exp(-1/2\varepsilon)} .$$

Characteristics: the equation has a turning point at x = 0; the solution has a corner layer in the turning-point region.

For this problem, the results obtained outside the turning point region are shown in table 3.7.7.

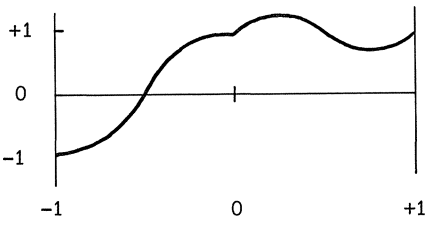


Fig. 3.7.3
The solution of problem (3.7.8).

METHOD	ε	h = 1/7		h = 1/14		h = 1/28
		e	r	e	r	e
LOB1	10^{-1}	1.88(-2)	2.0	4.63(-3)	2.0	1.15(-3)
EFWR	10^{-3}	2.50(-2)	2.0	6.15(-3)	2.0	1.53(-3)
	10^{-5}	2.51(-2)	2.0	6.17(-3)	2.0	1.53(-3)
	10^{-10}	2.51(-2)	2.0	6.17(-3)	2.0	1.54(-3)
LOB2	10^{-1}	1.90(-4)	3.9	1.26(-5)	4.0	7.95(-7)
EFWR	10^{-3}	3.59(-4)	4.0	2.25(-5)	4.2	1.23(-6)
	10^{-5}	4.19(-4)	4.0	2.65(-5)	4.0	1.66(-6)
	10^{-10}	4.20(-4)	4.0	2.65(-5)	4.0	1.66(-6)
LOB3	10^{-1}	1.13(-6)	5.9	1.83(-8)	6.0	2.89(-10)
EFWR	10^{-3}	5.44(-6)	4.9	1.85(-7)	4.5	7.96(-9)
	10^{-5}	3.01(-6)	5.7	5.90(-8)	5.9	9.92(-10)
	10^{-10}	3.01(-6)	5.7	5.91(-8)	5.9	9.92(-10)
LOB1	10^{-1}	1.88(-2)	2.0	4.63(-3)	2.0	1.15(-3)
GAL	10^{-3}	2.73(-2)	1.8	8.10(-3)	2.2	1.96(-3)
	10^{-5}	2.98(-2)	2.0	7.28(-3)	2.0	1.83(-3)
	10^{-10}	2.99(-2)	2.0	7.24(-3)	2.0	1.80(-3)
LOB2	10^{-1}	1.90(-4)	3.9	1.26(-5)	4.0	7.95(-7)
GAL	10^{-3}	1.12(-2)	4.2	6.04(-4)	3.2	6.70(-5)
	10^{-5}	1.97(-3)	1.4	7.47(-4)	0.3	5.95(-4)
	10^{-10}	2.04(-3)	2.0	5.10(-4)	2.0	1.28(-4)
LOB3	10^{-1}	1.13(-6)	5.9	1.83(-8)	6.0	2.89(-10)
GAL	10^{-3}	4.31(-3)	7.6	2.29(-5)	6.8	2.02(-7)
	10^{-5}	8.17(-4)	-0.2	9.71(-4)	0.4	7.52(-4)
	10^{-10}	8.65(-5)	5.1	2.47(-6)	4.0	1.50(-7)

Table 3.7.7

Problem (3.7.8). The error outside the turning-point region, $e = \|y - y_h\|_{\pi_0,\infty}$, was measured on the equidistant grid $\Pi_0 = \{-1 = x_0 < x_1 < \ldots < x_7 = 1\}$.

CONCLUSIONS

Examining the numerical results given in this section we arrive at the following observations:

1. In almost all cases EFWR methods yield more accurate results than GAL methods. This is also the case when the solutions are smooth over the whole interval (examples 3.7.1 and 3.7.3).

2. For large ε/h ratios (say $\varepsilon/h > f$), EFWR methods yield about the same results as GAL methods.

3. The order of convergence for EFWR methods, as determined in section 3.1 and 3.6, is strictly attained; this means that no better estimates can be found.

4. For problems with a turning point no uniform ε-convergence is obtained by EFWR methods.

5. For problems with smooth solutions, the order of convergence of the GAL methods decreases for small values of ε. The pointwise error then appears to be $\mathcal{O}(h^k)$ for even k and $\mathcal{O}(h^{k+1})$ for odd k.

CHAPTER IV

NONLINEAR PROBLEMS

In this chapter the methods developed in the previous chapters are applied to nonlinear problems. These problems are of interest since they cover more practical situations. They also give us the opportunity to show the advantages of exponentially fitted methods, because (in contrast to linear problems) the region where the solution may vary rapidly, not only depends on the equation but also on the boundary conditions. This means that, if a classical numerical method is to be used, a careful analysis is required for each particular problem, in order to determine where the mesh should be refined.

In section 1 some basic facts and definitions are given. In section 2 a convergence theorem is derived and the techniques used to solve the nonlinear problems are explained. In the third section some numerical experiments are treated and in section 4 we give an ALGOL 68 prelude which contains routines for the solution of singularly perturbed two-point boundary-value problems.

4.1. INTRODUCTION

We consider the nonlinear problem

$$\text{(4.1.1.a)} \qquad N y \equiv -\varepsilon y'' - F(x,y,y') = 0 \text{ on } I = [a,b],$$

$$\text{(4.1.1.b)} \qquad y(a) = \alpha,\ y(b) = \beta,\ 0 < \varepsilon \le \varepsilon_0.$$

For this type of problem a rich variety of phenomena is possible in the limit as $\varepsilon \to 0$. This is illustrated in WASOW [1970] who pointed out the "capriciousness" of these problems.

In general the existence of a solution of (4.1.1) cannot be guaranteed. It is well known, for example, that the problem

$$\text{(4.1.2)} \qquad \begin{aligned} &\varepsilon y'' + y' + (y')^3 = 0 \\ &y(0) = \alpha,\ y(1) = \beta,\ \alpha \ne \beta, \end{aligned}$$

has no solution for ε sufficiently small, even though the solution exists for large ε (O'MALLEY [1974], p.116). This shows that asymptotic solutions for $\varepsilon \to 0$ are available only for restricted classes of problems (4.1.1) and that possibilities for obtaining numerical approximations to solutions of problems like (4.1.1) are also limited.

Thus we can consider only a restricted subclass of problems (4.1.1). In particular, for the problems that we solve numerically, we make the following assumptions:

A1. $F(x,y,y')$ is such that there exists an isolated solution y_0 of $F(x,y,y') = 0$.

A2. There exists an $\varepsilon_0 > 0$ and a family of isolated solutions $\{y(x;\varepsilon)\}_{0<\varepsilon\leq\varepsilon_0}$ to the problem (4.1.1).

A3. The functions $y_0(x)$ and $y(x;\varepsilon)$ are such that

$$\lim_{\varepsilon\to 0} \max_{x\in I\setminus R} |y(x;\varepsilon)-y_0(x)| = 0$$

$$\lim_{\varepsilon\to 0} \max_{x\in I\setminus R} |y'(x;\varepsilon)-y_0'(x)| = 0$$

uniformly in ε,

where R is a closed subinterval of I, independent of ε.

This subinterval R will contain the boundary-layer or turning-point regions. For $\varepsilon < h$ we strive for an accurate approximation to $y(x;\varepsilon)$ on $I\setminus R$ only.

CODDINGTON & LEVINSON [1952] proved that the assumptions A2-A3 hold if the following conditions are satisfied:

B1. Equation (4.1.1) is quasilinear; i.e. it can be written in the form

$$\varepsilon y'' + F_1(x,y)y' + F_2(x,y) = 0.$$

B2. $F_1(\cdot,y)$, $F_2(\cdot,y) \in C^1[a,b]$ for y in a neighbourhood of y_0 which includes the points (a,α) and (b,β).

B3. $|F_1(x,y)| \geq K > 0$.

B4. Assumption A1 holds with

$$y_0(a) = \alpha \text{ if } F_1(x,y) < 0, \text{ or,}$$
$$y_0(b) = \beta \text{ if } F_1(x,y) > 0.$$

Since 1952, progress had been made by many people (see HOWES [1976] and references therein) in refining the conditions for problem (4.1.1) to satisfy the assumptions A1-A3. However, we shall mention only a result by DORR, PARTER & SHAMPINE [1973] which complements that of CODDINGTON & LEVINSON [1952]. This result is:

If the following conditions are satisfied: B1, $|F_2(x,y)| < M$, B3 *cnd* A2, *then the assumptions* A_1 *and* A_3 *hold; moreover,* y_0 *is such that* B4 *also holds.*

We note that under condition B3, the region R is restricted to a single boundary layer.

In order to apply the exponentially fitted methods that were discussed in chapter 3, to the nonlinear problem, we consider (4.1.1) in its variational form. A function $y \in H^1(a,b)$ is called a solution of (4.1.1) in the weak sense, if it satisfies the variational equation

$$\begin{cases} (Ny,v) = \varepsilon(y',v') - (F(\cdot,y,y'),v) = 0 \qquad \forall v \in H_0^1(a,b) \\ y(a) = \alpha,\ y(b) = \beta. \end{cases} \tag{4.1.3}$$

Denoting the dual space of $H_0^1(a,b)$ by $H^{-1}(a,b)$, *we assume that* F *is such that* $F(\cdot,y,y') \in H^{-1}(a,b)$ *for* $y \in H^1(a,b)$. Now, by eq. (4.1.3), we may extend the meaning of N, considering N as an operator $N: H^1(a,b) \to H^{-1}(a,b)$. Thus we write (Ny,v) for the nonlinear analogue of B(y,v) in eq. (3.1.4).

We introduce the following property of N (cf. CIARLET et al. [1969]):

DEFINITION

The operator $N: H^1(a,b) \to H^{-1}(a,b)$, defined by (4.1.3) is called *strictly monotone* if there is a $C > 0$ such that

$$C\|y-z\|_1^2 \le (Ny-Nz,y-z) \qquad \forall y,z \in H_0^1(a,b).$$

It is obvious that, if a solution of the variational equation (4.1.3) exists, then it is unique if N is strictly monotone.

LEMMA 4.1.1. *The operator* $N: H^1(a,b) \to H^{-1}(a,b)$ *associated with* (4.1.1) *is strictly monotone if*

$$-\frac{\partial}{\partial y}F + \frac{1}{2}\frac{d}{dx}\frac{\partial}{\partial y'}F \ge \gamma > -\varepsilon\left(\frac{\pi}{b-a}\right)^2. \tag{4.1.4}$$

PROOF. See BAKKER [1976], p.22.

EXAMPLE 4.1.1. Applying the preceding lemma to the linear operator defined in (1.1.1), we see that this operator is strictly monotone, independently of the value of ε, if

$$(4.1.5) \qquad -g(x) + \frac{1}{2}f'(x) \geq 0.$$

Substituting, for example, the coefficients of equation (1.1.12) into this inequality, we obtain

$$N_c y \equiv \varepsilon y'' + xy' + cy = 0.$$

We see that N_c is strictly monotone if $c < \frac{1}{2}$. Comparing this with the result from section 1.1, we see that, for equation (1.1.12), condition (4.1.5) is equivalent to the absence of classical turning points.

Strict monotonicity can be used to establish convergence for classical Galerkin methods. However, for weighted residual methods, where not $V_h \subset S_h$, we have to introduce the more general concept of strict coercivity with respect to the two subspaces.

DEFINITION

Let S and V be two Banach spaces with norms $\|\cdot\|_S$ and $\|\cdot\|_V$ and let V' denote the dual space of V. The (nonlinear) operator N: S → V' is *strictly coercive with respect to* S *and* V if there is a $C > 0$, such that

$$\forall y,z \in S \quad \exists v \in V \quad v \neq 0 \quad C\|y-z\|_S\|v\|_V \leq (Ny-Nz,v).$$

It is obvious that any solution $y \in S$ (if it exists) of the variational problem

$$(Ny,v) = (f,v) \qquad \forall v \in V$$

is unique if N is strictly coercive with respect to S and V.

REMARK. If $S = V = H_0^1(a,b)$, then strict monotonicity implies strict coercivity with respect to S and V.

4.2. APPROXIMATION OF NONLINEAR PROBLEMS

We solve the nonlinear equation (4.1.1) by a variant of the Newton-Kantorovich method. Referring to RALL [1969] or KRASNOSEL'SKII et al. [1972] for details about this method, we construct a sequence $\{y_m\}$ of approximate solutions to (4.1.1) as follows. It is assumed that $F(x,y,y')$ is continuous in x, $a \le x \le b$, and twice continuously differentiable with respect to y and y', so that N is a twice continuously differentiable operator from $C^2[a,b]$ into $C[a,b]$. The first two Fréchet derivatives of N at y_m are

$$(4.2.1) \qquad N'(y_m) = -\varepsilon\left(\frac{d}{dx}\right)^2 - F_y(x,y_m,y_m')I - F_{y'}(x,y_m,y_m')\left(\frac{d}{dx}\right)$$

and

$$(4.2.2) \qquad \begin{aligned} N''(y_m) = &- F_{yy}(x,y_m,y_m')II - 2F_{yy'}(x,y_m,y_m')\left(\frac{d}{dx}\right)I - \\ &- F_{y'y'}(x,y_m,y_m')\left(\frac{d}{dx}\right)\left(\frac{d}{dx}\right), \end{aligned}$$

where I is the identity operator. Setting

$$(4.2.3) \qquad u_m(x) = y_{m+1}(x) - y_m(x),$$

where y_m and y_{m+1} satisfy the boundary conditions of (4.1.1.b), we arrive at the *linear boundary value problem* for Newton-Kantorovich iteration

$$(4.2.4) \qquad \begin{aligned} &N'(y_m)u_m = -Ny_m, \\ &u_m(a) = u_m(b) = 0. \end{aligned}$$

For the generation of the Newton sequence $\{y_m(x)\}$, we add $N'(y_m)y_m$ to both sides of the equation and we solve the sequence of linear problems

$$(4.2.5) \quad \begin{aligned} &N'(y_m)y_{m+1} = F(x,y_m,y_m') - F_y(x,y_m,y_m')y_m - F_{y'}(x,y_m,y_m')y_m' \overset{\text{def}}{\equiv} R(y_m), \\ &\qquad y_{m+1}(a) = \alpha, \; y_{m+1}(b) = \beta. \end{aligned}$$

Each problem is exactly of the type treated in chapters 2 and 3. Generally, the exact solution of these equations is impossible and we must resort to the approximate solution. In effect, therefore, the successive approximations actually employed are not those of the Newton-Kantorovich method. The only thing we can do is to derive a "better" approximation $\tilde{y}_{m+1}$ from an approximation y_m via the discretization of $N'(y_m)$. Applying any of the methods developed in the previous chapters to equation (4.2.5), we get the iterative process

$$(4.2.6) \qquad N_m'(\tilde{y}_m)\tilde{y}_{m+1} = R(\tilde{y}_m),$$

where $\{N_m'\}$ is a sequence of discrete operators approximating N' and $\tilde{y}_{m+1}$ is the solution of the discretized problem.

Let y_{m+1} be the exact solution of

$$(4.2.7) \qquad N'(\tilde{y}_m)y_{m+1} = R(\tilde{y}_m),$$

then we shall first assume that

$$(4.2.8) \qquad \|\tilde{y}_{m+1} - y_{m+1}\|_1 \le q\|\tilde{y}_m - y_{m+1}\|_1$$

where $0 \le q < 1$.

It is clear from the previous chapters that $\{N_m'\}$ can be constructed in such a way that q is arbitrarily small. Such a sequence, for which the discrete operators $\{N_m'\}$ should be of increasing accuracy, is obtained by refining the partition Π during the iteration process or by taking higher order methods.

Practical rules for the convergence of the Newton sequence $\{\tilde{y}_m\}$ to the solution of problem (4.1.1) are hard to give. In fact, it depends on the problem as well as on the choice of the initial estimate $\tilde{y}_0(x)$. However, in certain cases the following modified Kantorovich theorem can be applied to obtain a convergence criterion.

<u>THEOREM 4.2.1</u>. *Let $\beta(q)$ be the smaller root of the quadratic equation*

$$2(1-\beta)^2 = (1+q)\beta + 2q.$$

Starting from $\tilde{y}_0$ it is assumed that $[N'(\tilde{y}_0)]^{-1}$ exists and constants B and H can be calculated such that

$$\|[N'(\tilde{y}_0)]^{-1}\| \le B\ ,$$
$$\|[N'(\tilde{y}_0)]^{-1}\,N\tilde{y}_0\| \le H.$$

If $\|N''(y)\| \le K$ in some closed ball $S(\tilde{y}_0,R)$ around $\tilde{y}_0$ with radius R, and if

$$B\,K\,H \le \frac{\beta(q)}{1+q}\,,$$

$$\frac{1+q}{\beta(q)}\,H \le R,$$

then the successive approximations $\{\tilde{y}_m\}_{m=0,1,\ldots}$ *defined by* (4.2.6) *converge to a solution* y *of* (4.1.1) *which exists in* $S(\tilde{y}_0,R)$.

PROOF. See KRASNOSEL'SKII et al. [1972] pp.157-160.

REMARK. Since $0 \le q < 1$ we see that $\frac{1}{2} \ge \beta > 0$. In particular, for $q = 0$ $(\beta=\frac{1}{2})$ this theorem is identical with the Kantorovich theorem.

If we keep the discretization method and the partition fixed, then N_m' is independent of m, which we denote by $N_h'(\cdot) = N_m'(\cdot)$ for $m = 0,1,\ldots$. Thus

$$N_h'(\tilde{y}_m) = N_m'(\tilde{y}_m), \qquad m = 0,1,2,\ldots,$$

depends only on $\tilde{y}_m$. We have then to solve the sequence of linear problems

$$(4.2.9) \qquad N_h'(\approx{y}_m)\approx{y}_{m+1} = R(\approx{y}_m)$$

rather than (4.2.6). In this case (4.2.8) is not true and we obtain a Newton-sequence $\{\approx{y}_m\}$ such that $\lim_{m\to\infty} \approx{y}_m = y_h$, where $y_h \in S_h$, if it exists, is the solution of the nonlinear problem

$$(4.2.10) \qquad N_h y_h \equiv N_h'(y_h)y_h - R(y_h) = 0.$$

THEOREM 4.2.2. *Let the error estimates of a weighted residual method* (3.2.2), *for any linear problem of the form* (3.2.1), *be*

$$(4.2.11) \qquad \|y-y_h\|_1 = \mathcal{O}(h^k),\ \|y-y_h\|_0 = \mathcal{O}(h^{k+1}) \ \textit{ and } \ \|y-y_h\|_{\pi,\infty} = \mathcal{O}(h^{2k}).$$

Let N *be strictly coercive with respect to* S_h *and* V_h *and let* N_h' *satisfy the conditions of theorem* 4.2.1, *then the iterative process* (4.2.9) *converges to a solution* y_h, *and the error estimates* (4.2.11) *also hold for the nonlinear problem* (4.1.3).

PROOF. By the weighted residual method and the Newton-Kantorovich iteration, $y_h \in S_h \subset H^1(a,b)$ is determined such that

$$(N'(y_h)y_h,v) = (R(y_h),v) \qquad \forall v \in V_h.$$

The solution y of (4.1.3) satisfies

$$(N'(y)y,v) = (R(y),v) \qquad \forall v \in V.$$

We introduce $u_h \in S_h$, an auxiliary approximate solution, that satisfies

$$(N'(y)u_h,v) = (R(y),v) \qquad \forall v \in V_h.$$

This u_h is the approximate solution of a linear problem, so that

$$\|y - u_h\|_1 = \mathcal{O}(h^k).$$

By the strict coercivity of N with respect of S_h and V_h, there exist a $v \in V_h \subset V$ and a $\sigma > 0$, such that

$$\sigma\|u_h - y_h\|_1\|v\|_V \leq |(Nu_h-Ny_h,v)|$$

and

$$\begin{aligned}
|(Nu_h-Ny_h,v)| &= |(Nu_h,v)| = |(N'(u_h)u_h-R(u_h),v)| \\
&= |([N'(u_h)-N'(y)]u_h - [R(u_h) - R(y)],v)| \\
&= |(N'(y)y-N'(y)u_h-Ny+Nu_h,v)| \\
&= |(N'(y)e_h-N'(y+\theta e_h)e_h,v)| \\
&= |([N'(y)-N'(y+\theta e_h)]e_h,v)| \\
&= |(N''(y+\theta^* e_h)\theta e_h e_h,v)| \leq K\|e_h\|_1^2\|v\|_V,
\end{aligned}$$

where $0 \leq \theta(x)$, $\theta^*(x) \leq 1$ and $e_h = y - u_h$.
Hence

$$\sigma\|u_h-y_h\|_1\|v\|_V \leq |(Nu_h-Ny_h,v)| \leq K\|e_h\|_1^2\|v\|_V.$$

Therefore

$$(4.2.12)\qquad \|u_h - y_h\|_1 \le \frac{K}{\sigma}\|e_h\|_1^2 = \mathcal{O}(h^{2k})$$

and

$$\|y - y_h\|_1 \le \|y - u_h\|_1 + \|u_h - y_h\|_1 \le \mathcal{O}(h^k)$$

By Poincaré's inequality it follows that $\|u_h - y_h\|_{0,\infty} < C\,h^{2k}$, and so we also obtain

$$\|y - y_h\|_0 = \mathcal{O}(h^{k+1}) \text{ and } \|y - y_h\|_{\pi,\infty} = \mathcal{O}(h^{2k}). \qquad \square$$

Replacing N by N_h, we can apply theorem 4.2.1 to the process (4.2.9). Now $q = 0$ and equation (4.2.9) describes a genuine Newton-Kantorovich process. Hence convergence is quadratic (see e.g. KRASNOSEL'SKII [1972] p.144). The quadratic convergence suggests a strategy for choosing the order of a method during the integration process. We first iterate by a first order method until convergence is obtained; then $\|y - y_h\|_{\pi,\infty} = \mathcal{O}(h)$. Assuming that $\|y - y_h\|_1$ is small enough, we need only a single iteration step by a second order method to obtain $\|y - y_h\|_{\pi,\infty} = \mathcal{O}(h^2)$ and one iteration step more by a fourth order method to obtain $\|y - y_h\|_{\pi,\infty} = \mathcal{O}(h^4)$.

To start the Newton-Kantorovich series of approximations, it is important to have available a sufficiently accurate initial approximation. However, in particular for small values of ε, it may be difficult to determine the global character of a solution beforehand. A convenient way to solve this problem is by the Davidenko principle. We assume that there exists an ε_0 for which the problem (4.1.1) has a smooth solution. For this (rather large) ε_0 an initial guess at the solution is made and the problem is solved approximately. The approximation thus obtained can be used as an initial guess for the solution with a smaller value of ε. If this process is executed with successively smaller values of ε we call it a *Newton-Kantorovich-Davidenko process*.

In general, for a fixed partition Π, this process still does not guarantee convergence to a solution of (4.2.10) as $\varepsilon \to 0$. The possible lack of a good representation in a turning-point region can mean that no function in S_h can be found, which is close enough to the solution y to be a feasible initial estimate for the Newton-Kantorovich process. In this case, inevitably, a proper mesh-refinement is required. However, exponentially fitted weighted residual methods are not sensitive to errors in the down-stream direction. This means that often convergence outside the turning-point region can still be achieved, even without an accurate representation in the turning-point region.

4.3. NUMERICAL RESULTS.

In this section we show four examples of nonlinear problems of type (4.1.1) and we comment on their numerical solution. We use three different methods of discretization: the exponentially fitted finite difference method (3.5.12), method "A", and the exponentially fitted weighted residual methods (3.5.15)-(3.5.23) with $k = 1$ (method "B") and $k = 2$ (method "C"). Asymptotically for $\varepsilon \to 0$, the pointwise convergence rates of these methods are 1,2 and 4 respectively. The approximate solutions are compared with either the exact solution, or the asymptotic solution or a numerical solution on a much finer mesh.

The programs were written in ALGOL 68 and executed on a CDC CYBER 73/28, using the CDC ALGOL 68-compiler version 1.0.9. The main routines are listed in section 4.4. The machine precision is approximately 10^{-14}.

EXAMPLE 4.3.1. We consider the boundary-value problem

$$(4.3.1) \qquad \begin{array}{l} \varepsilon y'' + e^{y}y' - \frac{\pi}{2}\sin(\frac{\pi x}{2})\; e^{2y} = 0, \qquad 0 \le x \le 1, \\ y(0) = \alpha,\ y(1) = 0. \end{array}$$

The asymptotic solution for $\varepsilon \to 0$ of this problem is (O'MALLEY [1974] p.123)

$$(4.3.2) \qquad y(x) = -\log[(1+\cos(\frac{\pi x}{2}))(1-e^{-x/(2\varepsilon)}+\frac{1}{2}e^{-\alpha}e^{-x/(2\varepsilon)}] + \mathcal{O}(\varepsilon).$$

The problem is quasilinear and it satisfies the conditions B1-B4 of section 4.1. Hence, the methods described in section 4.2 can be used to obtain a numerical approximation. With $\alpha = 0$, the solution exhibits a simple boundary layer near $x = 0$. For this value of α, the problem was solved numerically for various equidistant partitions of $[0,1]$, viz. for $N = 128$, 64, 32, 16 and 8 subintervals. The solution with 128 subintervals was been used as a reference solution and compared with the asymptotic solution. The numerical solution is computed for $\varepsilon = 10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}, 10^{-12}$; the same ε- sequence was used for the Newton-Kantorovich-Davidenko process. Each solution was obtained by iteration with method A until convergence was obtained; thereafter the methods B and C were applied once. The initial approximation was $y \equiv 0$. To perform the whole iteration process, at most 31 iteration steps were necessary.

ε	10^{-1}	10^{-2}	10^{-4}	10^{-8}	10^{-12}
e	6.63(-2)	1.72(-2)	9.04(-4)	2.01(-8)	8.74(-11)

Table 4.3.1

The difference between the numerical reference solution $y_{1/128}$ and the asymptotic approximation (4.3.2).

$$e = \| y_{1/128} - y_{asymp} \|_{\pi,\infty},$$
$$\Pi = \{ i/128 \mid i = 0,1,2,\ldots,128 \}.$$

The smoothly varying behaviour of the solution outside the boundary layer allows us to check the order of accuracy of methods A, B and C. The results are listed in table 4.3.2.a and 4.3.2.b. We see that method A shows almost uniform convergence of order 1 and convergence of order 2 for $h/\varepsilon \to 0$ (cf. theorem 2.4.1). Methods B and C are not uniformly convergent, but, in general, they are more accurate than method A. Moreover they show convergence of order 2 (respectively 4) both for $\varepsilon << h$ and for $h << \varepsilon$.

method	ε	e_8	${}_8r_{16}$	e_{16}	${}_{16}r_{32}$	e_{32}	${}_{32}r_{64}$	e_{64}
A	10^{-1}	8.26(-3)	1.87	2.26(-3)	1.96	5.82(-4)	1.99	1.47(-4)
	10^{-2}	4.98(-2)	1.17	2.22(-2)	0.58	1.49(-2)	0.84	8.35(-3)
	10^{-4}	6.68(-2)	0.90	3.58(-2)	0.96	1.84(-2)	0.99	9.28(-3)
	10^{-8}	6.70(-2)	0.90	3.60(-2)	0.95	1.86(-2)	0.97	9.48(-3)
B	10^{-1}	6.10(-3)	2.12	1.40(-3)	2.04	3.41(-4)	2.00	8.53(-5)
	10^{-2}	3.97(-3)	-1.17	8.91(-3)	-2.31	4.45(-2)	2.02	1.10(-2)
	10^{-4}	2.78(-3)	2.04	6.77(-4)	1.98	1.72(-4)	0.74	1.03(-4)
	10^{-8}	2.78(-3)	2.04	6.76(-4)	1.99	1.71(-4)	1.99	4.30(-5)
C	10^{-1}	2.28(-4)	2.58	1.73(-5)	3.99	1.09(-6)	4.09	6.43(-8)
	10^{-2}	2.05(-2)	-1.58	6.11(-2)	4.05	3.69(-3)	3.09	4.32(-4)
	10^{-4}	9.34(-6)	-0.26	1.12(-5)	-2.01	4.52(-5)	-1.99	1.78(-4)
	10^{-8}	9.38(-6)	4.05	5.67(-7)	4.69	2.22(-8)	4.26	1.26(-9)

Table 4.3.2.a

The pointwise errors and observed convergence rates associated with problem (4.3.1).

$$e_N = \|y_{1/N} - y_{1/128}\|_{\Pi,\infty},$$

$$\Pi = \{i/N \mid i = 0,1,\dots,N\},$$

$${}_Nr_{2N} = {}^2\log(e_N/e_{2N}).$$

method	ε	e_8	${}_8r_{16}$	e_{16}	${}_{16}r_{32}$	e_{32}	${}_{32}r_{64}$	e_{64}
A	10^{-1}	8.26(-3)	1.87	2.26(-3)	1.96	5.82(-4)	1.99	1.47(-4)
	10^{-2}	4.98(-2)	1.45	1.82(-2)	1.75	5.42(-3)	1.91	1.44(-3)
	10^{-4}	6.68(-2)	0.99	3.36(-2)	1.00	1.68(-2)	1.01	8.32(-3)
	10^{-8}	6.70(-2)	0.99	3.37(-2)	0.99	1.69(-2)	0.99	8.49(-3)
B	10^{-1}	6.10(-3)	2.12	1.40(-3)	2.04	3.41(-4)	2.01	8.48(-5)
	10^{-2}	3.97(-3)	1.73	1.20(-3)	-1.16	2.69(-3)	3.66	2.13(-4)
	10^{-4}	2.78(-3)	2.06	6.65(-4)	1.96	1.71(-4)	1.98	4.34(-5)
	10^{-8}	2.78(-3)	2.07	6.63(-4)	1.96	1.71(-4)	1.99	4.29(-5)
C	10^{-1}	2.28(-4)	3.96	1.47(-5)	3.98	9.30(-7)	4.10	5.42(-8)
	10^{-2}	2.05(-2)	2.98	2.59(-3)	0.65	1.65(-3)	8.67	4.04(-6)
	10^{-4}	9.34(-6)	4.17	5.21(-7)	5.03	1.59(-8)	4.25	8.35(-10)
	10^{-8}	9.38(-6)	4.16	5.26(-7)	5.03	1.61(-8)	4.15	9.04(-10)

Table 4.3.2.b.

The pointwise errors and convergence rates, observed on a fixed mesh Π_0, which avoids the boundary layer.

$$e_N = \|y_{1/N} - y_{1/128}\|_{\pi_0,\infty},$$

$$\Pi_0 = \{i/8 \mid i = 0,1,\dots,8\},$$

$${}_N r_{2N} = {}^2\log(e_N/e_{2N}).$$

EXAMPLE 4.3.2. We consider the equation

$$(4.3.3)\qquad \varepsilon y'' + (y')^2 = 1;$$

the general solution of which is

$$(4.3.4)\qquad y = A + \varepsilon \log\cosh\left(\frac{x-B}{\varepsilon}\right).$$

The limit solution as $\varepsilon \to 0$ is

$$y_0 = A + |x - B|.$$

The problem originates from PEARSON [1968b] and is also discussed by WASOW [1970]. The problem exhibits a corner layer at $x = B$ and satisfies conditions A1-A3 of section 4.1. Therefore, we can apply the methods described in section 4.2 to obtain a numerical solution outside the corner layer. For the non-quasilinear equation (4.3.3), method A requires some difference approximation of y' to construct a linearized problem (4.2.9). Since all possible difference quotients show their own particular properties, we do not use method A. We start iteration with method B until convergence is attained, thereafter method C had to be applied once or twice. A straight line between the boundary values is used as an initial approximation.

With $A = 1.0$ and $B = 0.745$ and for $\varepsilon = 10^{-1}$, 10^{-2}, 10^{-4} and 10^{-8}, the problem was solved on $[0,1]$, the boundary conditions being prescribed by (4.3.4). The errors in the corner layer and in the smooth part of the solution were measured separately by use of the norms $\|\cdot\|_{\pi,\infty}$ and $\|\cdot\|_{\pi_0,\infty}$, where Π denotes the equidistant partition of N intervals and $\Pi_0 = \{x \in \Pi \mid x < 0.7 \vee x > 0.8\}$. The results are listed in table 4.3.3.

It appears that outside the corner layer an accurate approximation is obtained on a mesh that is not at all adapted to the particular properties of the solution. Since the limit solution outside the corner layer consists of linear pieces only, method C did not yield essentially better results than method B.

	$\|y - y_h\|_{\pi,\infty}$			$\|y - y_h\|_{\pi_0,\infty}$		
ε \ N	15	30	45	15	30	45
10^{-1}	1.81(-3)	4.39(-4)	1.84(-4)	1.07(-3)	2.52(-4)	1.35(-4)
10^{-2}	6.87(-3)	5.88(-3)	2.24(-3)	5.27(-4)	8.53(-7)	2.21(-5)
10^{-4}	1.52(-2)	2.16(-2)	1.20(-2)	7.15(-6)	9.75(-6)	1.40(-9)
10^{-8}	1.53(-2)	1.45(-2)	3.25(-2)	7.22(-10)	4.62(-10)	7.18(-13)

Table 4.3.3

The errors observed inside and outside the turning-point region.

$$\Pi = \{i/N \mid i = 0,1,2,\dots,N\},$$
$$\Pi_0 = \{x \in \Pi \mid x < 0.7\} \cup \{x \in \Pi \mid x > 0.8\}.$$

EXAMPLE 4.3.3. We consider the problem

$$(4.3.5) \qquad \varepsilon y'' + yy' - y = 0, \qquad 0 \le x \le 1,$$
$$y(0) = \alpha,\ y(1) = \beta.$$

Various aspects of this classical problem have been treated by e.g. COLE [1968, pp.29-38], O'MALLEY [1968, pp.389-390], PEARSON [1968b, p.356], DORR [1970a, p.307], WASOW [1970] and DORR et al. [1973, pp.57-63]. Asymptotic expressions for $\varepsilon \to 0$ are derived for the solution in COLE [1968].

The character of the solution depends on α and β and it may involve boundary layers, corner layers or a shock layer. For various values of α and β a sketch of the solution is given in figure 4.3.1.

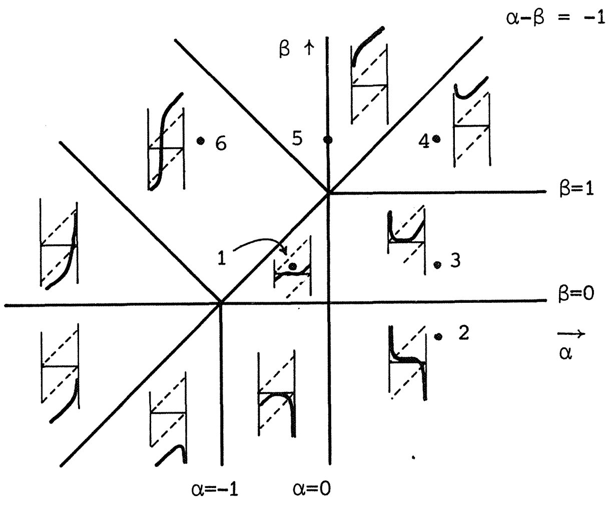


Fig. 4.3.1

The behaviour of the solution of 4.3.5 for small values of ε and for different values of α and β.

For various values of ε the solution was computed with the following boundary conditions:

problem number	α	β	Remarks
1.	-1/3	1/3	corner layers at x = 1/3 and x = 2/3.
2.	1	-1/3	boundary layers at x = 0 and x = 1.
3.	1	1/3	boundary layer at x = 0; corner layer at x = 2/3.
4.	1	3/2	boundary layer at x = 0.
5.	0	3/2	boundary layer at x = 0.
6.	-7/6	3/2	shock layer at x = 1/3.

The Newton-Kantorovich-Davidenko process was started with the linear function between the boundary values as an initial guess. The ε- sequence used was $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-8}, 10^{-12}\}$. All different kinds of global behaviour, known from the asymptotic analysis, were recovered by the numerical method on an equidistant mesh of 16 or 32 subintervals. The deviation from the limit solution for ε → 0 is given in the table 4.3.4. The difference from a reference solution, computed on a mesh of 48 subintervals, is shown in table 4.3.5.

	$\|y_{1/32} - y_{\text{lim}}\|_{\pi,\infty}$			
problem \ ε	10^{-2}	10^{-4}	10^{-8}	10^{-12}
1	4.77(-2)	2.00(-3)	7.73(-3)	7.74(-3)
2	3.46(-1)	5.05(-3)	5.06(-5)	1.81(-5)
3	3.46(-1)	5.04(-3)	7.73(-3)	7.74(-3)
4	6.99(-2)	4.31(-5)	2.13(-11)	2.24(-13)
5	1.50(-1)	1.30(-4)	1.82(-4)	5.01(-5)
6	3.00(-1)	2.07	3.41	3.18

	$\|y_{1/32} - y_{\text{lim}}\|_{\pi_0,\infty}$			
problem \ ε	10^{-2}	10^{-4}	10^{-8}	10^{-12}
1	1.53(-2)	1.17(-7)	1.78(-14)	1.42(-14)
2	7.60(-2)	5.99(-7)	1.17(-14)	2.34(-21)
3	7.63(-2)	6.49(-7)	1.87(-14)	1.87(-14)
4	1.34(-3)	1.85(-13)	2.31(-13)	1.99(-13)
5	1.34(-3)	1.74(-13)	1.88(-13)	2.27(-13)
6	1.38(-3)	1.07(-10)	1.21(-13)	9.24(-14)

Table 4.3.4

The difference between the numerical solution and exact limit solution for ε → 0. The pointwise error has been observed on Π (the whole interval) and on Π_0 (the smooth part of the solution).

$$\Pi = \{i/32 \mid i = 0,1,2,\ldots,32\};$$
$$\Pi_0 = \{i/32 \mid i = 2,3,7,8,9,13,14\}.$$

problem	N \ ε	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-8}
1	16	9.75(-8)	4.95(-5)	9.03(-4)	4.44(-5)	1.78(-8)
	32	4.93(-9)	5.51(-7)	5.95(-4)	6.40(-8)	1.60(-14)
2	16	5.63(-6)	1.22(-3)	1.37(-3)	9.90(-6)	1.35(-12)
	32	4.82(-7)	1.49(-4)	4.98(-4)	2.65(-7)	1.17(-14)
3	16	6.29(-6)	1.22(-3)	1.17(-3)	4.44(-5)	1.78(-12)
	32	4.87(-7)	1.49(-4)	5.95(-4)	2.90(-7)	1.57(-14)
4	16	1.30(-5)	1.77(-3)	5.29(-8)	1.24(-11)	6.39(-14)
	32	5.89(-7)	2.28(-3)	4.61(-8)	7.11(-14)	9.95(-14)
5	16	5.06(-6)	1.86(-3)	1.53(-7)	4.86(-10)	1.56(-13)
	32	2.51(-7)	2.28(-3)	6.05(-8)	3.59(-13)	1.56(-13)
6	16	5.58(-5)	1.09(-1)	1.69(-3)	1.66(-5)	4.89(-1)
	32	2.99(-6)	2.16(-3)	1.10(-8)	1.07(-10)	9.59(-14)

Table 4.3.5

The difference between the solution on an equidistant mesh of 48 points (reference solution) and equidistant meshes of 16 or 32 points, measured outside the rapidly varying regions:

$$\|y_{1/48} - y_{1/N}\|_{\pi_0,\infty},$$

$$\Pi_0 = \{i/16 \mid i = 2,3,7,8,9,13,14\}.$$

EXAMPLE 4.3.4. This problem describes the shock wave in a one-dimensional nozzle flow (PEARSON [1968b]). The Navier-Stokes equations reduce to the single equation

$$\varepsilon A y y'' - \left[\frac{1+\gamma}{2} - \varepsilon A'\right] y y' + \frac{y'}{y} + \frac{A'}{A}\left(1 - \frac{\gamma-1}{2} y^2\right) = 0, \qquad 0 \le x \le 1, \tag{4.3.8}$$

where γ = 1.4 the ratio of specific heats,

$\varepsilon = 4\gamma/3\mathrm{Re}$,

Re = Reynolds number.

We use the same additional data as mentioned in PEARSON [1968b], viz.

$$
\begin{aligned}
y(0) &= 0.9129, \\
y(1) &= 0.375\ , \\
A &= 1 + x^2, \\
\varepsilon &= 4.792\ 10^{-8}.
\end{aligned}
\tag{4.3.9}
$$

The linearized form of the problem reads

$$
\begin{aligned}
&\varepsilon(Ay'_{n+1})' + (y_n^{-2}-1.2)y'_{n+1} - (\tfrac{A'}{A}(y_n^{-2}+0.2)+2y'_ny_n^{-3})y_{n+1} = \\
&\qquad = -2(\tfrac{A'}{A}+y'_ny_n^{-1})y_n^{-1}.
\end{aligned}
\tag{4.3.10}
$$

The location of the turning point is that value of x for which $y(x;\varepsilon) = 1/\sqrt{1.2}$ and depends on the value of ε. Numerical computations indicate that, for $\varepsilon = 0.5$, the solution is a monotonic decreasing function; for $\varepsilon = 0.1$ it has a maximum near $x = 0.25$ and a turning point near $x = 0.4$. As $\varepsilon \to 0$, both the turning point and the maximum move to the right. Both points approach $x = 0.63$ as $\varepsilon \to 10^{-6}$ (fig. 4.3.2). Because of the moving turning point, condition A2 of section 4.1 is not satisfied unless a rather large region R is assumed.

For $\varepsilon = 4.792\ 10^{-8}$, the problem was solved by Newton-Kantorovich-Davidenko iteration. A straight line between the boundary values was used as an initial approximation. The ε-sequence chosen was

$$
\{0.5,\ 0.1,\ 0.05,\ 0.01,\ 5.0(-3),\ 1.0(-3),\ 5.0(-4),\ 1.0(-4),\ 1.0(-6),\ 4.792(-8)\}.
$$

The problem was solved on an equidistant grid Π_1 of 120 points and on a non-equidistant grid Π_2 of 210 points (see fig. 4.3.2). For Π_2, the smallest mesh size was still considerably larger than the final shock layer (whose width is approx. 10^{-7} (PEARSON [1968b])).

As a convergence criterion, the condition $\|y_{n+1}-y_n\|_{\pi,\infty} < 1.0(-7)$ was used, where $\Pi \subset \Pi_j$ is an arbitrary subset of gridpoints, which contains at least 2/3 of the gridpoints in Π_j $(j = 1,2)$. The Newton-Kantorovich-Davidenko process appeared to be sensitive to the convergence criterion; to obtain convergence in the case of the coarser grid Π_1, additional ε-values were inserted by an automatic device, viz. $\varepsilon = 0.03,\ 0.02,\ 0.015,\ 0.0125,\ 0.01125$.

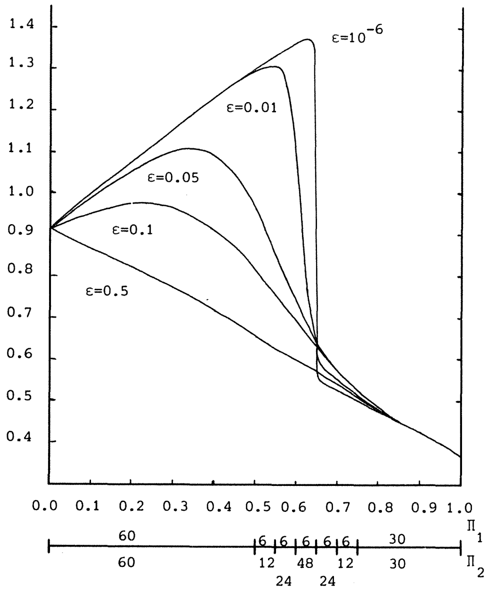


Fig. 4.3.2

The solution of the problem (4.3.8-9) for various values of ε. Below: the number of mesh-intervals in the different subregions of $[0,1]$ for the non-equidistant partition Π_2:

$$
\begin{aligned}
h &= 1/120 \quad \text{if} \quad 0.00 < x < 0.50 \text{ or } 0.75 < x < 1.00,\\
h &= 1/240 \quad \text{if} \quad 0.50 < x < 0.55 \text{ or } 0.70 < x < 0.75,\\
h &= 1/480 \quad \text{if} \quad 0.55 < x < 0.60 \text{ or } 0.65 < x < 0.70,\\
h &= 1/960 \quad \text{if} \quad 0.60 < x < 0.65.
\end{aligned}
$$

To obtain convergence on the grid Π_2, however, the given ε-sequence was sufficient and the number of iteration steps was 36 (for method A) + 3 (method B) + 2 (method C). The location of the shock layer was determined to within an interval of length approx. 0.006 (i.e. 5 mesh-intervals). Furthermore, the methods gave an accurate approximation of both smooth parts of the solution. Inside the turning-point region the approximation failed and a properly adapted mesh would have been required to obtain an accurate representation in this region. The accuracy of the numerical approximation of the smooth parts was determined by comparing it with an approximation on a finer mesh. Outside the region (0.6250, 0.6375) the pointwise error on Π_2 was approximately 10^{-3} for method A, 10^{-5} for method B and 10^{-7} for method C.

4.4. ALGOL 68 ROUTINES

In this section we give the basic ALGOL 68 routines that were used to obtain the results in section 4.3. The main routines are WOSD and EFGAL. These routines solve a linear problem of the form

$$(4.4.1) \qquad \begin{array}{ll} (-c_0y')' + c_1y' + c_2y = c_3, & \text{on } a \le x \le b, \\ y(a) = \alpha,\ y(b) = \beta. & \end{array}$$

They can also be used to perform one step in the Newton-Kantorovich process for a nonlinear problem. The approximate solution is computed on a partition that is specified by the user. To define the problem, the headings of WOSD and EFGAL contain

```
ref [ ] real XX,YY,
proc (real, real, real) [ ] real EQTN.
```

XX[0: upb XX] should contain the partition of [a,b], i.e.

$$a = XX[0] < XX[1] < \dots < XX[\underline{upb}\ XX] = b.$$

Upon entry YY[0] and YY[upb XX] should specify the boundary conditions

$$YY[0] = \alpha,\ YY[\underline{upb}\ XX] = \beta.$$

The other elements YY[i], $1 \le i \le$ upb XX - 1, should contain an initial estimate of $y(x_i)$. (These values are irrelevant in the case of a linear problem.) Upon exit YY[i] contains the approximate solution at x = XX[i]; i = 0,1,2,..., upb XX. The coefficients c_0, c_1, c_2, c_3, that may depend on x, y and y', are communicated to WOSD or EFGAL by the ALGOL 68 routine EQTN. This routine should deliver the [1:4] real:

$$(-c_0(x,y,y'),\ c_1(x,y,y'),\ c_2(x,y,y'),\ c_3(x,y,y')).$$

For the computation of this [ ] real, the values of x, y and y' are communicated to EQTN by its 3 parameters respectively.

WOSD applies the exponentially fitted finite difference method (3.5.12) on a possibly non-uniform partition. EFGAL applies a classical Galerkin method or the exponentially fitted weighted residual method as described in section 3.5. The particular method used by EFGAL depends on the parameter METHOD. This parameter refers to a set of method-defining coefficients that are calculated beforehand by the procedure METHOD. The coefficients computed by the procedure METHOD depend on the integer parameter CODE.

METHOD (CODE), CODE = +1, +2, +3, +4, delivers the coefficients for the exponentially fitted method described in eqs. (3.5.15)-(3.5.23), where k = CODE. This set of coefficients causes EFGAL to solve the problem by this method, or (if z in (3.5.22-23) is small, i.e. $^2\log|z|^2 < k+1$) by the classical Galerkin method.

METHOD (0) delivers an empty set of coefficients and causes EFGAL to reduce to WOSD.

METHOD (CODE), CODE = -1, -2, -3, -4, causes EFGAL to solve the problem by the classical Galerkin method (i.e. without exponential fitting), using the efficient implementation given by (3.1.31-32), where k = -CODE.

Some auxiliary modes, operators and procedures are used: the modes vector, matrix, tridiamat (a tridiagonal matrix) and method (a structure with references to method-defining coefficients) are introduced. The operators * and ich are introduced; both operators work on two vectors: * computes the scalar product of the two vectors; ich interchanges the corresponding elements of two vectors. The procedure TRIDSOL solves a linear system with a tridiagonal coefficient matrix; for a description of the Gaussian elimination process and a discussion of its stability properties for two-point boundary-value problems see BABUŠKA [1972].

```
#
EFGAL:
´BEGIN´  # WOSD AND EFGAL #

 ´MODE´ ´VECTOR´ = ´REF´ [ ] ´REAL´;
 ´MODE´ ´MATRIX´ = ´REF´ [,] ´REAL´;
 ´MODE´ ´TRIDIAMAT´ = ´STRUCT´ (´VECTOR´ SUB,DIA,SUP);
 ´MODE´ ´METHOD´ = ´STRUCT´(´VECTOR´ SUBN,W,SPW,PHI,
                            ´MATRIX´ WCOF,CSPW,COEF,COEI,CWWI,PHID);
 ´PRIO´ ´ICH´ = 4;

 #SCALAR PRODUCT#
 ´OP´ * = (´VECTOR´ A,B) ´REAL´:
 (´REAL´ S:= 0;
  ´FOR´ I ´FROM´ ´LWB´ A ´TO´ ´UPB´ A
  ´DO´ S +:= A[I]*B[I] ´OD´; S);

 #INTERCHANGE#
 ´OP´ ´ICH´ = (´VECTOR´ A,B) ´VOID´:
 ´FOR´ I ´FROM´ ´LWB´ A ´TO´ ´UPB´ A
 ´DO´ ´REAL´ S = A[I]; A[I]:= B[I]; B[I]:= S ´OD´;

 #SOLUTION TRIDIAGONAL SYSTEM#
 ´PROC´ TRIDSOL := (´TRIDIAMAT´ MAT, ´VECTOR´ F) ´VECTOR´ :
 ´BEGIN´ #FOR A MATRIX OF POSITIVE TYPE#
         ´VECTOR´ A #[1:N  ]# = DIA ´OF´ MAT,
                  B #[1:N-1]# = SUP ´OF´ MAT,
                  C #[1:N-1]# = SUB ´OF´ MAT;
         ´INT´    N = ´UPB´ F;  ´INT´  I:= 1;
         ´REAL´   P,G:=  F[1];

         ´FOR´ J ´FROM´ 2 ´TO´ N
         ´DO´ A[J]-:= B[I] * (P:= C[I]/A[I]);
             G:= F[I:=J] -:= G * P
         ´OD´;
         F[N]:= G /:= A[N];
         ´FOR´ J ´FROM´ N-1 ´BY´ -1 ´TO´ 1
         ´DO´ G:= (F[J]-:= B[J]*G) /:= A[J] ´OD´;
         F
 ´END´ # TRIDSOL #;
#
```

```
#

 ´PROC´ WOSD = (´VECTOR´ XX,YY, ´PROC´(´REAL´,´REAL´,´REAL´)
                [ ]´REAL´ EQTN) ´VOID´:
 ´BEGIN´ ´INT´ N = ´UPB´ XX;
         [1:4] ´REAL´ EVAL,
         [0:N] ´REAL´ SUB,DIA,SUP;
         ´VECTOR´     RHS = YY;
         ´REF´ ´REAL´ EE  = EVAL[1], FF= EVAL[2],
                      GG  = EVAL[3], RR= EVAL[4];

         #THE FUNCTION M, DEFINED BY EQ.(2.4.8)#
         ´PROC´ M = (´REAL´ A) ´REAL´:
         ´IF´ ´REAL´ X, W:= ´ABS´ A; W < 0.2
         ´THEN´ W*:= W; (((( W - 9.9 ) * W + 99.0 ) * W
                - 1039.5 ) * W + 15592.5) * A / 46777.5
         ´ELSE´ X:= (W-1.0)/W;
                (W < 18.0 ! X+:= 2.0/(EXP(W + W)-1.0));
                (A >  0.0 !  X ! -X )
         ´FI´ # M #;

         ´REAL´ XK,H,K,EH,EK,KH,MM,YK,
                XH:=XX[1],YH:=YY[1],YM:=YY[0];
         H:= XH - XX[0];
         DIA[0]:= DIA[N]:= 1;
         SUB[N]:= SUP[0]:= 0;

         ´FOR´ I ´TO´ N - 1
         ´DO´ XK:= XX[I+1]; YK:= YY[I+1];
              K := XK - XH; KH:=  K + H ;
              EVAL:= EQTN(XH,YH,(YK-YM)/KH);
              EE *:= 2.0; EH:= EE/H; EK:= EE/K;
              MM  := M(( FF*EE<0 ! FF/EH ! FF/EK));
              KH +:= (K-H)*MM;   MM*:= FF;
              SUB[I]:= EH - FF + MM;
              DIA[I]:=-EH - EK - MM - MM + GG * KH;
              SUP[I]:= EK + FF + MM;
              RHS[I]:= RR * KH;
              XH:= XK; H:= K; YM:= YH; YH:= YK
         ´OD´;
         TRIDSOL((SUB,DIA[´AT´1],SUP[´AT´1]),RHS[´AT´1])
 ´END´ # WOSD #;
#
```

```
#


'PROC' METHOD = ('INT' CODE) 'METHOD':
'IF' CODE = 0
'THEN' ('NIL','NIL','NIL','NIL','NIL','NIL','NIL','NIL','NIL','NIL')
'ELSE'
       'INT' AC= 'ABS' CODE;
       'INT' NC= AC + 1;
       'HEAP' [1:NC,1:NC] 'REAL' WCOF,CSPW,COEF,COEI,CWWI,PHID,
       'HEAP' [1:NC]      'REAL' SUBN,W   ,SPW ,PHI;

       [,] 'REAL' PHIS =
       #THE COEFFICIENTS OF THE POLYNOMIALS CAPITAL PHI,EQ.(3.1.21)#
       'CASE' AC
       'IN' 'BEGIN' SUBN:= ( 0, 1);
                  (( 1, -1,  0),
                   ( 0,  1,  0))
            'END',
            'BEGIN' SUBN:= ( 0, .5, 1);
                  (( 1, -3,  2),
                   ( 0,  4, -4),
                   ( 0, -1,  2))
            'END',
            'BEGIN' 'REAL' A,B,C:=SQRT(5);
                   B:= (5+C)/10; A:= 0.2/B;
                   SUBN:= ( 0,  A,  B,  1);
                   C*:= 5; B:= (C+5)/2 ; A:= 25/B;
                  (( 1, -6,  10, -5),
                   ( 0,  B,-B-C,  C),
                   ( 0, -A, A+C, -C),
                   ( 0,  1,  -5,  5))
            'END',
            'BEGIN' 'REAL'  A, B:= 1/7, D:= (7+SQRT(21))/14,
                   P:= -3/49, Q:= 3/112; A:= B/D;
                   SUBN:= ( 0, A, .5, D, 1 );
                  (( 1, -10  ,     30   ,   -35     ,14  ),
                   ( 0, -D/P , (1+3*D)/P, (-3-2*D)/P, 2/P),
                   ( 0, -B/Q , (1+ B)/Q,    -2/Q    , 1/Q),
                   ( 0, -A/P , (1+3*A)/P, (-3-2*A)/P, 2/P),
                   ( 0, -1   ,      9   ,   -21     , 14 ))
            'END'
       'ESAC' # PHIS #;
#
```

```
#

        #CONSTRUCTION OF METHOD-DEFINING COEFFICIENTS#
        'FOR' I 'TO' NC
        'DO' SPW[I]:= SUBN[I] *( PHI[I]:= PHIS[I,2] );
             W[I]  :=
             ('REAL'S:=0;'FOR' J 'TO' NC 'DO' S+:=PHIS[I,J]/J 'OD';S);
             'FOR' K 'TO' NC
             'DO' COEF[K,I]:= ('REAL' S:= AC*PHIS[I,NC], SK:=SUBN[K];
                              'FOR' J 'FROM' AC-1 'BY' -1 'TO' 1
                              'DO'( S *:=SK )+:= J*PHIS[I,J+1] 'OD'; S
                              );
                  CWWI[K,I]:= 'IF' K = 1
                              'THEN' SUBN[I]*PHIS[I,3]
                              'ELSE' COEF[K,I]*SUBN[I]/SUBN[K]
                              'FI';
                  PHID[K,I]:= 2*PHIS[I,3]*PHIS[K,1]+PHIS[I,2]*PHIS[K,2]
             'OD';
             'IF' I /= 1 'THEN' CWWI[I,I] -:= 1/SUBN[I] 'FI'
        'OD';

        'FOR' I 'TO' NC
        'DO' SPW[I] *:= W[1]/W[I];
             'FOR' K 'TO' NC
             'DO' 'REAL' C   = COEF[K,I];
                  WCOF[K,I] := C*W[K];
                  COEI[K,I] := C/W[I];
                  CWWI[K,I]/:=   W[I];
                  CSPW[I,K] := COEF[1,K]*SPW[I]
             'OD'
        'OD' #CONSTRUCTION COEFFICIENTS#;

        'IF'   CODE > 0
        'THEN' (SUBN,W, SPW,  PHI, WCOF, CSPW,COEF,COEI, CWWI, PHID)
        'ELSE' (SUBN,W,'NIL','NIL',WCOF,'NIL',COEF,COEI,'NIL','NIL')
        'FI'
'FI' #PROC METHOD# ;
#
```

```
#

'PROC' EFGAL = ('METHOD'METHOD, 'VECTOR' XX,YY,
               'PROC'('REAL','REAL','REAL')[]'REAL'EQTN) 'VOID':
'IF' SUBN 'OF' METHOD :=: 'VECTOR' ('NIL')
'THEN' WOSD(XX, YY, EQTN)
'ELSE'  'VECTOR' SUBN = SUBN'OF'METHOD,  W    = W   'OF'METHOD,
                 SPW  = SPW 'OF'METHOD,  PHI  = PHI 'OF'METHOD,
        'MATRIX' WCOF = WCOF'OF'METHOD,  CSPW = CSPW'OF'METHOD,
                 COEF = COEF'OF'METHOD,  COEI = COEI'OF'METHOD,
                 CWWI = CWWI'OF'METHOD,  PHID = PHID'OF'METHOD;

   'BOOL' EF = (PHID'OF'METHOD:/=:'MATRIX'('NIL'));
   'INT'  NC = 'UPB' SUBN, NR = 'UPB' XX;
   'INT'  AC = NC - 1;

   [1:NC, 1:4]'REAL'EVALS,
   [1: 4,1:NC]'REAL'   WW,
   [1:   NR+1]'REAL'  SUB,DIA,SUP,
   [1:NC,0:NC]'REAL'     A;

   'PROC' ('INT','INT') 'REAL' CC =
   'IF'   AC > 2
   'THEN' ('INT' I,J)'REAL': A[I,J] - A[I,2:AC]*A[2:AC,J]
   'ELIF' AC = 2
   'THEN' ('INT' I,J)'REAL': A[I,J] - A[I,2]   *A[2,J]
   'ELSE' ('INT' I,J)'REAL': A[I,J]
   'FI';

   'VECTOR' RHS = YY['AT'1],
            EVALL=EVALS[1,], EVALR=EVALS[NC,],
            WA=WW[1,], WB=WW[2,], WC=WW[3,], WD=WW[4,];
   'REF' 'REAL' WA1=WA[1],  WB1=WB[1],  WC1=WC[1],  WD1=WD[1],
            EVALL1=EVALL[1], EVALL2=EVALL[2], EVALL3=EVALL[3],
            EVALR1=EVALR[1], EVALR2=EVALR[2], EVALR3=EVALR[3];

   'BOOL' POST:='FALSE',PRE:='FALSE',TWO:='FALSE';
   'INT'  I1,IN;
   'REAL' X := XX[0], Y := YY[0],
          XH:= XX[1], YH:= YY[1];
   'REAL' H := XH-X , Y1:=(YH-Y)/H,
          HH,XHH,YHH,Y1H,PE,PO,PW,
          ALPHA:= 0.0, RHS1:= Y,
          DIAR := 0.0, RHSR:= 0.0,
          CRIT := SQRT('REAL':(2**NC));

#
```

```
#
    ´FOR´ N ´TO´ NR
    ´DO´ ´IF´   N = 1
        ´THEN´ EVALL:= EQTN(X,Y,Y1);
               ( EF ! PO:= EVALL2*H/EVALL1 )
        ´ELSE´ X    := XH;       Y  := YH;
               XH   := XHH;      YH := YHH;
               H    := HH;       Y1 := Y1H;
               EVALL:= (TWO ! TWO:=´FALSE´; EQTN(X,Y,Y1) ! EVALR )
        ´FI´;
        EVALR  := EQTN(XH,YH,
        ´IF´   N = NR
        ´THEN´ Y1
        ´ELSE´ XHH  := XX[N+1]; YHH:= YY[N+1];
               HH   := XHH -XH; Y1H:=(YHH-YH)/HH;
               (TWO:= ´ABS´(Y1H-Y1)>0.1 ! Y1 ! 0.5*(Y1H+Y1) )
        ´FI´                  );

        ´FOR´ I ´FROM´ 2 ´TO´ AC
        ´DO´  EVALS[I,]:=(´REAL´  C  =  H*SUBN[I];
                          EQTN(X+C,Y+C*Y1,Y1)    )
        ´OD´;

        ´IF´ EF
        ´THEN´ PRE := CRIT < -( PE:= PO);
               POST:= CRIT <  ( PO:= EVALR2*H/EVALR1);

               ALPHA:=
                  ´IF´   POST ´EQ´ PRE
                  ´THEN´ 0.0
                  ´ELIF´ POST
                  ´THEN´ ((PW:= EVALR3*H/EVALR2)<-CRIT ! 0.0 ! PO-PW )
                  ´ELSE´ ((PW:= EVALL3*H/EVALL2)> CRIT ! 0.0 ! PW-PE )
                  ´FI´;
               ( PRE := ALPHA>CRIT ! ´SKIP´ ! POST := ´FALSE´ )
        ´FI´;

        ´FOR´ I ´TO´ NC
        ´DO´ ´REF´[]´REAL´ EVAL= EVALS[(POST!NC+1-I!I),];
            WW[,I] :=
            (EVAL[1]/H,(POST!-EVAL[2]!EVAL[2]),EVAL[3]*H,EVAL[4]*H)
        ´OD´;

#
```

```
#
        #CONSTRUCTION OF ELEMENT MATRIX (3.1.24) AND VECTOR (3.1.25)#
        ´IF´ PRE
        ´THEN´ ´REAL´ AW:= ALPHA * ALPHA;
               ´REAL´ MU:=(ALPHA > 50.0 ¦ 0.0
                    ¦ ALPHA * AW * (´REAL´ C=EXP(-ALPHA); C/(1.0-C)));
               ´REAL´ ZZ:= (A[1,0]:= ( ALPHA*WD1+PHI*WD )/
                    ( AW *:= W[1] ));

               ´FOR´ I ´FROM´ 2 ´TO´ NC
               ´DO´ A[I,0]:= WD[I] - SPW[I]*(ZZ-WD1) ´OD´;

               ´FOR´ J ´TO´ NC
               ´DO´ ´REAL´ ZZ:= (J=1 ¦ ALPHA*WC1 + PHI*WC ¦ 0.0);
                    ´FOR´ K ´TO´ NC
                    ´DO´ ZZ +:= MU*WCOF[K,J]*WA[K]
                            +   PHID[K,J]*(ALPHA*WA[K]+WB[K])
                    ´OD´;
                    A[1,J]:= (ZZ /:= AW);

                    ´FOR´ I ´FROM´ 2 ´TO´ NC
                    ´DO´ ´REAL´ Z:= COEF[I,J]*WB[I] + CSPW[I,J]*WB1;
                         ´FOR´ K ´TO´ NC
                         ´DO´ Z -:= WCOF[K,J]*CWWI[K,I]*WA[K] ´OD´;
                         A[I,J]:= (J=I ¦ Z+WC[I] ¦ Z ) - SPW[I] *
                                  (J=1 ¦ ZZ- WC1 ¦ ZZ)
                    ´OD´
               ´OD´
        ´ELSE´ ´FOR´ I ´TO´ NC
               ´DO´ ´FOR´ J ´TO´ NC
                    ´DO´ ´REAL´ Z:= COEF[I,J]*WB[I];
                         ´FOR´ K ´TO´ NC
                         ´DO´ Z -:= WCOF[K,J]*COEI[K,I]*WA[K] ´OD´;
                         A[I,J]:= (J=I ¦ Z+WC[I] ¦ Z )
                    ´OD´;
                    A[I,0]:= WD[I]
               ´OD´
        ´FI´ #ELEMENT MATRIX AND VECTOR CONSTRUCTION#;
#
```

```
#

        #STATIC CONDENSATION#
        'IF' AC>2
        'THEN' 'FOR' J 'FROM' 2 'TO' AC
               'DO' 'INT' JP1= J+1; 'REAL' SI,S:= 'ABS' A[J,J];
                    'INT' PJ:= J;
                    'FOR' I 'FROM' JP1 'TO' AC
                    'DO' ((SI:='ABS'A[I,J]) >S ! S:=SI; PJ:=I ) 'OD';
                    'IF'J /= PJ 'THEN' A[PJ,] 'ICH' A[J,]'FI'; S:= A[J,J];
                    'FOR' I 'FROM' JP1 'TO' AC
                    'DO' SI:= A[I,J]/S;
                         'FOR' K 'FROM' 0 'TO' NC
                         'DO' A[I,K] -:= A[J,K]*SI 'OD'
                    'OD'
               'OD';
               'FOR' J 'FROM' AC 'BY' -1 'TO' 2
               'DO' 'REAL' SI  = A[J,J]; 'REAL' AJ0 = A[J, 0]/:=SI,
                           AJ1 = A[J,1]/:= SI, AJNC = A[J,NC]/:=SI;
                    'FOR' I 'FROM' J-1 'BY' -1 'TO' 2
                    'DO' 'REAL' SI= A[I,J]; A[I, 0]-:= AJ0 *SI;
                         A[I,1] -:= AJ1*SI; A[I,NC]-:= AJNC*SI
                    'OD'
               'OD'
        'ELIF' AC=2
        'THEN' 'REAL' SI = A[2,2];
               'FOR' K 'FROM' 0 'TO' NC 'DO' A[2,K] /:= SI 'OD'
        'FI' #STATIC CONDENSATION# ;

        (POST ! I1:=NC; IN:=1 ! I1:= 1; IN:=NC);
        DIA[N]:= CC(I1,I1) + DIAR; SUP[N]:= CC(I1,IN);
        SUB[N]:= CC(IN,I1);        DIAR  := CC(IN,IN);
        RHS[N]:= CC(I1, 0) + RHSR; RHSR  := CC(IN, 0)
   'OD';

   RHS[1]:= RHS1;
   DIA[1]:= DIA[NR+1]:=1.0;
   SUP[1]:= SUB[NR` ]:=0.0;
   TRIDSOL((SUB,DIA,SUP),RHS)
'FI' # EFGAL #;

'PR' PROG 'PR'
'SKIP'
'END'

#
```

REFERENCES

ABRAHAMSSON, L.R., [1975], *A priori estimates for solutions of singular perturbations with a turning point*, Report No.56, Dept. Comp. Sc., Uppsala University.

ABRAHAMSSON, L.R., H.B. KELLER & H.O. KREISS [1974], *Difference approximations for singular perturbations of systems of ordinary differential equations*, Numer. Math. 22 (1974) pp.367-391.

ABRAMOWITZ, M., & I.A. STEGUN [1965], *Handbook of mathematical functions*, Dover Publ., New York, 5th printing, 1968.

ACKERBERG, R.C., & R.E. O'MALLEY, JR. [1970], *Boundary layer problems exhibiting resonance*, Stud. Appl. Math. 49 (1970) pp.277-295.

AXELSSON, O., [1969], *A class of A-stable methods*, BIT 9 (1969) pp.185-199.

AZIZ, A.K., ed. [1972], *The mathematical foundations of the finite element method with applications to partial differential equations*, Academic Press, New York, 1972.

BABUŠKA, I., [1972], *Numerical stability in problems of linear algebra*, SIAM J. Numer. Anal. 9 (1972) pp.53-77.

BABUŠKA, I., & A.K. AZIZ [1972], *Survey lectures on the mathematical foundations of the finite element method*, in: Aziz ed. [1972] pp.1-359.

BAKKER, M., [1976], *Numerical solution of mildly nonlinear two-point boundary value problems by means of Galerkin's method*, Report NW 27/76, Math. Centrum, Amsterdam.

CIARLET, P.G., & P.A. RAVIART [1972], *General Lagrange and Hermite interpolation in R^n with applications to the finite element methods*, Arch. Rat. Mech. Anal. 46 (1972) pp.177-199.

CIARLET, P.G., M.H. SCHULTZ & R.S. VARGA [1969], *Numerical methods of high-order accuracy for nonlinear boundary value problems*, V. *Monotone operator theory*, Numer. Math. 13 (1969) pp.51-77.

CODDINGTON, E.A., & N. LEVINSON [1952], *A boundary value problem for a nonlinear differential equation with a small parameter*, Proc. Amer. Math. Soc. 3 (1952) pp.73-81.

COLE, J.D., [1968], *Perturbation methods in applied mathematics*, Blaisdell, Waltham, Mass. 1968.

DAVIS, P.J., [1963], *Interpolation and approximation*, Blaisdell Publ., New York, 1963.

DAVIS, P.J., & P. RABINOWITZ [1967], *Numerical integration*, Blaisdell, Waltham, Mass., 1972.

DE BOOR, C., & B. SWARTZ [1973], *Collocation at Gaussian points*, SIAM J. Numer. Anal. 10 (1973) pp.582-606.

DORR, F.W., [1970a], *The numerical solution of singular perturbations of boundary-value problems*, SIAM J. Numer. Anal. 7 (1970) pp.281-313.

DORR. F.W., [1970b], *Some examples of singular perturbation problems with turning points*, SIAM J. Math. Anal. 1 (1970) pp.141-146.

DORR, F.W., S.V. PARTER & L.F. SHAMPINE [1973], *Applications of the maximum principle to singular perturbation problems*, SIAM Review 15 (1973) pp.43-88.

DOUGLAS, J., JR. & T. DUPONT [1974], *Galerkin approximations for the two-point boundary problem using continuous piecewise polynomial spaces*, Numer. Math. 22 (1974) pp.99-109.

ECKHAUS, W., [1973], *Matched asymptotic expansions and singular perturbations*, North-Holland Publ. Comp., Amsterdam, 1973.

EHLE, B.L., [1968], *High order A-stable methods for the numerical solution of systems of D.E.'s*, BIT 8 (1968) pp.276-278.

GUDERLEY, K.G., [1975], *A unified view of some methods for stiff two-point boundary value problems*, SIAM Review 17 (1975) pp.416-442.

HEMKER, P.W., [1974], *A method of weighted one-sided differences for stiff boundary-value problems with turning points*, Report NW9/74, Math. Centrum, Amsterdam.

HENRICI, P., [1962], *Discrete variable methods in ordinary differential equations*, Wiley, New York, 1962.

HOWES, F.A., [1976], *Singular perturbations and differential inequalities*, Memoirs of the A.M.S., Vol 5, Issue 1, Number 168.

HULME, B.L., [1972a], *One-step piecewise polynomial Galerkin methods for initial value problems*, Math. Comp. 26 (1972) pp.415-426.

HULME, B.L., [1972b], *Discrete Galerkin and related one-step methods for ordinary differential equations*, Math. Comp. 26 (1972) pp.881-891.

IL'IN, A.M., [1969], *Differencing scheme for a differential equation with a small parameter affecting the highest derivative*, Math. Notes Acad. Sc. USSR 6 (1969) pp.596-602.

KELLER, H.B., [1974], *Accurate difference methods for nonlinear two point boundary value problems*, SIAM J. Numer. Anal. 11 (1974) pp.305-320.

KRASNOSEL'SKII, M.A., G.M. VAINIKKO, P.P. ZABREIKO, Ya. B. RUTITSKII & V. Ya. STETSENKO [1972], *Approximate solution of operator equations*, Wolters-Noordhoff. Publ., Groningen, 1972.

KREISS, H.O., [1973], *Difference approximations for singular perturbations of systems of ordinary differential equations*, in: Topics in numerical analysis (J. Miller ed.), Royal Irish Academy, 1973.

KREISS, H.O., [1974], *Numerical solution of singular perturbation problems*, in: Stiff differential systems (R.A. Willoughby ed.), Plenum Press, 1974.

KREISS, H.O., & S.V. PARTER [1974], *Remarks on singular perturbations with turning points*, SIAM J. Math. Anal. 5 (1974) pp.230-251.

O'MALLEY, R.E., JR. [1968], *Topics in singular perturbations*, Adv. Math. 2 (1968) pp.365-470.

O'MALLEY, R.E., JR. [1970], *On boundary value problems for a singularly perturbed equation with a turning point*, SIAM J. Math. Anal. 1 (1970) pp.479-490.

O'MALLEY, R.E., JR. [1974], *Introduction to singular perturbations*, Academic Press, New York, 1974.

PEARSON, C.E., [1968a], *On a differential equation of boundary layer type*, J. Math. Phys. 47 (1968) pp.134-154.

PEARSON, C.E., [1968b], *On nonlinear ordinary differential equations of boundary layer type*, J. Math. Phys 47 (1968) pp.351-358.

PROTTER, M.H., & H.F. WEINBERGER [1967], *Maximum principles in differential equations*, Prentice-Hall, Englewood Cliffs, N.J., 1967.

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

RUSSELL, R.D., [1975], *A comparison of collocation and finite differences for two-point boundary-value problems*, To be published.

RUSSELL, R.D., & L.F. SHAMPINE [1972], *A collocation method for boundary value problems*, Numer. Math. 19 (1972) pp.1-28.

SIROVICH, L., [1971], *Techniques of asymptotic analysis*, Springer, New York, 1971.

VARAH, J.M., [1974], *On the condition of piecewise polynomial finite element bases*, Techn. Report 74-02, Dept. of Comp. Sc., Univ. of B.C., Vancouver.

WASOW, W., [1965], *Asymptotic expansions for ordinary differential equations*, Interscience, New York, 1965.

WASOW, W., [1970], *The capriciousness of singular perturbations*, Nieuw Arch. Wisk. 18 (1970) pp.190-210.

WEISS, R., [1974], *The application of implicit Runge-Kutta and collocation methods to boundary-value problems*, Math. Comp. 28 (1974) pp.449-464.

WHITTAKER, E.T., & G.N. WATSON [1946], *A course of modern analysis*, (4th ed.), Cambridge Univ. Press, 1946.

YOSIDA, K., [1960], *Lectures on differential and integral equations*, Interscience, New York, 1960.

YOSIDA, K., [1965], *Functional analysis*, Springer, Berlin, (4-th ed.) 1974.

OTHER TITLES IN THE SERIES MATHEMATICAL CENTRE TRACTS

A leaflet containing an order-form and abstracts of all publications mentioned below is available at the Mathematisch Centrum, Tweede Boerhaavestraat 49, Amsterdam-1005, The Netherlands. Orders should be sent to the same address.

MCT 1 T. VAN DER WALT, *Fixed and almost fixed points*, 1963. ISBN 90 6196 002 9.

MCT 2 A.R. BLOEMENA, *Sampling from a graph*, 1964. ISBN 90 6196 003 7.

MCT 3 G. DE LEVE, *Generalized Markovian decision processes, part I: Model and method*, 1964. ISBN 90 6196 004 5.

MCT 4 G. DE LEVE, *Generalized Markovian decision processes, part II: Probabilistic background*, 1964. ISBN 90 6196 005 3.

MCT 5 G. DE LEVE, H.C. TIJMS & P.J. WEEDA, *Generalized Markovian decision processes, Applications*, 1970. ISBN 90 6196 051 7.

MCT 6 M.A. MAURICE, *Compact ordered spaces*, 1964. ISBN 90 6196 006 1.

MCT 7 W.R. VAN ZWET, *Convex transformations of random variables*, 1964. ISBN 90 6196 007 X.

MCT 8 J.A. ZONNEVELD, *Automatic numerical integration*, 1964. ISBN 90 6196 008 8.

MCT 9 P.C. BAAYEN, *Universal morphisms*, 1964. ISBN 90 6196 009 6.

MCT 10 E.M. DE JAGER, *Applications of distributions in mathematical physics*, 1964. ISBN 90 6196 010 X.

MCT 11 A.B. PAALMAN-DE MIRANDA, *Topological semigroups*, 1964. ISBN 90 6196 011 8.

MCT 12 J.A.TH.M. VAN BERCKEL, H. BRANDT CORSTIUS, R.J. MOKKEN & A. VAN WIJNGAARDEN, *Formal properties of newspaper Dutch*, 1965. ISBN 90 6196 013 4.

MCT 13 H.A. LAUWERIER, *Asymptotic expansions*, 1966, out of print; replaced by MCT 54 and 67.

MCT 14 H.A. LAUWERIER, *Calculus of variations in mathematical physics*, 1966. ISBN 90 6196 020 7.

MCT 15 R. DOORNBOS, *Slippage tests*, 1966. ISBN 90 6196 021 5.

MCT 16 J.W. DE BAKKER, *Formal definition of programming languages with an application to the definition of ALGOL 60*, 1967. ISBN 90 6196 022 3.

MCT 17 R.P. VAN DE RIET, *Formula manipulation in ALGOL 60, part 1*, 1968. ISBN 90 6196 025 8.

MCT 18 R.P. VAN DE RIET, *Formula manipulation in ALGOL 60, part 2*, 1968. ISBN 90 6196 038 X.

MCT 19 J. VAN DER SLOT, *Some properties related to compactness*, 1968. ISBN 90 6196 026 6.

MCT 20 P.J. VAN DER HOUWEN, *Finite difference methods for solving partial differential equations*, 1968. ISBN 90 6196 027 4.

MCT 21 E. WATTEL, *The compactness operator in set theory and topology*, 1968. ISBN 90 6196 028 2.

MCT 22 T.J. DEKKER, *ALGOL 60 procedures in numerical algebra, part 1*, 1968. ISBN 90 6196 029 0.

MCT 23 T.J. DEKKER & W. HOFFMANN, *ALGOL 60 procedures in numerical algebra, part 2*, 1968. ISBN 90 6196 030 4.

MCT 24 J.W. DE BAKKER, *Recursive procedures*, 1971. ISBN 90 6196 060 6.

MCT 25 E.R. PAERL, *Representations of the Lorentz group and projective geometry*, 1969. ISBN 90 6196 039 8.

MCT 26 EUROPEAN MEETING 1968, *Selected statistical papers, part I*, 1968. ISBN 90 6196 031 2.

MCT 27 EUROPEAN MEETING 1968, *Selected statistical papers, part II*, 1969. ISBN 90 6196 040 1.

MCT 28 J. OOSTERHOFF, *Combination of one-sided statistical tests*, 1969. ISBN 90 6196 041 X.

MCT 29 J. VERHOEFF, *Error detecting decimal codes*, 1969. ISBN 90 6196 042 8.

MCT 30 H. BRANDT CORSTIUS, *Excercises in computational linguistics*, 1970. ISBN 90 6196 052 5.

MCT 31 W. MOLENAAR, *Approximations to the Poisson, binomial and hypergeometric distribution functions*, 1970. ISBN 90 6196 053 3.

MCT 32 L. DE HAAN, *On regular variation and its application to the weak convergence of sample extremes*, 1970. ISBN 90 6196 054 1.

MCT 33 F.W. STEUTEL, *Preservation of infinite divisibility under mixing and related topics*, 1970. ISBN 90 6196 061 4.

MCT 34 I. JUHÁSZ, A. VERBEEK & N.S. KROONENBERG, *Cardinal functions in topology*, 1971. ISBN 90 6196 062 2.

MCT 35 M.H. VAN EMDEN, *An analysis of complexity*, 1971. ISBN 90 6196 063 0.

MCT 36 J. GRASMAN, *On the birth of boundary layers*, 1971. ISBN 90 6196 064 9.

MCT 37 J.W. DE BAKKER, G.A. BLAAUW, A.J.W. DUIJVESTIJN, E.W. DIJKSTRA, P.J. VAN DER HOUWEN, G.A.M. KAMSTEEG-KEMPER, F.E.J. KRUSEMAN ARETZ, W.L. VAN DER POEL, J.P. SCHAAP-KRUSEMAN, M.V. WILKES & G. ZOUTENDIJK, *MC-25 Informatica Symposium*, 1971. ISBN 90 6196 065 7.

MCT 38 W.A. VERLOREN VAN THEMAAT, *Automatic analysis of Dutch compound words*, 1971. ISBN 90 6196 073 8.

MCT 39 H. BAVINCK, *Jacobi series and approximation*, 1972. ISBN 90 6196 074 6.

MCT 40 H.C. TIJMS, *Analysis of (s,S) inventory models*, 1972. ISBN 90 6196 075 4.

MCT 41 A. VERBEEK, *Superextensions of topological spaces*, 1972. ISBN 90 6196 076 2.

MCT 42 W. VERVAAT, *Success epochs in Bernoulli trials (with applications in number theory)*, 1972. ISBN 90 6196 077 0.

MCT 43 F.H. RUYMGAART, *Asymptotic theory of rank tests for independence*, 1973. ISBN 90 6196 081 9.

MCT 44 H. BART, *Meromorphic operator valued functions*, 1973. ISBN 90 6196 082 7.

MCT 45 A.A. BALKEMA, *Monotone transformations and limit laws*, 1973. ISBN 90 6196 083 5.

MCT 46 R.P. VAN DE RIET, *ABC ALGOL, A portable language for formula manipulation systems, part 1: The language*, 1973. ISBN 90 6196 084 3.

MCT 47 R.P. VAN DE RIET, *ABC ALGOL, A portable language for formula manipulation systems, part 2: The compiler*, 1973. ISBN 90 6196 085 1.

MCT 48 F.E.J. KRUSEMAN ARETZ, P.J.W. TEN HAGEN & H.L. OUDSHOORN, *An ALGOL 60 compiler in ALGOL 60, Text of the MC-compiler for the EL-X8*, 1973. ISBN 90 6196 086 X.

MCT 49 H. KOK, *Connected orderable spaces*, 1974. ISBN 90 6196 088 6.

MCT 50 A. VAN WIJNGAARDEN, B.J. MAILLOUX, J.E.L. PECK, C.H.A. KOSTER, M. SINTZOFF, C.H. LINDSEY, L.G.L.T. MEERTENS & R.G. FISKER (Eds), *Revised report on the algorithmic language ALGOL 68*, 1976. ISBN 90 6196 089 4.

MCT 51 A. HORDIJK, *Dynamic programming and Markov potential theory*, 1974. ISBN 90 6196 095 9.

MCT 52 P.C. BAAYEN (ed.), *Topological structures*, 1974. ISBN 90 6196 096 7.

MCT 53 M.J. FABER, *Metrizability in generalized ordered spaces*, 1974. ISBN 90 6196 097 5.

MCT 54 H.A. LAUWERIER, *Asymptotic analysis, part 1*, 1974. ISBN 90 6196 098 3.

MCT 55 M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 1: Theory of designs, finite geometry and coding theory*, 1974. ISBN 90 6196 099 1.

MCT 56 M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry*, 1974. ISBN 90 6196 100 9.

MCT 57 M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 3: Combinatorial group theory*, 1974. ISBN 90 6196 101 7.

MCT 58 W. ALBERS, *Asymptotic expansions and the deficiency concept in statistics*, 1975. ISBN 90 6196 102 5.

MCT 59 J.L. MIJNHEER, *Sample path properties of stable processes*, 1975. ISBN 90 6196 107 6.

MCT 60 F. GÖBEL, *Queueing models involving buffers*, 1975. ISBN 90 6196 108 4.

* MCT 61 P. VAN EMDE BOAS, *Abstract resource-bound classes, part 1.* ISBN 90 6196 109 2.

* MCT 62 P. VAN EMDE BOAS, *Abstract resource-bound classes, part 2.* ISBN 90 6196 110 6.

MCT 63 J.W. DE BAKKER (ed.), *Foundations of computer science*, 1975. ISBN 90 6196 111 4.

MCT 64 W.J. DE SCHIPPER, *Symmetric closed categories*, 1975. ISBN 90 6196 112 2.

MCT 65 J. DE VRIES, *Topological transformation groups 1 A categorical approach*, 1975. ISBN 90 6196 113 0.

MCT 66 H.G.J. PIJLS, *Locally convex algebras in spectral theory and eigenfunction expansions*, 1976. ISBN 90 6196 114 9.

* MCT 67 H.A. LAUWERIER, *Asymptotic analysis, part 2.*
ISBN 90 6196 119 X.

MCT 68 P.P.N. DE GROEN, *Singularly perturbed differential operators of second order*, 1976. ISBN 90 6196 120 3.

MCT 69 J.K. LENSTRA, *Sequencing by enumerative methods*, 1977.
ISBN 90 6196 125 4.

MCT 70 W.P. DE ROEVER JR., *Recursive program schemes: semantics and proof theory*, 1976. ISBN 90 6196 127 0.

MCT 71 J.A.E.E. VAN NUNEN, *Contracting Markov decision processes*, 1976.
ISBN 90 6196 129 7.

MCT 72 J.K.M. JANSEN, *Simple periodic and nonperiodic Lamé functions and their applications in the theory of conical waveguides*,1977.
ISBN 90 6196 130 0.

* MCT 73 D.M.R. LEIVANT, *Absoluteness of intuitionistic logic.*
ISBN 90 6196 122 x.

MCT 74 H.J.J. TE RIELE, *A theoretical and computational study of generalized aliquot sequences*, 1976. ISBN 90 6196 131 9.

MCT 75 A.E. BROUWER, *Treelike spaces and related connected topological spaces*, 1977. ISBN 90 6196 132 7.

MCT 76 M. REM, *Associons and the closure statement*, 1976. ISBN 90 6196 135 1.

* MCT 77 W.C.M. KALLENBERG, *Asymptotic optimality of likelihood ratio tests in exponential families*, ISBN 90 6196 134 3.

MCT 78 E. DE JONGE, A.C.M. VAN ROOIJ, *Introduction to Riesz spaces*, 1977.
ISBN 90 6196 133 5.

MCT 79 M.C.A. VAN ZUIJLEN, *Empirical distributions and rankstatistics*, 1977.
ISBN 90 6196 145 9.

MCT 80 P.W. HEMKER, *A numerical study of stiff two-point boundary problems*, 1977. ISBN 90 6196 146 7.

MCT 81 K.R. APT & J.W. DE BAKKER (eds), *Foundations of computer science II*, part I, 1976. ISBN 90 6196 140 8.

MCT 82 K.R. APT & J.W. DE BAKKER (eds), *Foundations of computer science II*, part II, 1976. ISBN 90 6196 141 6.

* MCT 83 L.S. VAN BENTEM JUTTING, *Checking Landau's "Grundlagen" in the automath system*, ISBN 90 6196 147 5.

MCT 84 H.L.L. BUSARD, *The translation of the elements of Euclid from the Arabic into Latin by Hermann ofCarinthia (?) books vii-xii*, 1977.
ISBN 90 6196 148 3.

MCT 85 J. VAN MILL, *Supercompactness and Wallman spaces*, 1977.
ISBN 90 6196 151 3.

* MCT 86 S.G. VAN DER MEULEN & M. VELDHORST, *Torrix I*,
ISBN 90 6196 152 1.

* MCT 87 S.G. VAN DER MEULEN & M. VELDHORST, *Torrix II*,
ISBN 90 6196 153 x.

* MCT 88 A. SCHRIJVER, *Matroids and linking systems*,
ISBN 90 6196 154 8.

An asterisk before the number means "to appear".