

A Numerical Method for Solving Singular Boundary Value Problems

Bertil Gustafsson

Uppsala University Department of Computer Sciences

Received April 15, 1973

Summary. A numerical method is treated for solving singular boundary value problems with solutions that can be represented as series expansions on a subinterval near the singularity. A regular boundary value problem is derived on the remaining interval, for which a difference method is used. Convergence theorems are given for general schemes and for schemes of positive type for second order equations.

1. Introduction

Consider the differential equation

$$Lu \equiv \frac{d^n u}{dx^n} + f_{n-1}(x) \frac{d^{n-1} u}{dx^{n-1}} + \dots + f_0(x) u = F(x) \quad (1.1)$$

on the interval $0 < x \leq 1$. By letting one or more coefficients $f_i(x)$ be infinite at $x=0$ the equation becomes singular there. Since there are singular components in the solution we demand that u and its derivatives up to a given order p shall exist and be bounded for $0 \leq x \leq 1$. Under this restriction it is assumed that a unique solution is determined by the boundary conditions

$$\sum_{j=0}^p A_{lj} \left[\frac{d^j u}{dx^j} \right]_{x=0} + \sum_{j=0}^{n-1} B_{lj} \left[\frac{d^j u}{dx^j} \right]_{x=1} = g_l, \quad l = 1, 2, \dots, s, \quad (1.2)$$

where s may be less than n .

The method treated in this paper consists of a series expansion at a small interval near $x=0$ and a difference method at the rest of the interval. For that purpose we assume that there is a positive constant δ such that the smooth part of the general solution to (1.1) can be written

$$u(x) = \sum_{i=1}^s \alpha_i R_i(x) + R_{s+1}(x), \quad s \leq n \quad (1.3)$$

for $0 \leq x \leq \delta$, where $R_1(x), \dots, R_s(x)$ are linearly independent solutions to $Lu=0$, and where $R_{s+1}(x)$ is a particular solution to (1.1). $R_i(x)$ consists of one or more terms of the form $\phi(x) \sum_{h=0}^{\infty} a_h x^h$, where $\phi(x)$ is an elementary function and the a_h 's are recursively defined (see the example in Sec. 2). The derivation of $R_i(x)$ in general is treated in [2]. It is assumed that $d^p R_i(x)/dx^p$ exists at $x=0$, with p

defined above. The boundary conditions could make sense without this last restriction, but for a well posed problem, $R_i(x)$ can always be defined such that it is fulfilled. This will be explained in the next section where a simple example is treated.

Our method is briefly described as follows: We first calculate approximations of $R_i(x)$. Then using (1.3), we formulate a new regular problem on the interval $[\delta, 1]$ with boundary conditions given at $x = \delta$. We then solve this problem by a difference approximation, solve for the coefficients α_i and calculate u on $[0, \delta]$ using (1.3).

If a difference approximation is used on the whole interval $[0, 1]$ the convergence rate will be very poor. Jamet [4] studied the equation

$$\frac{d^2 u}{dx^2} + \frac{\sigma}{x} \frac{du}{dx} - \tau x = 0, \quad 0 < \sigma < 1, \tau \geq 0, \tag{1.4}$$

with boundary conditions $u(0) = 1, u(1) = 0$, and gave an error estimate $\sim h^{1-\tau}$ for the standard three point approximation. Ciarlet *et al.* [1] later developed a Ritz-Galerkin method with an error estimate $\sim h^{2-\tau}$ for a slightly generalized problem.

We will show that with a centered r -order accurate difference approximation on $[\delta, 1]$ we get a convergence rate $\sim h^r$ on the whole interval provided $R_i(x)$ are calculated with r -order accuracy.

2. Formulation of the Regular Problem

We will first study the Eq. (1.4).

The general solution is obtained by formally differentiating

$$u(x) \equiv x^m \sum_{k=0}^{\infty} a_k x^k, \quad a_0 \neq 0.$$

Possible values of m are determined from the indicial equation

$$m(m - 1 + \sigma) = 0.$$

Accordingly, for $\sigma \neq 1$ the general solution can be written

$$u(x) = \alpha_1 \sum_{k=0}^{\infty} a_{1k} x^k + \alpha_2 x^{1-\sigma} \sum_{k=0}^{\infty} a_{2k} x^k \tag{2.1}$$

where the coefficients are determined recursively from

$$a_{1k} = \frac{\tau a_{1,k-2}}{k(k-1+\sigma)} \left. \vphantom{a_{1k}} \right\} k = 2, 3, \dots \tag{2.2a}$$

$$a_{2k} = \frac{\tau a_{2,k-2}}{k(k+1-\sigma)} \tag{2.2b}$$

$$a_{i0} = 1, \quad a_{i1} = 0, \quad i = 1, 2.$$

For $\sigma = 1$ we get the general solution of the form

$$u = \alpha_1 \sum_{k=0}^{\infty} a_{1k} x^k + \alpha_2 \left(\ln x \sum_{k=0}^{\infty} a_{1k} x^k + \sum_{k=2}^{\infty} b_k x^k \right) \tag{2.3}$$

with a_{1k} defined by (2.2a) and where b_k are uniquely determined by the a_{1k} 's.

For this particular example, the series expansions are valid on the whole interval $(0,1]$; in general we need the expansions on $(0, \delta]$ only.

If $\sigma \geq 1$ we must have $\alpha_2 = 0$, and only one boundary condition can be prescribed. If $0 < \sigma < 1$ both components of the solution are bounded, two boundary conditions must be prescribed, but cannot contain $\left[\frac{du}{dx} \right]_{x=0}$.

Consider now the inhomogenous problem

$$\frac{d^2 u}{dx^2} + \frac{\sigma}{x} \frac{du}{dx} = -x^{1-\sigma} \cos x - (2-\sigma)x^{-\sigma} \sin x. \tag{2.4}$$

The general solution to (2.4) for $\sigma \neq 1$ is $u = \alpha_1 + \alpha_2 x^{1-\sigma} + x^{1-\sigma} \cos x$ where the first two terms are solutions to the homogenous problem. If $1 < \sigma < 3$, the solution is bounded iff $\alpha_2 = -1$, therefore it can be written $u = \alpha_1 + x^{1-\sigma}(\cos x - 1)$, which is on the form (1.4) with $s = 1$. $\left[\frac{du}{dx} \right]_{x=0}$ could formally be entered into the boundary conditions, but can always be substituted by the constant 0.

These arguments can easily be generalized to more general cases, and therefore the assumptions made in the previous section are reasonable.

We will transform the problem into a problem on the interval $[\delta, 1]$, and therefore boundary conditions must be derived at $x = \delta$. To achieve that, the differentiated forms of the expansion (1.3) are first substituted for $u^{(i)}(0)$ in (1.2). (From now on the notation $u^{(i)}(\xi)$ will be used for $\left[\frac{d^i u}{dx^i} \right]_{x=\xi}$.) This gives a linear system of equations for $u^{(j)}(1)$, $j = 0, \dots, n$, and the unknown coefficients α_i . The differentiated forms of (1.3)

$$\begin{aligned} R_1(\delta)\alpha_1 + \dots + R_s(\delta)\alpha_s &= u(\delta) - R_{s+1}(\delta) \\ \vdots & \\ R_1^{(n-1)}(\delta)\alpha_1 + \dots + R_s^{(n-1)}(\delta)\alpha_s &= u^{(n-1)}(\delta) - R_{s+1}^{(n-1)}(\delta) \end{aligned} \tag{2.5}$$

make it possible to express the α_i 's as linear combinations of the $u^{(j)}(\delta)$'s. This is true because the n -order Wronskian for the fundamental solutions to $Lu = 0$ is nonsingular at $x = \delta$, and we can pick s equations out of (2.5) with a nonsingular matrix:

$$\begin{aligned} R_1^{(v_1)}(\delta)\alpha_1 + \dots + R_s^{(v_1)}(\delta)\alpha_s &= u^{(v_1)}(\delta) - R_{s+1}^{(v_1)}(\delta) \\ \vdots & \\ R_1^{(v_s)}(\delta)\alpha_1 + \dots + R_s^{(v_s)}(\delta)\alpha_s &= u^{(v_s)}(\delta) - R_{s+1}^{(v_s)}(\delta). \end{aligned} \tag{2.6}$$

The so obtained expressions for α_i are plugged into (1.2) (with (1.3) substituted for u) and into the rest of the Eq. (2.5). In that way n linear equations

$$B_1 \begin{bmatrix} u(\delta) \\ \vdots \\ u^{(n-1)}(\delta) \\ u(1) \\ \vdots \\ u^{(n-1)}(1) \end{bmatrix} = \hat{G}_1 \tag{2.7}$$

are obtained, where the elements of the matrix B_1 and the vector \hat{G}_1 depend on $R_i^{(j)}(0)$, $R_i^{(j)}(\delta)$.

When the solution to (1.1), (2.7) is obtained on $[\delta, 1]$ the solution on $[0, \delta]$ is obtained by solving (2.6) for $\hat{\alpha} = (\alpha_1, \dots, \alpha_s)^T$ and then using (1.3).

We will make the assumption that the original problem has a unique solution with the assumptions made in sec. 1 fulfilled. From the arguments above it is then clear that this solution is also obtained as solution to the problem derived above.

3. A Stability Theorem for the Regular Problem

We will first concern ourselves with the question of how to get a numerical solution to the regular problem (1.1), (2.7). The theory by Grigorieff [3] and Kreiss [5] for difference methods can be directly applied here. We assume that the reader is familiar with the latter paper. We will look in some detail upon the way of treating the boundary conditions and the accuracy thereby obtained when using centered difference operators for approximating L .

Let h be the steplength and define gridpoints $x_j = jh + \delta - rh$, $j = 0, 1, \dots, N$, where r is an integer and depends on the width of the difference operator L_h . Using the notation $v_j = v(x_j)$, $E v_j = v_{j+1}$, $hD_+ v_j = E - I$, $hD_- = I - E$, we make the following assumption on the approximation $L_h v_j = F_j$:

Assumption 3.1. That part of L_h which approximates d^n/dx^n can be written

$$QD_+^n \equiv \sum_{j=-\nu}^{\mu} \gamma_j E^j D_+^n \quad (3.1)$$

where the γ_j 's are independent of x, h . Furthermore the lower order terms of L_h do not use more points than QD_+^n does. If $\mu = -\nu$ then L_h is called compact.

This assumption is not restrictive since all operators used in practice to our knowledge fulfill it. For example the standard second order approximation $D_+ D_-$ to d^2/dx^2 can be written $E^{-1} D_+^2$ and the corresponding fourth order operator $D_+ D_- - \frac{h^2}{12} D_+^2 D_-^2$ is equivalent to $\left(-\frac{1}{12} E^{-2} + \frac{14}{12} E^{-1} - \frac{1}{12} I\right) D_+^2$. The discrete boundary conditions are written in the form

$$B_{lh} v = \tilde{g}_l, \quad l = 0, 1, \dots \quad (3.2)$$

where B_{lh} contains difference operators of order l and lower. Those boundary conditions which are of order $\geq n$ can be written on the form (cf. [5])

$$\sum_{j=0}^{N-n} H_{jl} (D_+^n v_j) = \tilde{g}_l, \quad l \geq n \quad (3.3)$$

where the equations are linearly independent, and where \tilde{g}_l denotes the lower order terms.

We want to study the solution y_j to

$$Q y_j = 0, \quad j = \nu, \dots, N - n - \mu \quad (3.4)$$

$$\sum_j H_{jl} y_j = \tilde{g}_l, \quad l \geq n \quad (3.5)$$

and define the characteristic equation corresponding to (3.4) by

$$\sum_{j=-v}^{\mu} \gamma_j \kappa^j = 0. \tag{3.6}$$

With the norm defined by $\|v\| = \max_j |v_j|$, where the maximum is taken over all points where v is defined, we state the following theorem which is a special version of the results in [5].

Theorem 3.1. Assume that the roots of (3.6) satisfy $|\kappa_k| \neq 1$, and that the solution of (3.4), (3.5) fulfills

$$\|y\| \leq \text{const} \max_{l \geq n} |\tilde{g}_l|. \tag{3.7}$$

If the error $w = u - v$ is a solution of

$$L_h w = h^r G \tag{3.8}$$

$$B_{lh} w = \begin{cases} h^r g_l, & l \leq n-1 \\ h^{r+n-l-1} g_l, & l \geq n \end{cases} \tag{3.9}$$

then it can be estimated by

$$\|w\| \leq \text{const} h^r \left(\|G\| + \sum_{l \geq 0} |g_l| \right). \tag{3.10}$$

In particular (3.10) is always fulfilled if L_h is compact.

Proof. This theorem is proved in [5] for the case that all κ_k 's are distinct. We refer to that paper and conclude that the only additional difficulty with multiple κ_k 's is in the construction of a gridfunction z_j such that $D_+^n z_j = y_j$ where y satisfies (3.4), (3.5), (3.7) and where $\|D_+^l z\| \leq \text{const} h^{n-l} \max_{l \geq n} |\tilde{g}_l|$ for $l \leq n-1$. Every other step in the proof is independent of the multiplicity of the κ_k 's.

By assumption

$$y_j = \sum_{|\kappa_k| < 1} \left(\sum_{i=0}^{i_k-1} \sigma_{ki} j^i \right) \kappa_k^j + \sum_{|\kappa_k| > 1} \left(\sum_{i=0}^{i_k-1} \sigma_{ki} (j-N+n)^i \right) \kappa_k^{j-N+n} \tag{3.11}$$

where

$$|\sigma_{ki}| \leq \text{const} \max_{l \geq n} |\tilde{g}_l|$$

(0^0 is here defined as 1).

For every given polynomial $\sum_{i=0}^p a_i j^i$ it is possible to construct another polynomial $\sum_{i=0}^p b_i j^i$ such that

$$D_+ h(\kappa - 1)^{-1} \sum_{i=0}^p b_i j^i \kappa^j = \sum_{i=0}^p a_i j^i \kappa^j.$$

By expanding the factors $(j+1)^i$ into a binomial series the coefficients b_i are obtained recursively by

$$\begin{aligned} b_p &= a_p \\ b_{p-i} &= a_{p-i} - (\kappa - 1)^{-1} \left[\binom{p}{i} b_p + \binom{p}{i-1} b_{p-1} + \dots + \binom{p}{1} b_{p-i} \right], \quad i = 1, 2, \dots, p. \end{aligned}$$

This procedure can then be repeated up to order n , and if $a_i = \sigma_{k,i}$, we denote the so obtained coefficients with $\tilde{\sigma}_{k,i}$. z then takes the form

$$z_j = h^n \sum_{|\kappa_k| < 1} \left(\sum_{i=0}^{i_k-1} \tilde{\sigma}_{k,i} j^i \right) (\kappa_k - 1)^{-n} \kappa_k^j + h^n \sum_{|\kappa_k| > 1} \left(\sum_{i=0}^{i_k-1} \tilde{\sigma}_{k,i} (j - N + n)^i \right) (\kappa_k - 1)^{-n} \kappa_k^{j-N+n},$$

and it is obvious that our requirements on z are fulfilled.

4. Centered Symmetric Compact Difference Methods

We will first look at second order methods, and separate between even and odd n .

For n even the difference operator L_h is defined by substituting all differential operators according to

$$\frac{d^l}{dx^l} \rightarrow \begin{cases} (D_+ D_-)^{l/2} & \text{for } l \text{ even} \\ D_0 (D_+ D_-)^{(l-1)/2} & \text{for } l \text{ odd} \end{cases} \tag{4.1}$$

where $2hD_0 = E - E^{-1}$.

We get

$$L_h v_j = (D_+ D_-)^{n/2} v_j + f_{n-1}(x_j) D_0 (D_+ D_-)^{(n-2)/2} v_j + \dots + f_0(x_j) v_j = F(x_j), \quad j = n/2, \dots, N - n/2.$$

The grid is located so that

$$\begin{aligned} \delta &= x_{n-1} \\ 1 &= Nh - [(q+1)/2]h \quad ([x] = \text{integer part of } x) \end{aligned} \tag{4.3}$$

where $u^{(q)}(1)$ is the highest order derivative occurring at the right.

The boundary conditions could now be prescribed by simply substituting centered second order approximations for all u -derivatives occurring in (2.7). However, there is a simpler way of treating the left boundary, avoiding the computing of $R_+^{(l)}(\delta)$, $l \geq 1$. Instead of differentiating (1.3) we can directly use difference operators and get the difference analogue of (2.5) and of (2.6):

$$\begin{aligned} D_+^{n_1} R_1(x_0) \alpha_1 + \dots + D_+^{n_s} R_s(x_0) \alpha_s &= D_+^{n_1} (u_0 - R_{s+1}(x_0)) \\ \vdots & \\ D_+^{n_1} R_1(x_0) \alpha_1 + \dots + D_+^{n_s} R_s(x_0) \alpha_s &= D_+^{n_1} (u_0 - R_{s+1}(x_0)). \end{aligned} \tag{4.4}$$

The matrix of this system is nonsingular since it converges to the matrix of the system (2.6) when $h \rightarrow 0$. The complete set of boundary conditions is now derived in exactly the same way as (2.7) was derived, we write this new set as

$$B_2 \begin{bmatrix} u(x_0) \\ \vdots \\ D_+^{n-1} u(x_0) \\ u(1) \\ \vdots \\ u^{(n-1)}(1) \end{bmatrix} = \hat{G}_2. \tag{4.5}$$

(4.5) has infinite accuracy as long as $D_+^l R_i(0)$, $D_+^l R_i(x_0)$ are computed exactly. To obtain the boundary conditions for the difference approximation we substitute

$$u^{(l)}(1) \rightarrow \begin{cases} (D_+ D_-)^{l/2} v_{N-[(p+1)/2]} & \text{for } l \text{ even} \\ D_0 (D_+ D_-)^{(l-1)/2} v_{N-[(p+1)/2]} & \text{for } l \text{ odd} \end{cases} \quad (4.6)$$

in (4.5).

The problem treated in Section 6 serves as an illustration of the derivation of the boundary conditions.

It should be noted that if only odd order derivatives $u^{(l)}(1)$ occur, accuracy is gained by locating $x=1$ in the middle of an interval and using compact difference operators. It causes trouble however, if one wants to use Richardson extrapolation with halving of the stepsize.

The approximation $\tilde{R}_i(x)$ of $R_i(x)$ is obtained by truncating the corresponding series expansion $\sum_{k=0}^{\infty} a_{ik} x^k$, $x = x_0, x_1, \dots, x_{n-1}$. $R_i^{(l)}(0)$ can be obtained exactly.

When v_0, v_1, \dots, v_N have been solved for, the solution on $[0, \delta]$ is obtained by solving (4.4) with $R_i(x_j)$ substituted by $\tilde{R}_i(x_j)$, and then using (1.3).

To define L_h for n odd we substitute

$$\left[\frac{d^l u}{dx^l} \right]_{x=x_j+h/2} \rightarrow \begin{cases} \frac{1}{2} (D_+ D_-)^{l/2} (I + E) v_j & \text{for } l \text{ even} \\ D_+ (D_+ D_-)^{(l-1)/2} v_j & \text{for } l \text{ odd} \end{cases} \quad (4.7)$$

and get

$$L_h v_j = D_+ (D_+ D_-)^{(n-1)/2} v_j + f_{n-1}(x_j + h/2) \frac{1}{2} (D_+ D_-)^{(n-2)/2} (I + E) v_j + \dots + f_0(x_j + h/2) v_j = F(x_j + h/2), \quad j = (n-1)/2, \dots, N-1 - (n-1)/2. \quad (4.8)$$

If $f_i(x)$ and $F(x)$ are known only at gridpoints, there values in between are taken as meanvalues. Since the difference operators are properly centered at $x_j + h/2$ it is clear that L_h has second order accuracy.

The location of the grid is again defined by (4.3), and the boundary conditions as above.

Compact approximations of arbitrary high order accuracy can always be achieved by differentiating the differential equation one or more times. Consider for example the equation (1.4) and the second order approximation of $u^{(2)}(x_j)$

$$D_+ D_- u(x_j) = u^{(2)}(x_j) + \frac{h^2}{12} u^{(4)}(x_j) + \mathcal{O}(h^4). \quad (4.9)$$

By differentiating (1.4) we obtain

$$u^{(4)}(x_j) = \left[\frac{\sigma}{x_j^2} (\sigma + 2) + \tau \right] u^{(2)}(x_j) + \left[\frac{\sigma}{x_j} \left(-\frac{\sigma}{x_j^2} - \frac{2}{x_j^2} - \tau \right) \right] u^{(1)}(x_j). \quad (4.10)$$

By using second order approximations for $u^{(2)}(x_j)$ and $u^{(1)}(x_j)$, and substituting the right hand side of (4.10) for $u^{(4)}(x_j)$ in (4.9), a fourth order approximation of $u^{(2)}(x_j)$ is obtained, which uses no more than 3 points. For $u^{(1)}(x_j)$ one can do in the same way, and if fourth order approximations are used even in the boundary conditions, a fourth order compact approximation is obtained.

In this way compact approximations with an r order accuracy can be constructed for all even r and arbitrary n .

We can now prove the convergence theorem for the class of methods we have described above:

Theorem 4.1. Assume that $\tilde{R}_i(x_j)$ are calculated as described above with

$$|\tilde{R}_i(x_j) - R_i(x_j)| \leq \mathcal{O}(h^r), \quad j = 0, 1, \dots, n - 1, i = 1, 2, \dots, s + 1, \quad (4.11)$$

and that $u(x), f_i(x), F(x)$ are sufficiently smooth. (This last assumption is used throughout the rest of the paper.) Then there are constants $K_1(\delta), K_2(\delta)$ depending on δ only, such that the solution to the r -order method described above satisfies

$$\|u - v\| \leq K_1(\delta) h^r \quad (4.12)$$

and

$$\max_{0 \leq x \leq \delta} |u - v| \leq K_2(\delta) h^r. \quad (4.13)$$

Proof. The boundary conditions are derived by solving (4.4) for $\alpha_1, \dots, \alpha_s$ and plugging these into the rest of the equations. Since the $R_i(x)$'s are infinitely differentiable at $x = \delta$, (4.11) implies $|D_+^l \tilde{R}_i(x_0) - D_+^l R_i(x_0)| \leq \mathcal{O}(h^r)$, and therefore the elements of the inverse

$$\begin{bmatrix} D_+^{r_1} R_1(x_0) & \dots & D_+^{r_1} R_s(x_0) \\ \vdots & & \vdots \\ D_+^{r_s} R_1(x_0) & \dots & D_+^{r_s} R_s(x_0) \end{bmatrix}^{-1}$$

have errors less than $\mathcal{O}(h^r)$. If r -order approximations are used for $u^{(l)}(1)$ and $R_i^{(l)}(0)$ are computed exactly, then it is clear that the boundary conditions have r -order accuracy and (3.8), (3.9) are valid, G, g_l being bounded functions. Since L_h is compact, (4.12) follows from Theorem 3.1.

Approximations $\tilde{\alpha}_i$ of α_i are obtained from (4.4) with $R_i(x_j)$ substituted by $\tilde{R}_i(x_j)$ and u_j by v_j , and we have

$$\max_i |\tilde{\alpha}_i - \alpha_i| \leq \mathcal{O}(h^r).$$

With $v(x)$ defined for $0 \leq x \leq \delta$ by

$$v(x) = \sum_{i=1}^s \tilde{\alpha}_i \tilde{R}_i(x) + \tilde{R}_{s+1}(x),$$

(4.13) is immediately obtained, and the theorem is proved.

5. Higher Order Symmetric Non Compact Difference Methods

If we want higher order approximations without using differentiated forms of (1.1), non compact operators must be used. Even order differential operators can be formally written as

$$\begin{aligned} \frac{\partial^{2l}}{\partial x^{2l}} &= (D_+ D_-)^l \sum_{k=0}^{\infty} \beta_k (h^2 D_+ D_-)^k \\ &\equiv D_r^{(2l)} + (D_+ D_-)^l \sum_{k=M+1}^{\infty} \beta_k (h^2 D_+ D_-)^k \end{aligned} \quad (5.1)$$

where $D_r^{(2l)}$ denotes a symmetric $r = 2(M + 1)$ order approximation.

Odd order operators can be written

$$\begin{aligned} \frac{\partial^{2l+1}}{\partial x^{2l+1}} &= E^{-1} D_+ (D_+ D_-)^l \sum_{k=0}^{\infty} \beta_k (h^2 D_+ D_-)^k \\ &\equiv D_r^{(2l+1)} + E^{-1} D_+ (D_+ D_-)^l \sum_{k=M+1}^{\infty} \beta_k (h^2 D_+ D_-)^k. \end{aligned} \tag{5.2}$$

We have

Lemma 5.1. $D_r^{(2l)}$ and $D_r^{(2l+1)}$ fulfill the root condition ($|\kappa_j| \neq 1$) of Theorem 3.1 for any r and l .

Proof. We apply (5.1) to the function $e^{i2\pi\omega x}$ and obtain with $\xi = \pi\omega h$

$$(2\xi)^{2l} = (2 \sin \xi)^{2l} \sum_{k=0}^{\infty} \beta_k (-1)^k (2 \sin \xi)^{2k}.$$

If $\theta = \sin \xi$, then

$$(\arcsin \theta)^{2l} = \theta^{2l} \sum_{k=0}^{\infty} \beta_k (-1)^k (2\theta)^{2k}. \tag{5.3}$$

Since the coefficients in the power series for arcsin are all positive, β_k must have alternating signs and $\beta_0 = 1$.

The characteristic equation (3.6) corresponding to $D_r^{(2l)}$ is

$$\sum_{k=0}^M \beta_k (\kappa - 2 + \kappa^{-1})^k = 0.$$

If $\kappa = e^{i\theta}$, $-\pi < \theta \leq \pi$, then this equation goes over into

$$\sum_{k=0}^M \beta_k (-1)^k (2 \sin(\theta/2))^{2k} = 0.$$

Since β_k have alternating signs and $\beta_0 = 1$, this is impossible for any real θ , and the lemma is proved for $D_r^{(2l)}$. For odd order operators we get analogously the β_k 's defined by

$$(\arcsin \theta)^{2l+1} = \theta^{2l+1} \sum_{k=0}^{\infty} \beta_k (-1)^k (2\theta)^{2k} \tag{5.4}$$

and precisely as before we conclude that D_r^{2l+1} fulfills the root condition.

L_h is now defined by substituting the r -order approximation $D_r^{(l)}$ for $\partial^l/\partial x^l$ as in Section 4 and the grid is fixed by

$$\begin{aligned} \delta &= x_{n-1+M} \\ 1 &= (N - [(q+1)/2] - M)h & \text{if } q > 0 \\ 1 &= Nh & \text{if } q = 0. \end{aligned} \tag{5.5}$$

($u^{(q)}$ is the highest order derivative occurring in the boundary conditions at $x = 1$.) The operator Q defined by (3.1) is symmetric, i.e. $\gamma_{-\nu+j} = \gamma_{\mu-j}$, $j = 0, 1, \dots, (\mu + \nu)/2$ where $\mu + \nu$ is even. Therefore, if κ_k is a root of (3.6), then κ_k^{-1} is also a root, which means that the number of κ_k 's inside the unit circle is equal to the

number outside. Since the stability of the two point boundary value problem is equivalent to the stability of both half line problems (cf. [5, Thm 3.5]), it then follows that M extra boundary conditions must be given at each boundary.

We first consider the right half line problem and define the extra boundary conditions at $x = \delta$ by supplementing (4.4) with

$$\sum_{i=1}^s D_+^l R_i(x_0) \alpha_i - D_+^l u_0 = D_+^l R_{s+1}(x_0), \quad l = n, \dots, n + M - 1.$$

Theorem 5.1. The approximation defined above with the condition $v \in L_2(\delta, \infty)$ is stable. Therefore if $\tilde{R}_i(x_j), j = 0, 1, \dots, n + M - 1; i = 1, 2, \dots, s + 1$ are calculated as described in Sec. 4 with (4.11) satisfied, and if the right boundary conditions are stable with an accuracy according to (3.9), then the estimates (4.12), (4.13) are valid.

Proof. The solution to (3.4) is

$$y_j = \sum_{|\kappa_k| < 1} \left(\sum_{i=0}^{i_k-1} \sigma_{k i} \tilde{r}^i \right) \kappa_k^j,$$

where i_k denotes the multiplicity of κ_k and where

$$\sum_{|\kappa_k| < 1} i_k = M.$$

The boundary conditions are $y_0 = \tilde{g}_n, hD_+ y_0 = \tilde{g}_{n+1}, \dots$, where \tilde{g}_i denote lower order terms, and therefore y_0, y_1, \dots , can be solved for explicitly. The determination of the $\sigma_{k i}$'s is now equivalent to the determination of the interpolating function for the points $(0, y_0), (1, y_1), \dots$ with $e^{\lambda_1 x}, x e^{\lambda_1 x}, \dots, e^{\lambda_k x}, \dots$, as basis functions, where $\kappa_i = e^{\lambda_i}$. But this set of basis functions is linearly independent on any x -interval and satisfies the Haar-condition. Hence, the interpolating function is uniquely determined and the estimate (3.7) is valid.

By the same arguments as in the proof of Theorem 4.1 it is easily seen that the accuracy of the boundary conditions is according to (3.9) (actually higher order accuracy than necessary is obtained) and the theorem is proved.

We will now give examples of stable boundary conditions to the right.

Theorem 5.2. The left half line problem is stable for $r = 4$ with the extra boundary condition defined by a 3rd order accurate approximation of $Lu = F$ at $x = x_{N-n/2}$ which uses no other points than x_{N-n-2}, \dots, x_N . (By Theorem 3.1 the accuracy is sufficient to give an overall 4th order convergence rate.)

Proof. From (5.3), (5.4) we get $\beta_0 = 1, \beta_1 = -n/24$ for the 4th order approximation of $\partial^n / \partial x^n$. The characteristic equation (3.6) therefore is

$$1 - \frac{n}{24} (\kappa - 2 + \kappa^{-1}) = 0$$

or equivalently

$$\kappa^2 - \frac{2(12+n)}{n} \kappa + 1 = 0 \tag{5.6}$$

which has the root

$$\kappa_1 = (12+n + \sqrt{144+24n})/n > 1. \tag{5.7}$$

The 3rd order approximation is defined by

$$(D_+ D_-)^{n/2} \left(I - \frac{n}{24} h^2 D_-^2 \right) v_{N-[(n+1)/2]}$$

for n even, and by

$$E^{-1} D_+ (D_+ D_-)^{(n-1)/2} \left(I - \frac{n}{24} h^2 D_-^2 \right) v_{N-[(n+1)/2]}$$

for n odd. Since the general solution in $L_2(-\infty, 1)$ can be written $v_j = \sigma \kappa_1^{j-N}$ we have only to show that the characteristic equation corresponding to the operator $I - \frac{n}{24} h^2 D_-^2$

$$\left(1 - \frac{n}{24} \right) \kappa^2 + \frac{n}{12} \kappa - \frac{n}{24} = 0 \tag{5.8}$$

is not satisfied by κ_1 for any $n > 0$. When multiplying (5.8) by $24/n$ and adding it to (5.6) we get the equation

$$\kappa^2 - \kappa = 0$$

which is not satisfied by κ_1 , and the theorem is proved.

We will also treat the method of extrapolating at the boundary:

Theorem 5.3. Assume that r and n are such that (3.6) has no multiple roots. Then the left half line problem is stable if the extra boundary conditions are defined by

$$(hD_-)^{r-l} D_-^n v_N = 0, \quad l = 1, \dots, M. \tag{5.9}$$

In particular, the problem is stable for all $n > 0$ if $r = 4$ or $r = 6$. (The accuracy of (5.9) is again sufficient by Theorem 3.1.)

Proof. By assumption the general solution to (3.4) in $L_2(-\infty, 1)$ can be written

$$v_j = \sum_{k=1}^M \sigma_k \kappa_k^{j-N}, \quad |\kappa_k| > 1.$$

Therefore there is a non trivial solution satisfying (5.9) if and only if

$$\text{Det} \begin{bmatrix} (1-\kappa_1)^{n+r-1} & (1-\kappa_2)^{n+r-1} & \dots & (1-\kappa_M)^{n+r-1} \\ (1-\kappa_1)^{n+r-2} & & & \vdots \\ \vdots & & & \vdots \\ (1-\kappa_1)^{n+r-M} & \dots & \dots & (1-\kappa_M)^{n+r-M} \end{bmatrix} = 0.$$

Since $\kappa_k \neq 1$ this is equivalent to

$$\text{Det} \begin{bmatrix} (1-\kappa_1)^{M-1} & (1-\kappa_2)^{M-1} & \dots & (1-\kappa_M)^{M-1} \\ (1-\kappa_1)^{M-2} & & & \vdots \\ \vdots & & & \vdots \\ 1 & \dots & \dots & 1 \end{bmatrix} = \pm \prod_{k>i} (\kappa_k - \kappa_i) = 0$$

which is impossible.

If $r=4$, then $M=1$ and there can be no multiple roots. If $r=6$ there are multiple roots \varkappa if and only if there are multiple roots η to

$$\beta_0 + \beta_1\eta + \beta_2\eta^2 = 0.$$

This is clear by the fact that the equation

$$\varkappa - 2 + \varkappa^{-1} = \eta$$

has multiple roots only on the unit circle.

From (5.3), (5.4) we get $\beta_2 = \frac{n}{72} \left(n + \frac{22}{5} \right)$, hence the condition for double roots is

$$\frac{n}{72} \left(n + \frac{22}{5} \right) - \frac{n^2}{48^2} = 0.$$

This is impossible for $n > 0$, which proves the theorem.

6. Approximations of Positive Type to Second Order Equations

In this section we consider the particular equation (1.4) with $0 < \sigma < 1$.

The general solution is given in Section 2 by (2.1), (2.2). The problem is well posed with the boundary conditions

$$\begin{aligned} u(0) &= A \\ u(1) &= B. \end{aligned} \tag{6.1}$$

The boundary conditions as described in Section 4 are derived from the system

$$\begin{aligned} R_1(0)\alpha_1 + R_2(0)\alpha_2 &= A \\ u(1) &= B \\ R_1(x_0)\alpha_1 + R_2(x_0)\alpha_2 &= u(x_0) \\ D_+ R_1(x_0)\alpha_1 + D_+ R_2(x_0)\alpha_2 &= D_+ u(x_0). \end{aligned} \tag{6.2}$$

From the last two equations (which correspond to (4.4)) we obtain

$$\begin{aligned} \alpha_1 &= \frac{u(x_0)D_+ R_2(x_0) - R_2(x_0)D_+ u(x_0)}{R_1(x_0)D_+ R_2(x_0) - R_2(x_0)D_+ R_1(x_0)} \\ \alpha_2 &= \frac{R_1(x_0)D_+ u(x_0) - u(x_0)D_+ R_1(x_0)}{R_1(x_0)D_+ R_2(x_0) - R_2(x_0)D_+ R_1(x_0)} \end{aligned}$$

and since $R_1(0) = 1, R_2(0) = 0$ the discrete boundary conditions will be

$$v_0 - \frac{\tilde{R}_2(x_0)}{\tilde{R}_2(x_1)} v_1 = \frac{A}{\tilde{R}_2(x_1)} [\tilde{R}_2(x_1)\tilde{R}_1(x_0) - \tilde{R}_2(x_0)\tilde{R}_1(x_1)] \tag{6.3}$$

$$u_N = B \tag{6.4}$$

where $x_1 = \delta, x_N = 1$.

Using the standard second order approximation

$$L_\varepsilon v_j = D_+ D_- v_j + \frac{\sigma}{x_j} D_0 v_j - \tau v_j = 0, \quad j = 1, 2, \dots, N-1 \tag{6.5}$$

no more boundary conditions are needed. The theory from the previous sections can now be applied, and we can immediately show h^2 -convergence provided $\tilde{R}_1(x), \tilde{R}_2(x)$ are calculated sufficiently accurate. The δ -dependence in the error estimate is obtained by considering the truncation error of $L_h u$ which contains terms proportional to $h^2 u^{(4)}(x)$ and to $h^2 x^{-1} u^{(3)}(x)$. From the representation (2.1) we therefore get

Lemma 6.1. There is a constant c such that the solution $u(x)$ to (1.4), (6.1) satisfies

$$|L_h u(x_j)| \leq c h^2 x_j^{-3-\sigma}, \quad j = 1, 2, \dots, N-1. \tag{6.6}$$

The error estimate will therefore be proportional to $\delta^{-3-\sigma} h^2$. The aim of this section is to sharpen this to a $\delta^{-1-\sigma} h^2$ estimate. In order to do that, we will use some of the results in [4]. We start with two lemmas given there:

Lemma 6.2. The operator L_h is of positive type and therefore fulfills the maximum principle:

$$L_h v_j \geq 0 \Rightarrow \max_{2 \leq j \leq N-1} v_j \leq \max(0, v_1, v_N).$$

Lemma 6.3. Let d be a constant with $d > 1 + 16\tau^2$. Then

$$L_h(x_j^{-1-\sigma} - d) > x_j^{-3-\sigma}/4 \quad \text{for} \quad 12h \leq x_j \leq 1-h. \tag{6.7}$$

In the same way as in [4] we also define

$$z(x) \equiv 4c h^2 (x^{-1-\sigma} - \delta^{-1-\sigma}) - |w(\delta)|$$

where c is defined by (6.6) and $w(x) = u(x) - v(x)$. Then by (6.6),

$$\begin{aligned} L_h z(x) &> c h^2 x^{-3-\sigma} - L_h |w(\delta)| \\ &= c h^2 x^{-3-\sigma} + \tau |w(\delta)| \geq |L_h u(x)| = |L_h w(x)| \end{aligned}$$

and therefore $L_h(z(x) \pm w(x)) \geq 0$. By the maximum principle we have $z(x) \pm w(x) \leq \max(0, -|w(\delta)| \pm w(\delta), z(1)) = 0$. Hence

$$|w(x)| \leq |z(x)| \leq |z(1)| < 4c h^2 \delta^{-1-\sigma} + w(\delta) \tag{6.8}$$

for $x = x_2, x_3, \dots, x_N$.

So, what still remains is an estimate of $w(\delta) = w_1$.

We define

$$\tilde{R}_i(x) = \sum_{k=0}^{M-1} a_{i,k} x^{k+\mu_i}, \quad i = 1, 2;$$

$$\mu_1 = 0; \mu_2 = 1 - \sigma,$$

and use these as approximations of $R_i(x)$. From (2.2) it is easily seen that

$$|R_i(x) - \tilde{R}_i(x)| \leq \text{const } \delta^M, \quad i = 1, 2, x = x_0, x_1. \tag{6.9}$$

Taking $L_h v_1 = 0$ and (6.3) together constitutes a new operator, acting on v_1, v_2 . We denote this by L_h^B and prove:

Lemma 6.4. There are constants c_1, c_2 independent of δ and h such that

$$|L_h^B w_1| \leq c_1 \delta^{M-1} h^{-1} + c_2 \delta^{-3-\sigma} h^2. \tag{6.10}$$

Proof. We have

$$L_h^B w_1 = L_h^B u_1 = D_+ D_- u_1 + \sigma \delta^{-1} D_0 u_1 - \tau u_1 + h^{-2} T(h) \tag{6.11}$$

where

$$T(h) = \left(1 + \frac{\sigma h}{2\delta}\right) \left[\frac{\tilde{R}_2(x_0)}{\tilde{R}_2(x_1)} u_1 + A \left\{ \tilde{R}_1(x_0) - \frac{\tilde{R}_2(x_0)}{\tilde{R}_2(x_1)} \cdot \tilde{R}_1(x_1) \right\} \right. \\ \left. - \left(\frac{R_2(x_0)}{R_2(x_1)} u_1 + A \left\{ R_1(x_0) - \frac{R_2(x_0)}{R_2(x_1)} R_1(x_1) \right\} \right) \right].$$

We first estimate

$$\frac{R_2(x_0)}{R_2(x_1)} - \frac{\tilde{R}_2(x_0)}{\tilde{R}_2(x_1)} = \frac{R_2(x_0) - \tilde{R}_2(x_0) \left(1 - R_2^{-1}(x_1) x_1^{1-\sigma} \sum_{k=M}^{\infty} a_{2k} x_1^k\right)^{-1}}{R_2(x_1)} \\ = \frac{(\delta - h)^{1-\sigma} (-a_{2M} M h \delta^{M-1} + \mathcal{O}(M^2 h^2 \delta^{M-2} + \delta^{M+1}))}{\delta^{1-\sigma} (1 + \mathcal{O}(\delta))} \\ = \mathcal{O}(a_{2M} M h \delta^{M-1}) \leq \mathcal{O}(h \delta^{M-1}),$$

the last inequality following from (2.2b) for sufficiently large M .

In the same way we obtain

$$R_1(x_0) - \frac{R_2(x_0)}{R_2(x_1)} R_1(x_1) - \tilde{R}_1(x_0) + \frac{\tilde{R}_2(x_0)}{\tilde{R}_2(x_1)} \tilde{R}_1(x_1) \\ \leq \mathcal{O}(a_{1M} M h \delta^{M-1} + a_{2M} M h \delta^{M-1}) \leq \mathcal{O}(h \delta^{M-1}).$$

The first part of (6.11) is estimated analogously to (6.6) by const $h^2 \delta^{-3-\sigma}$, and the lemma is proved.

We can now prove the main result of this section:

Theorem 6.1. There is a constant c_3 independent of δ and h such that the error $w = u - v$ satisfies

$$\|w\| \leq c_3 \max(\delta^M, h^2 \delta^{-1-\sigma}) \tag{6.12}$$

for $h \leq h_0, h_0 > 0$.

Therefore

$$\|w\| \leq c_3 h^2 \delta^{-1-\sigma} \tag{6.13}$$

if M is chosen such that

$$M \geq \frac{2 \log h}{\log \delta} - 1 - \sigma. \tag{6.14}$$

Proof. From (6.10) we get

$$\left(1 - \frac{\sigma}{2} \delta^{-1} h\right) w_2 = \left[2 + \tau h^2 - \left(1 - \frac{\sigma}{2} \delta^{-1} h\right) \frac{\tilde{R}_2(x_0)}{\tilde{R}_2(x_1)}\right] w_1 + h^2 \mathcal{O}(h^{-1} \delta^{M-1} + h^2 \delta^{-3-\sigma}).$$

Since

$$\frac{\tilde{R}_2(x_0)}{\tilde{R}_2(x_1)} = 1 + \mathcal{O}(h \delta^{-1})$$

we have

$$w_2 = \gamma w_1 + \mathcal{O}(h \delta^{M-1} + h^4 \delta^{-3-\sigma}) \tag{6.15}$$

with

$$\gamma = 1 + \left(1 - \frac{\sigma}{2} \delta^{-1} h\right)^{-1} [\tau h^2 + (1 - \sigma) h \delta^{-1} + \mathcal{O}(h^2 \delta^{-2})] > 1$$

for h sufficiently small.

We now use the first inequality of (6.8) and get

$$\begin{aligned} |w_2| &\leq |z_2| = |4c h^2 [(\delta + h)^{-1-\sigma} - \delta^{-1-\sigma}] - |w_1|| \\ &= |4c h^2 \delta^{-1-\sigma} [-(1 + \sigma) h \delta^{-1} + \mathcal{O}(h^2 \delta^{-3-\sigma})] - |w_1|| \\ &= |w_1| + \mathcal{O}(h^3 \delta^{-2-\sigma}). \end{aligned}$$

Therefore, as a consequence of Lemma 6.4 and (6.15)

$$h \delta^{-1} |w_1| = \text{const} \max(h \delta^{M-1}, h^4 \delta^{-3-\sigma}, h^3 \delta^{-2-\sigma})$$

or equivalently

$$|w_1| \leq \text{const} \max(\delta^M, h^2 \delta^{-1-\sigma}).$$

When taking (6.8) into account, the first part of the theorem is proved. The second part follows immediately from the inequality

$$\delta^M \leq h^2 \delta^{-1-\sigma}.$$

(6.14) shows that the number of terms in the expansions can be kept very low. For example, if $\delta = 0.1$ then a three times finer mesh demands only one more term in $R_i(x)$.

7. Numerical Experiments

The numerical experiments (performed by Tom Smedsaas) were made on the Eq. (2.4) with $\sigma = 1/2$ and boundary conditions $u(0) = 0, u(1) = \cos 1$, which has the solution $u = x^{1-\sigma} \cos x$. In order to make the experiment realistic, the coefficients in the series expansion of $\cos x$ were generated recursively, with the number of terms determined by (6.14). Three approximations were run:

Scheme 1: Compact second order approximation according to Sec. 4.

Scheme 2: Compact fourth order approximation according to Sec. 4.

Scheme 3: Non-compact fourth order approximation according to Sec. 5.

As extra boundary condition at the right boundary for Scheme 3 a compact fourth order approximation of Lu was used at $x = x_{N-1}$. To make it 3rd order accurate only, the term $h^3/10$ was added to $F(x_{N-1})$.

The first three tables below show the maximum error $w_{\max} = \max_{0 \leq j \leq N} |u_j - v_j|$, where the discretization is made over the whole x -axis such that $x_0 = 0, x_N = 1$. (The difference scheme is of course used on the subinterval $[\delta, 1]$ only.)

Table 1. w_{\max} for Scheme 1

N	δ		
	0.1	0.2	0.4
40	$7.7 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$1.8 \cdot 10^{-5}$
80	$1.7 \cdot 10^{-4}$	$3.8 \cdot 10^{-5}$	$4.2 \cdot 10^{-6}$
160	$4.0 \cdot 10^{-5}$	$9.1 \cdot 10^{-6}$	$1.0 \cdot 10^{-6}$
320	$9.6 \cdot 10^{-6}$	$2.2 \cdot 10^{-6}$	$2.5 \cdot 10^{-7}$
640	$2.4 \cdot 10^{-6}$	$5.5 \cdot 10^{-7}$	$6.1 \cdot 10^{-8}$

Table 2. w_{\max} for Scheme 2

N	δ		
	0.1	0.2	0.4
40	$1.5 \cdot 10^{-5}$	$7.9 \cdot 10^{-7}$	$3.7 \cdot 10^{-8}$
80	$7.2 \cdot 10^{-7}$	$4.4 \cdot 10^{-8}$	$2.1 \cdot 10^{-9}$
160	$4.0 \cdot 10^{-8}$	$2.6 \cdot 10^{-9}$	$1.3 \cdot 10^{-10}$
320	$2.3 \cdot 10^{-9}$	$1.6 \cdot 10^{-10}$	$8.7 \cdot 10^{-12}$
640	$1.4 \cdot 10^{-10}$	$1.3 \cdot 10^{-11}$	$3.6 \cdot 10^{-12}$

Table 3. w_{\max} for Scheme 3

N	δ		
	0.1	0.2	0.4
40	$8.2 \cdot 10^{-5}$	$3.9 \cdot 10^{-6}$	$1.8 \cdot 10^{-7}$
80	$3.6 \cdot 10^{-6}$	$2.1 \cdot 10^{-7}$	$1.0 \cdot 10^{-8}$
160	$2.0 \cdot 10^{-7}$	$1.2 \cdot 10^{-8}$	$6.2 \cdot 10^{-10}$
320	$1.2 \cdot 10^{-8}$	$7.6 \cdot 10^{-10}$	$3.7 \cdot 10^{-11}$
640	$7.1 \cdot 10^{-10}$	$4.2 \cdot 10^{-11}$	$2.5 \cdot 10^{-12}$

There is obviously an h^2 -convergence in the first case, and an h^4 -convergence in the other two, except for $\delta=0.4$ and large N which is due to rounding errors (an IBM 370/155 computer was used).

The following table shows the $\delta^{-1-\sigma}$ -dependence of the error as was derived for Scheme 1 in Sec. 6.

Table 4. w_{\max} for Scheme 1, $N = 640$

δ	0.025	0.050	0.100	0.200
w_{\max}	$2.8 \cdot 10^{-5}$	$8.4 \cdot 10^{-6}$	$2.4 \cdot 10^{-6}$	$5.5 \cdot 10^{-7}$

Acknowledgement. I want to thank Professor Heinz-Otto Kreiss who read and criticized the first version of the manuscript.

References

1. Ciarlet, P. G., Natterer, F., Varga, R. S.: Numerical methods of high-order accuracy for singular nonlinear boundary value problems. *Numer. Math.* **15**, 87–99 (1970)
2. Coddington, E. A., Levinson, N.: *Theory of ordinary differential equations*. McGraw-Hill Book Co. 1955
3. Grigorieff, R. D.: Die Konvergenz des Rand- und Eigenwertproblems linearer gewöhnlicher Differenzgleichungen. *Num. Math.* **15**, 15–48 (1970)
4. Jamet, P.: On the convergence of finite-difference approximations to one-dimensional singular boundary-value problems. *Num. Math.* **14**, 355–378 (1970)
5. Kreiss, H. O.: Difference approximations for boundary and eigenvalue problems for ordinary differential equations. *Math. Comp.* **26**, 605–624 (1972)

Bertil Gustafsson
Dept. of Computer Sciences
Sturegatan 4B 2tr
S-752-23 Uppsala, Sweden