13

Boundary value problems for ODEs

13.1 Introduction

In the previous chapter we discussed numerical methods for initial value problems in which all the associated side conditions for a system of differential equations are prescribed at the same point. Now we go on to consider problems where these conditions specify values at more than one point. Typically we require the solution on an interval $[a, b]$, and some conditions are given at a, and the rest at b, although more complicated situations are possible, involving three or more points.

We shall begin with the simplest case, of a second-order equation with one condition given at a and one at b. This problem is sufficient to introduce the basic ideas, and is of a type which arises quite often in practice.

We then go on to discuss the shooting method for the solution of more general problems.

13.2 A model problem

The simplest two-point boundary problem involves the second-order differential equation

$$-y'' + r(x)y = f(x)\,, \qquad a < x < b\,, \tag{13.1}$$

with the boundary conditions

$$y(a) = A\,, \quad y(b) = B\,, \tag{13.2}$$

where A and B are given real numbers. We shall assume that r and f are given real-valued functions, defined and continuous on the bounded closed interval $[a, b]$ of the real line, and that

$$r(x) \geq 0\,, \quad a \leq x \leq b\,.$$

https://doi.org/10.1017/CBO9780511801181.014 Published online by Cambridge University Press

The reason for this condition will appear later, in Theorem 13.4.

We shall construct a numerical approximation to the solution on a uniform mesh of points

$$x_j = a + jh\,, \quad j = 0, 1, \ldots, n\,, \qquad h = (b-a)/n\,, \qquad n \geq 2\,,$$

so that $x_0 = a$, $x_n = b$. The second derivative is approximated using the second central difference defined below.

Definition 13.1 *The central difference δy of y is defined by*

$$\delta y(x_j) = y(x_j + \tfrac{1}{2}h) - y(x_j - \tfrac{1}{2}h)\,.$$

Higher-order differences are defined recursively by

$$\delta^{m+1} y(x_j) = \delta[\delta^m y(x_j)] = \delta^m y(x_j + \tfrac{1}{2}h) - \delta^m y(x_j - \tfrac{1}{2}h)\,.$$

In particular, the second central difference may be written

$$\begin{aligned} \delta^2 y(x_j) &= \delta y(x_j + \tfrac{1}{2}h) - \delta y(x_j - \tfrac{1}{2}h) \\ &= y(x_j + h) - 2y(x_j) + y(x_j - h)\,. \end{aligned}$$

Theorem 13.1 *(i) Suppose that $y \in \mathrm{C}^4[x-h, x+h]$, i.e., that y has continuous fourth derivative on the interval $[x-h, x+h]$. Then, there exists a number ξ in $(x-h, x+h)$ such that*

$$\frac{\delta^2 y(x)}{h^2} = y''(x) + \tfrac{1}{12} h^2 y^{\mathsf{iv}}(\xi)\,.$$

(ii) Suppose that $y \in \mathrm{C}^6[x-h, x+h]$; then, there exists a number η in $(x-h, x+h)$ such that

$$\frac{\delta^2 y(x)}{h^2} = y''(x) + \tfrac{1}{12} h^2 y^{\mathsf{iv}}(x) + \tfrac{1}{360} h^4 y^{\mathsf{vi}}(\eta)\,. \tag{13.3}$$

Proof (i) Taylor's Theorem shows that there exist numbers ξ_1 and ξ_2 in the intervals $(x-h, x)$ and $(x, x+h)$, respectively, such that

$$\left.\begin{aligned} y(x-h) &= y(x) - hy'(x) + \tfrac{1}{2}h^2 y''(x) - \tfrac{1}{6}h^3 y'''(x) + \tfrac{1}{24}h^4 y^{\mathsf{iv}}(\xi_1)\,, \\ y(x+h) &= y(x) + hy'(x) + \tfrac{1}{2}h^2 y''(x) + \tfrac{1}{6}h^3 y'''(x) + \tfrac{1}{24}h^4 y^{\mathsf{iv}}(\xi_2)\,. \end{aligned}\right\} \tag{13.4}$$

Since y^{iv} is continuous on $[x-h, x+h]$, there is a number ξ in (ξ_1, ξ_2), and thus also in $(x-h, x+h)$, such that

$$\tfrac{1}{2}(y^{\mathsf{iv}}(\xi_1) + y^{\mathsf{iv}}(\xi_2)) = y^{\mathsf{iv}}(\xi)\,.$$

https://doi.org/10.1017/CBO9780511801181.014 Published online by Cambridge University Press

The required result is now obtained by adding the two equations (13.4) and dividing by h^2.

(ii) The proof is completely analogous, and is left to the reader as an exercise. (See Exercise 1.) □

We can now use the central difference approximation to construct the numerical solution. Writing Y_j for the numerical approximation to $y(x_j)$, we approximate the differential equation by

$$-\frac{\delta^2 Y_j}{h^2} + r_j Y_j = f_j\,, \quad j = 1, 2, \ldots, n-1\,, \tag{13.5}$$

where we have used the notation $r_j = r(x_j)$, $f_j = f(x_j)$. Now, (13.5) is a system of $n-1$ linear algebraic equations for the $n-1$ unknowns Y_j, $j = 1, 2, \ldots, n-1$, with the boundary conditions specifying the values of Y_0 and Y_n,

$$Y_0 = A\,, \qquad Y_n = B\,. \tag{13.6}$$

The system may be written in matrix form as

$$M\mathbf{Y} = \boldsymbol{g}\,,$$

where $\mathbf{Y}, \boldsymbol{g} \in \mathbb{R}^{n-1}$ and, for $n \geq 4$, the matrix $M \in \mathbb{R}^{(n-1)\times(n-1)}$ is tridiagonal. Here $\mathbf{Y} = (Y_1, \ldots, Y_{n-1})^{\mathrm{T}}$, the nonzero elements of M are

$$M_{jj} = 2/h^2 + r_j\,, \qquad M_{j\,j-1} = M_{j\,j+1} = -1/h^2\,, \tag{13.7}$$

and the elements of the column vector $\boldsymbol{g}$ on the right-hand side are

$$g_1 = f_1 + A/h^2,\ \ g_{n-1} = f_{n-1} + B/h^2,\ \ g_j = f_j\,,\ \ j = 2, 3, \ldots, n-2\,.$$

Note how the known boundary values Y_0 and Y_n have been transferred to the right-hand side, and appear in the first and last elements of $\boldsymbol{g}$. The solution of this system is very easy, using the algorithm for tridiagonal matrices described in Section 3.3. Using the fact that $r(x) \geq 0$, we see that the off-diagonal elements of M are negative, the diagonal elements are positive, and in each row the diagonal element is at least as large as the sum of absolute values of the off-diagonal elements. Theorem 3.4 implies that no row interchanges are needed in the calculation, and that the matrix M is nonsingular. The calculation is therefore very straightforward and efficient, and requires very little computational time, even for a mesh which may contain several hundred points.

https://doi.org/10.1017/CBO9780511801181.014 Published online by Cambridge University Press

13.3 Error analysis

Having obtained the numerical solution we must now analyse its accuracy. In the same way as for initial value problems, we begin by finding the truncation error.

Definition 13.2 *The* **truncation error** *of the central difference approximation to the problem (13.1) is*

$$T_j = -\frac{\delta^2 y(x_j)}{h^2} + r_j y(x_j) - f_j\,, \qquad j = 1, 2, \dots, n-1\,,$$

where y is the exact solution of (13.1), (13.2).

Theorem 13.2 *Suppose that the solution y to the boundary value problem (13.1), (13.2) has a continuous fourth derivative on $[a, b]$. Then, the truncation error may be written*

$$T_j = -\tfrac{1}{12} h^2 y^{iv}(\xi_j)\,, \tag{13.8}$$

for some value of ξ_j in the interval (x_{j-1}, x_{j+1}), $j = 1, 2, \dots, n-1$. The truncation error is bounded by T, where

$$|T_j| \le T = \tfrac{1}{12} h^2 M_4\,, \qquad j = 1, 2, \dots, n-1\,,$$

and

$$M_4 = \max_{x \in [a,b]} |y^{iv}(x)|\,. \tag{13.9}$$

Proof The expression for T_j follows from the substitution of the expression for $\delta^2 y(x_j)$ given by Theorem 13.1 into the definition of T_j, and the use of the fact that y is the solution of the differential equation. The proof of the bound for T_j is then immediate; since y^{iv} is known to be continuous on $[a, b]$ it is bounded on $[a, b]$, so M_4 exists. □

In order to simplify writing, we define

$$L(u_j) = -\frac{\delta^2 u_j}{h^2} + r_j u_j\,, \qquad j = 1, 2, \dots, n-1,$$

for any set of real numbers $\{u_0, u_1, \dots, u_n\}$. The **global error** in the numerical solution is defined by

$$e_j = y(x_j) - Y_j\,, \qquad j = 0, 1, \dots, n\,.$$

Now, $y(x_j)$ and Y_j satisfy

$$\begin{aligned} L(y(x_j)) &= f_j + T_j\,, \qquad j = 1, 2, \ldots, n-1\,, \\ L(Y_j) &= f_j\,, \qquad\qquad j = 1, 2, \ldots, n-1\,, \end{aligned}$$

from the definition of truncation error and (13.5); hence, by subtraction,

$$L(e_j) = T_j\,, \qquad j = 1, 2, \ldots, n-1\,,$$

with the boundary conditions $e_0 = e_n = 0$. We must now use the bound on T_j to derive a bound on the error e_j. This will be achieved by means of the following theorem.

Theorem 13.3 (**Maximum Principle**) *Let a_j, b_j, c_j, $j = 0, 1, \ldots, n$, be positive real numbers such that $b_j \geq a_j + c_j$, and suppose that u_j, $j = 0, 1, \ldots, n$, are real numbers such that*

$$-a_j u_{j-1} + b_j u_j - c_j u_{j+1} \leq 0\,, \qquad j = 1, 2, \ldots, n-1\,.$$

Then, $u_j \leq K$, $j = 0, 1, \ldots, n$, where $K = \max\{u_0, u_n, 0\}$.

Proof Let $u_r = \max\{u_0, u_1, \ldots, u_n\}$; then if $r = 0$, $r = n$, or $u_r \leq 0$ the result is trivial. Suppose then that $1 \leq r \leq n-1$, and that $u_r > 0$. Since u_r is the maximum of the u_j, we know that

$$u_r \geq u_{r-1}\,, \qquad u_r \geq u_{r+1}\,.$$

Hence

$$\begin{aligned} b_r u_r &\leq a_r u_{r-1} + c_r u_{r+1} \\ &\leq a_r u_r + c_r u_r \\ &\leq b_r u_r\,, \end{aligned}$$

since $u_r > 0$. This means that equality holds throughout, so that $u_{r-1} = u_r = u_{r+1}$. We can then apply the same argument to both u_{r-1} and u_{r+1}, continuing until we find that either $u_r = u_n$ or $u_r = u_0$. Thus, in this case $u_0 = u_n = \max\{u_0, u_1, \ldots, u_n\}$, as required. □

Theorem 13.4 *Suppose that the solution y of the boundary value problem (13.1), (13.2) has a continuous fourth derivative on $[a, b]$, and that Y_j, $j = 0, 1, \ldots, n$, is the solution of the central difference approximation (13.5), (13.6). Then,*

$$\max_{0 \leq j \leq n} |y(x_j) - Y_j| \leq \tfrac{1}{96} h^2 (b-a)^2 M_4\,. \tag{13.10}$$

https://doi.org/10.1017/CBO9780511801181.014 Published online by Cambridge University Press

Proof Let $e_j = y(x_j) - Y_j$. We have already seen that $L(e_j) = T_j$, $j = 1, 2, \ldots, n-1$. Defining

$$\varphi_j = C\left\{(2j-n)^2h^2 - n^2h^2\right\}, \qquad j = 0, 1, \ldots, n, \tag{13.11}$$

where C is a constant, we see that

$$\begin{aligned} L(\varphi_j) &= -C\left\{(2j-2-n)^2 - 2(2j-n)^2 + (2j+2-n)^2\right\} + r_j\varphi_j \\ &= -8C + r_j\varphi_j, \qquad j = 1, 2, \ldots, n-1. \end{aligned}$$

Hence

$$L(e_j + \varphi_j) = T_j - 8C + r_j\varphi_j, \qquad j = 1, 2, \ldots, n-1.$$

If we choose $C = T/8$ with $T = \frac{1}{12}h^2M_4$, we see that $L(e_j + \varphi_j) \le 0$, since $|T_j| \le T$, $r_j \ge 0$ and $\varphi_j \le 0$, and L satisfies the conditions of the Maximum Principle. Now,

$$e_0 + \varphi_0 = e_n + \varphi_n = 0,$$

so that, according to Theorem 13.3, $e_j + \varphi_j \le 0$ for $j = 0, 1, \ldots, n$. However, $-Cn^2h^2 \le \varphi_j \le 0$, so we have the result

$$e_j \le Cn^2h^2 = \tfrac{1}{8}(b-a)^2T = \tfrac{1}{96}h^2(b-a)^2M_4, \qquad j = 0, 1, \ldots, n.$$

By applying the same argument to $L(-e_j + \varphi_j)$ we find that

$$-e_j \le \tfrac{1}{96}h^2(b-a)^2M_4, \qquad j = 0, 1, \ldots, n.$$

Combining these upper bounds for e_j and $-e_j$ gives the required result.

□

The function φ defined by (13.11) is called a **comparison function**. An alternative proof of Theorem 13.4, based on the properties of monotone matrices, can be given by using the result in Exercise 2. Notice that the condition $r(x) \ge 0$ is used in the application of the Maximum Principle in the above proof.

This theorem shows that, provided the solution y has a continuous fourth derivative, the numerical method is **convergent**, that is

$$\max_{0 \le j \le n} |y(x_j) - Y_j| \to 0 \qquad \text{as } n \to \infty$$

(or, equivalently, as $h = (b-a)/n \to 0$). This means that we can obtain any required accuracy by choosing n sufficiently large.

13.4 Boundary conditions involving a derivative

The same differential equation (13.1) may be associated with boundary conditions involving the first derivative of the solution. Suppose, for example, that we are given real numbers $\alpha > 0$, A and B. Consider the differential equation (13.1) together with the boundary conditions

$$y'(a) - \alpha y(a) = A\,, \qquad y(b) = B\,. \tag{13.12}$$

The condition at $x = a$ may be approximated in various ways; we shall introduce an extra mesh point x_{-1} outside the interval and use the approximate version

$$\frac{Y_1 - Y_{-1}}{2h} - \alpha Y_0 = A\,.$$

This gives

$$Y_{-1} = Y_1 - 2h\alpha Y_0 - 2hA\,.$$

Writing the same central difference approximation (13.5) as before, but now for $j = 0, 1, \ldots, n-1$, we can eliminate the extra unknown Y_{-1} from the equation at $j = 0$ to give

$$\left[\frac{2(1+\alpha h)}{h^2} + r_0\right] Y_0 - \frac{2}{h^2} Y_1 = f_0 - \frac{2}{h} A\,.$$

Together with (13.5), for $j = 1, 2, \ldots, n-1$, we now have a system of n equations for the unknowns Y_j, $j = 0, 1, \ldots, n-1$. There are one more equation and one more unknown than before, but the new matrix is still tridiagonal, and also diagonally dominant because of the condition $\alpha > 0$. The computation is again very straightforward.

Theorem 13.5 *Suppose that $y \in \mathrm{C}^3[x-h, x+h]$; then, there exists a real number χ in $(x-h, x+h)$ such that*

$$\frac{y(x+h) - y(x-h)}{2h} = y'(x) + \tfrac{1}{6}h^2 y'''(\chi)\,. \tag{13.13}$$

Proof Taylor's Theorem shows that there exist $\chi_1 \in (x-h, x)$ and $\chi_2 \in (x, x+h)$ such that

$$\begin{aligned} y(x-h) &= y(x) - hy'(x) + \tfrac{1}{2}h^2 y''(x) - \tfrac{1}{6}h^3 y'''(\chi_1)\,, \\ y(x+h) &= y(x) + hy'(x) + \tfrac{1}{2}h^2 y''(x) + \tfrac{1}{6}h^3 y'''(\chi_2)\,. \end{aligned}$$

We subtract the first equality from the second, and the result follows as in the proof of Theorem 13.1. □

Note that the approximation to $y'(x)$ at $x = x_0$ may be written

$$\frac{\frac{1}{2}[\delta y(x_0 + \frac{1}{2}h) + \delta y(x_0 - \frac{1}{2}h)]}{h}.$$

For $j = 1, 2, \ldots, n-1$, we define the truncation error T_j as in Definition 13.2. In addition, since we shall now also incur an error in the approximation of the boundary condition at $x = a$, we define

$$T_0 = \left[\frac{2(1+\alpha h)}{h^2} + r_0\right] y(0) - \frac{2}{h^2} y(h) - f_0 + \frac{2}{h} A\,.$$

The aim of our next result is to quantify the size of the truncation error in terms of the mesh size h.

Theorem 13.6 *Suppose that the solution y to the boundary value problem (13.1), (13.2) has a continuous fourth derivative on the closed interval $[a-h, b]$. Then, the truncation error of the central difference approximation to (13.1) with boundary conditions (13.12) may be written*

$$\begin{aligned} T_j &= -\tfrac{1}{12} h^2 y^{\mathrm{iv}}(\xi_j)\,, \qquad j = 1, 2, \ldots, n-1\,, \\ T_0 &= -\tfrac{1}{12} h^2 y^{\mathrm{iv}}(\xi_0) - \tfrac{1}{3} h y'''(\chi)\,, \end{aligned}$$

for some value of ξ_j in the interval (x_{j-1}, x_{j+1}), $1 \le j \le n-1$, and some value χ in the interval (x_{-1}, x_1) where $x_{-1} = a - h$.

Proof For $j = 1, 2, \ldots, n-1$, this is the same result as in Theorem 13.2. When $j = 0$, we find that

$$\begin{aligned} T_0 &= \left[\frac{2(1+\alpha h)}{h^2} + r_0\right] y(0) - \frac{2}{h^2} y(h) - f_0 + \frac{2}{h} A \\ &= -\frac{y(h) - 2y(0) + y(-h)}{h^2} + r(0)y(0) - f(0) \\ &\quad -\frac{2}{h}\left[\frac{y(h) - y(-h)}{2h} - \alpha y(0) - A\right] \\ &= -\tfrac{1}{12} h^2 y^{\mathrm{iv}}(\xi_0) - \tfrac{2}{h}\,\tfrac{1}{6} h^2 y'''(\chi)\,, \end{aligned}$$

where we have used Theorem 13.5. □

Theorem 13.7 *Suppose that the solution y of (13.1) with the boundary conditions (13.12) has a continuous fourth derivative on the interval $[a-h, b]$; then, the numerical solution obtained from the central difference* *approximation satisfies*

$$\max_{0\le j\le n} |y(x_j) - Y_j| \le h^2 \left\{ \tfrac{1}{24}(b-a)^2 M_4 + \tfrac{1}{6}(b-a)M_3 \right\} .$$

Proof The proof is very similar to that of Theorem 13.4, but requires the use of a more complicated comparison function φ_j. Let us define

$$\begin{aligned} L^*(u_j) &= -\frac{\delta^2 u_j}{h^2} + r_j u_j \,, \qquad j = 1, 2, \dots, n-1 \,, \\ L^*(u_0) &= \left[\frac{2(1+\alpha h)}{h^2} + r_0 \right] u_0 - \frac{2}{h^2} u_1 \,, \end{aligned}$$

for any set of real numbers $\{u_0, u_1, \dots, u_n\}$, and let

$$\varphi_j = C j^2 h^2 + D j h + E \,, \qquad j = 0, 1, \dots, n \,,$$

where C, D and E are constants to be determined. Then, with $e_j = y(x_j) - Y_j$, as in the proof of Theorem 13.4, we see that

$$L^*(e_j) = T_j \,, \qquad j = 0, 1, \dots, n-1 \,.$$

A simple calculation shows that

$$\begin{aligned} L^*(\varphi_j) &= -2C + r_j \varphi_j \,, \qquad j = 1, 2, \dots, n-1 \,, \\ L^*(\varphi_0) &= -2C - 2D/h + [2\alpha/h + r_0] E \,. \end{aligned}$$

Hence

$$\begin{aligned} L^*(e_j + \varphi_j) &= -\tfrac{1}{12} h^2 y^{iv}(\xi_j) - 2C + r_j \varphi_j \,, \quad j = 1, 2, \dots, n-1 \,, \\ L^*(e_0 + \varphi_0) &= -\tfrac{1}{12} h^2 y^{iv}(\xi_0) - \tfrac{1}{3} h y'''(\chi) \\ &\quad -2C - 2D/h + [2\alpha/h + r_0] E \,. \end{aligned}$$

If we now choose

$$C = \tfrac{1}{24} h^2 M_4 \,, \qquad D = \tfrac{1}{6} h^2 M_3 \,, \qquad E = -C(b-a)^2 - D(b-a) \,,$$

it is easy to check that

$$\begin{aligned} \varphi_j &\le 0 \,, \qquad j = 0, 1, \dots, n \,, \\ L^*(e_j + \varphi_j) &\le 0 \,, \qquad j = 0, 1, \dots, n-1 \,. \end{aligned}$$

The Maximum Principle then applies, and we deduce that

$$e_j + \varphi_j \le \max\{e_0 + \varphi_0, e_n + \varphi_n, 0\} \,, \qquad j = 0, 1, \dots, n \,.$$

We see at once that $e_n = \varphi_n = 0$ and $\varphi_0 \le 0$, but in this case e_0 is not zero. Therefore, all we can conclude for the moment is that

$$e_j + \varphi_j \le \max\{e_0 + \varphi_0, 0\} \,, \qquad j = 0, 1, \dots, n \,. \tag{13.14}$$

https://doi.org/10.1017/CBO9780511801181.014 Published online by Cambridge University Press

In particular,

$$e_1 + \varphi_1 \leq \max\{e_0 + \varphi_0, 0\}\,. \tag{13.15}$$

However, $L^*(e_0 + \varphi_0) \leq 0$; thus, by the definition of $L^*(e_0 + \varphi_0)$,

$$e_0 + \varphi_0 \leq \frac{2}{2(1 + \alpha h) + h^2 r_0}\,(e_1 + \varphi_1)\,.$$

On writing $\delta = 2/(2(1 + \alpha h) + h^2 r_0)$ and noting that, since $\alpha > 0$ and $r_0 \geq 0$, we have $0 < \delta < 1$, it follows that

$$e_0 + \varphi_0 \leq \delta(e_1 + \varphi_1)\,. \tag{13.16}$$

Inserting this inequality into the left-hand side of (13.15), we find that

$$e_0 + \varphi_0 \leq \max\{\delta(e_0 + \varphi_0), 0\}\,.$$

If $e_0+\varphi_0$ were positive, this inequality and the fact that $0 < \delta < 1$ would imply $e_0 + \varphi_0 \leq 0$, leading to a contradiction. Therefore, $e_0 + \varphi_0 \leq 0$. Returning with this information to (13.14), we conclude that $e_j + \varphi_j \leq 0$ for $j = 0, 1, \ldots, n$, and the rest of the proof then follows as in the proof of Theorem 13.1. □

13.5 The general self-adjoint problem

The general self-adjoint boundary value problem is

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(p(x)\frac{\mathrm{d}y}{\mathrm{d}x}\right) + r(x)y = f(x)\,, \qquad a < x < b\,, \tag{13.17}$$

where r and f are real-valued functions, defined and continuous on $[a, b]$, p is a real-valued continuously differentiable function on $[a, b]$, $r(x) \geq 0$ and $p(x) \geq c_0 > 0$. We shall consider only the case where the boundary conditions prescribe the values of y at each end,

$$y(a) = A\,, \qquad y(b) = B\,. \tag{13.18}$$

The central difference approximation to the equation (13.17) may be written

$$-\frac{\delta(p_j\,\delta Y_j)}{h^2} + r_j Y_j = f_j\,, \qquad j = 1, 2, \ldots, n-1\,,$$

or, in detail,

$$-\frac{p_{j+1/2}(Y_{j+1} - Y_j) - p_{j-1/2}(Y_j - Y_{j-1})}{h^2} + r_j Y_j = f_j\,, \tag{13.19}$$

for $j = 1, 2, \ldots, n$, and is supplemented by the boundary conditions

$$Y_0 = A\,, \qquad Y_n = B\,. \tag{13.20}$$

It is easy to see that this represents a system of linear equations for the unknowns $Y_1, Y_2, \ldots, Y_{n-1}$, and that the matrix of the system is tridiagonal and diagonally dominant, just as it was in the special case (13.1), which corresponds to $p(x) \equiv 1$. The solution of the system is therefore a very simple matter.

Next, we consider the error analysis of the difference scheme (13.19), (13.20). We begin by quantifying the size of the truncation error

$$T_j = -\frac{\delta(p_j\,\delta y(x_j))}{h^2} + r_j y(x_j) - f_j\,, \qquad j = 1, 2, \ldots, n-1\,,$$

in terms of the mesh size h.

Lemma 13.1 *Suppose that $p \in \mathrm{C}^3[a,b]$ and $y \in \mathrm{C}^4[a,b]$. The truncation error T_j of the central difference approximation (13.19) then satisfies*

$$|T_j| \le T = \tfrac{1}{24} h^4 \max_{x\in[a,b]} \left\{ |(py')'''(x)| + |p'y'''(x)| + 2|py^{\mathrm{iv}}(x)| \right\},$$

for $j = 1, 2, \ldots, n-1$.

Proof By expanding in Taylor series as we have done before, we find that

$$\begin{aligned} p_{j+1/2}[y(x_{j+1}) - y(x_j)] &= p_{j+1/2}[hy'_{j+1/2} + \tfrac{1}{24}h^3 y'''(\xi_1)]\,, \\ p_{j-1/2}[y(x_j) - y(x_{j-1})] &= p_{j-1/2}[hy'_{j-1/2} + \tfrac{1}{24}h^3 y'''(\xi_2)]\,, \end{aligned}$$

where $\xi_1 \in (x_j, x_{j+1})$ and $\xi_2 \in (x_{j-1}, x_j)$. The first term in the difference of these expressions gives, in the same way,

$$h[p_{j+1/2}y'(x_{j+1/2}) - p_{j-1/2}y'(x_{j-1/2})] = h[h(py')'(x_j) + \tfrac{1}{24}h^3 (py')'''(\xi_3)]$$

where $\xi_3 \in (x_{j-1/2}, x_{j+1/2})$. For the other term we can write

$$\begin{aligned} &\tfrac{1}{24}h^3 |p_{j+1/2}y'''(\xi_1) - p_{j-1/2}y'''(\xi_2)| \\ &\qquad = \tfrac{1}{24}h^3 |(p_{j+1/2} - p_{j-1/2})y'''(\xi_1) + p_{j-1/2}[y'''(\xi_1) - y'''(\xi_2)]| \\ &\qquad \le \tfrac{1}{24}h^3 \left\{ |hp'(\xi_4)y'''(\xi_1)| + |p_{j-1/2} 2h y^{\mathrm{iv}}(\xi_5)| \right\}, \end{aligned}$$

since $|\xi_1 - \xi_2| < 2h$. Here, $\xi_4 \in (x_{j-1/2}, x_{j+1/2})$ and ξ_5 lies between ξ_1 and ξ_2. The required bound follows immediately. □

As in the proof of Theorem 13.4, we can now derive a bound on the global error in the numerical solution in terms of the truncation error by using the Maximum Principle. The only difficulty in extending that theorem to the more general self-adjoint problem lies in the construction of a comparison function corresponding to (13.11). The general case requires some detailed analysis, which can be simplified under certain conditions on the function p, for example if p is monotonic.

Lemma 13.2 *Suppose that p and r are continuous functions defined on $[a, b]$, p is monotonic increasing on $[a, b]$, $p(x) \geq c_0 > 0$, $r(x) \geq 0$, and define*

$$L(u_j) = -\frac{\delta(p\,\delta u_j)}{h^2} + r_j u_j\,, \qquad j = 1, 2, \ldots, n-1\,,$$

for any set of real numbers $\{u_0, u_1, \ldots, u_n\}$. Further, let

$$\varphi_j = C(j^2 - n^2)h^2\,, \qquad j = 0, 1, \ldots, n\,,$$

where C is a positive constant. Then,

$$L(\varphi_j) \leq -2c_0 C\,, \quad j = 1, 2, \ldots, n-1\,.$$

Proof It follows from the definition that

$$\begin{aligned} L(\varphi_j) &= -p_{j+1/2}C(2j+1) + p_{j-1/2}C(2j-1) + C(j^2-n^2)h^2 r_j \\ &= -C[(p_{j+1/2} + p_{j-1/2}) + 2j(p_{j+1/2} - p_{j-1/2}) + h^2(n^2 - j^2)r_j] \\ &\leq -2c_0 C\,, \end{aligned}$$

for $j = 1, 2, \ldots, n-1$, as required. □

Note that we have imposed various conditions on the problem, which are usually necessary, though some can be slightly relaxed. The condition in this lemma, that p should be monotonic increasing on $[a, b]$, is only needed to simplify the subsequent proof. The main result is true much more generally. We leave it as an exercise to derive the same result under the assumption that p is monotonic decreasing on $[a, b]$.

Theorem 13.8 *Suppose that p and r are continuous functions defined on $[a, b]$, p is monotonic increasing on $[a, b]$, $p(x) \geq c_0 > 0$, $r(x) \geq 0$. Assume further that the solution y of (13.17), (13.18) has a continuous fourth derivative on $[a, b]$, that p has a continuous third derivative, and* *that* Y_j, $j = 0, 1, \ldots, n$, *is the solution of the central difference approximation (13.19), (13.20). Then, with* T *as in Lemma 13.1,*

$$\max_{0 \le j \le n} |y(x_j) - Y_j| \le \tfrac{1}{2c_0} T\,. \tag{13.21}$$

Proof The proof of this theorem follows that of Theorem 13.4, using the bound from Lemma 13.1 on the truncation error and the comparison function φ_j from Lemma 13.2. The details are left as an exercise. □

13.6 The Sturm–Liouville eigenvalue problem

Suppose that r is a real-valued function, defined and continuous on the closed interval $[a, b]$, p is a real-valued function, defined and continuously differentiable on $[a, b]$, and $r(x) \ge 0$, $p(x) \ge c_0 > 0$ for all $x \in [a, b]$. The differential equation

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(p(x)\frac{\mathrm{d}y}{\mathrm{d}x}\right) + r(x)y = \lambda y\,, \qquad a < x < b\,, \tag{13.22}$$

with homogeneous boundary conditions $y(a) = y(b) = 0$, has only the trivial solution $y \equiv 0$, except for an infinite sequence of positive *eigenvalues* $\lambda = \lambda_m$, $m = 1, 2, \ldots$. We shall now consider a numerical method for finding these eigenvalues and the corresponding *eigenfunctions*, $y_{(m)}(x)$, $m = 1, 2, \ldots$.

In the simple case where $p(x) \equiv 1$ and $r(x) \equiv 0$ the solution to this problem is, of course, $\lambda_m = [m\pi/(b-a)]^2$, $y_{(m)}(x) = A \sin m\pi t$, $m = 1, 2, \ldots$, where A is a nonzero constant and $t = (x-a)/(b-a)$.

Using the same finite difference approximation as in the previous section, we obtain the equations

$$-\frac{p_{j+1/2}(Y_{j+1} - Y_j) - p_{j-1/2}(Y_j - Y_{j-1})}{h^2} + r_j Y_j = \Lambda Y_j\,,$$
$$j = 1, 2, \ldots, n-1\,.$$

Together with the boundary conditions $Y_0 = Y_n = 0$, this shows that Λ is an eigenvalue of a symmetric tridiagonal matrix M whose entries are

$$M_{jj} = \frac{p_{j+1/2} + p_{j-1/2}}{h^2} + r_j\,, \qquad 1 \le j \le n-1\,,$$
$$M_{j\,j-1} = -\frac{p_{j-1/2}}{h^2}\,,\ 2 \le j \le n\,, \qquad M_{j\,j+1} = -\frac{p_{j+1/2}}{h^2}\,,\ 1 \le j \le n-1\,,$$

https://doi.org/10.1017/CBO9780511801181.014 Published online by Cambridge University Press

and the approximate function values Y_j are the elements of the corresponding eigenvector. This algebraic eigenvalue problem is easily solved by the method described in Chapter 5.

The boundary value problems which we have discussed so far have all had a unique solution. The eigenvalue problem (13.22) has an infinite number of solutions, and the mesh used in the numerical computation has to be chosen to adequately represent the eigenfunctions required – the computation can obviously only find a finite number of them. The matrix M has $n-1$ eigenvalues and eigenvectors and, as we shall see, it will normally give a good approximation to the first few eigenvalues, $\lambda_1, \lambda_2, \ldots$, and a much less accurate approximation to λ_{n-1}.

To analyse the error in the eigenvalue we proceed as before, by defining the truncation error

$$T_j = -\frac{p_{j+1/2}(y_{j+1}-y_j) - p_{j-1/2}(y_j - y_{j-1})}{h^2} + r_j y_j - \lambda y_j\,, \qquad j = 1, 2, \ldots, n-1\,,$$

where $y_j = y(x_j)$. These equations can now be written

$$\begin{aligned} (M - \Lambda I)\mathbf{Y} &= \mathbf{0}\,, \\ (M - \lambda I)\boldsymbol{y} &= \mathbf{T}\,, \end{aligned}$$

where

$$\begin{aligned} \mathbf{Y} &= (Y_1, \ldots, Y_{n-1})^{\mathrm{T}}\,, \\ \boldsymbol{y} &= (y_1, \ldots, y_{n-1})^{\mathrm{T}}\,, \\ \mathbf{T} &= (T_1, \ldots, T_{n-1})^{\mathrm{T}}\,. \end{aligned}$$

Theorem 5.15 of Chapter 5 applies to this problem, and shows that one of the eigenvalues, Λ_m, of the matrix M satisfies

$$|\lambda_m - \Lambda_m| \leq \|\mathbf{T}\|_2 / \|\boldsymbol{y}\|_2\,. \tag{13.23}$$

In the simpler case where $p(x) \equiv 1$ and $r(x) \equiv 0$ the truncation error is

$$T_j = -\tfrac{1}{12}h^2 y^{iv}(\xi_j)\,, \qquad \xi_j \in (x_{j-1}, x_{j+1})\,,$$

so the numerical method has evaluated the eigenvalue with error less than

$$\tfrac{1}{12}h^2 \left\{ \sum_{j=1}^{n-1} [y^{iv}(\xi_j)]^2 \right\}^{1/2} \left\{ \sum_{j=1}^{n-1} [y(x_j)]^2 \right\}^{-1/2}.$$

Since the mth eigenfunction $y_{(m)}$ is given by

$$y(x) = y_{(m)}(x) = \sin(m\pi(x-a)/(b-a)), \qquad x \in (a,b),$$

we see that

$$y^{\mathsf{iv}}(x) = \left[\frac{m\pi}{b-a}\right]^4 y(x), \qquad x \in (a,b).$$

This shows that, for example, the error in the tenth eigenvalue, corresponding to $m = 10$, is likely to be about 10^4 times larger than the error in the first eigenvalue; more generally, to evaluate higher eigenvalues of the equation will require the use of a smaller interval h.

13.7 The shooting method

The methods we have described for the linear boundary value problem may be extended to nonlinear differential equations. We shall not discuss how this is done; instead, we shall describe an alternative approach, called the **shooting method**. We shall consider the nonlinear model problem

$$y'' = f(x,y), \quad a < x < b, \qquad y(a) = A, \quad y(b) = B,$$

where we assume that the function $f(x,y)$ is continuous and differentiable, and that

$$\frac{\partial f}{\partial y}(x,y) \geq 0, \qquad a < x < b, \quad y \in \mathbb{R}.$$

The central idea of the method is to replace the boundary value problem under consideration by an initial value problem of the form

$$y'' = f(x,y), \quad a < x \leq b, \qquad y(a) = A, \quad y'(a) = t,$$

where t is to be chosen in such a way that $y(b) = B$. This can be thought of as a problem of trying to determine the angle of inclination $\tan^{-1} t$ of a loaded gun, so that, when shot from height A at the point $x = a$, the bullet hits the target placed at height B at the point $x = b$. Hence the name, shooting method.

Once the boundary value problem has been transformed into such an 'equivalent' initial value problem, any of the methods for the numerical solution of initial value problems discussed in Chapter 12 can be applied to find a numerical solution. Thus, in particular, the costly exercise of solving a large system of nonlinear equations, arising from a direct finite difference approximation of the nonlinear boundary value problem, can be completely avoided.

If we write

$$y'(a) = t\,,$$

a numerical solution of the differential equation with the initial conditions $y(a) = A, y'(a) = t$ can be obtained by any of the methods of Chapter 12. This solution will depend on t, and we may write it as $y(x;t)$. In particular the value at $x = b$ will be a function of t,

$$y(b;t) = \psi(t)\,. \tag{13.24}$$

The solution of the nonlinear boundary value problem therefore reduces to the determination of the value of t for which the boundary condition at $x = b$ is also satisfied, *i.e.*,

$$\psi(t) - B = 0\,.$$

There are a number of well-known methods for the solution of equations of this form; Newton's method is an obvious example. Generally, we shall not, of course, have a closed form expression for the function $\psi(t)$, in general, but this is not necessary; all that is needed is a numerical algorithm to calculate the value of $\psi(t)$ for a given value of t, and this we have. To use Newton's method we shall also need to be able to calculate the value of $\psi'(t)$, and this is easily done.

The function $y(x;t)$ is defined, for all t, as the solution of the initial value problem

$$y''(x;t) = f(x, y(x;t))\,, \qquad y(a;t) = A\,, \quad y'(a;t) = t\,, \tag{13.25}$$

where $'$ and $''$ indicate differentiation with respect to the variable x. We can differentiate these throughout with respect to t, giving

$$\frac{\partial}{\partial t}y''(x;t) = \frac{\partial f}{\partial y}(x, y(x;t))\frac{\partial y}{\partial t}(x;t)\,, \quad \frac{\partial y}{\partial t}(a;t) = 0\,, \;\; \frac{\partial y'}{\partial t}(a;t) = 1\,.$$

Writing

$$w(x,t) = \frac{\partial y}{\partial t}(x;t)\,,$$

and interchanging the order of differentiation, we find that $w(x;t)$ may be obtained as the solution of the initial value problem

$$w''(x;t) = w(x;t)\,\frac{\partial f}{\partial y}(x, y(x;t))\,, \qquad w(a;t) = 0\,, \;\; w'(a;t) = 1\,. \tag{13.26}$$

By virtue of (13.24), the required derivative is then given by

$$\psi'(t) = w(b,t)\,.$$

To implement this method, it is convenient to solve the two initial value problems, (13.25) and (13.26), in tandem, by writing them as a system of four simultaneous first-order differential equations:

$$\left.\begin{array}{rcl} u_1'(x;t) & = & u_2(x;t)\,, \\ u_2'(x;t) & = & f(x,u_1(x;t))\,, \\ u_3'(x;t) & = & u_4(x;t)\,, \\ u_4'(x;t) & = & u_3(x;t)\,\frac{\partial f}{\partial u_1}(x,u_1(x;t))\,, \end{array}\right\} \qquad (13.27)$$

with the initial conditions

$$u_1(a;t) = A\,, \quad u_2(a;t) = t\,, \quad u_3(a;t) = 0\,, \quad u_4(a;t) = 1\,,$$

where $u_1(x;t)$ denotes $y(x;t)$, $u_3(x;y)$ signifies $w(x;t)$, and u_2 and u_4 are defined by $u_2 = u_1' = y'$ and $u_4 = u_3' = w'$.

Having obtained a numerical solution of this system of differential equations for some chosen value of t, $t^{(k)}$ say, Newton's method gives, as the next, improved, value for t,

$$t^{(k+1)} = t^{(k)} - \frac{\psi(t^{(k)}) - B}{\psi'(t^{(k)})} = t^{(k)} - \frac{u_1(b,t^{(k)}) - B}{u_3(b,t^{(k)})}\,, \quad k = 0,1,\ldots\,,$$

iterating until a certain number of decimal digits have converged.

Theorem 13.9 *Suppose that a numerical algorithm for the solution of the system of differential equations (13.27) gives the result $v_{i,j}(t)$, the numerical approximation to $u_i(x_j;t)$, $i = 1,2,3,4$, $j = 1,2,\ldots,n$, where the error satisfies*

$$\max_{1\le j\le n} |u_i(x_j;t) - v_{i,j}(t)| \le C(t)h^s\,, \quad i = 1,2,3,4\,,$$

for some $s > 0$; here $C(t)$ depends on bounds on the derivatives of y and $f(x,y)$, and on t. Suppose also that the Newton iteration is performed until

$$|v_{1,n}(t^{(k)}) - B| \le \varepsilon\,.$$

Then, $v_{1,j}(t^{(k)})$ is an approximation to the solution of the boundary value problem which satisfies

$$\max_{1\le j\le n} |y(x_j) - v_{1,j}(t^k)| \le 2C(t^{(k)})h^s + \varepsilon\,.$$

Proof Suppose that the solution of the system of differential equations with $t = t^{(k)}$ is $u_i(x; t^{(k)})$, $i = 1, 2, 3, 4$, and the corresponding numerical solution is $v_{i,j}(t^{(k)})$, $i = 1, 2, 3, 4$, $j = 1, 2, \ldots, n$; then

$$|u_i(x_j; t^{(k)}) - v_{i,j}(t^{(k)})| \le C(t^{(k)})h^s \, .$$

Moreover $|v_{1,n}(t^{(k)}) - B| \le \varepsilon$, so that

$$\begin{aligned} |u_1(b; t^{(k)}) - B| &\le |u_1(b; t^{(k)}) - v_{1,n}(t^{(k)})| + |v_{1,n}(t^{(k)}) - B| \\ &\le C(t^{(k)})h^s + \varepsilon \, . \end{aligned} \tag{13.28}$$

Let us write $\eta(x; t) = y(x) - u_1(x; t)$; by subtraction we see that

$$\begin{aligned} \eta''(x; t) &= y''(x) - u_1''(x; t) \\ &= f(x, y(x)) - f(x, u_1(x, t)) \\ &= \eta(x; t) \frac{\partial f}{\partial y}(x, \xi(x; t)) \, , \end{aligned}$$

where $\xi(x; t)$ lies between $u_1(x; t)$ and $y(x)$.

Suppose that $\eta'(a; t) > 0$; since $\eta(a; t) = 0$, there is some interval to the right of a in which $\eta(x; t) > 0$. Then, either $\eta(x; t) > 0$ for the whole of $(a, b]$, or there is a value c such that $a < c < b$ and $\eta(c; t) = 0$. In the latter case, $\eta'(x; t)$ must vanish at some point $x = d$ between a and c. However, in the interval $[a, d]$, $\eta(x; t) > 0$ and $\partial f/\partial y \ge 0$, so that $\eta''(x; t) > 0$. Consequently, in the interval $[a, d]$, $\eta'(x; t) > \eta'(a; t) > 0$, and we have a contradiction. Thus, $\eta(x; t) > 0$ for all $a < x \le b$. It then follows that $\eta''(x; t)$, and hence also $\eta'(x; t)$ are positive on the whole interval $[a, b]$, which means that $x \mapsto \eta(x; t)$ is monotonic increasing on $[a, b]$. If we had begun with the assumption that $\eta'(a; t) < 0$ an analogous argument shows that $x \mapsto \eta(x; t)$ would have been monotonic decreasing on $[a, b]$. It is left to the reader to discuss the trivial case when $\eta'(a, t) = 0$.

In any case,

$$|\eta(x; t)| \le |\eta(b; t)| \, , \qquad a \le x \le b \, ,$$

and therefore, since $y(b) = B$ and recalling (13.28),

$$|y(x) - u_1(x; t^{(k)})| \le |B - u_1(b; t^{(k)})| \le C(t^{(k)})h^s + \varepsilon \, .$$

Thus, finally,

$$\begin{aligned} |y(x_j) - v_{1,j}(t^{(k)})| &\le |y(x_j) - u_1(x_j; t^{(k)})| + |u_1(x_j; t^{(k)}) - v_{1,j}(t^{(k)})| \\ &\le C(t^{(k)})h^s + \varepsilon + C(t^{(k)})h^s \, , \qquad j = 1, 2, \ldots, n \, , \end{aligned}$$

and hence the desired bound. □

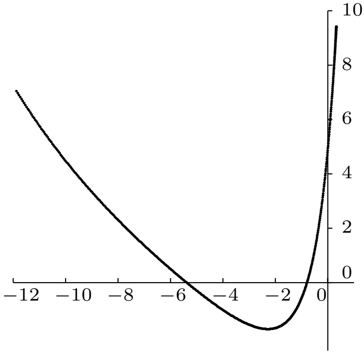

Fig. 13.1. The function $t \mapsto \psi(t)$.

The shooting method is an example of a technique which can be applied to much more general problems, including systems of differential equations of any order, with some boundary conditions specified at each end of the interval. The condition $\partial f/\partial y \geq 0$ is restrictive, and may often not be satisfied in practical problems. Note that if $f(x, y)$ is linear in y, of the form $f(x, y) = r(x)y + g(x)$, this condition is the same as the condition $r(x) \geq 0$ imposed on the model problem in (13.2). Perhaps the simplest example of a nonlinear two-point boundary value problem is

$$y'' = y^2 , \qquad y(-1) = y(1) = 1 , \tag{13.29}$$

where $\partial f/\partial y = 2y$, which does *not* satisfy the condition $\partial f/\partial y \geq 0$, $y \in \mathbb{R}$. In fact, problem (13.29) has two solutions, one of which is positive, and the other takes negative values around $x = 0$.

Figure 13.1 shows a graph of the corresponding function $t \mapsto \psi(t)$ defined in (13.24), over the range $-12 \leq t \leq 0$; outside this range the function ψ tends quite rapidly to $+\infty$. This shows clearly the two solutions to the boundary value problem, given by the two values of t at which $\psi(t) = 1$. The two solutions are displayed in Figure 13.2.

For the positive solution it is reasonable to suppose that the above proof could be modified so that it requires only that $\partial f/\partial y$ is positive for values of y in the neighbourhood of the solution, and the error bound would then hold, at least if h and ε were sufficiently small. The analysis of the error of the other solution, which takes negative values, will be much more difficult, as our proof relies heavily on the monotonicity of solutions of the linearised equation.

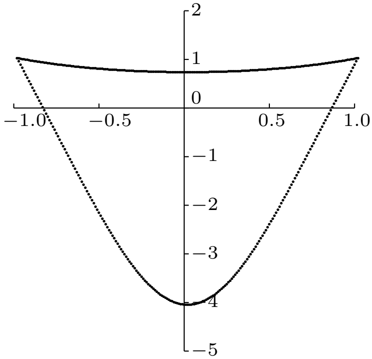

Fig. 13.2. The two solutions of the nonlinear boundary value problem (13.29).

13.8 Notes

The following books are standard texts on the subject of numerical approximation of boundary value problems:

- H.B. Keller, *Numerical Methods for Two-Point Boundary Value Problems,* Reprint of the 1968 original published by Blaisdell, Dover, New York, 1992.
- H.B. Keller, *Numerical Solution of Two-Point Boundary Value Problems*, SIAM, Philadelphia, fourth printing, 1990.

A more recent survey of the subject is found in

- U.M. Ascher, R.M.M. Mattheij and R.D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations,* Corrected reprint of the 1988 original, Classics in Applied Mathematics, 13, SIAM, Philadelphia, 1995.

In practical implementations of the shooting method into mathematical software (see, for example, Appendix A in the Ascher *et al.* book), the interval $[a, b]$ is subdivided into smaller intervals on each of which the shooting method is applied with appropriately chosen initial values. The 'initial' conditions on the subintervals are then simultaneously adjusted in order to satisfy the boundary conditions and appropriate continuity conditions at the points of the subdivision. From the practical viewpoint, this extension of the basic shooting method considered in this chapter is extremely important: the various difficulties which may arise in the implementation of the basic method (such as, for example, growth of the solution to the initial value problem over the interval $[a, b]$, leading to loss of accuracy in the solution of the equation $\psi(t) = B$) are discussed, for example, in Section 2.4 of the 1992 book by Keller.

Sturm–Liouville problems originated in a paper of Jacques Charles François Sturm: Sur les équations différentielles linéaires du second ordre, *J. Math. Pures Appl.* **1**, 106–186, 1836, in Joseph Liouville's newly founded journal. Sturm's paper was followed by a series of articles by Sturm and Liouville in subsequent volumes of the journal. They examined general linear second-order differential equations, the properties of their eigenvalues, the behaviour of the eigenfunctions and the series expansion of arbitrary functions in terms of these eigenfunctions. An extensive survey of the theory and numerical analysis of Sturm–Liouville problems can be found in

- John D. Pryce, *Numerical Solution for Sturm–Liouville Problems,* Oxford University Press Monographs in Numerical Analysis, Clarendon Press, Oxford, 1993.

See also Section 11.3, page 478, of the Ascher *et al.* book cited above.

Exercises

13.1 Suppose that $y \in \mathrm{C}^6[x-h, x+h]$; show that there exists a real number η in $(x-h, x+h)$ such that

$$\frac{\delta^2 y(x)}{h^2} = y''(x) + \tfrac{1}{12}h^2 y^{\mathrm{iv}}(x) + \tfrac{1}{360}h^4 y^{\mathrm{vi}}(\eta)\,.$$

13.2 Use Theorem 3.6 to show that the matrix M in (13.7) is monotone. Use the result of Exercise 4 to show that $\|M^{-1}\|_\infty \le \frac{1}{8}$.

13.3 On the interval $[a, b]$ the differential equation

$$-y'' + f(x)y = g(x)$$

is approximated by

$$-\frac{\delta^2 y_j}{h^2} + \beta_{-1}y_{j-1} + \beta_0 y_j + \beta_1 y_{j+1} = \beta_{-1}g_{j-1} + \beta_0 g_j + \beta_1 g_{j+1}\,,$$

where β_{-1}, β_0 and β_1 are constants. Assuming that the solution y has the appropriate number of continuous derivatives, show that the truncation error of this approximation may be written as follows:

https://doi.org/10.1017/CBO9780511801181.014 Published online by Cambridge University Press

(i) if $\beta_{-1} + \beta_0 + \beta_1 \neq 1$, then

$$T_j = (\beta_{-1} + \beta_0 + \beta_1) y''(x_j) + Z_j^{(0)} h\,,$$

where $|Z_j^{(0)}| \le (|\beta_{-1}| + |\beta_1|) M_3$;

(ii) if $\beta_{-1} + \beta_0 + \beta_1 = 1$ and $\beta_{-1} \neq \beta_1$, then

$$T_j = (\beta_1 - \beta_{-1}) h y'''(x_j) + Z_j^{(1)} h^2\,,$$

where $|Z_j^{(1)}| \le [\frac{1}{2}(|\beta_{-1}| + |\beta_1|) + \frac{1}{12}] M_4$;

(iii) if $\beta_{-1} + \beta_0 + \beta_1 = 1$, $\beta_1 = \beta_{-1}$ and $\beta_1 \neq \frac{1}{12}$, then

$$T_j = (\beta_1 - \tfrac{1}{12}) h^2 y^{iv}(x_j) + Z_j^{(2)} h^3\,,$$

where $|Z_j^{(2)}| \le [\frac{1}{12}|\beta_1| + \frac{1}{360}] M_6$;

(iv) if $\beta_{-1} = \beta_1 = \frac{1}{12}$ and $\beta_0 = \frac{5}{6}$, then

$$T_j = \tfrac{1}{240} h^4 y^{vi}(x_j) + Z_j^{(4)} h^6\,,$$

where $|Z_j^{(4)}| \le \frac{1}{60480} M_8$.

13.4 The approximation of Exercise 3 is used, with the values $\beta_1 = \beta_{-1} = 1/12$, $\beta_0 = 5/6$. Use Taylor's Theorem with integral remainder (Appendix, Theorem A.5) to show that the truncation error of this approximation may be written

$$T_j = \int_{-h}^{h} G(s) y^{vi}(x_j + s)\,\mathrm{d}s\,,$$

where

$$G(s) = (h - s)^5/5! - \tfrac{1}{12} h^2 (h - s)^3/3!\,, \qquad 0 \le s \le h\,,$$

with a similar expression for $-h \le s \le 0$. Show that $G(s) \le 0$ for all $s \in [-h, h]$, and hence use the Integral Mean Value Theorem to show that the truncation error can be expressed as

$$T_j = \frac{h^6}{240} y^{vi}(\xi)$$

for some value of ξ in $(x_j - h, x_j + h)$.

13.5 Suppose that the solution of (13.1), (13.2) has a continuous sixth derivative on $[a,b]$, and that Y_j is the solution of the approximation used in Exercise 4. Show that

$$|y(x_j) - Y_j| \le \tfrac{1}{2880} h^4 (b-a)^2 M_6\,, \qquad j = 0, \ldots, n\,,$$

provided that

$$h^2 r(x_j) \le 12\,, \qquad j = 1, \ldots, n-1\,.$$

13.6 Complete the proof of Theorem 13.7.

13.7 Show that the solution of the boundary value problem

$$-y'' + a^2 y = 0\,, \quad y(-1) = 1\,, \qquad y(1) = 1\,,$$

is

$$y(x) = \frac{\cosh ax}{\cosh a}\,.$$

Use the identity

$$\cosh(x+h) + \cosh(x-h) = 2\cosh x\, \cosh h$$

to verify that the solution of the difference approximation (13.5) to this problem is

$$Y_j = \frac{\cosh \vartheta x_j}{\cosh \vartheta}\,,$$

where

$$\vartheta = (1/h)\cosh^{-1}(1 + \tfrac{1}{2}a^2 h^2)\,.$$

By expanding in Taylor series, show that

$$Y_j = y(x_j) + \tfrac{1}{24} h^2 a^3 (\cosh ax \sinh a - x \sinh ax \cosh a)/(\cosh a)^2 \\ + \mathcal{O}(h^4)\,.$$

Verify that this result is consistent with Theorem 13.4 when h is small.

13.8 Carry out a similar analysis as in Exercise 7 for the boundary value problem

$$-y'' - a^2 y = 0\,, \qquad y(0) = 0\,, \quad y(1) = 1\,,$$

and explain why in this case Theorem 13.4 cannot be used. What restriction is required on the value of a?

13.9 The eigenvalue problem

$$-y'' = \lambda y\,, \qquad y(0) = y(1) = 0\,,$$

is approximated by

$$-\frac{Y_{j+1} - 2Y_j + Y_{j-1}}{h^2} = \mu Y_j\,,\ 1 \le j \le n-1\,, \quad Y_0 = Y_n = 0\,.$$

Show that the differential equation has solution $y = \sin m\pi x$, $\lambda = m^2\pi^2$ for any positive integer m. Show also that the difference approximation has solution $Y_j = \sin m\pi x_j,\ j = 0, 1, \ldots, n$, and give an expression for the corresponding value of μ. Use the fact that

$$1 - \cos\vartheta = \tfrac{1}{2}\vartheta^2 - \tfrac{1}{24}\xi\vartheta^4, \quad |\xi| \le 1\,,$$

to show that $|\lambda - \mu| \le m^4\pi^4 h^2/12$, and compare with the bound given by (13.23).

https://doi.org/10.1017/CBO9780511801181.014 Published online by Cambridge University Press