Chapter 10

Solution of Ordinary Differential Equations

10.1 Introduction

While, historically, finite-difference methods have been and remain the standard numerical technique for solving ordinary differential equations, newer alternative methods can be more effective in certain contexts. In particular we consider here methods founded on orthogonal expansions—the so-called *spectral* and *pseudospectral* methods—with special reference to methods based on expansions in Chebyshev polynomials.

In a typical finite-difference method, the unknown function $u(x)$ is represented by a table of numbers $\{y_0, y_1, \ldots, y_n\}$ approximating its values at a set of discrete points $\{x_0, x_1, \ldots, x_n\}$, so that $y_j \approx u(x_j)$. (The points are almost always equally spaced through the range of integration, so that $x_{j+1} - x_j = h$ for some small fixed h.)

In a spectral method, in contrast, the function $u(x)$ is represented by an infinite expansion $u(x) = \sum_k c_k \phi_k(x)$, where $\{\phi_k\}$ is a chosen sequence of prescribed basis functions. One then proceeds somehow to estimate as many as possible of the coefficients $\{c_k\}$, thus approximating $u(x)$ by a finite sum such as

$$u_n(x) = \sum_{k=0}^{n} c_k \phi_k(x). \tag{10.1}$$

One clear advantage that spectral methods have over finite-difference methods is that, once approximate spectral coefficients have been found, the approximate solution can immediately be evaluated at any point in the range of integration, whereas to evaluate a finite-difference solution at an intermediate point requires a further step of interpolation.

A pseudospectral method, at least according to some writers, is one in which $u(x)$ is still approximated by a function of the form $u_n(x)$ of (10.1), as in a spectral method, but this approximation is actually represented not by its coefficients but by its values $u_n(x_j)$ at a number ($n+1$ in this particular instance) of discrete points $\{x_j\}$. These points may be equally spaced, but equal spacing gives no advantages, and other spacings are frequently better.

The oldest and probably the most familiar spectral methods are based on the idea of Fourier series. Supposing for the moment, for convenience, that the independent variable x is confined to the interval $-\pi \le x \le \pi$, the technique is to assume that the unknown function has an expansion in the form

$$u(x) = \tfrac{1}{2}a_0 + \sum_{k=1}^{\infty}\{a_k \cos kx + b_k \sin kx\} \tag{10.2}$$

and to attempt to determine values for the coefficients $\{a_k, b_k\}$ such that the required differential equation and other conditions are satisfied.

Fourier methods may well be suitable when the problem is inherently periodic; for instance where the function $u(x)$ satisfies a second-order differential equation subject to the periodicity boundary conditions $u(\pi) = u(-\pi)$ and $u'(\pi) = u'(-\pi)$. If we have a second-order differential equation with the more usual boundary conditions $u(-\pi) = a$ and $u(\pi) = b$, however, with $a \neq b$, then obviously any finite partial sum

$$\tfrac{1}{2}a_0 + \sum_{k=1}^{n}\{a_k \cos kx + b_k \sin kx\}$$

of (10.2) is periodic and cannot satisfy both boundary conditions simultaneously; more importantly, very many terms are needed if this partial sum is to represent the function $u(x)$ at all closely near both ends of the range at the same time. It is better in such a case to take for $\{\phi_k\}$ a sequence of polynomials, so that the partial sum

$$\sum_{k=0}^{n} c_k \phi_k(x)$$

is a polynomial of degree n.

From now on, we shall suppose that the independent variable x is confined to the interval $-1 \leq x \leq 1$, so that a reasonable choice for $\phi_k(x)$ is the Chebyshev polynomial $T_k(x)$. The choice of basis is only the beginning of the story, however; we are still left with the task of determining the coefficients $\{c_k\}$.

10.2 A simple example

For the purposes of illustration, we shall consider the simple linear two-point boundary-value problem on the range $[-1, 1]$:

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}u(x) = f(x), \quad u(-1) = a, \quad u(+1) = b, \tag{10.3}$$

where the function f and the boundary values a and b are given.

We can start out in several ways, of which the two following are the simplest:

- We may write $u(x)$ directly in the form

$$u(x) = \sum_{k=0}^{\infty}{}' c_k T_k(x). \tag{10.4}$$

Using the result quoted in Problem 16 of Chapter 2, namely

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} T_k(x) = \sum_{\substack{r=0 \\ (k-r) \text{ even}}}^{k-2}{}' (k-r)k(k+r)T_r(x), \ (k \geq 2), \tag{10.5}$$

we express (10.3) in the form

$$\sum_{k=2}^{\infty} \sum_{\substack{r=0 \\ (k-r) \text{ even}}}^{k-2}{}' (k-r)k(k+r)c_k T_r(x) = f(x), \tag{10.6a}$$

$$\sum_{k=0}^{\infty}{}' (-1)^k c_k = a, \qquad \sum_{k=0}^{\infty}{}' c_k = b. \tag{10.6b}$$

- An alternative procedure, incorporating the boundary conditions in the representation itself, is to write $u(x)$ in the form

$$u(x) = (1-x^2) \sum_{k=0}^{\infty} \gamma_k U_k(x) + \frac{1-x}{2} a + \frac{1+x}{2} b. \tag{10.7}$$

Then, using a result given from Problem 12 of Chapter 3,

$$(1-x^2)U_k(x) = \tfrac{1}{2}(T_k(x) - T_{k+2}(x)),$$

together with (10.5), we get (10.3) in a single equation of the form

$$\tfrac{1}{2} \sum_{k=2}^{\infty} \sum_{\substack{r=0 \\ (k-r) \text{ even}}}^{k-2}{}' (k-r)k(k+r)(\gamma_k - \gamma_{k-2})T_r(x) = f(x). \tag{10.8}$$

We should be treading on dangerous ground, however, if we went ahead blindly with translating either of the above infinite systems of equations (10.6) or (10.8) into an algorithm, since non-trivial questions arise relating to convergence of the infinite series and the validity of differentiating term by term. In any event, since one can never calculate the whole of an infinite sequence of coefficients, the realistic approach is to accept the fact that we must perforce approximate $u(x)$ by a finite sum of terms, and to go on from there.

If we truncate the summation (10.4) or (10.7) after a finite number of terms, we obviously cannot in general satisfy (10.6a) or (10.8) throughout the range $-1 \leq x \leq 1$. We can, however, attempt to satisfy either equation approximately in some sense. We shall discuss two ways of doing this: *collocation* methods and *projection* or *tau* (τ) methods.

10.2.1 Collocation methods

Suppose that we approximate $u(x)$ by

$$u_n(x) := \sum_{k=0}^{n}{}' c_k T_k(x), \tag{10.9}$$

involving $n+1$ unknown coefficients $\{c_k\}$, then we may select $n-1$ points $\{x_1, \ldots, x_{n-1}\}$ in the range of integration and require $u_n(x)$ to satisfy the differential equation (10.3) at just these $n-1$ points, the so-called *collocation points*, in addition to obeying the boundary conditions. This requires us to solve just the system of $n+1$ linear equations

$$\sum_{k=2}^{n} \sum_{\substack{r=0 \\ (k-r)\text{ even}}}^{k-2}{}' (k-r)k(k+r)c_k T_r(x_j) = f(x_j),$$

$$j = 1, \ldots, n-1, \tag{10.10a}$$

$$\sum_{k=0}^{n}{}' (-1)^k c_k = a, \tag{10.10b}$$

$$\sum_{k=0}^{n}{}' c_k = b, \tag{10.10c}$$

These equations may be reduced to a simpler form, especially if the $n-1$ points are carefully chosen so that we can exploit discrete orthogonality properties of the Chebyshev polynomials. Suppose that we choose for our collocation points the zeros of $T_{n-1}(x)$, namely

$$x_j = \cos \frac{(j-\frac{1}{2})\pi}{n-1}. \tag{10.11}$$

Multiply (10.10a) by $2T_\ell(x_j)$, where ℓ is an integer with $0 \le \ell \le n-2$, and sum from $j = 1$ to $j = n-1$. We can then use the discrete orthogonality relations (4.42) (for $0 \le r \le n-2$, $0 \le \ell \le n-2$)

$$\sum_{j=1}^{n-1} T_r(x_j)T_\ell(x_j) = \begin{cases} n-1, & r = \ell = 0, \\ \frac{1}{2}(n-1), & r = \ell \neq 0, \\ 0, & \text{otherwise} \end{cases} \tag{10.12}$$

to deduce that

$$\sum_{\substack{k=\ell+2 \\ (k-\ell)\text{ even}}}^{n} (k-\ell)k(k+\ell)c_k = \frac{2}{n-1}\sum_{j=1}^{n-1} T_\ell(x_j)f(x_j), \quad \ell = 0, \ldots, n-2. \tag{10.13}$$

The matrix of coefficients on the left-hand side of equation (10.13) is upper triangular, with elements

$$\begin{pmatrix}
8 & & 64 & & 216 & & 512 & & 1000 & \cdots \\
& 24 & & 120 & & 336 & & 720 & & \cdots \\
& & 48 & & 192 & & 480 & & 960 & \cdots \\
& & & 80 & & 280 & & 648 & & \cdots \\
& & & & 120 & & 384 & & 840 & \cdots \\
& & & & & 168 & & 504 & & \cdots \\
& & & & & & 224 & & 640 & \cdots \\
& & & & & & & 288 & & \cdots \\
& & & & & & & & 360 & \cdots \\
& & & & & & & & & \ddots
\end{pmatrix}.$$

We now have the following algorithm for generating an approximate solution of the problem (10.3) by collocation.

1. Find the collocation points $x_j = \cos\frac{(j-\frac{1}{2})\pi}{n-1}$, for $j = 1, \ldots, n-1$, and evaluate $f(x_j)$;
2. Use the recurrence (1.3) to evaluate $T_\ell(x_j)$, for $\ell = 0, \ldots, n-2$ and $j = 1, \ldots, n-1$;
3. Use equations (10.13), in reverse order, to determine the coefficients $c_n, \ldots, c_3, c_2$ one by one;
4. Use the boundary conditions (10.10b), (10.10c) to determine c_0 and c_1.

This algorithm can be made considerably more efficient in the case where $n-1$ is a large power of 2, as we can then use a technique derived from the fast Fourier transform algorithm to compute the right-hand sides of (10.13) in $O(n \log n)$ operations, without going through step 2 above which requires $O(n^2)$ operations (see Section 4.7).

Alternatively, we approximate $u(x)$ by

$$u_n(x) := (1-x^2)\sum_{k=0}^{n-2}\gamma_k U_k(x) + \frac{1-x}{2}a + \frac{1+x}{2}b, \tag{10.14}$$

involving $n-1$ unknown coefficients $\{\gamma_k\}$ and satisfying the boundary conditions automatically. With the same $n-1$ collocation points $\{x_1, \ldots, x_{n-1}\}$ we now solve the system of $n-1$ linear equations

$$\sum_{k=2}^{n}\sideset{}{'}\sum_{\substack{r=0\\(k-r)\text{ even}}}^{k-2} \tfrac{1}{2}(k-r)k(k+r)(\gamma_k - \gamma_{k-2})T_r(x_j) = f(x_j) \tag{10.15}$$

(with $\gamma_n = \gamma_{n-1} = 0$). If the collocation points are again taken as the zeros of $T_{n-1}(x)$, discrete orthogonality gives the equations

$$\sum_{\substack{k=\ell+2\\(k-\ell)\text{ even}}}^{n} \tfrac{1}{2}(k-\ell)k(k+\ell)(\gamma_k - \gamma_{k-2}) = \frac{2}{n-1}\sum_{j=1}^{n-1} T_\ell(x_j) f(x_j) \tag{10.16}$$

or, equivalently,

$$\sum_{\substack{k=\ell\\(k-\ell)\text{ even}}}^{n-2} (\ell^2 - 3k^2 - 6k - 4)\gamma_k = \frac{2}{n-1}\sum_{j=1}^{n-1} T_\ell(x_j) f(x_j), \quad \ell = 0, \ldots, n-2. \tag{10.17}$$

The matrix of coefficients on the left-hand side of equation (10.17) is upper triangular, with elements

$$\begin{pmatrix}
-4 & & -28 & & -76 & & -148 & & -244 & \cdots \\
 & -12 & & -48 & & -108 & & -192 & & \cdots \\
 & & -24 & & -72 & & -144 & & -240 & \cdots \\
 & & & -40 & & -100 & & -184 & & \cdots \\
 & & & & -60 & & -132 & & -228 & \cdots \\
 & & & & & -84 & & -168 & & \cdots \\
 & & & & & & -112 & & -208 & \cdots \\
 & & & & & & & -144 & & \cdots \\
 & & & & & & & & -180 & \cdots \\
 & & & & & & & & & \ddots
\end{pmatrix}.$$

We then have the following algorithm.

1. Find the collocation points $x_j = \cos\frac{(j-\frac{1}{2})\pi}{n-1}$, for $j = 1, \ldots, n-1$, and evaluate $f(x_j)$;
2. Use the recurrence (1.3) to evaluate $T_\ell(x_j)$, for $\ell = 0, \ldots, n-2$ and $j = 1, \ldots, n-1$;
3. Solve equations (10.17), in reverse order, to determine the coefficients $\gamma_{n-2}, \ldots, \gamma_1, \gamma_0$ one by one.

EXAMPLE 10.1: Taking the differential equation

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}u(x) + 6\,|x| = 0, \quad u(\pm 1) = 0, \tag{10.18}$$

whose known solution is $u(x) = 1 - x^2\,|x|$, and applying the above method with $n = 10$, we get the results in Table 10.1, where we show the values of the exact and approximate solutions at the collocation points. (Both solutions are necessarily even.)

Table 10.1: Solution of (10.18) by collocation, $n = 10$

x_j	$u(x_j)$	$u_n(x_j)$
0.0000	1.0000	0.9422
±0.3420	0.9600	0.9183
±0.6427	0.7344	0.7129
±0.8660	0.3505	0.3419
±0.9848	0.0449	0.0440

10.2.2 Error of the collocation method

We may analyse the error of the preceding collocation algorithms by the use of *backward error analysis*.

Here we shall look only at the first of the two alternative representations of $u(x)$. Let $u(x)$ denote the true solution to the problem (10.3), let

$$u_n(x) := \sum_{k=0}^{n}{}' c_k T_k(x)$$

be the approximate solution obtained by collocation and let $f_n(x)$ be the second derivative of $u_n(x)$. Then $u_n(x)$ is itself the exact solution to a similar problem

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} u_n(x) = f_n(x), \quad u_n(-1) = a, \quad u_n(+1) = b. \tag{10.19}$$

Since equations (10.3) and (10.19) are both linear, and have the same boundary conditions, the error $e(x) := u(x) - u_n(x)$ must be the solution to the homogeneous boundary-value problem

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} e(x) = \delta f(x), \quad e(-1) = e(+1) = 0, \tag{10.20}$$

where

$$\delta f(x) := f(x) - f_n(x). \tag{10.21}$$

We can write down the solution of (10.20) in integral form

$$e(x) = \int_{-1}^{1} G(x, \xi) \delta f(\xi) \, \mathrm{d}\xi, \tag{10.22}$$

where $G(x,\xi)$ is the *Green's function* for this problem

$$G(x,\xi) = \begin{cases} -\frac{1}{2}(1-x)(1+\xi), & \xi \le x, \\ -\frac{1}{2}(1+x)(1-\xi), & \xi \ge x. \end{cases} \tag{10.23}$$

Equation (10.22) may be used to derive bounds on the error $e(x)$ from various possible norms of the difference $\delta f(x)$. In particular:

$$|e(x)| \le \tfrac{1}{2}(1-x^2)\,\|\delta f\|_\infty\,, \tag{10.24a}$$

$$|e(x)| \le \frac{1}{\sqrt{6}}(1-x^2)\,\|\delta f\|_2\,, \tag{10.24b}$$

$$|e(x)| \le \tfrac{1}{2}(1-x^2)\,\|\delta f\|_1\,. \tag{10.24c}$$

We know that $u_n(x)$ is a polynomial of degree n in x, so that its second derivative $f_n(x)$ must be a polynomial of degree $n-2$. The collocation equations (10.10a), however, tell us that

$$f_n(x_j) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}u_n(x_j) = f(x_j),$$

so that $f_n(x)$ coincides with $f(x)$ at the $n-1$ points $\{x_1,\ldots,x_{n-1}\}$. Therefore $f_n(x)$ must be the unique $(n-2)$nd degree polynomial interpolating $f(x)$ at the zeros of $T_{n-1}(x)$, and $\|\delta f\|$ must be the corresponding interpolation error.

Now we may apply one of the standard formulae for interpolation error, for instance (Davis 1961, Chapter 3):

- [Hermite] Assuming that $f(x)$ has an analytic continuation into the complex plane, we have

$$\delta f(x) = f(x) - f_n(x) = \frac{1}{2\pi\mathrm{i}}\oint_C \frac{T_{n-1}(x)f(z)\,\mathrm{d}z}{T_{n-1}(z)(z-x)} \tag{10.25}$$

  where C is a closed contour in the complex plane, encircling the interval $[-1,1]$ but enclosing no singularities of $f(z)$.

- [Cauchy] Assuming alternatively that $f(x)$ has $(n-1)$ continuous derivatives, we have

$$\delta f(x) = f(x) - f_n(x) = \frac{T_{n-1}(x)}{2^n(n-1)!}f^{(n-1)}(\xi) \tag{10.26}$$

  for some real ξ in the interval $-1 \le \xi \le 1$.

To take (10.25) a little further, suppose that the analytic continuation $f(z)$ of $f(x)$ is regular on and within the ellipse E_r defined in Section 1.4.1, with foci at $z = \pm 1$. Then we know from (1.50) that

$$|T_{n-1}(z)| \ge \frac{1}{2}(r^{n-1} - r^{1-n})$$

for every z on E_r, while $|T_{n-1}(x)| \leq 1$ for every x in $[-1, 1]$. Therefore

$$|\delta f(x)| \leq \frac{1}{\pi(r^{n-1} - r^{1-n})} \oint_{E_r} \frac{|f(z)|}{|z - x|} \, |\,\mathrm{d}z| = O(r^{-n}) \text{ as } n \to \infty. \qquad (10.27)$$

Applying (10.24), we deduce that the collocation solution $u_n(x)$ converges exponentially to the exact solution $u(x)$ in this case, as the number $n + 1$ of terms and the number $n - 1$ of collocation points increase.

10.2.3 Projection (tau) methods

Just as collocation methods are seen to be related to approximation by interpolation, so there are methods that are related to approximation by least squares or, more generally, by projection.

- Approximating $u(x)$ by

  $$u_n(x) := {\sum_{k=0}^{n}}' c_k T_k(x),$$

  as in (10.9), suppose now that we select $n-1$ independent test functions $\{\psi_1(x), \ldots, \psi_{n-1}(x)\}$ and a positive weight function $w(x)$, and solve for the $n + 1$ coefficients $\{c_k\}$ the system of $n + 1$ linear equations

  $$\begin{aligned} &\int_{-1}^{1} w(x) \left\{ \frac{\mathrm{d}^2}{\mathrm{d}x^2} u_n(x) - f(x) \right\} \psi_\ell(x) \, \mathrm{d}x \\ &= \int_{-1}^{1} w(x) \left\{ {\sum_{k=2}^{n}}' \sum_{\substack{r=0 \\ (k-r) \text{ even}}}^{k-2} (k-r)k(k+r) c_k T_r(x) - f(x) \right\} \psi_\ell(x) \, \mathrm{d}x \\ &\qquad\qquad = 0, \quad \ell = 1, \ldots, n-1, \\ &{\sum_{k=0}^{n}}' (-1)^k c_k = a, \\ &{\sum_{k=0}^{n}}' c_k = b, \end{aligned} \qquad (10.28)$$

  so that u_n satisfies the boundary conditions and the residual

  $$\frac{\mathrm{d}^2}{\mathrm{d}x^2} u_n(x) - f(x)$$

  is orthogonal to each of the $n - 1$ test functions $\psi_1(x), \ldots \psi_{n-1}(x)$ with respect to the weight $w(x)$.

If we take $\psi_\ell(x) = T_{\ell-1}(x)$ and $w(x) = \frac{2}{\pi}(1-x^2)^{-1/2}$, this is equivalent to saying that the residual may be represented in the form

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}u_n(x) - f(x) = \sum_{k=n}^{\infty} \tau_{k-1}T_{k-1}(x),$$

for some sequence of undetermined coefficients $\{\tau_k\}$. The method is for this reason often referred to as the *tau method*[1] (Ortiz 1969), although differing slightly from Lanczos's original tau method (see Section 10.3 below), in which the approximation $u_n(x)$ was represented simply as a sum of powers of x

$$u_n(x) = \sum_{k=0}^{n} a_k x^k.$$

In our case, we can use the orthogonality relations (4.9), (4.11) to reduce the first $n-1$ of these equations to

$$\sum_{\substack{k=\ell+2\\(k-\ell)\text{ even}}}^{n} (k-\ell)k(k+\ell)c_k = \frac{2}{\pi}\int_{-1}^{1} \frac{T_\ell(x)f(x)}{\sqrt{1-x^2}}\,\mathrm{d}x, \quad \ell = 0,\ldots,n-2. \tag{10.29}$$

The similarities between equations (10.29) and (10.13) are no coincidence. In fact, the right-hand sides of (10.13) are just what we obtain when we apply the Gauss–Chebyshev quadrature rule 1 of Theorem 8.4 (page 183) to the integrals on the right-hand sides of (10.29).

If we use this rule for evaluating the integrals, therefore, we get precisely the same algorithm as in the collocation method; in many contexts we may, however, have a better option of evaluating the integrals more accurately—or even exactly.

- If we use the alternative approximation

$$u_n(x) := (1-x^2)\sum_{k=0}^{n-2} \gamma_k U_k(x) + \frac{1-x}{2}a + \frac{1+x}{2}b,$$

as in (10.14), we are similarly led to the equations

$$\sum_{\substack{k=\ell\\(k-\ell)\text{ even}}}^{n-2} (\ell^2 - 3k^2 - 6k - 4)\gamma_k = \frac{2}{\pi}\int_{-1}^{1} \frac{T_\ell(x)f(x)}{\sqrt{1-x^2}}\,\mathrm{d}x, \quad \ell = 0,\ldots,n-2, \tag{10.30}$$

and the same final remarks apply.

[1] Compare the tau method for approximating rational functions described in Section 3.6.

EXAMPLE 10.2: Taking same differential equation (10.18) as previously, and now applying the above projection method with $n = 10$, we get the results in Table 10.2, where for convenience we show the values of the exact and approximate solutions at same points as in Table 10.1. It will be seen that the results are slightly more accurate.

Table 10.2: Solution of (10.18) by projection, $n = 10$

x_j	$u(x_j)$	$u_n(x_j)$
0.0000	1.0000	1.0017
±0.3420	0.9600	0.9592
±0.6427	0.7344	0.7347
±0.8660	0.3505	0.3503
±0.9848	0.0449	0.0449

10.2.4 Error of the preceding projection method

We may carry out a backward error analysis just as we did for the collocation method in Section 10.2.2.

As before, let $u(x)$ denote the true solution to the problem (10.3) and $u_n(x)$ the approximate solution, let $f_n(x) = \mathrm{d}^2 u_n(x)/\,\mathrm{d}x^2$ and $\delta f(x) = f(x) - f_n(x)$. Then the error bounds (10.24) still apply.

Assume now that the integrals in (10.29) are evaluated exactly. The function $f_n(x)$ will again be a polynomial of degree $n-2$ but this time, instead of interpolating $f(x)$ at collocation points, it is determined by the integral relations

$$\int_{-1}^{1} \frac{T_\ell(x)\delta f(x)}{\sqrt{1-x^2}}\,\mathrm{d}x = 0, \quad \ell = 0, \ldots, n-2. \tag{10.31}$$

In other words, $f_n(x)$ is a weighted least-squares approximation to $f(x)$ with respect to the weight $w(x) = (1-x^2)^{-1/2}$; that is to say, it is the truncated Chebyshev series expansion

$$f_n(x) = {\sum_{k=0}^{n-2}}' d_k T_k(x)$$

of $f(x)$.

Then, for instance, applying the results of Section 5.7, we can say that if the analytic continuation $f(z)$ of $f(x)$ is regular on and within the ellipse E_r then

$$|\delta f(x)| = \left| f(x) - (S_{n-2}^T f)(x) \right| \le \frac{M}{r^{n-2}(r-1)}, \tag{10.32}$$

so that again

$$|\delta f(x)| = O(r^{-n}) \text{ as } n \to \infty, \tag{10.33}$$

and the projection solution converges as n increases at the same rate as the collocation solution based on the zeros of T_{n-1}.

10.3 The original Lanczos tau (τ) method

The original 'tau method' for ordinary differential equations as described by Lanczos (1938) approximated the unknown function by an ordinary polynomial rather than by a truncated Chebyshev expansion — Chebyshev polynomials made their appearance only in the residual. We illustrate this by a simple example.

EXAMPLE 10.3: Consider the equation

$$\frac{\mathrm{d}u}{\mathrm{d}x} + 4xu = 0 \tag{10.34}$$

on the interval $-1 \le x \le 1$, with the condition $u(0) = 1$, to which the solution is

$$u(x) = \mathrm{e}^{-2x^2}.$$

If we try the approximation

$$u_6(x) = 1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + a_6 x^6, \tag{10.35}$$

which satisfies the given condition, we come up with the residual

$$\begin{aligned} \frac{\mathrm{d}u_6}{\mathrm{d}x} + 4xu_6 &= a_1 + (4 + 2a_2)x + (4a_1 + 3a_3)x^2 + (4a_2 + 4a_4)x^3 + \\ &\quad + (4a_3 + 5a_5)x^4 + (4a_4 + 6a_6)x^5 + 4a_5 x^6 + 4a_6 x^7. \end{aligned}$$

The conventional technique (Frobenius method) for dealing with this residual would be to ignore the last two terms (the highest powers of x), and to equate the remaining terms to zero. This gives

$$a_2 = -2, \quad a_4 = 2, \quad a_6 = -\tfrac{4}{3}, \tag{10.36}$$

with all the odd-order coefficients vanishing.

Lanczos's approach in the same case would have been to equate the residual to

$$\tau_6 T_6(x) + \tau_7 T_7(x).$$

This gives $\tau_6 = 0$, and the odd-order coefficients again vanishing, while

$$\tau_7 = -\frac{4}{139}, \quad a_2 = -\frac{264}{139}, \quad a_4 = \frac{208}{139}, \quad a_6 = -\frac{64}{139}. \tag{10.37}$$

Thus the conventional approach gives the approximate solution

$$u_6(x) = 1 - 2x^2 + 2x^4 - \tfrac{4}{3}x^6 \tag{10.38}$$

while Lanczos's method gives

$$u_6(x) = 1 - \frac{264x^2 - 208x^4 + 64x^6}{139}. \tag{10.39}$$

The improvement is clear—compare Figures 10.1 and 10.2.

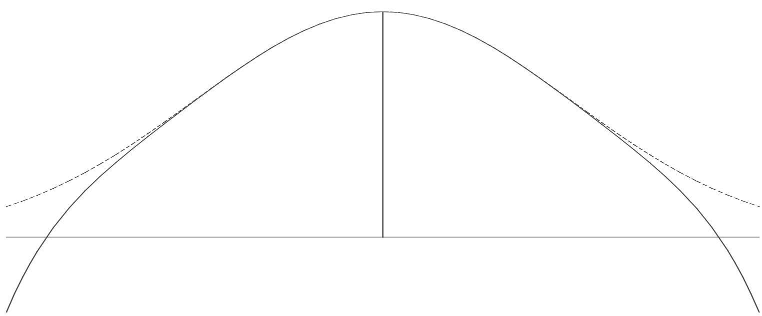

Figure 10.1: Power series solution (10.38) compared with true solution on $[-1, 1]$

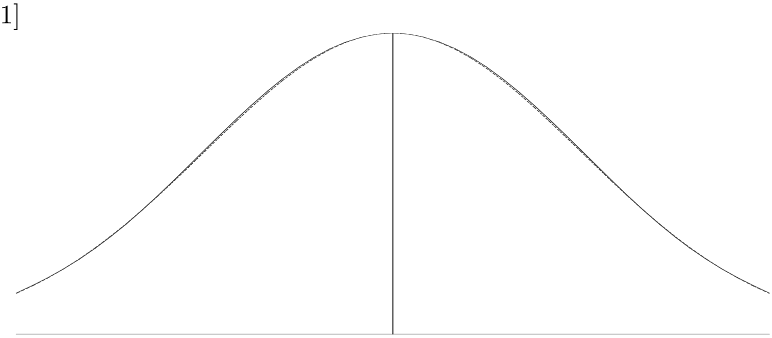

Figure 10.2: Lanczos tau solution (10.39) compared with true solution on $[-1, 1]$

10.4 A more general linear equation

The methods used to attack the simple equation of Section 10.2 may be applied with little alteration to the general linear two-point boundary-value problem

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}u(x) + q(x)\frac{\mathrm{d}}{\mathrm{d}x}u(x) + r(x)u(x) = f(x), \quad u(-1) = a, \quad u(+1) = b, \tag{10.40}$$

where $q(x)$ and $r(x)$ are given continuous functions of x.

Approximating $u(x)$ again by the finite sum (10.9)

$$u_n(x) = \sum_{k=0}^{n}{}' c_k T_k(x),$$

using the formula (10.5)

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}T_k(x) = \sum_{\substack{r=0 \\ (k-r) \text{ even}}}^{k-2}{}' (k-r)k(k+r)T_r(x), \ (k \geq 2),$$

for $\frac{\mathrm{d}^2}{\mathrm{d}x^2}T_k(x)$ and the formula (2.49)

$$\frac{\mathrm{d}}{\mathrm{d}x}T_k(x) = \sum_{\substack{r=0 \\ (k-r) \text{ odd}}}^{k-1}{}' 2kT_r(x), \ (k \geq 1)$$

for $\frac{\mathrm{d}}{\mathrm{d}x}T_k(x)$, we get linear equations similar to (10.6) but with the first equation (10.6a) replaced by

$$\begin{aligned} &\sum_{k=2}^{n} \sum_{\substack{r=0 \\ (k-r) \text{ even}}}^{k-2}{}' (k-r)k(k+r)c_k T_r(x) + \\ &\qquad + q(x) \sum_{k=1}^{n} \sum_{\substack{r=0 \\ (k-r) \text{ odd}}}^{k-1}{}' 2kc_k T_r(x) + \\ &\qquad\qquad + r(x) \sum_{k=0}^{n}{}' c_k T_k(x) = f(x). \end{aligned} \tag{10.41}$$

10.4.1 Collocation method

If we substitute $x = x_1, \ldots, x = x_{n-1}$ in (10.41), we again get a system of linear equations for the coefficients $c_0, \ldots, c_n$, which we can go on to solve directly.

If $q(x)$ and $r(x)$ are polynomials in x, and therefore can be expressed as sums of Chebyshev polynomials, we can go on to use the multiplication formula (2.38) or that quoted in Problem 4 of Chapter 2 to reduce the products $q(x)T_r(x)$ and $r(x)T_k(x)$ in (10.41) to simple sums of Chebyshev polynomials $T_k(x)$. We can then use discrete orthogonality as before to simplify the equations to some extent.

Whether this is possible or not, however, collocation methods for linear problems are straightforward to apply.

It should be noted that the error analysis in Section 10.2.2 does not extend to this general case. The reason why it breaks down is that, where previously we could say that $f_n(x)$ was an interpolating polynomial of degree $n-2$, we now have the more complicated expression

$$f_n(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}u_n(x) + q(x)\frac{\mathrm{d}}{\mathrm{d}x}u_n(x) + r(x)u_n(x).$$

We can therefore no longer appeal to formulae for polynomial interpolation error.

10.4.2 Projection method

If $q(x)$ and $r(x)$ are polynomials in x, and we can therefore proceed as in the collocation method to reduce $q(x)T_r(x)$ and $r(x)T_k(x)$ in (10.41) to simple sums of Chebyshev polynomials $T_k(x)$, then we can use integral orthogonality relations (multiplying by $\frac{2}{\pi}(1-x^2)^{-\frac{1}{2}}T_\ell(x)$ and integrating, for $\ell = 0, \ldots, n-2$) to derive a set of linear equations to solve for the coefficients $\{c_k\}$.

In more general circumstances, however, we may need either to approximate $q(x)$ and $r(x)$ by polynomials or to estimate the integrals

$$\int_{-1}^{1} \frac{T_l(x)q(x)T_k(x)}{\sqrt{1-x^2}}\,\mathrm{d}x, \qquad \int_{-1}^{1} \frac{T_l(x)r(x)T_k(x)}{\sqrt{1-x^2}}\,\mathrm{d}x$$

numerically.

10.5 Pseudospectral methods — another form of collocation

In the collocation method we discussed earlier in Section 10.2.1, we approximated $u(x)$ by the truncated Chebyshev expansion $u_n(x)$ of (10.9), an nth degree polynomial. Instead of representing this polynomial by its $n+1$ Chebyshev coefficients, suppose now that we represent it by its values at the two boundary points (x_0 and x_n) and at $n-1$ internal collocation points (x_1, $\ldots$, x_{n-1}); these $n+1$ values are exactly sufficient to define the polynomial uniquely. According to some writers, the coefficients yield a spectral and the values a pseudospectral representation of the polynomial. To make use of such a representation, we need formulae for the derivatives of such a polynomial in terms of these values.

10.5.1 Differentiation matrices

Suppose that we know the values of any nth degree polynomial $p(x)$ at $n+1$ points $x_0,\ldots,x_n$. Then these values determine the polynomial uniquely, and so determine the values of the derivatives $p'(x) = \mathrm{d}p(x)/\,\mathrm{d}x$ at the same $n+1$ points. Each such derivative can, in fact, be expressed as a fixed linear combination of the given function values, and the whole relationship written in matrix form:

$$\begin{pmatrix} p'(x_0) \\ \vdots \\ p'(x_n) \end{pmatrix} = \begin{pmatrix} d_{0,0} & \cdots & d_{0,n} \\ \vdots & \ddots & \vdots \\ d_{n,0} & \cdots & d_{n,n} \end{pmatrix} \begin{pmatrix} p(x_0) \\ \vdots \\ p(x_n) \end{pmatrix}. \tag{10.42}$$

We shall call $\mathbf{D} = \{d_{j,k}\}$ a *differentiation matrix*.

Suppose now that the points x_j are the $n+1$ zeros of some $(n+1)$st degree polynomial $P_{n+1}(x)$.

If for $k = 0,\ldots,n$ we let $p_k(x) = P_{n+1}(x)/(x - x_k)$, which is an nth degree polynomial since x_k is a zero of P_{n+1}, then we can show without much difficulty that

$$\begin{aligned} p_k(x_k) &= P'_{n+1}(x_k) \\ p_k(x_j) &= 0, \quad j \neq k \\ p'_k(x_k) &= \frac{1}{2}P''_{n+1}(x_k) \\ p'_k(x_j) &= \frac{P'_{n+1}(x_j)}{x_j - x_k}, \quad j \neq k. \end{aligned}$$

From this we can deduce (by setting $p(x) = p_k(x)$ in (10.42)) that the kth column of the differentiation matrix $\mathbf{D}$ must have elements

$$d_{k,k} = \frac{1}{2}\frac{P''_{n+1}(x_k)}{P'_{n+1}(x_k)}, \tag{10.43a}$$

$$d_{j,k} = \frac{P'_{n+1}(x_j)}{(x_j - x_k)P'_{n+1}(x_k)}, \quad j \neq k. \tag{10.43b}$$

Notice that if $p(x_0) = p(x_1) = \cdots = p(x_n) = 1$ then we must have $p(x) \equiv 1$ and $p'(x) \equiv 0$. It follows that each row of the matrix $\mathbf{D}$ must sum to zero, so that the matrix is singular, although it may not be easy to see this directly by looking at its elements.

Not only can we use the relationship

$$\begin{pmatrix} p'(x_0) \\ \vdots \\ p'(x_n) \end{pmatrix} = \mathbf{D} \begin{pmatrix} p(x_0) \\ \vdots \\ p(x_n) \end{pmatrix}$$

to connect the first derivatives with the function values, but we can repeat the process (since p' is an $(n-1)$st degree polynomial, which may be regarded as an nth degree polynomial with zero leading coefficient), to give us a similar relationship for the second derivatives,

$$\begin{pmatrix} p''(x_0) \\ \vdots \\ p'(x_n) \end{pmatrix} = \mathbf{D}^2 \begin{pmatrix} p(x_0) \\ \vdots \\ p(x_n) \end{pmatrix},$$

and so on.

10.5.2 Differentiation matrix for Chebyshev points

In particular, suppose that the $n+1$ points are the points $y_k = \cos\frac{k\pi}{n}$, which are the zeros of the polynomial $P_{n+1}(x) = (1-x^2)U_{n-1}(x)$ and the extrema in $[-1,1]$ of $T_n(x)$. Making the usual substitution $x = \cos\theta$ gives us

$$P_{n+1}(x) = \sin\theta\ \sin n\theta.$$

Differentiating with respect to x, we then have

$$P'_{n+1}(x) = -\frac{\cos\theta\ \sin n\theta + n\sin\theta\ \cos n\theta}{\sin\theta} \tag{10.44}$$

and

$$P''_{n+1}(x) = -\frac{\cos^2\theta\ \sin n\theta - n\sin\theta\ \cos\theta\ \cos n\theta + (1+n^2)\sin^2\theta\ \sin n\theta}{\sin^3\theta}. \tag{10.45}$$

However, we know that if $\theta_k = \frac{k\pi}{n}$ then $\sin n\theta_k = 0$ and $\cos n\theta_k = (-1)^k$. Therefore (10.44) gives us

$$P'_{n+1}(y_k) = \begin{cases} -(-1)^k n, & 0 < k < n, \\ -2n, & k = 0, \\ -2(-1)^n n, & k = n, \end{cases} \tag{10.46}$$

and (10.45) gives

$$P''_{n+1}(y_k) = \begin{cases} (-1)^k n\dfrac{y_k}{1-y_k^2}, & 0 < k < n, \\[2ex] -2n\dfrac{1+2n^2}{3}, & k = 0, \\[2ex] 2(-1)^n n\dfrac{1+2n^2}{3}, & k = n. \end{cases} \tag{10.47}$$

(In each case, the values for $k = 0$ and $k = n$ are obtained by proceeding carefully to the limit as $x \to 1$ and $x \to -1$, respectively.)

Substituting (10.46) and (10.47) in (10.43a) and (10.43b) gives the following as elements of the differentiation matrix $\mathbf{D}$:

$$
\begin{aligned}
d_{j,k} &= \frac{(-1)^{k-j}}{y_j - y_k}, \quad 0 < j \neq k < n, & d_{k,k} &= -\tfrac{1}{2}\frac{y_k}{1 - y_k^2}, \quad 0 < k < n, \\
d_{0,0} &= \tfrac{1}{6}(1 + 2n^2), & d_{n,n} &= -\tfrac{1}{6}(1 + 2n^2), \\
d_{0,k} &= 2\frac{(-1)^k}{1 - y_k}, \quad 0 < k < n, & d_{k,0} &= -\tfrac{1}{2}\frac{(-1)^k}{1 - y_k}, \quad 0 < k < n, \\
d_{k,n} &= \tfrac{1}{2}\frac{(-1)^{n-k}}{1 + y_k}, \quad 0 < k < n, & d_{n,k} &= -2\frac{(-1)^{n-k}}{1 + y_k}, \quad 0 < k < n, \\
d_{0,n} &= \tfrac{1}{2}(-1)^n, & d_{n,0} &= -\tfrac{1}{2}(-1)^n.
\end{aligned}
\tag{10.48}
$$

That is to say, we have $\mathbf{D} =$

$$
\begin{pmatrix}
\frac{1}{6}(1+2n^2) & -2\frac{1}{1-y_1} & 2\frac{1}{1-y_2} & \cdots & 2\frac{(-1)^{n-1}}{1-y_{n-1}} & \frac{1}{2}(-1)^n \\
\frac{1}{2}\frac{1}{1-y_1} & -\frac{1}{2}\frac{y_1}{1-y_1^2} & -\frac{1}{y_1-y_2} & \cdots & \frac{(-1)^{n-2}}{y_1-y_{n-1}} & \frac{1}{2}\frac{(-1)^{n-1}}{1+y_1} \\
-\frac{1}{2}\frac{1}{1-y_2} & -\frac{1}{y_2-y_1} & -\frac{1}{2}\frac{y_2}{1-y_2^2} & \cdots & \frac{(-1)^{n-3}}{y_2-y_{n-1}} & \frac{1}{2}\frac{(-1)^{n-2}}{1+y_2} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
-\frac{1}{2}\frac{(-1)^{n-1}}{1-y_{n-1}} & \frac{(-1)^{n-2}}{y_{n-1}-y_1} & \frac{(-1)^{n-3}}{y_{n-1}-y_2} & \cdots & -\frac{1}{2}\frac{y_{n-1}}{1-y_{n-1}^2} & -\frac{1}{2}\frac{1}{1+y_{n-1}} \\
-\frac{1}{2}(-1)^n & -2\frac{(-1)^{n-1}}{1+y_1} & -2\frac{(-1)^{n-2}}{1+y_2} & \cdots & 2\frac{1}{1+y_{n-1}} & -\frac{1}{6}(1+2n^2)
\end{pmatrix}.
\tag{10.49}
$$

For instance

$n = 1$ $(y_k = 1, -1)$

$$
\mathbf{D} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}, \quad \mathbf{D}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix};
$$

$n = 2\ \ (y_k = 1,\, 0,\, -1)$

$$\mathbf{D} = \begin{pmatrix} \frac{3}{2} & -2 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 2 & -\frac{3}{2} \end{pmatrix}, \quad \mathbf{D}^2 = \begin{pmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \\ 1 & -2 & 1 \end{pmatrix};$$

$n = 3\ \ (y_k = 1,\, \frac{1}{2},\, -\frac{1}{2},\, -1)$

$$\mathbf{D} = \begin{pmatrix} \frac{19}{6} & -4 & \frac{4}{3} & -\frac{1}{2} \\ 1 & -\frac{1}{3} & -1 & \frac{1}{3} \\ -\frac{1}{3} & 1 & \frac{1}{3} & -1 \\ \frac{1}{2} & -\frac{4}{3} & 4 & -\frac{19}{6} \end{pmatrix}, \quad \mathbf{D}^2 = \begin{pmatrix} \frac{16}{3} & -\frac{28}{3} & \frac{20}{3} & -\frac{8}{3} \\ \frac{10}{3} & -\frac{16}{3} & \frac{8}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{8}{3} & -\frac{16}{3} & \frac{10}{3} \\ -\frac{8}{3} & \frac{20}{3} & -\frac{28}{3} & \frac{16}{3} \end{pmatrix};$$

$n = 4\ \ (y_k = 1,\, 1/\sqrt{2},\, 0,\, -1/\sqrt{2},\, -1)$ (Fornberg 1996, p.164)[2]

$$\mathbf{D} = \begin{pmatrix} \frac{11}{2} & -4-2\sqrt{2} & 2 & -4+2\sqrt{2} & \frac{1}{2} \\ 1+1/\sqrt{2} & -1/\sqrt{2} & -\sqrt{2} & 1/\sqrt{2} & -1+1/\sqrt{2} \\ -\frac{1}{2} & \sqrt{2} & 0 & -\sqrt{2} & \frac{1}{2} \\ 1-1/\sqrt{2} & -1/\sqrt{2} & \sqrt{2} & 1/\sqrt{2} & -1-1/\sqrt{2} \\ -\frac{1}{2} & 4-2\sqrt{2} & -2 & 4+2\sqrt{2} & -\frac{11}{2} \end{pmatrix},$$

$$\mathbf{D}^2 = \begin{pmatrix} 17 & -20-6\sqrt{2} & 18 & -20+6\sqrt{2} & 5 \\ 5+3/\sqrt{2} & -14 & 6 & -2 & 5-3\sqrt{2} \\ -1 & 4 & -6 & 4 & -1 \\ 5-3/\sqrt{2} & -2 & 6 & -14 & 5+3\sqrt{2} \\ 5 & -20+6\sqrt{2} & 18 & -20-6\sqrt{2} & 17 \end{pmatrix}.$$

Notice that the matrices $\mathbf{D}$ and $\mathbf{D}^2$ are singular in each case, as expected.

10.5.3 Collocation using differentiation matrices

We return first to the simple example of (10.3)

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}u(x) = f(x), \quad u(-1) = a, \quad u(+1) = b, \tag{10.50}$$

[2]Fornberg numbers the nodes in the direction of increasing y_k—we have numbered them in order of increasing k and so of decreasing y_k.

and suppose that the collocation points $\{x_j\}$ are chosen so that $x_0 = +1$ and $x_n = -1$ (as will be the case if they are the Chebyshev points $\{y_j\}$ used in Section 10.5.2.)

We know that

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} u_n(x_j) = \sum_{k=0}^{n} (\mathbf{D}^2)_{j,k} u_n(x_k).$$

The collocation equations thus become

$$\sum_{k=0}^{n} (\mathbf{D}^2)_{j,k} u_n(x_k) = f(x_j), \; j = 1, \ldots, n-1, \quad u_n(x_n) = a, \quad u_n(x_0) = b. \tag{10.51}$$

Partition the matrices $\mathbf{D}$ and $\mathbf{D}^2$ as follows, by cutting off the first and last rows and columns:

$$\mathbf{D} = \left(\begin{array}{c|c|c} \cdot & \cdots & \cdot \\ \hline \mathbf{e}_0 & \mathbf{E} & \mathbf{e}_n \\ \hline \cdot & \cdots & \cdot \end{array} \right), \mathbf{D}^2 = \left(\begin{array}{c|c|c} \cdot & \cdots & \cdot \\ \hline \mathbf{e}_0^{(2)} & \mathbf{E}^{(2)} & \mathbf{e}_n^{(2)} \\ \hline \cdot & \cdots & \cdot \end{array} \right) \tag{10.52}$$

and let $\mathbf{u}$ and $\mathbf{f}$ denote the vectors

$$\mathbf{u} = \begin{pmatrix} u_n(x_1) \\ \vdots \\ u_n(x_{n-1}) \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_{n-1}) \end{pmatrix}.$$

The collocation equations (10.51) can then be written in matrix notation as

$$u_n(x_0)\mathbf{e}_0^{(2)} + \mathbf{E}^{(2)}\mathbf{u} + u_n(x_n)\mathbf{e}_n^{(2)} = \mathbf{f},$$

or

$$\mathbf{E}^{(2)}\mathbf{u} = \mathbf{f} - b\mathbf{e}_0^{(2)} - a\mathbf{e}_n^{(2)}, \tag{10.53}$$

since the values of $u_n(x_0)$ and $u_n(x_n)$ are given by the boundary conditions. We have only to solve (10.53) to obtain the remaining values of $u_n(x_k)$.

In order to obtain the corresponding Chebyshev coefficients, we can then make use of discrete orthogonality relationships, as in Section 6.3.2.

Similarly, in the case of the more general equation (10.40)

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}u(x) + q(x)\frac{\mathrm{d}}{\mathrm{d}x}u(x) + r(x)u(x) = f(x), \quad u(-1) = a, \quad u(+1) = b, \tag{10.54}$$

if we let $\mathbf{Q}$ and $\mathbf{R}$ denote diagonal matrices with elements $q(x_k)$ and $r(x_k)$ $(k = 1, \ldots, n-1)$, the collocation equations can be written as

$$(\mathbf{E}^{(2)} + \mathbf{QE} + \mathbf{R})\mathbf{u} = \mathbf{f} - b(\mathbf{e}_0^{(2)} + \mathbf{Qe}_0) - a(\mathbf{e}_n^{(2)} + \mathbf{Qe}_0). \tag{10.55}$$

10.6 Nonlinear equations

We mention briefly that the techniques discussed in the preceding Sections can sometimes be extended to include nonlinear equations.

To take one simple example, using pseudospectral methods and following the principles of (10.51) and (10.52), the problem

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}u(x) = f(u(x)), \quad u(-1) = u(+1) = 0, \tag{10.56}$$

where $f(u)$ is an arbitrary function of u, gives rise to the system of equations

$$\sum_{k=0}^{n} (\mathbf{D}^2)_{j,k} u_n(x_k) = f(u_n(x_j)),\ j = 1, \ldots, n-1, \quad u_n(x_n) = u_n(x_0) = 0, \tag{10.57}$$

or, in matrix terms,

$$\mathbf{E}^{(2)}\mathbf{u} = \mathbf{f}(\mathbf{u}), \tag{10.58}$$

where $\mathbf{f}(\mathbf{u})$ denotes the vector with elements $\{f(u_n(x_j))\}$.

Equations (10.58) may or may not have a unique solution. If they do, or if we can identify the solution we require, then we may be able to approach it by an iterative procedure. For instance:

simple iteration Assume that we have a good guess $\mathbf{u}^{(0)}$ at the required solution of (10.58). Then we can generate the iterates $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}$, and so on, by solving successive sets of linear equations

$$\mathbf{E}^{(2)}\mathbf{u}^{(k)} = \mathbf{f}(\mathbf{u}^{(k-1)}), \quad k = 1, 2, \ldots. \tag{10.59}$$

Newton iteration Provided that $f(u)$ is differentiable, let $\mathbf{f}'(\mathbf{u})$ denote the diagonal matrix with elements $\{f'(u_n(x_j))\}$, and again assume that we have a good guess $\mathbf{u}^{(0)}$ at the required solution. Then generate successive corrections $(\mathbf{u}^{(1)} - \mathbf{u}^{(0)})$, $(\mathbf{u}^{(2)} - \mathbf{u}^{(1)})$, and so on, by solving successive sets of linear equations

$$\left(\mathbf{E}^{(2)} - \mathbf{f}'(\mathbf{u}^{(k-1)})\right)(\mathbf{u}^{(k)} - \mathbf{u}^{(k-1)}) = \mathbf{f}(\mathbf{u}^{(k-1)}) - \mathbf{E}^{(2)}\mathbf{u}^{(k-1)}, \qquad k = 1, 2, \ldots. \tag{10.60}$$

There is no general guarantee that either iteration (10.59) or (10.60) will converge to a solution — this needs to be studied on a case-by-case basis. If both converge, however, then the Newton iteration is generally to be preferred, since its rate of convergence is ultimately quadratic.

10.7 Eigenvalue problems

Similar techniques to those of the preceding sections may be applied to eigenvalue problems in ordinary differential equations. One would not, of course, think of applying them to the simplest such problem

$$\frac{d^2}{dx^2}u(x) + \lambda u(x) = 0, \quad u(\pm 1) = 0, \tag{10.61}$$

since its solutions

$$u(x) = \sin \tfrac{1}{2}k\pi(x+1), \quad \lambda = (\tfrac{1}{2}k\pi)^2, \quad k = 1, 2, \ldots, \tag{10.62}$$

can be written down analytically. This problem nevertheless has its uses as a test case.

Less trivial is the slightly more general problem

$$\frac{d^2}{dx^2}u(x) + q(x)u(x) + \lambda u(x) = 0, \quad u(\pm 1) = 0, \tag{10.63}$$

where $q(x)$ is some prescribed function of x.

10.7.1 Collocation methods

If we approximate $u(x)$ in the form (10.9)

$$u_n(x) := \sum_{k=0}^{n}{}' c_k T_k(x), \tag{10.64}$$

and select the zeros (10.11) of $T_{n-1}(x)$

$$x_j = \cos \frac{(j - \frac{1}{2})\pi}{n-1}, \quad j = 1, \ldots, n-1, \tag{10.65}$$

as collocation points, then the collocation equations for (10.63) become

$$\sum_{k=2}^{n} \sum_{\substack{r=0 \\ (k-r)\text{ even}}}^{k-2}{}' (k-r)k(k+r)c_k T_r(x_j) +$$

$$+ (q(x_j) + \lambda) \sum_{k=0}^{n}{}' c_k T_k(x_j) = 0,$$

$$j = 1, \ldots, n-1, \tag{10.66a}$$

$$\sum_{k=0}^{n}{}' (-1)^k c_k = 0, \tag{10.66b}$$

$$\sum_{k=0}^{n}{}' c_k = 0. \tag{10.66c}$$

Using discrete orthogonality wherever possible, as before, (10.66a) gives us the equations

$$\sum_{\substack{k=\ell+2 \\ (k-\ell) \text{ even}}}^{n} (k-\ell)k(k+\ell)c_k + \frac{2}{n-1}\sum_{j=1}^{n-1}{\sum_{k=0}^{n}}' q(x_j)T_\ell(x_j)T_k(x_j)c_k + \\ + \lambda c_\ell = 0, \\ \ell = 0, \ldots, n-2. \tag{10.67}$$

Using (10.66b) and (10.66c), we can make the substitutions

$$c_n = -{\sum_{\substack{k=0 \\ (n-k) \text{ even}}}^{n-2}}' c_k, \tag{10.68a}$$

$$c_{n-1} = -{\sum_{\substack{k=0 \\ (n-k) \text{ odd}}}^{n-3}}' c_k \tag{10.68b}$$

to reduce (10.67) to the form of a standard matrix eigenvalue equation of the form

$$\mathbf{A}\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-2} \end{pmatrix} = \lambda \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-2} \end{pmatrix}, \tag{10.69}$$

which we may solve by standard algebraic techniques.

This will yield $n-1$ matrix eigenvalues $\{\lambda^{(j)}\}$, with corresponding eigenvectors $\{\mathbf{c}^{(j)}\}$. These eigenvalues should approximate the $n-1$ dominant eigenvalues of (10.63), with

$${\sum_{k=0}}' nc_k^{(j)}T_k(x)$$

approximating the eigenfunction $u^{(j)}(x)$ corresponding to the eigenvalue approximated by $\lambda^{(j)}$. (Equations (10.68) are used to obtain the coefficients $c_n^{(j)}$ and $c_{n-1}^{(j)}$.)

We illustrate this with the trivial example (10.61) in which $q(x) \equiv 0$, when (10.67) becomes

$$\sum_{\substack{k=\ell+2 \\ (k-\ell) \text{ even}}}^{n} (k-\ell)k(k+\ell)c_k + \lambda c_\ell = 0, \\ \ell = 0, \ldots, n-2. \tag{10.70}$$

Table 10.3: Eigenvalues of (10.70) and differential equation (10.61)

Matrix eigenvalues	O.D.E. eigenvalues
2.4674	2.4674
9.8696	9.8696
22.2069	22.2066
39.6873	39.4784
62.5951	61.6850
119.0980	88.8624
178.5341	120.9026
1991.3451	157.9136
3034.1964	199.8595

Results for $n = 10$ are shown in Table 10.3.

Note, incidentally, that if $q(x)$ is an even function of x, so that

$$\sum_{j=1}^{n-1} q(x_j) T_\ell(x_j) T_k(x_j) = 0, \quad k - l \text{ even},$$

we can separate the odd-order coefficients c_k from the even-order ones, thus halving the dimensions of the matrices that we have to deal with.

10.7.2 Collocation using the differentiation matrix

The methods of Section 10.5.3 may be applied equally well to eigenvalue problems. For instance, the problem posed in (10.63)

$$\frac{d^2}{dx^2} u(x) + q(x) u(x) + \lambda u(x) = 0, \quad u(\pm 1) = 0,$$

is the same as (10.54)

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} u(x) + q(x) \frac{\mathrm{d}}{\mathrm{d}x} u(x) + r(x) u(x) = f(x), \quad u(-1) = a, \quad u(+1) = b,$$

with $r(x) \equiv \lambda$, $f(x) \equiv 0$ and $a = b = 0$.

Thus equation (10.55) becomes

$$(\mathbf{E}^{(2)} + \mathbf{Q}\mathbf{E} + \lambda)\mathbf{u} = 0, \tag{10.71}$$

where

$$\mathbf{Q} = \begin{pmatrix} q(x_1) & & \\ & \ddots & \\ & & q(x_{n-1}) \end{pmatrix},$$

where $\mathbf{E}$ and $\mathbf{E}^{(2)}$ are defined as in (10.52) on page 250 and where

$$\mathbf{u} = \begin{pmatrix} u_n(x_1) \\ \vdots \\ u_n(x_{n-1}) \end{pmatrix}.$$

Equation (10.71) is another standard matrix eigenvalue equation.

The test case (10.61) gives the simple eigenvalue problem

$$\mathbf{E}^{(2)}\mathbf{u} + \lambda\mathbf{u} = 0. \tag{10.72}$$

We may use this to illustrate the accuracy of the computed eigenvalues. It will be seen from the last example in Section 10.5.2 that the matrix $\mathbf{E}^{(2)}$ will not always be symmetric, so that it could conceivably have some complex pairs of eigenvalues. However, Gottlieb & Lustman (1983) have shown that its eigenvalues are all real and negative in the case where the collocation points are taken as Chebyshev points (although not in the case where they are equally spaced). Whether the same is true for the matrix $\mathbf{E}^{(2)} + \mathbf{QE}$ depends on the size of the function $q(x)$.

Table 10.4: Eigenvalues of differentiation matrices and differential equation, Chebyshev abscissae

Matrix eigenvalues		O.D.E. eigenvalues
$n = 5$	$n = 10$	
−2.4668	−2.4674	2.4674
−9.6000	−9.8696	9.8696
−31.1332	−22.2060	22.2066
−40.0000	−39.5216	39.4784
	−60.7856	61.6850
	−97.9574	88.8624
	−110.8390	120.9026
	−486.2513	157.9136
	−503.3019	199.8595

We show in Table 10.4 the computed eigenvalues (which are indeed all real and negative) of the $(n-1)\times(n-1)$ matrix $\mathbf{E}^{(2)}$ corresponding to the Chebyshev collocation points $\{y_k\}$, for the cases $n = 5$ and $n = 10$, together with the dominant (smallest) eigenvalues of (10.61). We see that respectively three and five of these eigenvalues are quite closely approximated. For general values of n, in fact, the lowest $\lfloor 2n/\pi \rfloor$ eigenvalues are computed very accurately.

Table 10.5: Eigenvalues of differentiation matrices and differential equation, evenly spaced abscissae

Matrix eigenvalues		O.D.E. eigenvalues
$n = 5$	$n = 10$	
-2.4803	-2.4674	2.4674
-11.1871	-9.8715	9.8696
-15.7488	-22.3049	22.2066
-17.4587	-36.3672	39.4784
	-48.5199	61.6850
	$(-57.6718 \pm$	88.8624
	$\pm\ 45.9830$ i)	120.9026
	$(-58.2899 \pm$	157.9136
	$\pm\ 62.5821$ i)	199.8595

Table 10.5 displays the corresponding results when the Chebyshev points are replaced by points evenly spaced through the interval $[-1, 1]$. We see that not only are the lower eigenvalues less accurately computed, but higher eigenvalues can even occur in complex pairs.

The same phenomena are illustrated for $n = 40$ by Fornberg (1996, Figure 4.4-2).

10.8 Differential equations in one space and one time dimension

A general discussion of the application of Chebyshev polynomials to the solution of partial differential equations will be found in Chapter 11. A particular class of partial differential equations does, however, fall naturally within the scope of the present chapter—namely equations in two independent variables, the first of which (t, say) represents time and the second (x, say) runs over a finite fixed interval (which as usual we shall take to be the interval $[-1, 1]$). We may then try representing the solution at any fixed instant t of time in terms of Chebyshev polynomials in x.

For a specific example, we may consider the heat conduction equation

$$\frac{\partial}{\partial t}u(t,x) = q(x)\frac{\partial^2}{\partial x^2}u(t,x), \quad t \geq 0, \quad -1 \leq x \leq 1, \tag{10.73a}$$

where $q(x) > 0$, with the boundary conditions

$$u(t,-1) = u(t,+1) = 0, \quad t \geq 0, \tag{10.73b}$$

and the initial conditions

$$u(0,x) = u_0(x), \quad -1 \leq x \leq 1. \tag{10.73c}$$

10.8.1 Collocation methods

We try approximating $u(t,x)$ in the form

$$u_n(t,x) := \sum_{k=0}^{n}{}' c_k(t) T_k(x), \tag{10.74}$$

and again select the zeros (10.11) of $T_{n-1}(x)$

$$x_j = \cos\frac{(j-\frac{1}{2})\pi}{n-1}, \quad j = 1, \ldots, n-1, \tag{10.75}$$

as collocation points. The collocation equations for (10.73) become

$$\begin{aligned}
\sum_{k=0}^{n}{}' \frac{\mathrm{d}}{\mathrm{d}t} c_k(t) T_k(x_j) &= q(x_j) \sum_{k=2}^{n} \sum_{\substack{r=0\\(k-r)\text{ even}}}^{k-2}{}' (k-r)k(k+r) c_k(t) T_r(x_j), \\
& \qquad j = 1, \ldots, n-1, && (10.76a)\\
\sum_{k=0}^{n}{}' (-1)^k c_k(t) &= 0, && (10.76b)\\
\sum_{k=0}^{n}{}' c_k(t) &= 0. && (10.76c)
\end{aligned}$$

Using discrete orthogonality as before, equations (10.76a) give

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} c_\ell(t) &= \frac{2}{n-1} \sum_{k=2}^{n} \sum_{\substack{r=0\\(k-r)\text{ even}}}^{k-2}{}' (k-r)k(k+r) \sum_{j=1}^{n-1} q(x_j) T_r(x_j) T_\ell(x_j) c_k(t), \\
& \ell = 0, \ldots, n-2. && (10.77)
\end{aligned}$$

We again make the substitutions (10.68) for $c_{n-1}(t)$ and $c_n(t)$, giving us a system of linear differential equations for $c_0(t)$, …, $c_{n-2}(t)$, of the form

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} c_0(t) \\ c_1(t) \\ \vdots \\ c_{n-2}(t) \end{pmatrix} = \mathbf{A} \begin{pmatrix} c_0(t) \\ c_1(t) \\ \vdots \\ c_{n-2}(t) \end{pmatrix}. \tag{10.78}$$

The matrix $\mathbf{A}$ in (10.78) is not the same as the one appearing in (10.69); however, if we can find its eigenvectors $\{\mathbf{c}^{(j)}\}$ and eigenvalues $\{\lambda^{(j)}\}$, then we can write down the general solution of (10.78) as a linear combination of terms $\{\exp \lambda^{(j)} t\, \mathbf{c}^{(j)}\}$, and hence find the solution corresponding to the given initial conditions (10.73c).

We shall not discuss here the question of how well, if at all, the solution to this system of differential equations approximates the solution to the original partial differential equation (10.73). In particular, we shall not examine the possibility that some $\lambda^{(j)}$ have positive real parts, in which case the approximate solution would diverge exponentially with time and therefore be unstable and completely useless.

10.8.2 Collocation using the differentiation matrix

Once more, we have an alternative approach by way of differentiation matrices. The heat conduction problem (10.73)

$$\frac{\partial}{\partial t}u(t,x) = q(x)\frac{\partial^2}{\partial x^2}u(t,x), \quad u(t,\pm 1) = 0,$$

is another that can be derived from (10.54)

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}u(x) + q(x)\frac{\mathrm{d}}{\mathrm{d}x}u(x) + r(x)u(x) = f(x), \quad u(-1) = a, \quad u(+1) = b,$$

by replacing $u(x)$ and $\frac{\mathrm{d}}{\mathrm{d}x}$ by $u(t,x)$ and $\frac{\partial}{\partial x}$, and setting $q(x) \equiv r(x) \equiv 0$, $a = b = 0$ and

$$f(x) \equiv \frac{1}{q(x)}\frac{\partial}{\partial t}u(t,x).$$

In place of equation (10.55) we thus find the system of differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}(t) = \mathbf{Q}\mathbf{E}^{(2)}\mathbf{u}(t), \tag{10.79}$$

where

$$\mathbf{Q} = \begin{pmatrix} q(x_1) & & \\ & \ddots & \\ & & q(x_{n-1}) \end{pmatrix},$$

where $\mathbf{E}^{(2)}$ is defined as in (10.52) on page 250 and where

$$\mathbf{u}(t) = \begin{pmatrix} u_n(t,x_1) \\ \vdots \\ u_n(t,x_{n-1}) \end{pmatrix}.$$

As in the case of (10.78), we may write down the solution of this system of equations as a linear combination of terms $\{\exp \lambda^{(j)} t\, \mathbf{u}^{(j)}\}$, where the matrix $\mathbf{Q}\mathbf{E}^{(2)}$ has eigenvectors $\{\mathbf{u}^{(j)}\}$ and eigenvalues $\{\lambda^{(j)}\}$. In the special case where $q(x)$ is constant, $q(x) \equiv q > 0$ so that $\mathbf{Q}\mathbf{E}^{(2)} = q\mathbf{E}^{(2)}$, we recall from Section 10.7.2 that the eigenvalues will all be real and negative (Gottlieb & Lustman 1983), provided that collocation is at Chebyshev points; consequently all of these terms will decay exponentially with time and the approximate solution is stable.

10.9 Problems for Chapter 10

1. Show that the two collocation algorithms of Section 10.2.1 should lead to exactly the same result for any given value of n—likewise the two projection algorithms of Section 10.2.3.

2. (a) Consider the problem

   $$(1-x)y' = 1 \text{ on } [0,1], \quad y(0) = 0.$$

   Obtain polynomial approximations $y_n(x) = c_0 + c_1 + \cdots + c_n x^n$ to $y(x)$, of degrees $n = 1, 2, 3$, by including a term $\tau T_n^*(x)$ on the right-hand side of the equation. What is the true solution? Plot the errors in each case.

   (b) Consider the slightly modified problem

   $$(1-x)y' = 1 \text{ on } [0, \tfrac{3}{4}], \quad y(0) = 0.$$

   How do we apply the tau method to this problem? Repeat the exercise of the previous example using $\tau T_n^*(4x/3)$.

   (c) Repeat the exercises again for the intervals $[-1,1]$ (using $\tau T_n(x)$) and $[-\frac{3}{4}, \frac{3}{4}]$ (using $\tau T_n(4x/3)$). What effect does extending the interval have on the approximate solution and the size of the error?

3. Where necessary changing the independent variable so that the interval becomes $[-1,1]$, formulate and solve a (first-order) differentiation matrix approximation to one of the parts of Problem 2.

4. Obtain a numerical solution to the differential equation (10.18) by the standard finite-difference method

   $$(u_{j-1} - 2u_j + u_{j+1})/h^2 + 6\,|x_j| = 0, \quad u_{-n} = u_n = 0,$$

   where $h = 1/n$, $x_j = j/n$ ($|j| \leq n$) and u_j approximates $u(x_j)$.

   For $n = 10$, say, how does the solution compare with the Chebyshev solutions in Tables 10.1 and 10.2?

5. Verify formulae (10.44) and (10.45).

6. Justify the limiting values given in (10.46) and (10.47).

7. Investigate the application of Chebyshev and conventional finite-difference methods to the solution of the differential equation

$$(1+25x^2)^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} u(x) = 50(75x^2 - 1)u(x), \quad u(\pm 1) = \tfrac{1}{26},$$

whose exact solution is the function $1/(1+25x^2)$ used to illustrate the Runge phenomenon in Section 6.1.

8. Investigate similarly the non-linear equation

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} u(x) + \frac{1}{1+u(x)^2} = 0, \quad u(\pm 1) = 0.$$

9. Verify the eigenvalues quoted in Tables 10.4, 10.5.