In both cases, conduct the computations on a sequence of consecutively more fine grids (reduce the size h by a factor of two several times). Verify experimentally that the numerical solution converges to the exact solution (9.75) with the second order with respect to h.

3.* Investigate applicability of the shooting method to solving the boundary value problem:

$$y'' + a^2 y = 0, \quad 0 \le x \le 1,$$
$$y(0) = Y_0, \quad y(1) = Y_1,$$

which has a "$+$" sign instead of the "$-$" in the governing differential equation, but otherwise is identical to problem (9.74).

9.6 Saturation of Finite-Difference Methods by Smoothness

Previously, we explored the saturation of numerical methods by smoothness in the context of algebraic interpolation (piecewise polynomials, see Section 2.2.5, and splines, see Section 2.3.2). Very briefly, the idea is to see whether or not a given method of approximation fully utilizes all the information available, and thus attains the optimal accuracy limited only by the threshold of the unavoidable error. When the method introduces its own error threshold, which may only be larger than that of the unavoidable error and shall be attributed to the specific design, we say that the phenomenon of saturation takes place. For example, we have seen that the interpolation by means of algebraic polynomials on uniform grids saturates, whereas the interpolation by means of trigonometric or Chebyshev polynomials does not, see Sections 3.1.3, 3.2.4, and 3.2.7. In the current section, we will use a number of very simple examples to demonstrate that the approximations by means of finite-difference schemes are, generally speaking, also prone to saturation by smoothness.

Before we continue, let us note that in the context of finite differences the term "saturation" may sometimes acquire an alternative meaning. Namely, saturation theorems are proven as to what maximum order of accuracy may a (stable) scheme have if it approximates a particular equation or class of equations on a given fixed stencil.[4] Results of this type are typically established for partial differential equations, see, e.g., [EO80, Ise82]. Some very simple conclusions can already be drawn based on the method of undetermined coefficients (see Section 10.2.2). Note that this alternative notion of saturation is similar to the one we have previously introduced in this book (see Section 2.2.5) in the sense that it also discusses certain accuracy limits. The difference is, however, that these accuracy limits pertain to a particular class of discretization methods (schemes on a given stencil), whereas previously

[4]The stencil of a difference scheme is a set of grid nodes, on which the finite-difference operator is built that approximates the original differential operator.

we discussed accuracy limits that are not related to specific methods and are rather accounted for by the loss of information in the course of discretization.

Let us now consider the following simple boundary value problem for a second order ordinary differential equation:

$$u'' = f(x), \quad 0 \le x \le 1, \quad u(0) = 0, \quad u(1) = 0, \tag{9.81}$$

where the right-hand side $f(x)$ is assumed given.

We introduce a uniform grid on the interval $[0,1]$:

$$x_n = nh, \quad n = 0,1,\ldots,N, \quad Nh = 1,$$

and approximate problem (9.81) using central differences:

$$\begin{gathered} \frac{u_{n+1} - 2u_n + u_{n-1}}{h^2} = f_n \equiv f(x_n), \quad n = 1,2,\ldots,N-1, \\ u_0 = 0, \quad u_N = 0. \end{gathered} \tag{9.82}$$

Provided that the solution $u = u(x)$ of problem (9.81) is sufficiently smooth, or more precisely, provided that its fourth derivative $u^{(4)}(x)$ is bounded for $0 \le x \le 1$, the approximation (9.82) is second-oder accurate, see formula (9.20a). In this case, if the scheme (9.82) is stable, then it will converge with the rate $\mathcal{O}(h^2)$.

However, for such a simple difference scheme as (9.82) one can easily study the convergence directly, i.e., without using Theorem 9.1. A study of that type will be particularly instrumental because on one hand the regularity of the solution may not always be sufficient to guarantee consistency, and on the other hand, it will allow one to see whether or not the convergence accelerates for the functions that are smoother than those minimally required for obtaining $\mathcal{O}(h^2)$.

Note that the degree of regularity of the solution $u(x)$ to problem (9.81) is immediately determined by that of the right-hand side $f(x)$. Namely, the solution $u(x)$ will always have two additional derivatives. It will therefore be convenient to use different right-hand sides $f(x)$ with different degree of regularity, and to investigate directly the convergence properties of scheme (9.82). In doing so, we will analyze both the case when the regularity is formally insufficient to guarantee the second order convergence, and the opposite case when the regularity is "excessive" for that purpose. In the latter case we will, in fact, see that the convergence still remains only second order with respect to h, which implies saturation.

Let us first consider a discontinuous right-hand side:

$$f(x) = \begin{cases} 0, & 0 \le x \le \frac{1}{2}, \\ 1, & \frac{1}{2} < x \le 1. \end{cases} \tag{9.83}$$

On each of the two sub-intervals: $[0,1/2]$ and $[1/2,1]$, the solution can be found as a combination of the general solution to the homogeneous equation and a particular solution to the inhomogeneous equation. The latter is equal to zero on $[0,1/2]$ and on

$[1/2,1]$ it is easily obtained using undetermined coefficients. Therefore, the overall solution of problem (9.81), (9.83) can be found in the form:

$$u(x)=\begin{cases} c_1+c_2x, & 0\le x\le \frac{1}{2},\\ c_3+c_4x+\frac{1}{2}x^2, & \frac{1}{2}<x\le 1,\end{cases} \tag{9.84}$$

where the constants c_1, c_2, c_3, and c_4 are to be chosen so that to satisfy the boundary conditions $u(0)=u(1)=1$ and the continuity requirements:

$$u\left(\frac{1}{2}-0\right)=u\left(\frac{1}{2}+0\right),\quad u'\left(\frac{1}{2}-0\right)=u'\left(\frac{1}{2}+0\right). \tag{9.85}$$

Altogether this yields:

$$\begin{aligned} c_1&=0, & c_3+c_4+\frac{1}{2}&=0,\\ c_1+\frac{c_2}{2}-c_3-\frac{c_4}{2}-\frac{1}{8}&=0, & c_2-c_4-\frac{1}{2}&=0.\end{aligned} \tag{9.86}$$

Solving equations (9.86) we find:

$$c_1=0,\quad c_2=-\frac{1}{8},\quad c_3=\frac{1}{8},\quad c_4=-\frac{5}{8}, \tag{9.87}$$

so that

$$u(x)=\begin{cases} -\frac{1}{8}x, & 0\le x\le \frac{1}{2},\\ \frac{1}{8}-\frac{5}{8}x+\frac{1}{2}x^2, & \frac{1}{2}<x\le 1.\end{cases} \tag{9.88}$$

In the finite-difference case, instead of (9.83) we have:

$$f_n=\begin{cases} 0, & n=0,1,\ldots,\frac{N}{2},\\ 1, & n=\frac{N}{2}+1,\frac{N}{2}+2,\ldots,N.\end{cases} \tag{9.89}$$

Accordingly, the solution is to be sought for in the form:

$$u_n=\begin{cases} c_1+c_2(nh), & n=0,1,\ldots,\frac{N}{2}+1,\\ c_3+c_4(nh)+\frac{1}{2}(nh)^2, & n=\frac{N}{2},\frac{N}{2}+1,\ldots,N,\end{cases} \tag{9.90}$$

where on each sub-interval we have a combination of the general solution to the homogeneous difference equation and a particular solution of the inhomogeneous difference equation (obtained by the method of undetermined coefficients). Notice that unlike in the continuous case (9.84), the two grid sub-intervals in formula (9.90) overlap across the entire cell $[N/2,N/2+1]$ (we are assuming that N is even). Therefore, the constants c_1, c_2, c_3, and c_4 in (9.90) are to be determined from the boundary conditions at the endpoints of the interval $[0,1]$: $u_0=u_N=0$, and from the matching conditions in the middle that are given simply as [cf. formula (9.85)]:

$$\begin{aligned} c_1+c_2\left(\frac{N}{2}h\right)&=c_3+c_4\left(\frac{N}{2}h\right)+\frac{1}{2}\left(\frac{N}{2}h\right)^2,\\ c_1+c_2\left(\frac{N}{2}h+h\right)&=c_3+c_4\left(\frac{N}{2}h+h\right)+\frac{1}{2}\left(\frac{N}{2}h+h\right)^2.\end{aligned} \tag{9.91}$$

Altogether this yields:

$$\begin{aligned} c_1 &= 0, & c_3 + c_4 + \frac{1}{2} &= 0, \\ c_1 + \frac{c_2}{2} - c_3 - \frac{c_4}{2} - \frac{1}{8} &= 0, & c_2 - c_4 - \frac{1}{2} - \frac{h}{2} &= 0, \end{aligned} \tag{9.92}$$

where the last equation of system (9.92) was obtained by subtracting the first equation of (9.91) from the second equation of (9.91) and subsequently dividing by h.

Notice that system (9.92) which characterizes the finite-difference case is almost identical to system (9.86) which characterizes the continuous case, except that there is an $\mathscr{O}(h)$ discrepancy in the fourth equation. Accordingly, there is also an $\mathscr{O}(h)$ difference in the values of the constants [cf. formula (9.87)]:

$$c_1 = 0, \quad c_2 = -\frac{1}{8} + \frac{h}{4}, \quad c_3 = \frac{1}{8} + \frac{h}{4}, \quad c_4 = -\frac{5}{8} - \frac{h}{4},$$

so that the solution to problem (9.82), (9.89) is given by:

$$u_n = \begin{cases} -\frac{1}{8}(nh) + \frac{h}{4}(nh), & n = 0, 1, \ldots, \frac{N}{2} + 1, \\ \frac{1}{8} - \frac{5}{8}(nh) + \frac{1}{2}(nh)^2 + \frac{h}{4}(1 - nh), & n = \frac{N}{2}, \frac{N}{2} + 1, \ldots, N. \end{cases} \tag{9.93}$$

By comparing formulae (9.88) and (9.93), where $nh = x_n$, we conclude that

$$\| [u]_h - u^{(h)} \| = \max_n |u(x_n) - u_n| = \mathscr{O}(h),$$

i.e., that the solution of the finite-difference problem (9.82), (9.89) converges to the solution of the differential problem (9.81), (9.83) with the first order with respect to h. Note that scheme (9.82), (9.89) falls short of the second order convergence because the solution of the differential problem (9.82), (9.89) is not sufficiently smooth.

Instead of the discontinuous right-hand side (9.83) let us now consider a continuous function with discontinuous first derivative:

$$f(x) = \begin{cases} -x, & 0 \le x \le \frac{1}{2}, \\ x - 1, & \frac{1}{2} < x \le 1. \end{cases} \tag{9.94}$$

Solution to problem (9.81), (9.94) can be found in the form:

$$u(x) = \begin{cases} c_1 + c_2 x - \frac{1}{6}x^3, & 0 \le x \le \frac{1}{2}, \\ c_3 + c_4 x + \frac{1}{6}x^3 - \frac{1}{2}x^2, & \frac{1}{2} < x \le 1, \end{cases}$$

where the constants c_1, c_2, c_3, and c_4 are again to be chosen so that to satisfy the boundary conditions $u(0) = u(1) = 1$ and the continuity requirements (9.85):

$$\begin{aligned} c_1 &= 0, & c_3 + c_4 - \frac{1}{3} &= 0, \\ c_1 + \frac{c_2}{2} - \frac{1}{48} - c_3 - \frac{c_4}{2} + \frac{5}{48} &= 0, & c_2 - \frac{1}{8} - c_4 + \frac{3}{8} &= 0. \end{aligned} \tag{9.95}$$

Solving equations (9.95) we find:

$$c_1 = 0, \quad c_2 = \frac{1}{8}, \quad c_3 = -\frac{1}{24}, \quad c_4 = \frac{3}{8}, \tag{9.96}$$

so that

$$u(x) = \begin{cases} \frac{1}{8}x - \frac{1}{6}x^3, & 0 \le x \le \frac{1}{2}, \\ -\frac{1}{24} + \frac{3}{8}x + \frac{1}{6}x^3 - \frac{1}{2}x^2, & \frac{1}{2} < x \le 1. \end{cases} \tag{9.97}$$

In the discrete case, instead of (9.94) we write:

$$f_n = \begin{cases} -(nh), & n = 0, 1, \ldots, \frac{N}{2}, \\ (nh) - 1, & n = \frac{N}{2} + 1, \frac{N}{2} + 2, \ldots, N, \end{cases} \tag{9.98}$$

and then look for the solution u_n to problem (9.82), (9.98) in the form:

$$u_n = \begin{cases} c_1 + c_2(nh) - \frac{1}{6}(nh)^3, & n = 0, 1, \ldots, \frac{N}{2} + 1, \\ c_3 + c_4(nh) + \frac{1}{6}(nh)^3 - \frac{1}{2}(nh)^2, & n = \frac{N}{2}, \frac{N}{2} + 1, \ldots, N. \end{cases}$$

For the matching conditions in the middle we now have [cf. formulae (9.91)]:

$$\begin{aligned} c_1 + c_2\left(\frac{N}{2}h\right) - \frac{1}{6}\left(\frac{N}{2}h\right)^3 &= c_3 + c_4\left(\frac{N}{2}h\right) + \frac{1}{6}\left(\frac{N}{2}h\right)^3 - \frac{1}{2}\left(\frac{N}{2}h\right)^2, \\ c_1 + c_2\left(\frac{N}{2}h + h\right) - \frac{1}{6}\left(\frac{N}{2}h + h\right)^3 &= c_3 + c_4\left(\frac{N}{2}h + h\right) \\ &\quad + \frac{1}{6}\left(\frac{N}{2}h + h\right)^3 - \frac{1}{2}\left(\frac{N}{2}h + h\right)^2, \end{aligned} \tag{9.99}$$

and consequently:

$$\begin{aligned} c_1 &= 0, & c_3 + c_4 - \frac{1}{3} &= 0, \\ c_1 + \frac{c_2}{2} - \frac{1}{48} - c_3 - \frac{c_4}{2} + \frac{5}{48} &= 0, & c_2 - \frac{1}{8} - c_4 + \frac{3}{8} - \frac{h^2}{3} &= 0, \end{aligned} \tag{9.100}$$

where the last equation of (9.100) was obtained by subtracting the first equation of (9.99) from the second equation of (9.99) and subsequently dividing by h.

Solving equations (9.100) we obtain [cf. formula (9.96)]:

$$c_1 = 0, \quad c_2 = \frac{1}{8} + \frac{h^2}{6}, \quad c_3 = -\frac{1}{24} + \frac{h^2}{6}, \quad c_4 = \frac{3}{8} - \frac{h^2}{6},$$

and

$$u_n = \begin{cases} \frac{1}{8}(nh) - \frac{1}{6}(nh)^3 + \frac{h^2}{6}(nh), & n = 0, 1, \ldots, \frac{N}{2} + 1, \\ -\frac{1}{24} + \frac{3}{8}(nh) + \frac{1}{6}(nh)^3 - \frac{1}{2}(nh)^2 & \\ \qquad + \frac{h^2}{6}(1 - nh), & n = \frac{N}{2}, \frac{N}{2} + 1, \ldots, N. \end{cases} \tag{9.101}$$

It is clear that the error between the continuous solution (9.97) and the discrete solution (9.101) is estimated as

$$\|[u]_h - u^{(h)}\| = \max_n |u(x_n) - u_n| = \mathcal{O}(h^2),$$

which means that the solution of the finite-difference problem (9.82), (9.98) converges to the solution of the differential problem (9.81), (9.94) with the second order with respect to h. Note that second order convergence is attained here even though the degree of regularity of the solution — third derivative is discontinuous — is formally insufficient to guarantee second order accuracy (consistency).

In much the same way one can analyze the case when the right-hand side $f(x)$ has one continuous derivative (the so-called C^1 space of functions), for example:

$$f(x) = \begin{cases} -(x-\frac{1}{2})^2, & 0 \le x \le \frac{1}{2}, \\ (x-\frac{1}{2})^2, & \frac{1}{2} < x \le 1. \end{cases} \tag{9.102}$$

For problem (9.81), (9.102), it is also possible to prove the second order convergence, which is the subject of Exercise 1 at the end of the section.

The foregoing examples demonstrate that the rate of finite-difference convergence depends on the regularity of the solution to the underlying continuous problem. It is therefore interesting to see what happens when the regularity increases beyond C^1.

Consider the right-hand side in the form of a quadratic polynomial:

$$f(x) = x(x-1). \tag{9.103}$$

This function is obviously infinitely differentiable (C^∞ space), and so is the solution $u = u(x)$ of problem (9.81), (9.103), which is given by:

$$u(x) = \frac{1}{12}x + \frac{1}{12}x^4 - \frac{1}{6}x^3. \tag{9.104}$$

Scheme (9.82) with the right-hand side

$$f_n = nh(nh-1), \quad n = 0, 1, \ldots, N, \tag{9.105}$$

approximates problem (9.81), (9.103) with second order accuracy. The solution of the finite-difference problem (9.82), (9.105) can be found in the form:

$$u_n = \underbrace{c_1 + c_2(nh)}_{u_n^{(\mathrm{g})}} + \underbrace{(nh)^2(A(nh)^2 + B(nh) + C)}_{u_n^{(\mathrm{p})}}, \tag{9.106}$$

where $u_n^{(\mathrm{g})}$ is the general solution to the homogeneous equation and $u_n^{(\mathrm{p})}$ is a particular solution to the inhomogeneous equation. The values of A, B, and C are to be found

using the method of undetermined coefficients:

$$A\frac{((n+1)h)^4-2(nh)^4+((n-1)h)^4}{h^2} + B\frac{((n+1)h)^3-2(nh)^3+((n-1)h)^3}{h^2} + C\frac{((n+1)h)^2-2(nh)^2+((n-1)h)^2}{h^2} = (nh)^2-(nh),$$

which yields:

$$A(12(nh)^2+2h^2)+B(6nh)+2C=(nh)^2-(nh)$$

and accordingly,

$$A=\frac{1}{12},\quad B=-\frac{1}{6},\quad C=-Ah^2=-\frac{1}{12}h^2. \tag{9.107}$$

For the constants c_1 and c_2 we substitute the expression (9.106) and the already available coefficients A, B, and C of (9.107) into the boundary conditions of (9.82) and write:

$$u_0=c_1=0,\quad u_N=c_2+A+B+C=0.$$

Consequently,

$$c_1=0 \quad \text{and} \quad c_2=\frac{1}{12}+\frac{1}{12}h^2,$$

so that for the overall solution u_n of problem (9.81), (9.105) we obtain:

$$u_n=\frac{1}{12}(nh)+\frac{1}{12}h^2(nh)+(nh)^2\left(\frac{1}{12}(nh)^2-\frac{1}{6}(nh)-\frac{1}{12}h^2\right). \tag{9.108}$$

Comparing the continuous solution $u(x)$ given by (9.104) with the discrete solution u_n given by (9.108) we conclude that

$$\|[u]_h-u^{(h)}\|=\max_n|u(x_n)-u_n|=\mathcal{O}(h^2),$$

which implies that notwithstanding the infinite smoothness of the right-hand side $f(x)$ of (9.103) and that of the solution $u(x)$, scheme (9.82), (9.105) still shows only second order convergence. This is a manifestation of the phenomenon of *saturation by smoothness*. The rate of decay of the approximation error is determined by the specific approximation method employed on a given grid, and does not reach the level of the pertinent unavoidable error.

To demonstrate that the previous observation is not accidental, let us consider another example of an infinitely differentiable (C^∞) right-hand side:

$$f(x)=\sin(\pi x). \tag{9.109}$$

The solution of problem (9.81), (9.109) is given by:

$$u(x) = -\frac{1}{\pi^2}\sin(\pi x). \tag{9.110}$$

The discrete right-hand side that corresponds to (9.109) is:

$$f_n = \sin(\pi n h), \quad n = 0, 1, \ldots, N, \tag{9.111}$$

and the solution to the finite-difference problem (9.82), (9.111) is to be sought for in the form $u_n = A\sin(\pi n h) + B\cos(\pi n h)$ with the undetermined coefficients A and B, which eventually yields:

$$u_n = -\frac{h^2}{4\sin^2\frac{\pi h}{2}}\sin(\pi n h). \tag{9.112}$$

The error between the continuous solution given by (9.110) and the discrete solution given by (9.112) is easy to estimate provided that the grid size is small, $h \ll 1$:

$$\begin{aligned}
\|[u]_h - u^{(h)}\| = \max_n |u(x_n) - u_n| &= \left|\frac{1}{\pi^2} - \frac{h^2}{4\sin^2\frac{\pi h}{2}}\right| \\
&\approx \left|\frac{1}{\pi^2} - \frac{h^2}{4\left[\frac{\pi h}{2} - \frac{1}{6}\left(\frac{\pi h}{2}\right)^3\right]^2}\right| \approx \left|\frac{1}{\pi^2} - \frac{h^2}{4\left[\frac{\pi h}{2} - \frac{1}{3}\left(\frac{\pi h}{2}\right)^4\right]}\right| \\
&\approx \left|\frac{1}{\pi^2} - \frac{h^2}{4\left(\frac{\pi h}{2}\right)^2}\left[1 + \frac{1}{3}\left(\frac{\pi h}{2}\right)^2\right]\right| = \frac{1}{\pi^2}\frac{1}{3}\left(\frac{\pi h}{2}\right)^2 = \mathcal{O}(h^2).
\end{aligned}$$

This, again, corroborates the effect of saturation, as the convergence of the scheme (9.82), (9.111) is only second order in spite of the infinite smoothness of the data.

In general, all finite-difference methods are prone to saturation. This includes the methods for solving ordinary differential equations described in this chapter, as well as the methods for partial differential equations described in Chapter 10. There are, however, other methods for the numerical solution of differential equations. For example, the so-called spectral methods described briefly in Section 9.7 do not saturate and exhibit convergence rates that self-adjust to the regularity of the corresponding solution (similarly to how the error of the trigonometric interpolation adjusts to the smoothness of the interpolated function, see Theorem 3.5 on page 68). The literature on the subject of spectral methods is vast, and we can refer the reader, e.g., to the monographs [GO77, CHQZ88, CHQZ06], and textbooks [Boy01, HGG06].

Exercise

1. Consider scheme (9.82) with the right-hand side:

$$f_n = \begin{cases} -(nh - \frac{1}{2})^2, & n = 0, 1, 2, \ldots, \frac{N}{2}, \\ (nh - \frac{1}{2})^2, & \frac{N}{2}, \frac{N}{2}+1, \ldots, N. \end{cases}$$

This scheme approximates problem (9.81), (9.102). Obtain the finite-difference solution in closed form and prove second order convergence.

9.7 The Notion of Spectral Methods

In this section, we only provide one particular example of a spectral method. Namely, we solve a simple boundary value problem using a Fourier-based technique. Our specific goal is to demonstrate that alternative discrete approximations to differential equations can be obtained that, unlike finite-difference methods, will not suffer from the saturation by smoothness (see Section 9.6). A comprehensive account of spectral methods can be found, e.g., in the monographs [GO77, CHQZ88, CHQZ06], as well as in the textbooks [Boy01, HGG06]. The material of this section is based on the analysis of Chapter 3 and can be skipped during the first reading.

Consider the same boundary value problem as in Section 9.6:

$$u'' = f(x), \quad 0 \le x \le 1, \quad u(0) = 0, \quad u(1) = 0, \tag{9.113}$$

where the right-hand side $f(x)$ is assumed given. In this section, we will not approximate problem (9.113) on the grid using finite differences. We will rather look for an approximate solution to problem (9.113) in the form of a trigonometric polynomial.

Trigonometric polynomials were introduced and studied in Chapter 3. Let us formally extend both the unknown solution $u = u(x)$ and the right-hand side $f = f(x)$ to the interval $[-1, 1]$ antisymmetrically, i.e., $u(-x) = -u(x)$ and $f(-x) = -f(x)$, so that the resulting functions are odd. We can then represent the solution $u(x)$ of problem (9.113) approximately as a trigonometric polynomial:

$$u^{(n)}(x) = \sum_{k=1}^{n+1} B_k \sin(\pi k x) \tag{9.114}$$

with the coefficients B_k to be determined. Note that according to Theorem 3.3 (see page 66), the polynomial (9.114), which is a linear combination of the sine functions only, is suited specifically for representing the odd functions. Note also that for any choice of the coefficients B_k the polynomial $u^{(n)}(x)$ of (9.114) satisfies the boundary conditions of problem (9.113) exactly.

Let us now introduce the same grid (of dimension $n+1$) as we used in Section 3.1:

$$x_m = \frac{1}{n+1} m + \frac{1}{2(n+1)}, \quad m = 0, 1, \ldots, n, \tag{9.115}$$

and interpolate the given function $f(x)$ on this grid by means of the trigonometric polynomial with $n+1$ terms:

$$f^{(n)}(x) = \sum_{k=1}^{n+1} b_k \sin(\pi k x). \tag{9.116}$$

The coefficients of the polynomial (9.116) are given by:

$$\begin{aligned} b_k &= \frac{2}{n+1}\sum_{m=0}^{n} f(x_m)\sin k\left(\frac{\pi}{n+1}m+\frac{\pi}{2(n+1)}\right), \quad k=1,2,\ldots,n, \\ b_{n+1} &= \frac{1}{n+1}\sum_{m=0}^{n} f(x_m)(-1)^m. \end{aligned} \tag{9.117}$$

To approximate the differential equation $u''=f$ of (9.113), we require that the second derivative of the approximate solution $u^{(n)}(x)$:

$$\frac{d^2}{dx^2}u^{(n)}(x) = -\pi^2\sum_{k=1}^{n+1} B_k k^2 \sin(\pi k x) \tag{9.118}$$

coincide with the interpolant of the right-hand side $f^{(n)}(x)$ at every node x_m of the grid (9.115):

$$\frac{d^2}{dx^2}u^{(n)}(x_m) = f^{(n)}(x_m), \quad m=0,1,\ldots,n. \tag{9.119}$$

Note that both the interpolant $f^{(n)}(x)$ given by formula (9.116) and the derivative $\frac{d^2}{dx^2}u^{(n)}(x)$ given by formula (9.118) are sine trigonometric polynomials of the same order $n+1$. According to formula (9.119), they coincide at x_m for all $m=0,1,\ldots,n$. Therefore, due to the uniqueness of the trigonometric interpolating polynomial (see Theorem 3.1 on page 62), these two polynomials are, in fact, the same everywhere on the interval $0\le x\le 1$. Consequently, their coefficients are identically equal:

$$-\pi^2 k^2 B_k = b_k, \quad k=1,2,\ldots,n+1. \tag{9.120}$$

Equalities (9.120) allow one to find B_k provided that b_k are known.

Consider a particular example analyzed in the end of Section 9.6:

$$f(x) = \sin(\pi x). \tag{9.121}$$

The exact solution of problem (9.113), (9.121) is given by:

$$u(x) = -\frac{1}{\pi^2}\sin(\pi x). \tag{9.122}$$

According to formulae (9.117), the coefficients b_k that correspond to the right-hand side $f(x)$ given by (9.121) are:

$$b_1 = 1 \text{ and } b_k = 0, \ \ k=2,3,\ldots,n+1.$$

Consequently, relations (9.120) imply that

$$B_1 = -\frac{1}{\pi^2} \text{ and } B_k = 0, \ \ k=2,3,\ldots,n+1.$$

Therefore,

$$u^{(n)}(x) = -\frac{1}{\pi^2}\sin(\pi x). \tag{9.123}$$

By comparing formulae (9.123) and (9.122), we conclude that the approximate method based on enforcing the differential equation $u'' = f$ via the finite system of equalities (9.119) reconstructs the exact solution of problem (9.113), (9.121). The error is therefore equal to zero. Of course, one should not expect that this ideal behavior of the error will hold in general. The foregoing particular result only takes place because of the specific choice of the right-hand side (9.121). However, in a variety of other cases one can obtain a rapid decay of the error as n increases.

Consider the odd function $f(-x) = -f(x)$ obtained on the interval $[-1,1]$ by extending the right-hand side of problem (9.113) antisymmetrically from the interval $[0,1]$. Assume that this function can also be translated along the entire real axis:

$$\forall x \in [2l+1, 2(l+1)+1]: \; f(x) = f(x-2(l+1)), \quad l = 0, 1, \pm 2, \pm 3, \ldots$$

so that the resulting periodic function with the period $L = 2$ be smooth. More precisely, we require that the function $f(x)$ constructed this way possess continuous derivatives of order up to $r > 0$ everywhere, and a square integrable derivative of order $r+1$:

$$\int_{-1}^{1} \left[f^{(r+1)}(x) \right]^2 dx < \infty.$$

Clearly the function $f(x) = \sin(\pi x)$, see formula (9.121), satisfies these requirements. Another example which, unlike (9.121), leads to a full infinite Fourier expansion is $f(x) = \sin(\pi \sin(\pi x))$. Both functions are periodic with the period $\boldsymbol{L} = 2$ and infinitely smooth everywhere ($r = \infty$).

Let us represent $f(x)$ as the sum of its sine Fourier series:

$$f(x) = \sum_{k=1}^{\infty} \beta_k \sin(k\pi x), \tag{9.124}$$

where the coefficients β_k are defined as:

$$\beta_k = 2 \int_0^1 f(x) \sin(k\pi x) dx. \tag{9.125}$$

The series (9.124) converges to the function $f(x)$ uniformly and absolutely. The rate of convergence was obtained when proving Theorem 3.5, see pages 68–71. Namely, if we define the partial sum $S_n(x)$ and the remainder $\delta S_n(x)$ of the series (9.124) as done in Section 3.1.3:

$$S_n(x) = \sum_{k=1}^{n+1} \beta_k \sin(k\pi x), \quad \delta S_n(x) = \sum_{k=n+2}^{\infty} \beta_k \sin(k\pi x), \tag{9.126}$$

then

$$|f(x) - S_n(x)| = |\delta S_n(x)| \le \frac{\zeta_n}{n^{r+\frac{1}{2}}}, \tag{9.127}$$

where ζ_n is a numerical sequence such that $\zeta_n = o(1)$, i.e., $\lim_{n\to\infty} \zeta_n = 0$. Substituting the expressions:

$$f(x_m) = S_n(x_m) + \delta S_n(x_m), \quad m = 0, 1, \ldots, n,$$

into the definition (9.117) of the coefficients b_k we obtain:

$$\begin{aligned} b_k &= \underbrace{\frac{2}{n+1}\sum_{m=0}^{n} S_n(x_m)\sin(\pi k x_m)}_{\beta_k} + \underbrace{\frac{2}{n+1}\sum_{m=0}^{n} \delta S_n(x_m)\sin(\pi k x_m)}_{\delta\beta_k}, \quad k = 1,2,\ldots,n, \\ b_{n+1} &= \underbrace{\frac{1}{n+1}\sum_{m=0}^{n} S_n(x_m)(-1)^m}_{\beta_{n+1}} + \underbrace{\frac{1}{n+1}\sum_{m=0}^{n} \delta S_n(x_m)(-1)^m}_{\delta\beta_{n+1}}. \end{aligned} \tag{9.128}$$

The first sum on the right-hand side of each equality (9.128) is indeed equal to the genuine Fourier coefficient β_k of (9.125), $k = 1,2,\ldots,n+1$, because the partial sum $S_n(x)$ given by (9.126) coincides with its own trigonometric interpolating polynomial[5] for all $0 \le x \le 1$. As for the "corrections" to the coefficients, $\delta\beta_k$, they come from the remainder $\delta S_n(x)$ and their magnitudes can be easily estimated using inequality (9.127) and formulae (9.117):

$$|\delta\beta_k| \le 2\frac{\zeta_n}{n^{r+1/2}}, \quad k = 1,2,\ldots,n+1. \tag{9.129}$$

Let us now consider the exact solution $u = u(x)$ of problem (9.113). Given the assumptions made regarding the right-hand side $f = f(x)$, the solution u is also a smooth odd periodic function with the period $L = 2$. It can be represented as its own Fourier series:

$$u(x) = \sum_{k=1}^{\infty} \gamma_k \sin(k\pi x), \tag{9.130}$$

where the coefficients γ_k are given by:

$$\gamma_k = 2\int_0^1 u(x)\sin(k\pi x)dx. \tag{9.131}$$

Series (9.130) converges uniformly. Moreover, the same argument based on the periodicity and smoothness implies that the Fourier series for the derivatives $u'(x)$ and $u''(x)$ also converge uniformly.[6] Consequently, series (9.130) can be differentiated (at least) twice termwise:

$$u''(x) = -\pi^2 \sum_{k=1}^{\infty} k^2 \gamma_k \sin(k\pi x). \tag{9.132}$$

Recall, we must enforce the equality $u'' = f$. Then by comparing the expansions (9.132) and (9.124) and using the orthogonality of the trigonometric system, we have:

$$\gamma_k = -\frac{1}{\pi^2 k^2}\beta_k, \quad k = 1,2,\ldots. \tag{9.133}$$

[5] Due to the uniqueness of the trigonometric interpolating polynomial, see Theorem 3.1.

[6] The first derivative $u'(x)$ will be an even function rather than odd.

Next, recall that the coefficients B_k of the approximate solution $u^{(n)}(x)$ defined by (9.114) are given by formula (9.120). Using the representation $b_k = \beta_k + \delta\beta_k$, see formula (9.128), and also employing relations (9.133), we obtain:

$$\begin{aligned} B_k &= -\frac{1}{\pi^2 k^2} b_k \\ &= -\frac{1}{\pi^2 k^2}\beta_k - \frac{1}{\pi^2 k^2}\delta\beta_k \\ &= \gamma_k - \frac{1}{\pi^2 k^2}\delta\beta_k, \quad k = 1,2,\ldots,n+1. \end{aligned} \tag{9.134}$$

Formula (9.134) will allow us to obtain an error estimate for the approximate solution $u^{(n)}(x)$. To do so, we first rewrite the Fourier series (9.130) for the exact solution $u(x)$ as its partial sum plus the remainder [cf. formula (9.126)]:

$$u(x) = \tilde{S}_n(x) + \delta\tilde{S}_n(x) = \sum_{k=1}^{n+1} \gamma_k \sin(k\pi x) + \sum_{k=n+2}^{\infty} \gamma_k \sin(k\pi x), \tag{9.135}$$

and obtain an estimate for the convergence rate [cf. formula (9.127)]:

$$|u(x) - \tilde{S}_n(x)| = |\delta\tilde{S}_n(x)| \le \frac{\eta_n}{n^{r+\frac{5}{2}}}, \tag{9.136}$$

where $\eta_n = o(1)$ as $n \longrightarrow \infty$. Note that according to the formulae (9.136) and (9.127), the series (9.130) converges faster than the series (9.124), with the rates $o\left(n^{-\left(r+\frac{5}{2}\right)}\right)$ and $o\left(n^{-\left(r+\frac{1}{2}\right)}\right)$, respectively. The reason is that if the right-hand side $f = f(x)$ of problem (9.113) has r continuous derivatives and a square integrable derivative of order $r+1$, then the solution $u = u(x)$ to this problem would normally have $r+2$ continuous derivatives and a square integrable derivative of order $r+3$.

Next, using equalities (9.114), (9.134), and (9.135) and estimates (9.127) and (9.136), we can write $\forall x \in [0,1]$:

$$\begin{aligned} |u(x) - u^{(n)}(x)| &= \left| \tilde{S}_n(x) + \delta\tilde{S}_n(x) - \sum_{k=1}^{n+1} B_k \sin(\pi k x) \right| \\ &= \left| \tilde{S}_n(x) + \delta\tilde{S}_n(x) - \sum_{k=1}^{n+1} \gamma_k \sin(\pi k x) + \sum_{k=1}^{n+1} \frac{\delta\beta_k}{\pi^2 k^2} \sin(\pi k x) \right| \\ &= \left| \delta\tilde{S}_n(x) + \sum_{k=1}^{n+1} \frac{\delta\beta_k}{\pi^2 k^2} \sin(\pi k x) \right| \le |\delta\tilde{S}_n(x)| + \sum_{k=1}^{n+1} |\delta\beta_k| \\ &\le \frac{\eta_n}{n^{r+\frac{5}{2}}} + \frac{\zeta_n}{n^{r-\frac{1}{2}}} \le \frac{\sigma_n}{n^{r-\frac{1}{2}}}, \end{aligned} \tag{9.137}$$

where σ_n is another infinitesimal sequence: $\sigma_n = o(1)$ as $n \longrightarrow \infty$. The key distinctive feature of error estimate (9.137) is that it provides for a more rapid convergence in the case when the right-hand side $f(x)$ that drives the problem has higher

regularity. In other words, similarly to the original trigonometric interpolation (see Section 3.1.3), the foregoing method of obtaining an approximate solution to problem (9.113) *does not get saturated by smoothness*. Indeed, the approximation error self-adjusts to the regularity of the data without us having to change anything in the algorithm. Moreover, if the right-hand side $f(x)$ of problem (9.113) has continuous periodic derivatives of all orders, then according to estimate (9.137) the method *will converge with a spectral rate,* i.e., faster than any inverse power of n. For that reason, methods of this type are referred to as spectral methods in the literature.

Note that the simple Fourier-based spectral method that we have outlined in this section will only work for smooth periodic functions, i.e., for the functions that withstand smooth periodic extensions. There are many examples of smooth right-hand sides that do not satisfy this constraint, for example the quadratic function $f(x) = x(x-1)$ used in Section 9.6, see formula (9.103). However, a spectral method can be built for problem (9.113) with this right-hand side as well. In this case, it will be convenient to look for a solution as a linear combination of Chebyshev polynomials, rather than in the form of a trigonometric polynomial (9.114). This approach is similar to Chebyshev-based interpolations discussed in Section 3.2.

Note also that in this section we enforced the differential equation of (9.113) by requiring that the two trigonometric polynomials, $\frac{d^2}{dx^2}u^{(n)}(x)$ and $f^{(n)}(x)$, coincide at the nodes of the grid (9.115), see equalities (9.119). In the context of spectral methods, the points x_m given by (9.115) are often referred to as the collocation points, and the corresponding methods are known as the spectral collocation methods. Alternatively, one can use Galerkin approximations for building spectral methods. The Galerkin method is a very useful and general technique that has many applications in numerical analysis and beyond; we briefly describe it in Section 12.2.3 when discussing finite elements.

Similarly to any other method of approximation, one generally needs to analyze accuracy and stability when designing spectral methods. Over the recent years, a number of efficient spectral methods have been developed for solving a wide variety of initial and boundary value problems for ordinary and partial differential equations. For further detail, we refer the reader to [GO77,CHQZ88,Boy01,CHQZ06,HGG06].

Exercise

1. Solve problem (9.113) with the right-hand side $f(x) = \sin(\pi \sin(\pi x))$ on a computer using the Fourier collocation method described in this section. Alternatively, apply the second order difference method of Section 9.6. Demonstrate experimentally the difference in convergence rates.