Spectral Methods for Differential Problems

C. I. GHEORGHIU
"T. Popoviciu" Institute of Numerical Analysis, Cluj-Napoca, R O M A N I A

January 20, 2005-May 15, 2007[1]

[1]This work was partly supported by grant 2-CEx06-11-96/19.09.2006

Contents

List of Figures

Preface

If it works once, it is a trick;
if it works twice, it is a method;
if it works a hundred of times, it is a very good family of algorithms.
John P. Boyd, SIAM Rev., 46(2004)

The aim of this work is to emphasize the capabilities of spectral and pseudospectral methods in solving boundary value problems for differential and partial differential equations as well as in solving initial-boundary value problems for parabolic and hyperbolic equations. Both linear and genuinely nonlinear problems are taken into account. The class of linear boundary value problems include singularly perturbed problems as well as eigenvalue problems.

Our intention is to provide techniques that cater for a broad diversity of the problems mentioned above.

We believe that hardly any topic in modern mathematics fails to inspire numerical analysts. Consequently, a numerical analyst has to be an open minded scientist ready to borrow from a wide range of mathematical knowledge as well as from computer science. In this respect we also believe that the professional software design is just as challenging as theorem-proving.

The book is not oriented to formal reasoning, which means the well known sequence of axioms, theorem, proof, corollary, etc. Instead, it displays rigorously the most important qualities as well as drawbacks of spectral methods in the context of numerical methods devoted to solve boundary value and eigenvalue problems for differential equations as well as initial-boundary value problems for partial differential equations.

Introduction

Because of being extremely accurate, spectral methods have been intensively studied in the past decades. Mainly three types of spectral methods can be identified, namely, *collocation, tau* and *Galerkin.* The choice of the type of method depends essentially on the application. Collocation methods are suited to nonlinear problems or having complicated coefficients, while Galerkin methods have the advantage of a more convenient analysis and optimal error estimates. The tau method is applicable in the case of complicated (even nonlinear) boundary conditions, where Galerkin approach would be impossible and the collocation extremely tedious.

In any of these cases, the *standard approach,* where the *trial* (shape) and *test* functions simply span a certain family of functions (polynomials), has significant disadvantages. First of all, the matrices resulting in the discretization process have an increased condition number, and thus computational rounding errors deteriorate the expected theoretical exponential accuracy. Moreover, the discretization matrices are generally fully populated, and so efficient algebraic solvers are difficult to apply.

These disadvantages are more obvious when solving fourth order problems, where stability and numerical accuracy are lost when applying higher order approximations.

Several attempts were made in order to try to circumvent these inconveniences of the standard approach. All these attempts are based on the fairly large flexibility in the choice of trial and test functions. In fact, using various weight functions, they are constructed in order to incorporate as much boundary data as possible and, at the same time, to reduce the condition number and the bandwidth of matrices. In this respect we mention the papers of Cabos [27], Dongarra, Straughan and Walker [55], Hiegemann [112] or our contribution in some joint works with S. I. Pop [165], [164] for tau method; the papers of D. Funaro and W. Heinrichs [76], Heinrichs [103] and [104] or Hiegemann and Strauss [111] for the collocation variant; and the papers of Bjoerstad and Tjoestheim [16], Jie Shen [176] and [177] for Galerkin schemes, to quote but a few. All the above mentioned papers are dealing with methods in which the trial and test functions are based on Chebyshev polynomials. The monographs of Gottlieb and Orszag [90], Gottlieb, Hussaini and Orszag [93] and that of Canuto, Hussaini, Quarteroni and Zang [33] contain details about the spectral tau and Galerkin methods as well as about the collocation (pseudospectral) method. They con-

sider the basis of Chebyshev, Hermite and Legendre polynomials and Fourier and sinc functions in order to build up the test and trial functions. The well known monograph of J. P. Boyd [19], beyond very subtle observations about the performance and limitations of spectral methods, contains an exhaustive bibliography for spectral methods at the level of year 2000.

A more strange feature of spectral methods is the fact that, in some situations, they transform self-adjoint differential problems into non symmetric, i.e., non normal, discrete algebraic problems. We pay some attention to this aspect and observe that a proper choice of the trial and test functions can reduce significantly the non normality of the matrices involved in the approximation.

In order to carry out our numerical experiments we used exclusively the software system MATLAB. The textbook of Hunt, Lipsman and Rosenberg [118] is a useful guide to that. Particularly, to implement the pseudospectral derivatives we used the MATLAB codes provided by the paper of Weideman and Reddy, [204].

The writing of this book has benefited enormously from a lot of discussions with Dr. Sorin Iuliu Pop, presently at the T U Eindhoven, during the time he prepared his Ph. D. at the universities "Babes-Bolyai" Cluj-Napoca, Romania and Heidelberg , Germany.

Calin-Ioan Gheorghiu
June 20, 2007
Cluj-Napoca

Part I

The First Part

Chapter 1

Chebyshev polynomials

> *His courses were not voluminous, and he did not consider the quantity of knowledge delivered; rather, he aspired to elucidate some of the most important aspects of the problems he spoke on. These were lively, absorbing lectures; curious remarks on the significance and importance of certain problems and scientific methods were always abundant.*
>
> *A. M. Liapunov who attended Chebyshev's courses in late 1870 (see [26])*

In their monograph [71] Fox and Parker collected the underlying principles of the Chebyshev theory. The polynomials whose properties and applications will be discussed were introduced more than a century ago by the Russian mathematician P. L. Chebyshev (1821-1894). Chebyshev was the most eminent Russian mathematician of the nineteenth century. He was the author of more than 80 publications, covering approximation theory, probability theory, number theory, theory of mechanisms, as well as many problems of analysis and practical mathematics. His interest in mechanisms (as a boy he was fascinated by mechanical toys!) led him to the theory of the approximation of functions (see [181] P. 210 for a Note on the life of P. L. Chebyshev as well as the comprehensive article [26]). Their importance for numerical analysis was rediscovered around the middle of the last century by C. Lanczos (see [126]).

1.1 General properties

Let $\mathcal{P}_N$ be the space of algebraic polynomials of degree at most $N \in \mathbb{N},\ N > 0$, and the *weight function* $\omega : I = [-1, 1] \to \mathbb{R}_+$ defined by

$$\omega(x) := \frac{1}{\sqrt{1 - x^2}}.$$

Let us introduce the fundamental space $L^2_\omega(I)$ by

$$L^2_\omega(I) := \left\{ v : I \to \mathbb{R} |\ v \text{ Lebesgue measurable and } \|v\|_{0,\omega} < \infty \right\},$$

where the norm

$$\|v\|_\omega := \left(\int_{-1}^{1} |v(x)|^2 \omega(x)\, dx \right)^{\frac{1}{2}},$$

is induced by the *weighted scalar (inner) product*

$$(u,v)_\omega := \left(\int_{-1}^{1} u(x)\, v(x)\, \omega(x)\, dx \right). \tag{1.1}$$

Definition 1 *The polynomials $T_n(x)$, $n \in \mathbb{N}$, defined by*

$$T_n(x) := \cos(n\ \arccos(x)), \quad x \in [-1.1],$$

are called the Chebyshev polynomials of the first kind.

Remark 2 *[150] To establish a relationship between algebraic and trigonometric polynomials let us resort to the trigonometric identity*

$$\begin{aligned} \cos(n\theta) + i\sin(n\theta) &= (\cos\theta + i\sin\theta)^n = \\ &= \cos^n\theta + i\binom{n}{1}\cos^{n-1}\theta\cdot\sin\theta + i^2\binom{n}{2}\cos^{n-2}\theta\cdot\sin^2\theta + \dots . \end{aligned}$$

The terms on the right hand side involving even powers of sinθ are real while those with odd powers sinθ are imaginary. Besides, we know that $\sin^{2m}\theta = \left(1-\cos^2\theta\right)^m$, $m \in \mathbb{N}$. Consequently, for any natural n we can write

$$T_n(\cos\theta) := \cos(n\theta),$$

where $T_n(x) := \cos(n\ \arccos(x)) = \alpha_0^{(n)} + \alpha_1^{(n)}x + \ldots + \alpha_n^{(n)}x^n$ is the Chebyshev's polynomial of order (degree) n which is an algebraic polynomial of degree n with real coefficients. Obviously,

$$\begin{aligned} T_0(x) &= 1, \quad T_1(x) = x, \quad T_2(x) = 2x^2 - 1, \quad T_3(x) = 4x^3 - 3x, \\ T_4(x) &= 8x^4 - 8x^2 + 1, \dots . \end{aligned}$$

It follows that every even trigonometric polynomial

$$Q_n(\theta) := \frac{\alpha_0}{2} + \sum_{k=1}^{n} \alpha_k \cos(k\theta),$$

is transformed, with the aid of substitution $\theta = \arccos x$, into the corresponding algebraic polynomial of degree n

$$P_n(x) := Q_n(\arccos x) = \frac{\alpha_0}{2} + \sum_{k=1}^{n} \alpha_k \cos(k\ \arccos(x)).$$

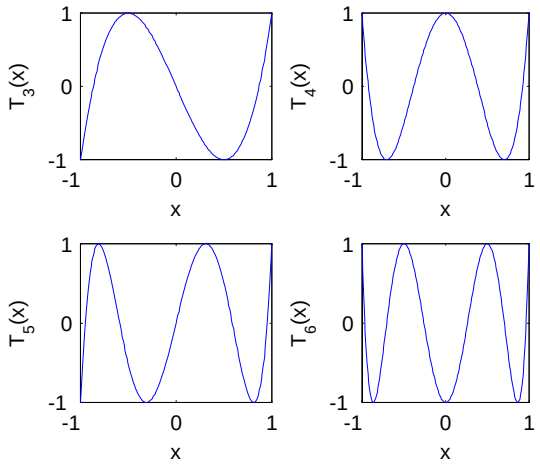

Figure 1.1: Some Chebyshev polynomials

This substitution specifies in fact a homeomorphic, continuous and one-to-one mapping of the closed interval $[0,\pi]$ *onto* $[-1,1]$. *It is important that, conversely, the substitution,* $x = cos\theta$, *transforms an arbitrary algebraic polynomial*

$$P_n(x) := a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n,$$

of degree n *into an even trigonometric polynomial*

$$Q_n(\theta) = P_n(\cos\theta) = \frac{\alpha_0}{2} + \sum_{k=1}^{n} \alpha_k \cos(k\theta),$$

where the coefficients α_k *depend on* P_n. *Indeed, we have*

$$\begin{aligned} \cos^m x &= \left(\frac{e^{ix}+e^{-ix}}{2}\right) = \frac{1}{2^m}\left(e^{imx} + \left(\frac{m}{1}\right)e^{i(m-2)x} + \ldots + e^{-imx}\right) = \\ &= \frac{1}{2^m}\left(\cos mx + \left(\frac{m}{1}\right)\cos(m-2)x + \ldots + \cos(-mx)\right). \end{aligned}$$

Here we should take into account that $\cos^m x$ *is a real function and therefore the last term in this chain of equalities is obtained from the preceding term by taking its real part. The imaginary part of* $\cos^m x$ *is automatically set to zero. Some Chebyshev polynomials are depicted in Fig. 1.1.*

Proposition 3 (Orthogonality) *The polynomials* $T_n(x)$ *are orthogonal, i.e.,*

$$(T_n, T_m)_{0,\omega} = \frac{\pi}{2} c_n \delta_{n,m}, \;\; m,n \in \mathbb{N},$$

where $\delta_{n,m}$ stands for the Kronecker delta symbol and, throughout this work, the coefficients c_n are defined by

$$c_n := \begin{cases} 0, & n < 0 \\ 2, & n = 0 \\ 1, & n \geq 1. \end{cases} \tag{1.2}$$

This fundamental property of Chebyshev polynomials, the recurrence relation

$$T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x), \ k > 0, \quad T_0(x) = 1, \quad T_1(x) = x, \tag{1.3}$$

as well as the estimations

$$\begin{aligned} |T_k(x)| &\leq 1, \quad |x| \leq 1, \quad T_k(\pm 1) = (\pm 1)^k, \\ |T_k'(x)| &\leq k^2, \quad |x| \leq 1, \quad T_k'(\pm 1) = (\pm 1)^k k^2, \end{aligned} \tag{1.4}$$

are direct consequences of the definition.

Remark 4 *As it is well known from the approximation theorem of Weierstrass, the set of orthogonal polynomials $\{T_n(x)\}_{n\in\mathbb{N}}$ is also complete in the space $L^2_\omega(I)$ and, consequently, each and every function u from this space can be expanded in a Chebyshev series as follows*

$$u(x) = \sum_{k=0}^{\infty} \widehat{u}_k \cdot T_k(x), \tag{1.5}$$

where the coefficients $\widehat{u}_k$ are

$$\widehat{u}_k = \frac{(u, T_k)_{0,\infty}}{\|T_k\|^2_{0,\omega}} = \frac{2}{\pi c_k} (u, T_k)_{0,\omega}.$$

Some other properties of Chebyshev polynomials are available, for instance, in the well known monographs Atkinson [10] and Raltson and Rabinowitz [171]. In [171], p.301, the following theorem is proved

Theorem 5 *Of all polynomials of degree r with coefficient of x^r equal to 1, the Chebyshev polynomial of degree r multiplied by $1/2^{r-1}$ oscillates with minimum maximum amplitude on the interval $[-1, 1]$.*

Due to this property the Chebyshev polynomials are sometimes called *equal-ripple polynomials.*

However, their importance in numerical analysis and in general, in scientific computation, is enormous and it appears in fairly surprising domains. For instance, in the monograph [39] p.162, a procedure currently in use for *accelerating the convergence of an iterative method,* making use of Chebyshev polynomials is considered.

Remark 6 ***Best approximation with Chebyshev polynomials*** *V. N. Murty shows in his paper [147] that there exists a unique best approximation of* $T_1(x)$ *with respect to linear space spanned by polynomials of odd degree* ≥ 3*, which is also a best approximation of* $T_1(x)$ *with respect to the linear space spanned by* $\{T_j(x)\}_{j=0,j\neq 1}^{n}$*. If* $n = 4k$ *or* $n = 4k-1$*, the extreme points of the deviation of* $T_1(x)$ *from its best approximation are* $2k$ *in number, whereas if* $n = 4k+1$ *or* $n = 4k+2$*, this number is* $2k+2$*.*

In the next section we try to introduce the Chebyshev polynomials in a more natural way. We advocate that the Fourier series is intimately connected with the Chebyshev series, and that some known convergence properties of the former provide valuable results for the latter.

1.2 Fourier and Chebyshev Series

> *The most important feature of Chebyshev series is that their convergence properties are not affected by the values of* $f(x)$ *or its derivatives at the boundaries* $x = \pm 1$ *but only by the smoothness of* $f(x)$ *and its derivatives throughout* $-1 \leq x \leq 1$*.*
>
> *Gottlieb and Orszag, [90], P. 28*

1.2.1 The trigonometric Fourier series

It is well known that the '*trigonometric polynomial*'

$$p_N(x) := \frac{1}{2}a_0 + \sum_{k=1}^{N}\left(a_k \cos kx + b_k \sin kx\right), \tag{1.6}$$

with

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx dx,\ b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx dx,$$

can be thought of as a *least square approximation* to $f(x)$ with respect to the unit weight function on $[-1, 1]$ (see Problem 6).

The Fourier series, obtained by letting $n \to \infty$ in (1.6), is apparently most valuable for the approximation of functions of period 2π. Indeed, for certain classes of such functions the series will converge for most values of x in the complete range $-\infty \leq x \leq +\infty$.

However, unless $f(x)$ and all its derivatives have the same values at $-\pi$ and π, there exists a 'terminal discontinuity' of some order at these points. The rate of convergence of the Fourier series, that is the rate of decrease of its coefficients, depends on the degree of smoothness of the function, measured by the order of the derivative which first becomes discontinuous at any point in the closed interval $[-\pi,\ \pi]$. Finally, we might be interested in a function defined

only in the range $[0, \pi]$, being then at liberty to extend its definition to the remainder of the periodic interval $[-\pi, 0]$ in any way we please.

It is worth noting that, integrating by parts in the expressions of a_k and b_k over $[0, \pi]$ we deduce that cosine series converge ultimately like k^{-2}, and sine series like k^{-1}, unless $f(x)$ has some special properties. If $f(0) = f(\pi) = 0$, we can show that sine series converges like k^{-3}, in general, the fastest possible rate for Fourier series.

1.2.2 The Chebyshev series

The terminal discontinuity of Fourier series of a non-periodic function can be avoided with the Chebyshev form of Fourier series. We consider the range $-1 \leq x \leq 1$ and make use of the change of variables

$$x = \cos\theta,$$

so that

$$f(x) = f(\cos\theta) = g(\theta). \tag{1.7}$$

The new function $g(\theta)$ is even and genuinely periodic, since $g(\theta) = g(\theta + 2\pi)$. Moreover, if $f(x)$ has a large numbers of derivatives in $[-1, 1]$, then $g(\theta)$ has similar properties in $[0, \pi]$. We should then expect the cosine Fourier series

$$g(\theta) = \frac{1}{2}a_0 + \sum_{k=1}^{N} a_k \cos k\theta, \; a_k = \frac{2}{\pi c_k}\int_{-1}^{1} g(\theta)\cos k\theta d\theta \tag{1.8}$$

to converge fairly rapidly. Interpreting (1.8) in terms of original variable x, we produce the following **Chebyshev series**

$$f(x) = a_0 + \sum_{k=1}^{\infty} a_k T_k(x), \; a_k = \frac{2}{\pi c_k}\int_{-1}^{1} \omega(x) f(x) T_k(x)\, dx, \; \omega(x) := \left(1 - x^2\right)^{-1/2}. \tag{1.9}$$

This series has the same convergence properties as the Fourier series for $f(x)$, with the advantage that the terminal discontinuities are eliminated. Elementary computations show that, for sufficiently smooth functions, the coefficients a_k have the order of magnitude $1/2^{k-1}(k!)$, considerably smaller for large k than the k^{-3} of the best Fourier series.

Remark 7 (Continuous least square approximation) *The expansion*

$$p_n(x) := \sum_{k=0}^{n} a_k T_k(x), \; a_k = \frac{2}{\pi c_k}\int_{-1}^{1} \omega(x) f(x) T_k(x)\, dx,$$

has the property that the error $e_n(x) := f(x) - p_n(x)$ satisfies the 'continuous' least square condition

$$S := \int_{-1}^{1} \omega(x) e_n^2(x) dx = \min.$$

The minimum value is given by

$$S_{\min} = \int_{-1}^{1} \omega(x) f^2(x) dx - \frac{1}{2}\pi \left(\sum_{k=0}^{n} c_k a_k^2 \right).$$

As $n \to \infty$, we produce the Chebyshev series, which has the same convergence properties as the Fourier series, but generally with a much faster rate of convergence.

1.2.3 Discrete least square approximation

We now move on to the *discrete case of least square approximation* in which the integrated mean square error over I, from the classical least square approximation, is replaced by a sum over a finite number of nodes, say $x_0,\ x_1, ..., x_N \in I$. The function $f(x),\ f : I \to \mathbb{R}$ is approximated by a polynomial $p(x)$ with the error $e(x) := f(x) - p(x)$ and find the polynomial $p(x)$ such that the sum

$$S := \sum_{k=0}^{N} \omega(x_k) e^2(x_k),$$

attains its minimum with respect to the position of the nodes x_k in $[-1, 1]$ and for a specified class of polynomials. We seek an expansion of the form

$$p_N(x) := \sum_{r=0}^{N} a_r \psi_r(x),$$

where the functions $\psi_r(x)$ are, at this stage, arbitrary members of some particular system (should that consist of polynomials, trigonometric functions, etc.). Conditions for a minimum are now expressed with respect to the coefficients $a_r,\ S = S(a_0, a_1, ..., a_r)$. They are

$$\partial S / \partial a_i = 0,\ i = 0, 1, 2, .., N$$

and they produce a set of linear algebraic equations for these quantities. The matrix involved is diagonal if the functions are chosen to satisfy the *discrete orthogonality conditions*

$$\sum_{k=0}^{N} \omega(x_k) \psi_r(x_k) \psi_s(x_k) = 0,\ r \neq s.$$

The corresponding coefficients a_r are then given by

$$a_r = \frac{\sum_{k=0}^{N} \omega(x_k) \psi_r(x_k) f(x_k)}{\sum_{k=0}^{N} \omega(x_k) \psi_r^2(x_k)},\ r = 0, 1, 2, ..., N,$$

and the minimum value of S is

$$S_{\min} = \sum_{k=0}^{N} \omega(x_k) \left\{ f^2(x_k) - \sum_{k=0}^{N} a_r^2 \psi_r^2(x_k) \right\}.$$

1.2.4 Chebyshev discrete least square approximation

Let's consider a particular case relevant for the Chebyshev theory.

In the trigonometric identity

$$\frac{1}{2}+\cos\theta+\cos 2\theta+...+\cos(N-1)\theta+\frac{1}{2}\cos N\theta=\frac{1}{2}\sin N\theta\cot\frac{\theta}{2}, \quad (1.10)$$

the right-hand side vanishes for $\theta=k\pi/N,\ k\in\mathbb{Z}$. Since

$$2\cos r\theta\cos s\theta=\cos\left(r+s\right)\theta+\cos\left(r-s\right)\theta, \quad (1.11)$$

it follows that the set of linearly independent functions $\psi_r\left(\theta\right)=\cos r\theta$ satisfy the discrete orthogonality conditions

$$\sum_{k=0}^{N}\frac{1}{\overline{c}_k}\psi_r\left(\theta_k\right)\psi_s\left(\theta_k\right)=0,\ r\neq s,\ \theta_k=k\pi/N, \quad (1.12)$$

where, throughout in this work, the coefficients $\overline{c}_k$ are defined by

$$\overline{c}_k:=\begin{cases}2,\ k=0,N,\\ 1,\ 1\leq k\leq N-1.\end{cases}$$

Further, we find from (1.10) and (1.11) that the normalization factors for these orthogonal functions are

$$\sum_{k=0}^{N}\frac{1}{\overline{c}_k}\psi_r^2\left(\theta_k\right)=\begin{cases}N/2,\ k=0,N,\\ N,\ 1\leq k\leq N-1.\end{cases} \quad (1.13)$$

Consequently, for the function $g\left(\theta\right),\ \theta\in[0,\pi]$, a trigonometric (Fourier) discrete least square approximation, over equally spaced nodes

$$\theta_k=k\pi/N,\ k=0,1,2,...,N,$$

is given by the 'interpolation' polynomial

$$p_N(\theta)=\sum_{r=0}^{N}\frac{1}{\overline{c}_r}a_r\cos r\theta,\ a_r=\sum_{k=0}^{N}\frac{2}{N\overline{c}_k}g\left(\theta_k\right)\cos r\theta_k,\ \theta_k=\frac{k\pi}{N}. \quad (1.14)$$

The corresponding *Chebyshev discrete least square approximation* follows immediately using (1.7). It reads

$$p_N(x)=\sum_{r=0}^{N}\frac{1}{\overline{c}_r}a_rT_r(x),\ a_r=\sum_{k=0}^{N}\frac{2}{N\overline{c}_k}f\left(x_k\right)T_r\left(x_k\right),\ x_k=\cos\left(\frac{k\pi}{N}\right). \quad (1.15)$$

Let us observe that the nodes x_k are not equally spaced in $[-1,1]$. The nodes

$$\theta_k=\frac{k\pi}{N},\ k=1,2,...,N-1,$$

are the *turning points* (extrema points) of $T_N\left(x\right)$ on $[-1,1]$ and they are called the *Chebyshev points of the second kind.*

Remark 8 *For the expansion (1.15), the error* $e_N(x) := f(x) - p_N(x)$ *satisfies the 'discrete' least square condition*

$$S := \sum_{k=0}^{N} \frac{1}{\overline{c}_k} e_N^2(x_k) = \min,$$

and

$$S_{\min} = \sum_{k=0}^{N} \frac{1}{\overline{c}_k} \left\{ f^2(x_k) - \sum_{r=0}^{N} a_r^2 T_r^2(x_k) \right\}.$$

1.2.5 Orthogonal polynomials least square approximation

We have to notice that, so far, we have not used the orthogonality properties of the Chebyshev polynomials, with respect to scalar product (1.1). Similar particular results can be found using this property. For the general properties of orthogonal polynomials we refer to the monographs [43] or [187].

Each and every set of such polynomials satisfies a *three-term recurrence relation*

$$\phi_{r+1}(x) = (\alpha_r x + \beta)\phi_r(x) + \gamma_{r-1}\phi_{r-1}(x), \tag{1.16}$$

with the coefficients

$$\alpha_r = \frac{A_{r+1}}{A_r},\ \gamma_{r-1} = -\frac{A_{r+1}}{A_r}\frac{A_{r-1}}{A_r}\frac{k_r}{k_{r-1}},$$

where A_r is the coefficient of x^r in $\phi_r(x)$, and

$$k_r = \int_{-1}^{1} \omega(x)\phi_r^2(x)\,dx.$$

Following Lanczos [126], we choose the normalization $k_r = 1$, and write (1.16) in the form

$$p_{r-1}\phi_{r-1}(x) + (-x + q_r)\phi_r(x) + p_r\phi_{r+1}(x) = 0, \tag{1.17}$$

with

$$p_r = \frac{A_r}{A_{r+1}},\quad q_r = -\beta_r p_r.$$

If we define $p_{-1}(x) := 0$ and choose the $N+1$ nodes $x_i,\ i = 0, 1, 2, ..., N$ so that they are the zeros of the orthogonal polynomial $\phi_{N+1}(x)$, we see that they are also the eigenvalues of the tridiagonal matrix $diag(p_{k-1}\ q_k\ p_k)$. The eigenvector corresponding to the eigenvalue x_k has the components $\phi_0(x_k),\ \phi_1(x_k),\ ..., \phi_N(x_k)$ and from the theory of symmetric matrices we know that the set of these vectors forms an independent orthonormal system. Each and every vector is *normalized to be a unit vector*, i.e.,

$$\lambda_k \sum_{r=0}^{N} \phi_r^2(x_k) = 1,$$

and the matrix

$$X = \begin{pmatrix} \lambda_0^{1/2}\phi_0(x_0) & \lambda_1^{1/2}\phi_0(x_1) & \dots & \lambda_N^{1/2}\phi_0(x_N) \\ \lambda_0^{1/2}\phi_1(x_0) & \lambda_1^{1/2}\phi_1(x_1) & \dots & \lambda_N^{1/2}\phi_1(x_N) \\ \dots & \dots & \dots & \dots \\ \lambda_0^{1/2}\phi_N(x_0) & \lambda_1^{1/2}\phi_N(x_1) & \dots & \lambda_N^{1/2}\phi_N(x_N) \end{pmatrix}$$

is orthogonal. It means $X \cdot X' = X' \cdot X = I_{N+1}$, which implies two more discrete conditions in addition to the normalization one. i.e.,

$$\begin{array}{c} \sum_{k=0}^{N} \lambda_k \phi_r^2(x_k) = 1,\ r = 0,1,2,...,N \\ \sum_{k=0}^{N} \lambda_k \phi_r(x_k)\,\phi_s(x_k) = 0,\ r \neq s. \end{array} \tag{1.18}$$

It follows that a solution of the least square problem in this case, with weights $\omega(x_k) = \lambda_k$, and the nodes taken as the $N+1$ zeros of $\phi_{N+1}(x)$, is given by

$$p_N(x) = \sum_{r=0}^{N} a_r \phi_r(x),\ a_r = \sum_{k=0}^{N} \lambda_k f(x_k)\,\phi_r(x_k). \tag{1.19}$$

For the Chebyshev case, using weight function $\omega(x) = \left(1-x^2\right)^{-1/2}$, we find

$$\phi_0(x) = \pi^{-1/2} T_0(x),\ \phi_r(x) = \left(\tfrac{1}{2}\pi\right)^{-1/2} T_r(x),\ r = 0,1,2,...$$

$$\lambda_k^{-1} = \tfrac{2}{\pi} \sum_{r=0}^{N} \tfrac{1}{c_k} T_r^2(x_k) = \tfrac{2}{\pi} \sum_{r=0}^{N} \tfrac{1}{c_k} \cos^2 r\theta_k,$$

$$\theta_k = \tfrac{2k+1}{N+1}\tfrac{\pi}{2},\ k = 0,1,2,...,N.$$

The trigonometric identity (1.10) leads to a very simple form of λ_k, namely

$$\lambda_k = \pi/(N+1),$$

and finally to

$$\begin{array}{c} p_N(x) = \sum_{r=0}^{N} \frac{1}{c_r} b_r T_r(x), \\ b_r = \frac{2}{N+1} \sum_{k=0}^{N} f(x_k)\, T_r(x_k),\ x_k = \cos\left(\frac{2k+1}{N+1}\frac{\pi}{2}\right),\ k = 0,1,2,...,N. \end{array} \tag{1.20}$$

Remark 9 *For the expansion (1.20), the error $e_N(x) := f(x) - p_N(x)$ satisfies the 'discrete' least square condition*

$$S := \sum_{k=0}^{N} e_N^2(x_k) = \min,$$

and

$$S_{\min} = \sum_{k=0}^{N} \left\{ f^2(x_k) - \sum_{r=0}^{N} a_r^2 T_r^2(x_k) \right\}.$$

It can be shown that the error $e_N(x)$ *satisfies the following minmax criterion for sufficiently smooth functions*

$$\max \left| e_N(x) / f^{(N+1)}(\xi) \right| = \min, \quad \xi \in (-1,1).$$

Remark 10 *The least square approximation polynomial* $p_N(x)$ *from* (1.20) *must agree with the Lagrangian interpolation polynomial*

$$p_N(x) = \sum_{k=0}^{N} l_k(x) f(x_k),$$

(see Appendix 1) which uses as nodes the Chebyshev points of the first kind $x_k = \cos\left(\frac{2k+1}{N+1}\frac{\pi}{2}\right)$, $k = 0, 1, 2, ..., N$. *These nodes are in fact the zeros of* $T_{N+1}(x)$.

Remark 11 *In* [71] *it is shown that for sufficiently well-behaved functions* $f(x)$ *the approximation formula* (1.20) *is slightly better than* (1.15).

1.2.6 Orthogonal polynomials and Gauss-type quadrature formulas

There exists an important connection between the weights λ_k of the orthogonal polynomial discrete least square approximation and the corresponding *Gauss type quadrature formulas.* First, we notice that *Lagrangian quadrature formula* (see Appendix 1) reads

$$\int_{-1}^{1} \omega(x) f(x) dx = \sum_{k=0}^{N} \mu_k f(x_k), \tag{1.21}$$

where

$$\mu_k = \int_{-1}^{1} \omega(x) l_k(x) dx.$$

The polynomial

$$p_N(x) = \sum_{k=0}^{N} l_k(x) f(x_k), \tag{1.22}$$

fits $f(x)$ exactly in the $N+1$ zeros of $\Pi(x)$ and has degree N. The formula (1.21) is exact for polynomials of degree N or less.

A Gauss quadrature formula has the form

$$\int_{-1}^{1} \omega(x) f(x) dx = \sum_{k=0}^{N} \nu_k f(x_k), \tag{1.23}$$

where the weights ν_k and abscissae x_k (quadrature nodes) are to be determined such that the formula should be exact for polynomials of as high a degree as

possible. Since there are $2N+2$ parameters in the above formula, we should expect to be able to make (1.23) exact for polynomials of degree $\leq 2N+1$.

To this end, we consider a system of polynomials $\phi_k(x)$, $k=0,1,2,...,N$ which satisfy the "continuous" orthogonality conditions

$$\int_{-1}^{1} \omega(x)\,\phi_r(x)\,\phi_s(x)\,dx = 0,\ r \neq s. \tag{1.24}$$

Suppose that $f(x)$ is a polynomial of degree $2N+1$ and write it in the form

$$f(x) = q_N(x)\,\phi_{N+1}(x) + r_N(x), \tag{1.25}$$

where the suffices indicate the degrees of the polynomial involved. Since $q_N(x)$ can be expressed as a linear combination of orthogonal polynomials $\phi_k(x)$, $k=0,1,2,...,N$, the orthogonality relations imply

$$\int_{-1}^{1} \omega(x)\,f(x)\,dx = \int_{-1}^{1} \omega(x)\,r_N(x)\,dx,$$

which by (1.21) is exactly, i.e.,

$$\int_{-1}^{1} \omega(x)\,f(x)\,dx = \sum_{k=0}^{N} \mu_k r_N(x_k),$$

for specified x_k and corresponding μ_k. If we choose x_k to be the zeros of $\phi_{N+1}(x)$, it follows from (1.25) that we obtained formally the required Gauss quadrature formula (1.23) with $\nu_k = \mu_k$. Now $r_N(x)$, as a polynomial of degree N can be represented exactly, due to (1.19), in the form

$$r_N(x) = \sum_{k=0}^{N} a_r \phi_r(x).$$

Consequently, we can write

$$\int_{-1}^{1} \omega(x)\,r_N(x)\,dx = \int_{-1}^{1} \omega(x) \left(\sum_{k=0}^{N} a_r \phi_r(x) \right) dx = a_0\phi_0 \int_{-1}^{1} \omega(x),$$

due to (1.24) with $r=0$. Moreover, the general solution of the least square problem (1.19) and in particular, the normalization condition, imply

$$a_0\phi_0 \int_{-1}^{1} \omega(x) = \sum_{k=0}^{N} \lambda_k f(x_k) \int_{-1}^{1} \omega(x)\,\phi_0^2 dx = \sum_{k=0}^{N} \lambda_k f(x_k),$$

or, more explicitly

$$\int_{-1}^{1} \omega(x)\,f(x)\,dx = \sum_{k=0}^{N} \lambda_k f(x_k).$$

It follows that the weights in Gauss quadrature formula (1.23), which is exact for polynomials of order $2N+1$, equal the weights λ_k of the discrete least square solution (1.19), and the nodes x_k are the zeros of the relevant orthogonal polynomial $\phi_{N+1}(x)$.

If, in particular,

$$\phi_0(x) := (\pi)^{-1/2} T_0(x), \quad \phi_r(x) := \left(\frac{1}{2}\pi\right)^{-1/2} T_r(x), \ r = 1, 2, ...$$

we get the *Gauss-Chebyshev quadrature formula, i.e.,*

$$\int_{-1}^{1} \omega(x) f(x) dx = \frac{\pi}{N+1} \sum_{k=0}^{N} f(x_k), \ x_k = \cos\left(\frac{2k+1}{N+1}\frac{\pi}{2}\right). \tag{1.26}$$

1.3 Chebyshev projection

Let us introduce the map $P_N : L^2_\omega(I) \to \mathcal{P}_N, \ I = [-1, 1]$,

$$P_N u(x) := \sum_{k=0}^{N} \widehat{u}_k \cdot T_k(x), \tag{1.27}$$

where the coefficients $\widehat{u}_k, \quad k = 1, 2, \ldots, N$ are defined in (1.5). Due to the orthogonality properties of Chebyshev polynomials, $P_N u(x)$ represents the **orthogonal projection** of function u onto $\mathcal{P}_N$ with respect to scalar product (1.1). Consequently, we can write

$$(P_N u(x), v(x))_\omega = (u(x), v(x))_\omega, \quad \forall v \in \mathcal{P}_N. \tag{1.28}$$

More than that, due to the completeness of the set of Chebyshev polynomials, the following limit holds:

$$\|u - P_N u\|_\omega \to 0 \quad as \quad N \to \infty.$$

Remark 12 *A lot of results concerning the general theory of approximation by polynomials are available in Chapter 9 of [33]. We extract from this source only the results we strictly use.*

The quantity $u - P_N u$ is called *truncation error* and for it we have the following estimate.

Lemma 13 *For each and every $u \in H^s_\omega(I), \ s \in \mathbb{N}$, one has*

$$\|u - P_N u\|_\omega \le C N^{-s} \|u\|_{s,\omega}, \tag{1.29}$$

where the constant C is independent of N and u.

Remark 14 *There exists a more general result which reads*

$$\|u - P_N u\|_\omega \leq C\ \sigma_N(p)\, N^{-m} \sum_{k=0}^{m} \left\|u^{(k)}\right\|_\omega,$$

for a function u *that belongs to* $L^2_\omega(-1,1)$ *along with its distributional derivatives of order* m *and* $\sigma_N(p) = \begin{cases} 1,\ 1 < p < \infty, \\ 1 + \log N,\ p = 1 \text{ and } p = \infty. \end{cases}$

Remark 15 *Unfortunately, the approximation using the Chebyshev projection is optimal only with respect to the scalar product* $(\cdot,\cdot)_{0,\omega}$. *This statement is confirmed by the estimation*

$$\|u - P_N u\|_\omega \leq C N^{2l-s-\frac{1}{2}} \|u\|_{s,\omega},\ s \geq l \geq 1,$$

in which a supplementary quantity $\left(l - \frac{1}{2}\right)$ *appears in the power of* N. *To avoid this inconvenient Canuto et al. [33] [1988, Ch. 9,11] introduced orthogonal projections with respect to other scalar products.*

Remark 16 *If (1.5) is the Chebyshev series for* $u(x)$, *the same series for the derivative of* $u \in H^1_\omega(I)$, *has the form*

$$u'(x) = \sum_{k=0}^{\infty} \widehat{u}_k^{(1)} \cdot T_k(x), \tag{1.30}$$

where (see (1.56) in the Problem 10)

$$\widehat{u}_k^{(1)} = \frac{2}{c_k} \sum_{\substack{p = k+1 \\ p+k = odd}}^{\infty} p\widehat{u}_k.$$

Consequently,

$$P_N(u') = \sum_{k=0}^{N} \widehat{u}_k^{(1)} \cdot T_k(x),$$

but in applications is sometimes used the derivative of the projection, namely $(P_N u)'$, *which is called the 'Chebyshev-Galerkin derivative'.*

We end this section with some '*inverse inequalities*' concerning summability and differentiability for algebraic polynomials.

Lemma 17 *For each and every* $u \in \mathcal{P}_N$, *we have*

$$\begin{array}{l} \|u\|_{L^q_\omega(-1,1)} \leq C N^{2\left(\frac{1}{p}-\frac{1}{q}\right)} \|u\|_{L^p_\omega(-1,1)},\ 1 \leq p \leq q \leq \infty, \\ \left\|u^{(r)}\right\|_{L^p_\omega(-1,1)} \leq C N^{2r} \|u\|_{L^p_\omega(-1,1)},\ 2 \leq p \leq \infty,\ r \geq 1. \end{array} \tag{1.31}$$

1.4 Chebyshev interpolation

We re-write the results from Fourier and Chebyshev Series Section in a more formal way.

First, we observe that the quadrature formulas represent a way to connect the space $L^2_\omega(-1,1)$ with the space of polynomials of a specified degree. For the sake of precision, the *interpolation nodes* will be furnished by following *Chebyshev-Gauss quadrature formula (rule)*

$$\int_{-1}^{1} f(x)\,\omega(x)\,dx := \sum_{j=0}^{N} f(x_j)\,\omega_j,$$

where the choices for the nodes x_j and the weights ω_j lead to rules which have different orders of precision. The most frequently encountered rules are:

1. the *Chebyshev-Gauss* formula (CGauss)

$$x_j := \cos\frac{(2j+1)\,\pi}{2N+1} \quad and \quad \omega_j = \frac{\pi}{N+1}, \quad j = 0,1,2,\ldots,N \tag{1.32}$$

   The quadrature nodes are the roots of the Chebyshev polynomial T_{N+1} and the formula is exact for polynomials in $\mathcal{P}_{2N+1}$.

2. the *Chebyshev-Gauss-Radau* formula (CGaussR)

$$x_j := \cos\frac{2j\pi}{2N+1} \quad j = 0,1,2,\ldots,N \ and \ \omega_j = \begin{cases} \frac{\pi}{N+1}, & j = 0, \\ \frac{\pi}{2N+2}, & j = 1,2,\ldots,N. \end{cases} \tag{1.33}$$

   In this case, the order of precision is only $2N$.

3. the *Chebyshev-Gauss-Lobatto* formula (CGaussL)

$$x_j := \cos\frac{j\pi}{N} \quad j = 0,1,2,\ldots,N \quad and \quad \omega_j = \begin{cases} \frac{\pi}{2N}, & j = 0 \ and \ j = N, \\ \frac{\pi}{N}, & j = 1,2,\ldots,N-1. \end{cases} \tag{1.34}$$

   In this case, the order of precision diminishes to $2N-1$.

Corresponding to each and every formula above we introduce a *discrete scalar (inner) product* and a norm as follows:

$$(u,v)_N := \sum_{j=0}^{N} \omega_j u(x_j)\,v(x_j), \tag{1.35}$$

$$\|u\|_N := \left(\sum_{j=0}^{N} \omega_j u^2(x_j)\right)^{\frac{1}{2}}. \tag{1.36}$$

The next result is due to Quarteroni and Vali [169], Ch.5.

Lemma 18 *For the set of Chebyshev polynomials, there holds*

$$\|T_k\|_N = \|T_k\|_\omega\,,\ k=0,1,2,...,N-1,\ \|T_N\|_N = \begin{cases} \|T_N\|_\omega\,, & for\ CGauss \\ \sqrt{2}\,\|T_N\|_\omega\,, & for\ CGaussL. \end{cases}$$

Proof. The first two equalities are direct consequences of the order of precision of quadrature formulas. For the third, we can write

$$\|T_N\|_N^2 = \frac{\pi}{2N}\left(\cos^2 0 + \cos^2 \pi\right) + \frac{\pi}{N}\sum_{j=1}^{N-1}\cos^2 j\pi = \pi = 2\,\|T_N\|_\omega^2\,.$$

■

Let $I_N u \in \mathcal{P}_N$ the *interpolation polynomial* of order (degree) N corresponding to one of the above three sets of nodes x_k. It has the form

$$I_N u = \sum_{k=0}^{N} u_k T_k\,(x)\,, \tag{1.37}$$

where the coefficients are to be determined and are called the 'degrees of freedom' of u in the *transformed space (* called also *"phase" space).* For the *(CGaussL)* choice of nodes, using the discrete orthogonality and normality conditions (1.12), (1.13) we have

$$(I_N u, T_k)_N = \sum_{p=0}^{N} u_p\,(T_p, T_k)_N = \frac{\overline{c}_k \pi}{2} u_k. \tag{1.38}$$

But interpolation means

$$I_N u\,(x_j) = u\,(x_j)\,,\ j=0,1,2,...,N,$$

which implies

$$(I_N u, T_n)_N = (u, T_n)_N = \sum_{j=0}^{N} \frac{\pi}{\overline{c}_j N} u(x_j)\cos\frac{nj\pi}{N}. \tag{1.39}$$

The identities (1.38) and (1.39) lead to the 'discrete Chebyshev transform'

$$u_k = \frac{2}{\overline{c}_k N}\sum_{j=0}^{N}\frac{1}{\overline{c}_j} u(x_j)\cos\frac{kj\pi}{N},\ k=0,1,2,...,N. \tag{1.40}$$

Making use of this transformation, we can pass from the set of values of the function u in the nodes $(CGaussL)$, the so-called *physical space,* to the *transformed space.* The inverse transform reads

$$u(x_j) = \sum_{j=0}^{N} u_j \cos\frac{kj\pi}{N},\ j=0,1,2,...,N. \tag{1.41}$$

Due to their trigonometric structure, these two transformations can be carried out using FFT (fast Fourier transform-see [33] Appendix B, or [40] and [41]).

A direct consequence of the last lemma is the equivalence of the norms $\|\cdot\|_\omega$ and $\|\cdot\|_N$. Thus, in the $(CGaussL)$ case, for $u^N = \sum_{k=0}^{N} u_k T_k$ we can write

$$\left\|u^N\right\|_N^2 = \sum_{k=0}^{N} (u_k)^2 \|T_k\|_N^2 = \sum_{k=0}^{N-1} (u_k)^2 \|T_k\|_\omega^2 + 2 (u_N)^2 \|T_N\|_\omega^2 ,$$

and

$$\left\|u^N\right\|_\omega^2 = \sum_{k=0}^{N} (u_k)^2 \|T_k\|_\omega^2 .$$

Consequently, we get the sequence of inequalities

$$\left\|u^N\right\|_\omega \leq \left\|u^N\right\|_N \leq \sqrt{2} \left\|u^N\right\|_\omega .$$

For the Chebyshev interpolation, in each and every case, (CG), (CGR), (CGL), we have the following result (see [33], Ch. 9 and [169] Ch. 4):

Lemma 19 *If $u \in H_\omega^m(-1,1)$, $m \geq 1$, then the following estimation holds*

$$\|u - I_N u\|_\omega \leq C N^{-m} \|u\|_{m,\omega} , \tag{1.42}$$

and if $0 \leq l \leq m$, then a less sharp one holds, namely

$$\|u - I_N u\|_{l,\omega} \leq C N^{2l-m} \|u\|_{m,\omega} . \tag{1.43}$$

In $L_\omega^\infty(-1,1)$, we have the estimation

$$\|u - I_N u\|_{L_\omega^\infty} \leq C N^{2l-m} \|u\|_{m,\omega} . \tag{1.44}$$

1.4.1 Collocation derivative operator

Associated with an interpolator is the concept of a *collocation derivative (differentiation) operator* called also *Chebyshev collocation derivative* or even *pseudospectral derivative.* The idea is summarized in [184]. Suppose we know the value of a function at several points (nodes) and we want to approximate its derivative at those points. One way to do this is to find the polynomial that passes through all of data points, differentiate it analytically, and evaluate this derivative at the grid points.

In other words, the derivatives are approximated by exact differentiation of the interpolate.

Since interpolation and differentiation are linear operations, the process of obtaining approximations to the values of the derivative of a function at a set of points can be expressed as a matrix-vector multiplication. The matrices involved are called *pseudospectral differentiation (derivation) matrices* or simply *differentiation matrices.*

Thus, if $u := (u(x_0)\ u(x_1)\ ...u(x_N))^T$ is the vector of function values, and $u' := (u'(x_0)\ u'(x_1)\ ...u'(x_N))^T$ is the vector of approximate nodal derivatives, obtained by this idea, then there exists a matrix, say $D^{(1)}$, such that

$$u' = D^{(1)}u. \tag{1.45}$$

We will deduce the matrix $D^{(1)}$ and the next differentiation matrix $D^{(2)}$ defined by

$$u'' = D^{(2)}u. \tag{1.46}$$

To get the idea we proceed in the simplest way following closely the paper of Solomonoff and Turkel [183].

Thus, if

$$L_N(x) := \sum_{k=0}^{N} u(x_k)\, l_k(x), \tag{1.47}$$

is the Lagrangian interpolation polynomial, we construct the first differentiation matrix $D^{(1)}$ by analytically differentiating that. In particular, we shall explicitly construct $D^{(1)}$ by demanding that for *Lagrangian basis* $\{l_k(x)\}_{k=0}^{N}$, $l_k(x) \in \mathcal{P}_N$,

$$D^{(1)} l_k(x_j) = l_k'(x_j),\ j,k = 0,1,2,...,N,$$

i.e.

$$D^{(1)} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} l_k'(x_0) \\ \vdots \\ l_k'(x_k) \\ \vdots \\ l_k'(x_N) \end{pmatrix},$$

where 1 stands in the $k\ th$ row. Performing the multiplication, we get

$$d_{jk}^{(1)} = l_k'(x_j). \tag{1.48}$$

We have to evaluate explicitly the entries $d_{jk}^{(1)}$ in terms of the nodes x_k, $k = 0,1,2,...,N$. To this end, we rewrite the Lagrangian polynomials $l_k(x)$ in the form

$$l_k(x) := \frac{1}{\alpha_k} \Pi_{\substack{l=0 \\ l\neq k}}^{N} (x - x_l),\ \alpha_k := \Pi_{\substack{l=0 \\ l\neq k}}^{N} (x_k - x_l).$$

Taking, with a lot of care, the logarithm of $l_k(x)$ and differentiating, we obtain

$$l_k'(x) = l_k(x) \sum_{\substack{l=0 \\ l\neq k}}^{N} 1/(x - x_l). \tag{1.49}$$

This equality implies the diagonal elements

$$d_{kk}^{(1)}=\sum_{\substack{l=0\\l\neq k}}^{N}1/\left(x_k-x_l\right),\ k=1,2,...,N. \tag{1.50}$$

In order to evaluate (1.49) at $x=x_j$, $j\neq k$ we have to eliminate the 0/0 indetermination from the right hand side of that. We therefore write (1.49) as

$$l_k'\left(x\right)=l_k\left(x\right)/\left(x-x_j\right)+l_k\left(x\right)\sum_{\substack{l=0\\l\neq k,j}}^{N}1/\left(x-x_l\right).$$

Since $l_k\left(x_j\right)=0$ for $j\neq k$, we obtain that

$$l_k'\left(x_j\right)=\lim_{x\to x_j}\frac{l_k\left(x\right)}{\left(x-x_j\right)}.$$

Using the definition of $l_k\left(x\right)$, we get the off-diagonal elements, i.e.,

$$d_{jk}^{(1)}=\frac{1}{\alpha_k}\Pi_{\substack{l=0\\l\neq k,j}}^{N}\left(x_j-x_l\right)=\frac{\alpha_j}{\alpha_k\left(x_j-x_k\right)}. \tag{1.51}$$

It is sometimes preferable to express the entries of $D^{(1)}$, (1.50) and (1.51), in a different way. Let's denote by $\phi_{N+1}\left(x\right)$ the product $\Pi_{l=0}^{N}\left(x-x_l\right)$. Then we have successively

$$\begin{gathered}\phi_{N+1}'\left(x\right)=\sum_{k=0}^{N}\Pi_{\substack{l=0\\l\neq k}}^{N}\left(x-x_l\right),\\ \phi_{N+1}'\left(x_k\right)=\alpha_k,\\ \phi_{N+1}''\left(x_k\right)=2\alpha_k\sum_{\substack{l=0\\l\neq k}}^{N}1/\left(x_k-x_l\right),\end{gathered}$$

and eventually we can write

$$d_{jk}^{(1)}=\begin{cases}\frac{\alpha_j}{\alpha_k\left(x_j-x_k\right)}=\frac{\phi_{N+1}'\left(x_j\right)}{\phi_{N+1}'\left(x_k\right)\left(x_j-x_k\right)},\ j\neq k\\ \sum_{\substack{l=0\\l\neq k}}^{N}\frac{1}{\left(x_k-x_l\right)}=\frac{\phi_{N+1}''\left(x_k\right)}{2\phi_{N+1}'\left(x_k\right)},\ j=k.\end{cases} \tag{1.52}$$

Similarly, for the second derivative we write

$$D^{(2)}l_k\left(x_j\right)=l_k''\left(x_j\right),\ j,k=0,1,2,...,N,$$

and consequently

$$d_{jk}^{(2)}=\begin{cases}2d_{jk}^{(1)}\left[d_{jj}^{(1)}-\frac{1}{x_j-x_k}\right],\ j\neq k,\\ \left[d_{kk}^{(1)}\right]^2-\sum_{\substack{l=0\\l\neq k}}^{N}\frac{1}{\left(x_k-x_l\right)^2},\ j=k.\end{cases} \tag{1.53}$$

Remark 20 *In [206], a simple method for computing $n \times n$ pseudodifferential matrix of order p in $O\left(pn^2\right)$ operations for the case of quasi-polynomial approximation is carried out. The algorithm is based on recursions relations for the generation of finite difference formulas derived in [68]. The existence of efficient preconditioners for spectral differentiation matrices is considered in [72]. Simple upper bounds for the maximum norms of the inverse $\left(D^{(2)}\right)^{-1}$, corresponding to (CGaussL) points, are provided in [182]. In [189] it is shown that differentiating analytic functions using the pseudospectral Fourier or Chebyshev methods, the error committed decays to zero at an exponential rate.*

Remark 21 *The entries of the Chebyshev first derivative matrix can be found also in [93]. The gridpoints used by this matrix are x_j from* (1.34), *i.e., Chebyshev Gauss Lobato nodes. The entries $d_{jk}^{(1)}$ are*

$$\begin{aligned} d_{jk}^{(1)} &= \frac{\overline{c}_j}{\overline{c}_k} \frac{(-1)^{j+k}}{(x_j - x_k)}, \ j \neq k, \\ d_{jj}^{(1)} &= \frac{-x_j}{2\left(1-x_j^2\right)}, \ j \neq 0, N, \\ d_{00} &= -d_{NN} = \frac{2N^2+1}{6}. \end{aligned} \tag{1.54}$$

Remark 22 *The software suite provided in the paper of Weideman and Reddy [204] contains, among others, some codes (MATLAB $*.m$ functions) for carrying out the transformations (1.40) and (1.41), as well as for computing derivatives of arbitrary order corresponding to Chebyshev, Hermite, Laguerre, Fourier and sinc interpolators. It is observed that for the matrix $D^{(l)}$, which stands for the $l-th$ order derivative, is valid the recurrence relation*

$$D^{(l)} = \left(D^{(1)}\right)^l, \ l = 1, 2, 3, ...,$$

which is also suggested by (1.53). *The existence of this relation is a consequence of the barycentric form of the interpolator (see P. Henrici [110], P. 252). On the other hand, we have to observe that throughout this work we use standard notations, which means that interpolating polynomials are considered to have order N and sums to have lower limit $j = 0$ and upper limit N. Since MATLAB environment does not have a zero index the authors of these codes begin sums with $j = 1$ and consequently their notations involve polynomials of degree $N-1$. Thus, in formulas* (1.54) *instead of N they introduce $N - 1$. However, it is fairly important that, in these codes, the authors use extensively the vectorization capabilities as well as the built-in (compiled) functions of MATLAB avoiding at the same time nested loops and conditionals. Another important source for pseudospectral derivative matrices is the book of L. N. Trefethen [197].*

Remark 23 *For Chebyshev and for Lagrangian polynomials as well, projection (truncation) and interpolation do not commute, i.e., $(P_N u)' \neq P_N(u')$ and $(I_N u)' \neq I_N(u')$. The Chebyshev-Galerkin derivative $(P_N u)'$ and the pseudospectral derivative $(I_N u)'$ are asymptotically worse approximations of u' than*

$P_{N-1}(u')$ and $I_{N-1}(u')$, respectively, for functions with finite regularity (see Canuto et al. [33] Sect. 9.5.2. and [93]).

Remark 24 (Computational cost) *First, we consider the cost associated with the matrix $D^{(l)}$. Thus, N^2 operations are requested to compute α_j. Given α_j, another $2N^2$ is required to find the off-diagonal elements. N^2 operations are required to find all the diagonal elements from (1.52). Hence, it requires $4N^2$ operations to construct the matrix $D^{(1)}$. Second, a matrix-vector multiplication takes N^2 operations and consequently the evaluation of u' in (1.45) would require $5N^2$ operations, which means asymptotically something of order $O(N^2)$. This operation seems to be a somewhat expensive one because this would take up most of CPU time if it were used in a numerical scheme to solve a typical PDE or ODE boundary value problem (the other computations take only $O(N)$ operations). Fortunately, the matrices of spectral differentiations have various regularities in them. It is reasonable to hope that they can be exploited. It is well known that certain methods using Fourier, Chebyshev or sinc basis functions can also be implemented using FFT. By applying this technique the matrix-vector multiplication (1.45) can be performed in $O(N \log N)$ operations rather than the $O\left(N^2\right)$ operations. However, our own experience, confirmed by [204], shows that there are situations where one might prefer the matrix approach of differentiation in spite of its inferior asymptotic operation count. Thus, for small values of N the matrix approach is in fact faster than the FFT approach. The efficiency of FFT algorithm depends on the fact that the integer N has to be a power of 2. More than that, the FFT algorithm places a limitation on the type of algorithm that can be used to solve linear systems of equations or eigenvalue problems that arise after discretization of the differential equations.*

1.5 Problems

(See Fox and Parker [71], Ch. 3, Practical Properties of Chebyshev Polynomials and Series)

1. Prove the recurrence relation (1.3) for Chebyshev polynomials, using the trigonometric identity

$$\cos((k+1)\theta) + \cos((k-1)\theta) = 2\cos\theta\cos(k\theta)$$

and the decomposition

$$T_k(x) = \cos(k\theta), \quad \theta = \arccos(x).$$

▲

2. In certain applications we need expressions for products like $T_r(x)T_s(x)$ and $x^r T_s(x)$. Show for the first that the following identity holds

$$T_r(x)T_s(x) = \frac{1}{2}\left[T_{r+s}(x) + T_{s-r}(x)\right].$$

For the second, we have to show first that

$$x^r = \frac{1}{2^{r-1}}\left[T_r(x) + \binom{r}{1}T_{r-2}(x) + \binom{r}{2}T_{r-4}(x) + \ldots\right],$$

and than we get

$$\begin{aligned} x^r T_s(x) &= \frac{1}{2^{r-1}}\left[T_r(x)\,T_s(x) + \binom{r}{1}T_{r-2}(x)\,T_s(x) + \ldots\right] = \\ &\frac{1}{2^r}\sum_{i=0}^{r}\binom{r}{i}T_{s-r+2i}(x). \end{aligned}$$

Show also that

$$T_r(T_s(x)) = T_s(T_r(x)) = T_{rs}(x).\blacktriangle$$

3. Show that for the indefinite integral we have

$$\begin{aligned} \int T_r(x)\,dx &= \frac{1}{2}\left\{\frac{1}{r+1}T_{r+1}(x) - \frac{1}{r-1}T_{r-1}(x)\right\},\ r \geq 2, (1.55) \\ \int T_0(x)\,dx &= T_1(x),\ \int T_1(x)\,dx = \frac{1}{4}\{T_0(x) + T_2(x)\}.\blacktriangle \end{aligned}$$

4. *The range* $0 \leq x \leq 1$. Any finite range, $a \leq y \leq b$, can be transformed to the basic range $-1 \leq x \leq 1$ with the change of variables

$$y := \frac{1}{2}\left[(b-a)\,x + (b+a)\right].$$

Following C. Lanczos [126], we write

$$T_r^*(x) := T_r(2x-1),$$

and all the properties of $T_r^*(x)$ can be deduced from those of $T_r(2x-1).\blacktriangle$

5. Show that the set of Chebyshev polynomials $T_0(x),\ T_1(x), ..., T_N(x)$ is a basis in $\mathcal{P}_N.\blacktriangle$

6. For the 'continuous' least square approximation, using 'trigonometric polynomial' (1.6), to a function $f(x),\ x \in [-\pi, \pi]$, show that

$$\min \int_{-\pi}^{\pi}\left[f(x) - p_N(x)\right]^2 dx = \int_{-\pi}^{\pi} f^2(x)\,dx - \pi\left\{\frac{1}{2}a_0^2 + \sum_{k=1}^{N}\left(a_k^2 + b_k^2\right)\right\}.\blacktriangle$$

7. Find in $[-\pi, \pi]$ the Fourier series for $f(x) = |x|$, and observe that it converges like k^{-2}. Find the similar series in the range $[-1, 1].\blacktriangle$

8. Justify the equality

$$(u - P_N u, v)_\omega = 0, \ \forall v \in \mathcal{P}_N,$$

where P_N is the projection operator defined in (1.27) .▲

9. Prove that the set of functions $\cos r\theta$, $r = 0, 1, ..., N$, are orthogonal under summation over the points

$$\theta_k = \frac{2k+1}{N+1}\frac{\pi}{2}, \ k = 0, 1, ..., N,$$

i.e., they satisfy (1.12) and hence find a discrete least square Fourier series different from (1.14).▲

10. Justify the formula for the derivative of a Chebyshev series (1.30) . *Hint* Let's differentiate first, term-by-term, the finite Chebyshev sum $p(x) = \sum_{r=0}^{n} \frac{1}{c_r} a_r T_r(x)$, to obtain $p'(x) = \sum_{r=0}^{n-1} \frac{1}{c_r} b_r T_r(x)$. We seek to compute the coefficients b_r in terms of a_r. To this end, we integrate $p'(x)$, using (1.55), to give

$$\begin{array}{c} \sum_{r=0}^{n} \frac{1}{c_r} a_r T_r(x) = \\ \frac{1}{2}\left\{a_0 T_0(x) + b_0 T_1(x) + \frac{1}{2} b_1 T_2(x) + \sum_{r=2}^{n-1} b_r \left[\frac{T_{r+1}(x)}{r+1} - \frac{T_{r-1}(x)}{r-1}\right]\right\}. \end{array}$$

By equating coefficients of $T_r(x)$ on each side we find

$$\begin{array}{c} a_r = \frac{1}{2r}\left(b_{r-1} - b_{r+1}\right), \ r = 1, 2, ..., n-2 \\ a_{n-1} = \frac{1}{2(n-1)} b_{n-2}, \ a_n = \frac{1}{2n} b_{n-1}. \end{array}$$

We then can calculate the coefficient b_r successively, for decreasing r, from the general recurrence relation

$$b_{r-1} = b_{r+1} + 2ra_r.$$

Consequently, we can write

$$\left\{\begin{array}{c} b_{n-1} = 2na_n, \ b_{n-2} = 2(n-1)a_{n-1}, \\ b_{n-3} = 2(n-2)a_{n-2} + 2na_n, \\ \cdots \\ b_1 = 4a_2 + 8a_4 + 12a_6 + ..., \\ b_0 = 2a_1 + 6a_3 + 10a_5 + ... \ . \end{array}\right.$$

Each sum above is finite and finishes at a_n or a_{n-1}. Whenever we differentiate term-by-term an infinite Chebyshev series, these sums are in fact infinite series which have the general expressions

$$\begin{array}{c} b_{2r} = \sum_{s=r}^{\infty} 2(2s+1) a_{2s+1}, \\ b_{2r+1} = \sum_{s=r}^{\infty} 2(2s+2) a_{2s+2}, r = 0, 1, 2, ... \ . \end{array} \ ▲ \qquad (1.56)$$

11. For any $v \in H_0^1(a,b)$, $(a,b) \subset \mathbb{R}$, prove the *Poincare inequality*

$$\pi^2 \int_a^b u^2 dx \leq (b-a)^2 \int_a^b (u')^2 dx.$$

Formally, we can write this in the form

$$\frac{\pi}{b-a} \|u\| \leq \|u\|_{1,0}. \tag{1.57}$$

Hint Expand $u(x)$ as well as $u'(x)$ in their respective Fourier series.▲

12. Express the function $f : [-1,1] \to \mathbb{R},\ f(x) = 1/\left(1+x+x^2\right)$ as a series of Chebyshev polynomials $\sum_{r=0}^{\infty} \frac{1}{c_r} a_r T_r(x)$. Try to estimate the error. Observe that the absolute values of the "degrees of freedom" a_r of $f(x)$ are decreasing with the increase of r. Find out the range L such that $a_L \neq 0$ and $a_r = 0,\ r > L$, i.e., the order of the smallest non vanishing coefficient. Using successively the transform (1.41) and the matrices of differentiation $D^{(l)}$, compute

$$\left(f^{(\nu)}(x_0)\ \ f^{(\nu)}(x_1)\ \ ...f^{(\nu)}(x_L)\right)^T,\ \nu = 0,1,2.$$

It is strongly recommended to set up a computing code, for instance a MATLAB m function (see also http://dip.sun.ac.za/˜weideman/research/differ.html).
Hint Equating each side of the identity

$$1 = \left(1+x+x^2\right) \sum_{r=0}^{\infty} \frac{1}{c_r} a_r T_r(x),$$

we find

$$\begin{cases} \frac{3}{4}a_0 + \frac{1}{2}a_1 + \frac{1}{4}a_2 = 1, \\ \frac{1}{2}a_0 + \frac{7}{4}a_1 + \frac{1}{2}a_2 + \frac{1}{4}a_3 = 0, \\ \frac{1}{4}a_{r-2} + \frac{1}{2}a_{r-1} + \frac{3}{2}a_r + \frac{1}{2}a_{r+1} + \frac{1}{4}a_{r+2} = 0,\ r = 2,3,... \end{cases}.$$

To solve this infinite set of linear algebraic equations we solve in fact successive subsets of equations, involving successive leading submatrices (square matrices) of the infinite matrix. We assume, without proof, that this process will converge whenever a convergent Chebyshev series exists.▲

13. Show that the matrix $D^{(2)}$ with the entries defined in (1.53) is singular. *Hint* $D^{(2)}v_0 = D^{(2)}v_1 = 0$, where $v_0 = (1\ 1\ ...1)^T$, $v_1 := (x_0\ x_1\ ...x_N)^T$, $v_0,\ v_1 \in \mathbb{R}^{N+1}$, see the definition of Lagrangian interpolation polynomial, (1.47).▲

14. [155], P. 702 Give a rule for computing the Chebyshev coefficients of the product $v(y)\,w(y)$ given that

$$v(y) := \sum_{n=0}^{\infty} a_n T_n(y),\ w(y) := \sum_{n=0}^{\infty} b_n T_n(y).$$

Hint. Let us define

$$\widetilde{T}_n(x) := \exp\left(i \cdot n \cos^{-1} x\right),\ |x| \leq 1,\ -\infty < n < \infty,\ i = \sqrt{-1}.$$

It follows that $2T_n(x) = \widetilde{T}_n(x) + \widetilde{T}_{-n}(x)$ and $\widetilde{T}_n(x)\widetilde{T}_m(x) = \widetilde{T}_{n+m}(x)$. With these we can rewrite the above expansions as

$$2v(y) := \sum_{n=-\infty}^{\infty} \widetilde{a}_n \widetilde{T}_n(y),\ 2w(y) := \sum_{n=-\infty}^{\infty} \widetilde{b}_n \widetilde{T}_n(y),$$

where $\widetilde{a}_n = c_{|n|}a_{|n|}$ and $\widetilde{b}_n = c_{|n|}b_{|n|}$ for $-\infty < n < \infty$. Therefore,

$$4v(y)w(y) = \sum_{n=-\infty}^{\infty} \widetilde{e}_n \widetilde{T}_n(y) = 2\sum_{n=0}^{\infty} e_n T_n(y),$$

where

$$e_n = \frac{1}{c_n} \sum_{m=-\infty}^{\infty} \widetilde{a}_{n-m}\widetilde{b}_m,\ \widetilde{e}_n = c_{|n|}e_{|n|}.$$

Consequently, the nth Chebyshev coefficient of $v(y)w(y)$ is $\frac{1}{2}e_n$ for $n \geq 0$.▲

15. Observe the Gibbs phenomenon for the map

$$f(x) = \begin{cases} 0,\ -2 \leq x \leq 0, \\ \cos(x),\ 0 \leq x \leq 2. \end{cases}$$

Hint Use the Fourier expansion

$$S_N(x) \quad := \quad = \frac{1}{4}\sin(2) + \sum_{n=1}^{N} \frac{1}{n^2\pi^2 - 4} \times$$
$$\left\{(-1)^{n+1} 2\sin(2)\cos\left(\frac{n\pi x}{2}\right) + n\pi\left[1 - (-1)^n \cos(2)\right]\sin\left(\frac{n\pi x}{2}\right)\right\}.$$▲

16. Provide the reason for the absence of a Gibbs phenomenon (effect) for the Chebyshev series of $f(x)$, $f : [-1,1] \to \mathbb{R}$, and its derivatives at $x = \pm 1$. *Hint* The map $F(\theta) := f(\cos\theta)$ satisfies $F^{(2p+1)}(0) = F^{(2p+1)}(\pi) = 0$ provided only that all derivatives of $f(x)$ of order at most $2p+1$ exist at $x = \pm 1$.▲

Chapter 2

Spectral methods for o. d. e.

"I have no satisfaction in formulas unless I feel their numerical magnitude".

Lord Kelvin

2.1 The idea behind the spectral methods

The *spectral methods (approximations)* try to approximate functions (solutions of differential equations, partial differential equations, etc.) by means of truncated series of orthogonal functions (polynomials) say, e_k, $k \in \mathbb{N}$. The well known Fourier series (for periodic problems), as well as series made up by Chebyshev or Legendre polynomials (for non-periodic problems), are examples of such series of orthogonal functions. Hermite polynomials and sinc functions are used to approximate on the real line and Laguerre polynomials to approximate on the half line.

Roughly speaking, a certain function $u(x)$ will be approximated by the finite sum

$$u^N(x) := \sum_{k=0}^{N} \widehat{u}_k \cdot e_k(x), \qquad N \in \mathbb{N},$$

where the real (sometimes complex !) coefficients $\widehat{u}_k$ are unknown.

A *spectral method* is characterized by a specific way to determine these coefficients.

We will shortly introduce the tree most important spectral methods by making use of a simple example.

Let us consider the two-point boundary value problem

$$\begin{cases} \mathcal{N}(u(x)) = f, \quad x \in (a,b) \sqsubset \mathbb{R}, \\ u(a) = u(b) = 0, \end{cases}$$

where $\mathcal{N}(\cdot)$ stands, generally, for a certain non-linear differential operator of a specified order, defined on an infinite dimensional space of functions.

The *Galerkin method,* (SG) for short, consists in the vanishing of the residue

$$R_N := \mathcal{N}\left(u^N\right) - f,$$

in a "weak sense", i.e.,

$$(SG) \quad \int_a^b w \cdot R_N \cdot e_k dx = 0, \quad k = 0, 1, 2, \ldots, N,$$

where $w(x)$ is a *weight function* associated with the orthogonality of the functions e_k. The applicability of this method strongly depends on the apriority fulfillment of the homogeneous boundary conditions by the functions e_k.

Whenever this is not the case (SG), method is modified as follows.

The $N+1$ unknown coefficients $\widehat{u}_k$ will be searched as the solution to the algebraic system

$$(ST) \quad \begin{cases} \int_a^b w \cdot R_N \cdot e_k dx = 0, \quad k = 0, 1, 2, \ldots, N-1, \\ \\ \sum_{k=0}^{N} \widehat{u}_k \cdot e_k(a) = \sum_{k=0}^{N} \widehat{u}_k \cdot e_k(b) = 0. \end{cases}$$

Thus we obtain the so called *tau method,* (ST) for short.

The *spectral collocation method,* (SC) for short, requires that the given equation is satisfied in the nodes of a certain grid, $\{x_k\}_{k=1,2,\ldots,N-1}$, $x_0 = a$, $x_N = b$, and the boundary conditions are enforced explicitly, i.e.,

$$(SC) \quad \begin{cases} \mathcal{N}\left(u^N(x_k)\right) - f(x_k) = 0, \quad x_k \in (a,b), \quad k = 1, 2, \ldots, N-1, \\ u^N(a) = u^N(b) = 0. \end{cases}$$

It is extremely important to underline the fact that the method does not use equidistant nodes, because as it is well known such nodes lead to ill-conditioning and *Runge's phenomenon.*

Remark 25 *Each and every formulation,* (SG), (ST) *and* (SC), *represents an algebraic system of equations- the first two for the unknowns* $\widehat{u}_0, \widehat{u}_1, \ldots, \widehat{u}_N$, *the third for* $u(x_0)$, $u(x_1), ..., u(x_N)$. *Thus, the first two methods belong to a more general class of methods, the so called* ***weighted residual methods.*** *The spectral collocation method, which does not belong to that, is also called the* ***pseudospectral method.*** *This method, unlike the finite difference and finite element methods, does not use local interpolants but use a single global interpolant instead.*

Remark 26 *Whenever the basis functions e_k are the Chebyshev polynomials, the methods (SG), (ST) and (SC) will be denoted respectively by (CG), (CT) and (CC). The same convention holds for the Fourier, Legendre and other classes of special polynomials.*

Remark 27 ***Approximation properties.*** *The spectral methods are particularly attractive due to the following approximation properties. The "distance" between the solution $u(x)$ of the above problem and its spectral approximation $u^N(x)$ is of order $1/N^s$, i.e.,*

$$\left\|u - u^N\right\| \leq \frac{C}{N^s},$$

where the exponent s depends only on the regularity (smoothness) of the solution $u(x)$. Moreover, if $u(x)$ is infinitely derivable, the above distance vanishes faster than any power of $1/N$, and this means spectral accuracy. This sharply contrasts with finite difference methods and finite element methods where a similar distance is of order $1/N^p$ with exponent p independent of the regularity of $u(x)$ but depending on the approximation scheme. In other words, while spectral methods use ***trial (shape)*** *and* ***test functions,*** *defined globally and very smooth, in finite elements methods these functions are defined only locally and are less smooth.*

Remark 28 *[33], [161]* ***Computational aspects.*** *The (SC) method solves the differential problems in the so called* ***physical space*** *which is a subset of $\mathbb{R}^{N+1}$ containing the nodal values $u(x_0)$, $u(x_1), ..., u(x_N)$ of the solution. The (SG) and (ST) methods solve the same problems in the so called* ***transformed space*** *which is again a subset of $\mathbb{R}^{N+1}$ containing the coefficients $\{\widehat{u}_k\}_{k=0,1,2,...,N}$. Each and every coefficient $\widehat{u}_k$ depends on all the values of $u(x)$ in the physical space. However, only a finite number of such coefficients can be calculated, with an accuracy depending on the smoothness of u, from a finite number of values of u. From the computational point of view this can be achieved by means of the Chebyshev discrete transform (1.40) .The inverse of this transform (1.41) is a map from transformed space onto physical space.*

2.2 General formulation for linear problems

Let us consider the following linear two-point boundary value problem

$$(LP) \quad \begin{cases} Lu = f, \quad in \quad (-1,1), \\ \qquad Bu = 0, \end{cases}$$

where L is a linear differential operator acting in a Hilbert space X and B stands for a set of linear differential operators defined on -1 and 1. If we introduce the domain of definition of the operator L as

$$D_B(L) := \{u \in X |\ Lu \in X \ \ and \ \ Bu = 0\},$$

and suppose that $D_B(L)$ is dense in X, such that $L : D_B(L) \subseteq X \to X$, the problem (LP) takes the functional form

$$(LP) \quad \begin{cases} u \in D_B(L), \\ Lu = f. \end{cases}$$

There exist criteria, such as Lax-Milgram lemma or inf-sup condition (criterion), which assure the fact that this problem is well defined. In order to obtain a numerical solution of (LP) we will approximate the operator L by a family of "discrete" operators $L_N,\ N \in \mathbb{N}$. Every "discrete" operator will be defined on a finite dimensional subspace X_N of X, in which the solution will be searched, and its codomain is a subspace Z of X.

The spectral approximation $u^N \in X_N$ of the solution $u \in X$ is obtained by imposing the vanishing of the projection of the "residual" $(L_N u - f)$ on a finite dimensional subspace Y_N of Z. Consequently, if we denote by Q_N the operator of orthogonal projection, $Q_N : Z \to Y_N$, the *nth order spectral approximation* u^N *of* u is defined by

$$(SA) \quad \begin{cases} u^N \in X_N, \\ Q_N\left(L_N u^N - f\right) = 0. \end{cases}$$

The *projection operator* Q_N will be defined with respect to a scalar product $(.,.)_N$ from Y_N as follows

$$\begin{cases} Q_N : Z \to Y_N, \\ (z - Q_N z, v)_N = 0, \quad \forall v \in Y_N. \end{cases}$$

Thus, the spectral approximation (SA) is equivalent with the "variational formulation"

$$(VA) \quad \begin{cases} u^N \in X_N, \\ \left(L_N u^N - f, v\right)_N = 0, \quad \forall v \in Y_N. \end{cases}$$

This equivalence justifies the name of "weighted residuals" used alternatively for the spectral methods. Schematically, the spectral methods look like this:

$$D_B(L) \subseteq X \xrightarrow{L_N} X,$$
$$X_N \subset X \xrightarrow{L_N} Z \subseteq X \xrightarrow{Q_N} Y_N \subset Z.$$

A particular choice of subspaces $Z,\ X_N,\ Y_N$, as well as of the scalar product $(.,.)_N$ (or projection operator Q_N), defines a specific spectral method. The space X_N is the space of *trial or shape functions* and the space Y_N is that of *test functions.*

2.3 Tau-spectral method

This method, discovered by C. Lanczos, [126] pp.464-469, is suitable for non-periodic problems with complicated boundary conditions. **The Chebyshev tau method (CT** for short) is characterized by the following choice:

- $$X := L^2_\omega(-1,1);$$

- $$X_N := \{v \in \mathcal{P}_N | \quad Bv = 0\}; \tag{2.1}$$

- $$Y_N := \mathcal{P}_{N-\beta}; \tag{2.2}$$

where β stands for the number of boundary conditions. The projection operator Q_N is the projection operator of X with respect to the scalar product (1.1), the "discrete" operators L_N coincide with L and the scalar product $(.,.)_N$ from Y_N remains as well the scalar product (1.1).

Practically, if we accept the family $\{T_k,\ k = 0,1,2,\ldots,N\}$ as a basis for the finite dimensional space X_N and a subset of this $\{T_k,\ k = 0,1,2,\ldots,N-\beta\}$ as the set of "test" functions in Y_N, the variational formulation (VA) corresponding to (LP) reads as follows:

$$(CT) \quad \begin{cases} \textit{find the coefficients } \widehat{u}_k \textit{ of } u^N(x) \in \mathcal{P}_N, \\ u^N(x) := \sum_{k=0}^{N} \widehat{u}_k \cdot T_k(x), \ \textit{such that} \\ \int_{-1}^{1} \left(Lu^N(x) - f(x)\right) T_k(x)\, \omega(x)\, dx = 0, \quad k = 0,1,2,\ldots N-\beta, \\ \textit{and} \quad \sum_{k=0}^{N} \widehat{u}_k B(T_k) = 0. \end{cases} \tag{2.3}$$

The equation (2.3) represents the projection of the equation (LP) onto the space $\mathcal{P}_{N-\beta}$ and the boundary conditions are explicitly imposed by the last β equations.

The stability and convergence of tau approximation is usually proved with the discrete form of *"inf-sup" condition* (see [11]), this tool being more suitable when the finite dimensional spaces X_N and Y_N are different one from the other. For linear differential operators the convergence results are fairly general. Thus, let L be the linear differential operator

$$Lu := u^{(m)} + \sum_{i=0}^{m-1} a_i u^{(i)}, \tag{2.4}$$

where $a_i \in L^2_\omega(-1,1)$ and let B be the bounded linear operator on the boundary which furnishes m supplementary conditions. The next result is proved in [172]:

Lemma 29 *If the coefficients (functions) a_i are polynomials and the homogeneous problem*

$$\begin{cases} Lu = 0, \quad in \quad (-1,1), \\ \qquad Bu = 0, \end{cases}$$

has only the null solution, then for large N the (CT) method leads to a unique solution, which converges to the unique solution of the non-homogeneous problem, with respect to the norm $\|\cdot\|_{m,\omega}$. The convergence error and the best approximation error of the solution in $\mathcal{P}_N$ with respect to the same norm have the same order of magnitude.

This result was extended to the case of more general coefficients, namely $a_i \in L^2_\omega(-1,1)$, by Cabos in his paper [27].

However, these results can not be extended for partial differential equations and consequently we consider an example where we alternatively use the "inf-sup" criterion (see [148] or [33] Ch. 10).

Lemma 30 (Existence and uniqueness) *Let $(W,\ \|\cdot\|_W)$ and $(V,\ \|\cdot\|_V)$ be two Hilbert spaces such that $W \sqsubseteq X$ and $V \sqsubseteq X$ the second inclusion being continuous i.e., there exists a constant C such that $\|v\|_V \leq C\|v\|$, $\forall v \in V$, and let $D_B(L)$ be dense in W. If there exists the real positive constants α and β such that the following three conditions with respect to the operator L hold*

$$\begin{cases} 0 < \sup\{(Lu,v) \mid u \in D_B(L)\}, \ \forall v \in V, \\ \alpha\|u\|_W \leq \sup\left\{\frac{(Lu,v)}{\|v\|_V} \mid v \in V, \ v \neq 0\right\}, \ \forall u \in D_B(L), \\ |(Lu,v)| \leq \beta\|u\|_W\|v\|_V, \quad \forall u \in D_B(L), \quad \forall v \in V, \end{cases} \tag{2.5}$$

then the non-homogeneous problem (2.4) has a unique solution (in a weak sense) which depends continuously on the right hand member f, i.e., there exists a positive constant C such that

$$\|u\|_W \leq C\|f\|.$$

Remark 31 *Just in case of elliptic (coercive) operators i.e.,*

$$\exists \alpha > 0 \text{ such that } \alpha\|u\| \leq (Lu,u), \ \forall u \in D_B(L),$$

the first two conditions above are fulfilled taking $V = W = X$. Similarly, the blondeness of bilinear form (Lu,v) implies automatically the third condition. Consequently, the Lax-Milgram lemma is a particular case of the above criterion (see our contribution [83] or the original reference [127]).

In order to prove the *numerical stability* of the method a *discrete form of the "inf-sup" criterion* is useful. The following lemma is proved in [11].

Lemma 32 (Numerical stability) *Let X_N and Y_N be two finite dimensional spaces of dimension N such that $X_N \sqsubseteq W$, $Y_N \sqsubseteq V$ and L satisfying the*

conditions from the above lemma. If there exists $\gamma > 0$, γ *independent of* N, *such that the discrete form of the "inf-sup" criterion holds, i.e.,*

$$\alpha \|u\|_W \leq \sup \left\{ \frac{(Lu, v)}{\|v\|_V} |\ v \in Y_N,\ v \neq 0 \right\},\ \forall u \in X_N,$$

then there exists $C > 0$, C *independent of* N, *for which*

$$\left\|u^N\right\|_W \leq C \|f\|. \tag{2.6}$$

The inequality (2.6) means exactly numerical stability of the numerical method.

In this case, the *convergence* of the numerical method can be established looking for a linear operator

$$R_N : D_B(L) \to X_N$$

which satisfies

$$\|u - R_N u\| \to 0, \quad as \quad N \to \infty, \quad \forall u \in D_B(L).$$

Then

$$\left\|u - u^N\right\|_W \to 0, \quad as \quad N \to \infty, \tag{2.7}$$

due to the inequality

$$\left\|u - u^N\right\|_W \leq \left(1 + \frac{\beta}{\gamma}\right) \|u - R_N u\|_W \tag{2.8}$$

(see for example [33], Ch. 10).

The limit (2.7) expresses precisely the convergence of the method.

Example 33 *Let us consider the fourth order boundary value problem*

$$\begin{cases} u^{(iv)} + \lambda^2 u = f, \quad x \in (-1, 1), \\ u(\pm 1) = u'(\pm 1) = 0, \end{cases}$$

where $\lambda \in \mathbb{R}$, $f \in L^2_\omega(-1, 1)$ *and* $Lu := u^{(iv)} + \lambda^2 u$,

$$L : D_B(L) \to L^2_\omega(-1, 1), \quad D_B(L) := \left\{ v \in H^4_\omega(-1, 1) \,|\, v(\pm 1) = v'(\pm 1) = 0 \right\}.$$

According to (2.3) the Chebyshev-tau solution of the problem has the form

$$u^N(x) := \sum_{k=0}^{N} \widehat{u}_k \cdot T_k(x),$$

where the coefficients $\{\widehat{u}_k\}_{k=0,1,2,\ldots,N}$ *solve the algebraic system*

$$\begin{cases} \int_{-1}^{1} \left[\left(u^N\right)^{(iv)}(x) + \lambda^2 u^N(x) - f(x) \right] T_k(x)\, \omega(x)\, dx = 0, \quad k = 0, 1, 2, \ldots, N-4, \\ u^N(\pm 1) = \left(u^N\right)'(\pm 1) = 0. \end{cases} \tag{2.9}$$

In this case $X_N := \{v \in \mathcal{P}_N |\ \ v(\pm 1) = v'(\pm 1) = 0\}$ *and* $Y_N = \mathcal{P}_{N-4}$. *The estimations (1.4) transform the boundary conditions into*

$$u^N(\pm 1) = \sum_{k=0}^{N} (\pm 1)^k \widehat{u}_k; \quad \left(u^N\right)'(\pm 1) = \sum_{k=0}^{N} (\pm 1)^k k^2 \widehat{u}_k.$$

The fourth order derivative can be written with respect to the system $T_k(x)$ *as*

$$\left(u^N\right)^{(iv)}(x) = \sum_{k=0}^{N-4} \widehat{u}_k^{(4)} T_k(x), \tag{2.10}$$

where

$$\widehat{u}_k^{(4)} = \frac{1}{c_k} \sum_{\substack{p=k+4 \\ p+k=even}}^{N} p\left[p^2\left(p^2-4\right)^2 - 3k^2p^4 + 3k^4p^2 - k^2\left(k^2-4\right)^2\right]\widehat{u}_p, \quad k = 0, 1, 2, \ldots, N-4.$$

Consequently, the final form of the system (2.9) reads as follows

$$\begin{cases} \sum_{k=0}^{N} (\pm 1)^k \widehat{u}_k = \sum_{k=0}^{N} (\pm 1)^k k^2 \widehat{u}_k = 0, \\ \frac{1}{c_k} \sum_{\substack{p=k+4 \\ p+k=even}}^{N} p\left[p^2\left(p^2-4\right)^2 - 3k^2p^4 + 3k^4p^2 - k^2\left(k^2-4\right)^2\right]\widehat{u}_p + \lambda^2 \widehat{u}_k = \widehat{f}_k, \\ k = 0, 1, 2, \ldots, N-4, \end{cases}$$

where

$$\widehat{f}_k = \frac{2}{\pi c_k} \int_{-1}^{1} f(x) T_k(x) \omega(x) dx, \quad k = 0, 1, 2, \ldots, N.$$

In order to establish the numerical stability of the method we have to mention first that the coercivity and the boundeness of the operator L *were proved by Maday [133]. This implies the validity of the three inequalities (2.5). It remains to verify the discrete form of "inf-sup" condition. To do this we define* $W := H_\omega^4(-1,1)$ *and* $V := L_\omega^2(-1,1)$ *along with the spaces* X_N *and* Y_N *defined above. As* $u^N \in \mathcal{P}_N$, $\left(u^N\right)^{(iv)} \in \mathcal{P}_{N-4}$, *and integrating two times by parts we get*

$$\begin{aligned} \left(Lu^N, \left(u^N\right)^{(iv)}\right)_{0,\omega} &= \int_{-1}^{1} \left[\left(u^N\right)^{(iv)}\right]^2 \omega dx + \lambda^2 \int_{-1}^{1} \left(u^N\right)^{(iv)} u^N \omega dx \\ &= \left\|\left(u^N\right)^{(iv)}\right\|_{0,\omega}^2 + \lambda^2 \int_{-1}^{1} \left(u^N\right)'' \left(u^N\omega\right)'' dx. \end{aligned} \tag{2.11}$$

From [133], Lemma 5.1, we use the inequality

$$301 \int_{-1}^{1} \phi'' (\omega\phi)'' dx \geq \int_{-1}^{1} \left(\phi''\right)^2 \omega dx, \quad \forall \phi \in H_{0,\omega}^2(-1,1). \tag{2.12}$$

The inequalities 2.11 and 2.12 lead to the inequality

$$\left(Lu^N, \left(u^N\right)^{(iv)}\right)_{0,\omega} \geq \left\|\left(u^N\right)^{(iv)}\right\|_{0,\omega}^2 + \frac{\lambda^2}{301} \left\|\left(u^N\right)''\right\|_{0,\omega}^2. \tag{2.13}$$

Due to the fact that $u^N \in H^4_\omega(-1,1) \cap H^2_{0,\omega}(-1,1)$ *the successive use of Poincare's inequality implies the existence of* $C > 0$, *independent of* u *such that*

$$\left\| \left(u^N\right)^{(iv)} \right\|_{0,\omega} \geq C \left\| u^N \right\|_{4,\omega}, \tag{2.14}$$

where $\|\cdot\|_{4,\omega}$ *stands for the norm in* $H^4_\omega(-1,1)$. *The inequalities (2.13) and (2.14) prove the discrete form of the "inf-sup" condition and consequently the numerical stability of the method. The convergence of this method is based on the inequality (2.8). To this end let us consider the algebraic polynomial* $P_{N,4}u \in \mathcal{P}_N$, *attached to the exact solution* u *of the problem. In [33] Ch. 9, the existence of this polynomial is proved for all* $u \in H^4_\omega(-1,1)$ *and the following two inequalities are established*

$$\begin{aligned} \left\| u - P_{N,4}u \right\|_{k,\omega} &\leq CN^{k-m} \left\| u \right\|_{m,\omega}, \ 0 \leq k \leq 4, \ m \geq 4, \\ \left| P_{N,4}u(\pm 1) \right| &= \left| \left(u - P_{N,4}u\right)(\pm 1) \right| \leq \left\| u - P_{N,4}u \right\|_\infty . \end{aligned} \tag{2.15}$$

Due to the fact that $u - P_{N,4}u \in H^4_\omega(-1,1)$ *the Sobolev inequality (see again [33], Appendix) implies*

$$\left\| u - P_{N,4}u \right\|_\infty \leq C \left\| u - P_{N,4}u \right\|_0^{\frac{1}{2}} \left\| u - P_{N,4}u \right\|_1^{\frac{1}{2}},$$

the norm $\|\cdot\|_p$ *being that of the unweighted space* $H^p(-1,1)$. *As the weight* $\omega(x) \geq 1, \ \forall x \in (-1,1)$ *we can write successively*

$$\begin{aligned} \left\| u - P_{N,4}u \right\|_p &\leq \left\| u - P_{N,4}u \right\|_{p,\omega}, \ \forall p \geq 0, \\ \left| P_{N,4}u(\pm 1) \right| &\leq C \left\| u - P_{N,4}u \right\|_{0,\omega}^{\frac{1}{2}} \left\| u - P_{N,4}u \right\|_{1,\omega}^{\frac{1}{2}}. \end{aligned}$$

The inequality (2.15) for $k = 0, 1$ *implies*

$$\left| P_{N,4}u(\pm 1) \right| \leq CN^{\frac{1}{2}-m} \left\| u \right\|_{m,\omega}, \tag{2.16}$$

and similarly we can obtain

$$\left| P_{N,4}u(\pm 1) \right| \leq CN^{\frac{3}{2}-m} \left\| u \right\|_{m,\omega}. \tag{2.17}$$

As the polynomial $P_{N,4}u$ *does not satisfy necessarily the homogeneous boundary conditions, we introduce a polynomial* $p \in \mathcal{P}_3$ *which interpolates the values of* $P_{N,4}u$ *and its derivative on* ± 1. *The coefficients of this polynomial depend linearly on the values* $P_{N,4}u(\pm 1)$ *and* $P_{N,4}u(\pm 1)$, *and consequently the quantity* $\|p\|^2_{4,\omega}$ *will be a quadratic form of these values. In this context the inequalities imply*

$$\left\| p \right\|^2_{4,\omega} \leq C \left(N^{\frac{3}{2}-m} \left\| u \right\|_{m,\omega} \right)^2, \tag{2.18}$$

the constant C *being independent of* p *and* u. *We can now define the operator* $R_N u \in X_N$ *as* $R_N u := P_{N,4}u - p$. *With (2.15) for* $k = 4$ *and (2.18) we can write successively*

$$\begin{aligned} \left\| u - R_N u \right\|_{4,\omega} &= \left\| u - P_{N,4}u + p \right\|_{4,\omega} \leq C_1 N^{4-m} \left\| u \right\|_{m,\omega} + C_2 N^{\frac{3}{2}-m} \left\| u \right\|_{m,\omega}, \\ \left\| u - R_N u \right\|_{4,\omega} &\leq CN^{4-m} \left\| u \right\|_{m,\omega}. \end{aligned}$$

The last inequality and (2.8) lead to the optimal error estimation, namely

$$\left\|u-u^N\right\|_{4,\omega} \leq CN^{4-m}\left\|u\right\|_{m,\omega}.$$

Example 34 *Let us consider the 1D Helmholtz problem*

$$\begin{cases} -u''+\lambda^2 u=f,\ x\in(-1,1) \\ u(\pm 1)=0. \end{cases} \tag{2.19}$$

We search a solution $u^N\in\mathcal{P}_N$ in the form

$$u^N=\sum_{k=0}^{N}\widehat{u}_k T_k.$$

Expressing the first and second order derivatives of u^N with the formulas (1.30) *and* (2.37), *after boundary conditions are imposed, we get algebraic system for $\widehat{u}_k$*

$$\begin{cases} \sum_{k=0}^{N}\widehat{u}_k(-1)^k=0, \\ \sum_{k=0}^{N}\widehat{u}_k=0, \\ -\frac{1}{c_k}\sum_{\substack{p=k+2 \\ p+k=even}}^{N} p\left(p^2-k^2\right)\widehat{u}_p+\lambda^2\widehat{u}_k=\widehat{f}_k,\ k=0,1,2,...,N-2. \end{cases} \tag{2.20}$$

Example 35 *The Chebyshev tau solution of the problem*

$$\begin{cases} u''+u=x^2+x,\ x\in(-1,1) \\ u(\pm 1)=0, \end{cases}$$

is depicted in Fig. 2.1. It is obtained using the MATLAB code `Chebyshev_tau.m` *from the Appendix MATLAB codes. When the order N was $N=128$, the precision attained the value* $3.8858e-016$ *!*

At the end of this section we have to observe that the tau method was exposed in its classical form. There exists also an *operator technique* (see [158], [61], [111] and [27]) as well as a *recursive technique* (see *[157]*). All these techniques are equivalent but sometimes there exists the possibility of a more efficient implementation from the point of view of numerical stability (see also [27]).

The applicability of this method for *singularly perturbed problem* was studied, for example, in [33] and in our previous paper [81].

The manoeuvrability of boundary conditions in the framework of tau method is a real advantage of this method. The possibility to obtain sparse matrices is another one. Anyway, its applicability in case of nonlinear problems and partial differential equations is quite tedious.

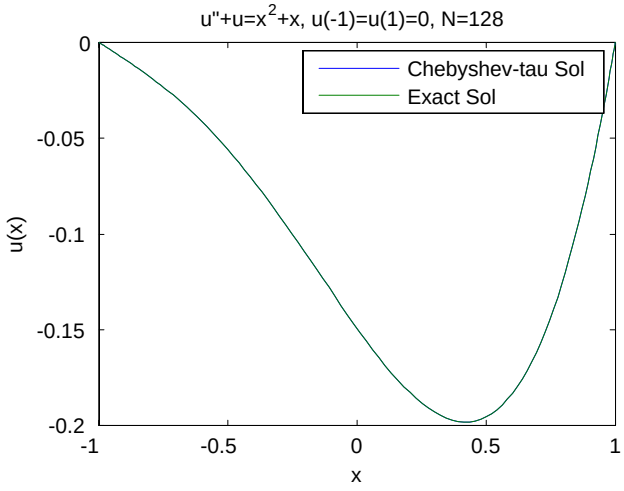

Figure 2.1: A Chebyshev tau solution

2.4 Collocation spectral methods (pseudospectral)

It seems that this category of spectral methods is the most frequently used in practical applications. The numerical approximation $u^N(x)$ of the solution $u(x)$ to the problem (LP) is searched again in a space of algebraic polynomials of degree N, but this space will be constructed such that the equation is satisfied in a specified number of points, called *collocation points,* of the interval $(-1,1)$.

When we make use of the *Chebyshev-Gauss-Lobatto interpolation nodes* (1.34), the method is called **Chebyshev collocation (CC** for short). We observe that for this method, in spite of the fact that the discrete operators $\{L_N,\ N \in \mathbb{N}^*\}$ act in $\mathcal{P}_N$, they can be different from the operator L.

The *"strong form"* of this method reads as follows:

$$(CC) \quad \left\{ \begin{array}{c} find \quad u^N \in \mathcal{P}_N \quad such\ that \\ L_N u^N(x_i) = f(x_i), \quad x_i \in (-1,1),\ i = 1,2,\ldots,N-1, \\ B_N u^N(x_i) = 0, \quad \forall x_i \in \{-1,1\} \end{array} \right. \tag{2.21}$$

The abscissas x_i are just the collocation points (1.34), and the discrete boundary operator B_N can be identical with B or constructed like L_N.

Hereafter we use the following two spaces:

- $X_N := \{v \in \mathcal{P}_N |\ \ Bv(x_k) = 0, \quad \forall x_k \in \{\pm 1\}\}$,
- $Y_N := \{v \in \mathcal{P}_N |\ \ v(x_k) = 0, \quad x_k \in \{\pm 1\}\}$.

If we use the *Lagrangian basis* $\{l_k(x)\}_{k=0}^N$, $l_k(x) \in \mathcal{P}_N$, associated with the (CGaussL) nodes, then we have

$$l_k(x_j) = \delta_{kj}, \quad k, j = 0, 1, 2, \ldots, N$$

where δ_{kj} stands for Kronecker symbol.

The space Y_N is generated by a subset of this basis made up by those polynomials which vanish in ± 1. For (CGaussL) formula this set is

$$\{l_k, \quad k = 1, 2, \ldots, N-1\}.$$

The projection operator Q_N is in fact the Lagrangian interpolation polynomial corresponding to nodes inside $(-1, 1)$ (just in case of (CGaussL): x_k, $k = 1, 2, \ldots, N-1$) and which vanish for the others, ± 1.

Consequently, we search a solution $u^N(x)$ of the form

$$u^N(x) := \sum_{k=0}^{N} u_j l_j(x), \tag{2.22}$$

with $u_j := u^N(x_j)$, $j = 0, 1, \ldots, N$, which satisfies (2.21), i.e.,

$$\begin{cases} \sum_{k=0}^{N} u_k L_N(l_k(x_j)) = f(x_j), \quad x_j \neq \pm 1, \\ \sum_{k=0}^{N} u_k B_N(l_k(x_j)) = 0, \quad x_j = \pm 1. \end{cases} \tag{2.23}$$

We have now to solve this system of algebraic equations for unknowns u_k, $k = 0, 1, \ldots, N$. Its matrix will be obtained by means of *pseudospectral derivative*. The strong collocation method is quite difficult to be analyzed from the theoretical point of view. The *weak form of Chebyshev collocation* was introduced by Canuto and Quarteroni in their work [31]. For the linear problem (LP) it reads

$$\begin{cases} u^N \in X_N, \\ \left(L_N u^N, l_k\right)_N = (f, l_k)_N, \ k = 0, 1, 2, ..., N, \end{cases} \tag{2.24}$$

or equivalently

$$\begin{cases} u^N \in X_N, \\ \left(L_N u^N, v\right)_N = (f, v)_N, \ \forall v \in Y_N, \end{cases} \tag{2.25}$$

where $(\cdot, \cdot)_N$ is the discrete scalar product defined by (1.35).

Remark 36 *The analysis of spectral collocation methods is based on a "inf-sup" type condition. The strong form of Chebyshev collocation was analyzed by Heinrichs in [106]. The weak form was considered by Canuto and Quarteroni in the above quoted paper and in [33]. In the special case where all boundary conditions are of Dirichlet type, i.e., $Bv := v$, the method is a particular case of Galerkin method ($X_N = Y_N$) and the Lax-Milgram lemma is the essential ingredient.*

Example 37 *Let us consider again the problem 1D Helmholtz problem (2.19). The strong form of (CC) method replaces the derivatives with the pseudodifferential matrices, i.e., the equations (2.21) or (2.23) are*

$$\left(-\left(D_N^{(1),C}\right)^2+\lambda^2 I_N\right)\cdot u=f, \tag{2.26}$$

where $D_N^{(1),C}$ is the differentiation matrix of order $N+1$, defined by (1.52) and corresponding to (CGaussL) nodes (1.34), I_N is the identity matrix of the same order,

$$u:=\left(u^N(x_0)\ \ u^N(x_1)...u^N(x_N)\right)^T \text{ and } f:=\left(f(x_0)\ \ f(x_1)...f(x_N)\right)^T.$$

Due to the Dirichlet boundary conditions, the values of solution are known at the end points ± 1 and we can eliminate these from the system (2.26). We carry out this operation by extracting the submatrices corresponding to rows and columns $1,2,...,N-1$. Consequently, to solve the problem (2.19) by (CC) method we have to solve the system of equations $1,2,...,N-1$ from (2.26). Two short observations are in order at this moment. First, the discrete operators L_N are in fact defined by

$$L_N:=\left(-\left(D_N^{(1),C}\right)^2+\lambda^2 I_N\right),\ N\in\mathbb{N},$$

and, second, an efficient technique, designed to introduce Robin (mixed) boundary conditions and based on Hermite interpolation, is available in [204], Sect. 4. In order to obtain the computational form of the weak (CC) method we start up with the familiar variational problem associated with (2.19), namely

$$\begin{cases} find\ u\in V\ such\ that \\ a_\omega(u,v)=(f,v)_\omega,\ \forall v\in W, \end{cases} \tag{2.27}$$

where the bilinear form $a_\omega(u,v)$ is defined

$$a_\omega(u,v):=\left(u',\frac{1}{\omega}(v\omega)'\right)_\omega+\lambda^2(u,v)_\omega,$$

and $V=W=H^1_{\omega,0}(-1,1)$. Our aim is to show that the weak (CC) method can be seen as Galerkin method in which the weight scalar product $(\cdot,\cdot)_\omega$ is replaced by discrete scalar product $(\cdot,\cdot)_N$. To this end we notice that the interpolating polynomial of a polynomial is the polynomial itself and consequently there does not exist an approximation process in computing the nodal values of the derivatives of this special type of functions. Consequently, to simplify the writing we denote $D:=D_N^{(1),C}$, and the equation (2.25) becomes

$$\left(-D^2u^N+\lambda^2u^N,v\right)_N=(f,v)_N,\ \forall v\in X_N.$$

For $u^N \in X_N \sqsubseteq \mathcal{P}_N$ one has $D^2u^N \in \mathcal{P}_{N-2}$ and for any $v \in X_N$, $\left(D^2u^N\right)v \in \mathcal{P}_{2N-2}$. Consequently, the quadrature formula (CGL), the integration by parts and boundary conditions imply

$$\left(-D^2u^N, v\right)_N = \sum_{j=0}^N \left(-D^2u^N \cdot v\right)(x_j)\cdot\omega_j = \int_{-1}^1 -D^2u^N(x)\,v(x)\,\omega(x)\,dx = \\ = \int_{-1}^1 \left(Du^N\right)(x)\,D(v\omega)(x)\,dx.$$

But, we can write

$$\int_{-1}^1 \left(Du^N\right)(x)\,D(v\omega)(x)\,dx = \left(Du^N, \frac{1}{\omega}D(v\omega)\right)_\omega,$$

and $v \in X_N$, so in fact $v(x) = \left(1-x^2\right)p(x)$, $p(x) \in \mathcal{P}_{N-2}$. In this situation we have successively

$$\frac{1}{\omega}D(v\omega) = Dv + v\frac{D\omega}{\omega} = Dv + v\frac{x\omega^3}{\omega} = Dv + xp \in \mathcal{P}_{N-1}.$$

More than that, one has

$$\left(Du^N, \frac{1}{\omega}D(v\omega)\right)_\omega = \left(Du^N, \frac{1}{\omega}D(v\omega)\right)_N,$$

due to the fact that $Du^N \cdot \frac{1}{\omega}D(v\omega) \in \mathcal{P}_{2N-2}$ and due to the order of accuracy of the (CGL) quadrature formula. All in all, the weak form of (CC) method reads

$$\begin{cases} find\ u^N \in X_N\ such\ that \\ a_N(u^N, v) = (f, v)_N,\ \forall v \in X_N, \end{cases} \tag{2.28}$$

where

$$a_N(u^N, v) := \left(Dv^N, \frac{1}{\omega}D(v\omega)\right)_N + \lambda^2\left(u^N, v\right)_N.$$

Remark 38 *Let us consider the partition*

$$-1 = x_0 < x_1 < x_2 < ... < x_{N-1} < x_N = 1, \tag{2.29}$$

of the interval $[-1,1]$ such that the nodes are symmetrically located around $x = 0$. Thus $x_j = -x_{N-j}$, $1 \leq j \leq N/2$, and there is a node at $x = 0$ if and only if N is even. Let also denote by $\widetilde{D}^{(2)} \in \mathbb{R}^{N-1} \times \mathbb{R}^{N-1}$ the second order differentiation matrix which is obtained by deleting the first and last rows and columns from $D^{(2)}$. This matrix corresponds to Dirichlet boundary conditions for second order differential operator. The matrix $\widetilde{D}^{(2)}$ is nonsingular and centrosymmetric (see the paper [4] for the definition of such matrices). It means that $\widetilde{D}^{(2)} = -R\widetilde{D}^{(2)}R$, where R is the permutation matrix with ones on the cross diagonal (bottom left to top right) and zero elsewhere. Its inverse $\left(\widetilde{D}^{(2)}\right)^{-1}$ has the same property of centrosymmetry and its norms, for three specific distributions of the collocation points, (2.29) *are studied in [182]. Some properties of centrosymmetric matrices are displayed in [4].*

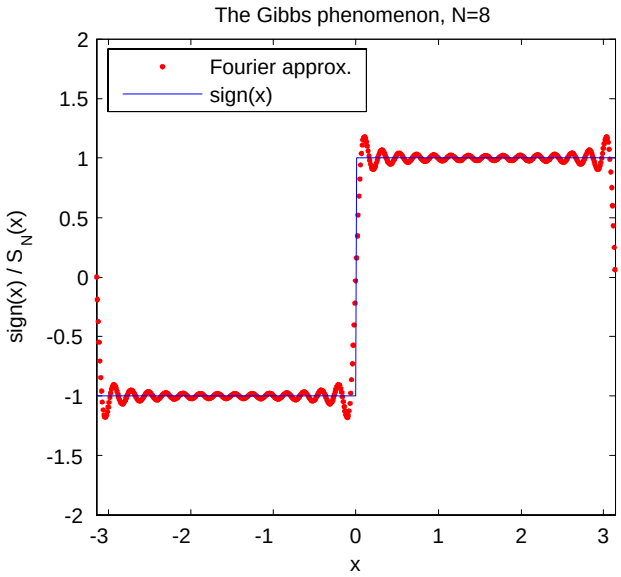

Figure 2.2: The Gibbs phenomenon

Remark 39 *The aliasing, sometimes called* ***Gibbs effect,*** *is the penalty that we endure for attempting to approximate a non-periodic solution with trigonometric polynomials (see [119], P. 141 or for a very intuitive explanation [154] P. 1122 and [156] in the context of spectral methods). It is well known that the spectral methods yield exponential convergence in the approximation of globally smooth functions. If a function has a local discontinuity, spectral accuracy is no longer manifested as the convergence is at most* $O(1)$ *in the* L_∞ *norm. We will not pay more attention to this phenomenon, but remark that several methods have been developed to deal with Gibbs effect. These methods roughly fall into two different categories; projection theories and direct-inverse theory. These methods which recover spectral accuracy from the Fourier data up to the discontinuity, i.e.,* ***the resolution of the Gibbs phenomenon,*** *are examined, for instance, in the paper of Jung and Shizgal, [122]. However, for the discontinuous map* $sign(x)$ *and its Fourier approximation*

$$S_N(x) = \frac{4}{\pi} \sum_{n=1}^{N} \frac{\sin[(2n-1)x]}{2n-1},$$

this effect is visualized in Fig. 2.2. The Fourier representation shows spurious oscillations near the discontinuity at $x = 0$ *and the domain boundaries* $x = \pm 1$.

Example 40 *Here, we solve a problem of the type which arises when dealing*

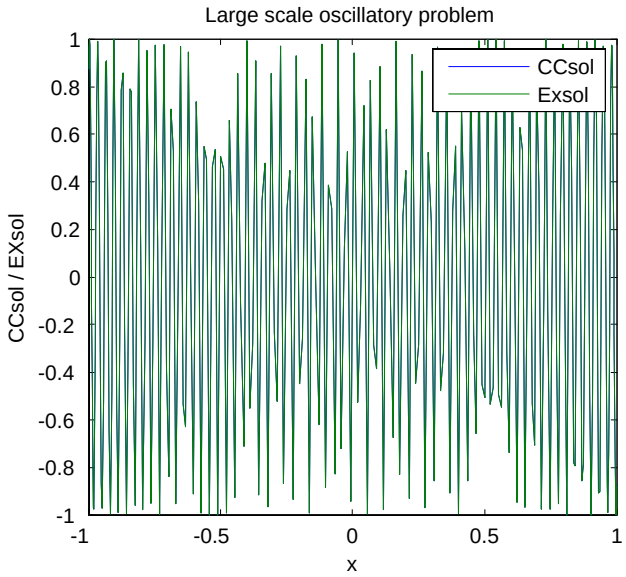

Figure 2.3: The solution to a large scale oscillatory problem

with a frequency domain equation for the vibrating string, namely

$$\begin{cases} u''(x) + \left(k^2 + 5\right) \cdot u(x) = 5 \cdot \sin(k \cdot x), \ 0 < x < 1, \\ u(c) = \sin(k \cdot c), \quad u(d) = \sin(k \cdot d). \end{cases}$$

For $c = -1$ *and* $d = 1$ *the problem is considered in the paper of Greengard and Rokhlin [98], P. 443. In order to demonstrate the performance of (CC) method on large-scale oscillatory cases, we solved the problem for* k *ranging from* 50 *to* 630 *and* N *between* 100 *and* 1200. *It was observed that reliable results can only be expected under the assumption that* ***scale resolution*** *is small, i.e., that* $k/N < 1$. *In this situation the error in approximating the exact solution* $u = \sin(k \cdot x)$ *is of order* $O\left(10^{-13}\right)$. *The case* $k = 200$ *and* $N = 256$ *is illustrated in Fig. 2.3.*

Example 41 *We solve the singular perturbation problem*

$$\begin{cases} \varepsilon \cdot u'' - u' = 0, \ -1 < x < 1, 0 < \varepsilon << 1, \\ u(-1) = 1, \ u(1) = 2. \end{cases}$$

The solution of this problem has an extremely sharp ***boundary layer*** *near the right end of the interval* $[-1\ 1]$, *causing severe numerical difficulties when standard algorithms are used (see for standard Galerkin the monograph of C. Johnson [121], P. 180). The strong (CC) method succeeded in solving this problem fairly accurate, i.e., the error with respect to the exact solution*

$$u(x, \varepsilon) = a + 3/2 - b \exp\left(\frac{x-1}{\varepsilon}\right), \ b = 1/\left(\exp(-2/\varepsilon) - 1\right), \ a = b + .5,$$

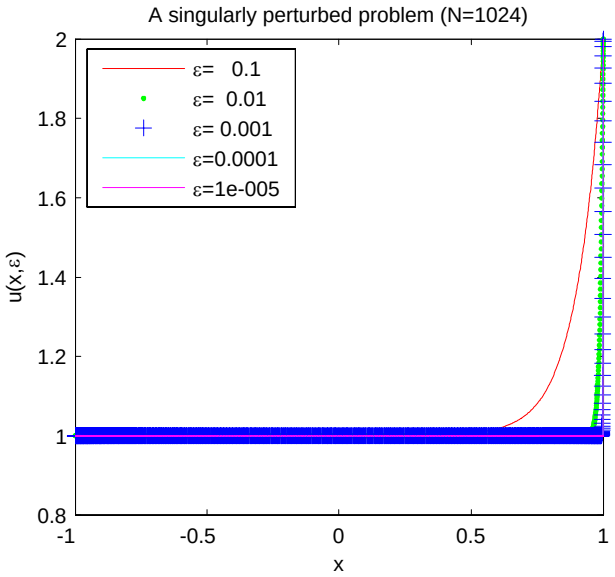

Figure 2.4: The solution to a singularly perturbed problem

is of order $O\left(10^{-12}\right)$ *for* $.1 \leq \varepsilon \leq 10^{-4}$ *but unfortunately drops to* $O\left(10^{-3}\right)$ *for* $\varepsilon = 10^{-5}$. *The solutions are depicted in Fig. 2.4.*

2.4.1 A class of nonlinear boundary value problems

In this section we consider boundary value problems of the form

$$\begin{cases} -u'' = \lambda F(u), \ 0 < x < 1, \ \lambda > 0, \\ u(0) = A, \quad u(1) = B, \end{cases} \tag{2.30}$$

where the nonlinear function $F(u)$ is assumed to have a power series representation.

The Green's function is well known and is given by

$$g(x,s) = \begin{cases} s(1-x), \ 0 \leq s \leq x, \\ x(1-s), \ x \leq s \leq 1. \end{cases}$$

Consequently, the nonlinear problem (2.30) can be represented in an integral form as

$$u(x) = \lambda \int_0^1 g(x,s) F(u(s)) ds + (1-x) A + xB.$$

This nonlinear integral equation was solved by a decomposition method in [45]. We shall show the capabilities of pseudospectral methods by solving some problems of the form (2.30).

Example 42 *(Troesch's problem [45] and [142]) In this example we consider the nonlinear boundary value problem,*

$$\begin{cases} u'' = \lambda \sinh(\lambda u), \ 0 \leq x \leq 1, \\ u(0) = 0, \quad u(1) = 1. \end{cases}$$

This problem can be reformulated as follows

$$\begin{cases} u' = v, \quad v' = \lambda \sinh(\lambda u), \ 0 \leq x \leq 1, \\ u(0) = 0, \quad u(1) = 1. \end{cases}$$

Its Jacobian matrix

$$J = \begin{pmatrix} 0 & k^2 \cosh(kx) \\ 1 & 0 \end{pmatrix},$$

is characterized by the following eigenvalues

$$\lambda_{1,2} = \pm k \sqrt{\cosh(kx)},$$

which, at the endpoints, become

$$\lambda(0) = \pm k, \quad \lambda(1) = \pm k \sqrt{\cosh(k)}.$$

It means that the matrix is highly non normal! However, for relatively low values of k, the eigenvalues are small, and the problem can be solved by conventional methods (finite differences or finite elements). On the other hand, for relatively large values of k, the eigenvalues are large, becoming $\lambda(1) = \pm 1049$ for $k = 10$. Thus, the use of some special techniques becomes desirable. We solved this problem by Chebyshev collocation technique and for the solution to nonlinear algebraic system

$$D_N^{(2),C} \cdot U = \lambda \sinh(\lambda U), \ U := (0, u(x_1), ..., u(x_{N-1}), 0)^T,$$

we employed the MATLAB code `fsolve`*. We observe that the problem was first translated into a homogenous one and consequently the first and the last component of U equal zero. For $N = 128$, and $k = 10$ the result is depicted in Fig. 2.5.The residual provided by* `fsolve` *equals* $6.260285523111991e-020$.

Example 43 *Let us solve a nonlinear two-point boundary value problem. The boundary value problem*

$$\begin{cases} u'' + u^3 = 0, \ 0 < x < L, \\ u(0) = u(L) = 0, \end{cases}$$

has a unique positive solution, $u(x) > 0$, $0 < x < L$, which represents from the physical point of view the average temperature in a reaction-diffusion process. In our previous paper [86] we solved this problem by a classical Galerkin method. A Chebyshev collocation solution is shown in Fig. 2.6 and it agrees fairly well with that quoted above (up to $O\left(10^{-8}\right)$). It is worth noting that the nonlinear system of algebraic equations obtained by Chebyshev collocation discretization was solved

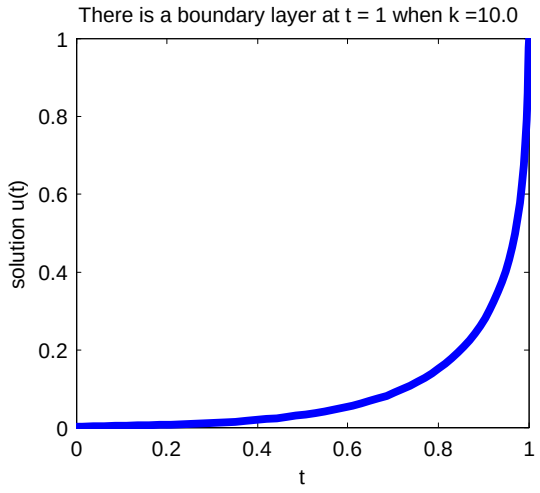

Figure 2.5: The solution to Troesch's problem

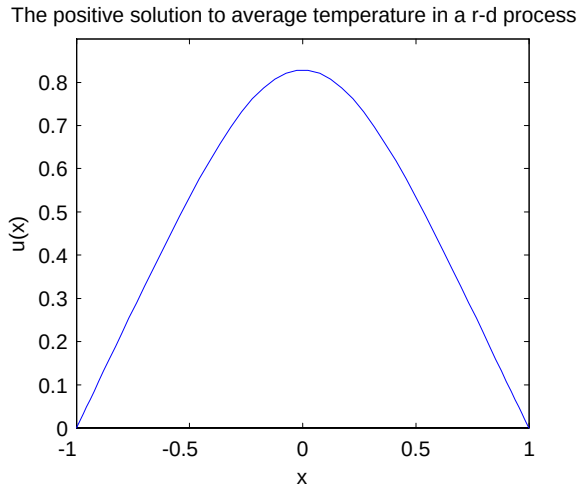

Figure 2.6: The positive solution to the problem of average temperature in a reaction-diffusion process

using the MATLAB code `fsolve` *and the initial guess (data) was* $u_0(x) := \left(1-x^2\right)^2$. *The iterative process was convergent. This fact was confirmed by the output argument exit flag which took the value* 1. *More than that, the residual of this process lowered to the value* $5.1091e-025$.

Remark 44 *It is known that, in general, the global methods are more sensitive to the boundary treatment than local methods. Solomonoff and Turkel examined in their paper [183] the influence of boundary conditions on the accuracy and stability of pseudospectral methods. They also consider the effect of the location of the collocation points on both the accuracy and the stability of the scheme and its effect on the allowable time step for an explicit time integration algorithm. Eventually they show that when Chebyshev points (nodes) are used to solve differential equations the error will be essentially uniform through the domain.*

Example 45 *(Bratu's problem [8], p.89) Consider the problem*

$$\begin{cases} -u'' = \lambda e^u, \ 0 < x < 1, \\ u(0) = u(1) = 0, \end{cases} \tag{2.31}$$

where $\lambda > 0$ *is a parameter. Substituting a function of the form*

$$u(x) = -2\log\left(\cosh((x-.5)\frac{\theta}{2})/\cosh\left(\frac{\theta}{4}\right)\right),$$

which satisfies the boundary conditions, into differential equation, we find that u *is a solution if*

$$\theta = \sqrt{2\lambda}\sinh\left(\frac{\theta}{4}\right).$$

This nonlinear algebraic equation for θ *has two, one or no solutions when* $\lambda < \lambda_c$, $\lambda = \lambda_c$, *or* $\lambda > \lambda_c$, *respectively. The critical value* λ_c *satisfies*

$$1 = \frac{1}{4}\sqrt{2\lambda_c}\sinh\left(\frac{\theta}{4}\right).$$

A diagram of bifurcation for this problem is also available in the monograph of Asher, Mattheij and Russell, [8], P. 491. If, for instance, $\lambda = 1$, *then we have two locally unique solutions whose initial slopes are* $s^* = 0.549$ *and* $s^* = 10.909$. *The numerical solution corresponding to the first slope is depicted in the Figure 2.7. The maximal error with respect to the exact solution equals* $1.718136466433151e-010$.

2.5 Spectral-Galerkin methods

They are in fact genuine Galerkin methods which use Chebyshev polynomials in order to construct the bases of finite dimensional spaces and consequently their theoretical analysis is well established. The difficulties appear whenever quite

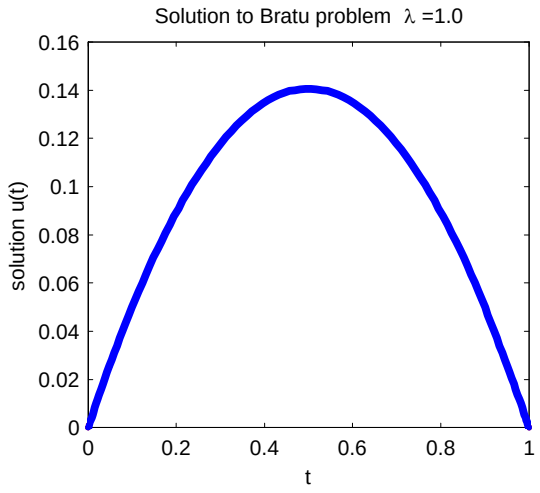

Figure 2.7: The solution to Bratu's problem $N = 128$

complicated boundary conditions have to be enforced into the basis functions. Thus with the exception of periodic boundary conditions the most studies confine to theoretical aspects. In our paper [164], we succeeded in overcoming some of these difficulties.

For the "classical" **Chebyshev-Galerkin method (CG** for short) we choose:

$$X := L^2_\omega(-1,1), \quad X_N = Y_N := \{v \in \mathcal{P}_N | \quad Bv = 0\}, \quad L_N = L,$$

and the projection operator Q_N represents the orthogonal projection with respect to scalar product (1.1).

The *discrete variational problem (formulation)* reads as follows:

$$(CG) \quad \begin{cases} find \;\; u^N \in X_N \;\; such\ that \\ \left(Lu^N, v\right) = (f, v) \quad \forall v \in X_N. \end{cases} \tag{2.32}$$

The applicability of (CG) method depends essentially on the existence of the basis functions which satisfy boundary conditions. Consequently, for a linear second order differential operator and homogeneous Dirichlet boundary conditions we will review the most known choices.

- In the monographs [90] and [33] for such problems is used the basis

$$\Phi_k := \begin{cases} T_k - T_0, \;\; k = even, \\ T_k - T_1, \;\; k = odd. \end{cases} \tag{2.33}$$

and consequently $X_N = span\{\Phi_2, \Phi_3, ..., \Phi_N\}$. A solution to 1D Helmholtz problem (2.19) has the form

$$u^N(x) := \sum_{k=2}^{N} u_k \Phi_k(x)$$

and then the variational equation (2.32) becomes

$$\lambda^2 \left(\sum_{k=2}^{N} u_k \Phi_k, \Phi_i \right)_\omega - \left(\sum_{k=2}^{N} u_k \Phi_k'', \Phi_i \right)_\omega = (f, \Phi_i)_\omega, \ i = 2, 3, ..., N,$$

which means a linear algebraic system

$$\mathbf{A} \cdot \mathbf{u} = \mathbf{f}, \tag{2.34}$$

with the entries $A_{ki} = \lambda^2 (\Phi_k, \Phi_i)_\omega - (\Phi_k'', \Phi_i)_\omega$, $k, i = 2, 3, ..., N$, of the matrix $\mathbf{A}$, $f_i = (f, \Phi_i)_\omega$, $\mathbf{f} = (f_2 ... f_N)^T$, $\mathbf{u} := (u_2 ... u_N)^T$.

- J. Shen introduced in [175] and [176] different basis functions using Legendre and respectively Chebyshev polynomials. In the later case he defined

$$\Phi_k (x) := T_k (x) - T_{k+2} (x), \ k = 0, 1, 2, ..., N. \tag{2.35}$$

The following result is due to Shen [176].

Lemma 46 *For $k, j = 0, 1, 2, ..., N$ the subsequent equalities hold*

$$a) \ \left(\Phi_k, \Phi_j'' \right)_\omega = \frac{\pi}{2} a_{kj}, \quad a_{kj} = \begin{cases} 4(k+1)(k+2), & j = k, \\ 8(k+1), & j = k+2, k+4, ..., \\ 0, & otherwise, \end{cases}$$

$$b) \ (\Phi_k, \Phi_j)_\omega = \frac{\pi}{2} b_{kj}, \quad b_{kj} = \begin{cases} c_k + 1, & j = k, \\ -1, & j = k \pm 2, \\ 0, & otherwise. \end{cases}$$

A solution to 1D Helmholtz problem (2.19) has again the form

$$u^N (x) := \sum_{k=0}^{N} u_k \Phi_k (x),$$

but $u^N (x) \in X_N$ defined with the Shen's basis functions (2.35), namely

$$X_N = Y_N := \{\Phi_0 (x), ..., \Phi_N (x)\}.$$

The variational equation (2.32) turns into the algebraic system

$$\left(-\mathbf{A} + \boldsymbol{\lambda}^2 \mathbf{B} \right) \cdot \mathbf{u} = \mathbf{f}, \tag{2.36}$$

where

$$\mathbf{A} = (a_{kj}), \ \mathbf{B} = (b_{kj}),$$

and

$$f_i = (f, \Phi_i)_\omega, \ \mathbf{f} = (f_0 ... f_N)^T, \ \mathbf{u} := (u_0 ... u_N)^T, \ k, j = 0, 1, 2, ..., N.$$

We will resort to our previous book [83] and to the monograph [33] in order to carry out the analysis of this problem.

2.6 Problems

1. If u^N is the polynomial

$$u^N := \sum_{k=0}^{N} \widehat{u}_k \cdot T_k(x),$$

then its second derivative is

$$\left(u^N\right)'' = \frac{2}{c_k} \sum_{k=0}^{N-2} \widehat{u}_k^{(2)} \cdot T_k(x), \tag{2.37}$$

where

$$\widehat{u}_k^{(2)} = \sum_{\substack{p=k+2 \\ p+k=even}}^{N} p(p-2)\,\widehat{u}_p,\ k = 0, 1, 2, ..., N-2.\blacktriangle$$

2. Justify the formula (2.10).▲

3. [103] Let us consider the first order problem

$$\begin{cases} u' = f,\ x \in (-1, 1) \\ u(-1) = 0, \end{cases}$$

and search for that a solution in the form $u(x) = (1+x)\sum_{p=0}^{N} a_p T_p(x)$. Show that $u'(x) = \sum_{p=0}^{N} r_p T_p(x)$ where

$$r_p = (p+1)\,a_p + \frac{1}{c_p} \sum_{k=p+1}^{N} (2k)\,a_k.\blacktriangle$$

4. [103] Show that for arbitrary constants a_p we have the equality

$$-\left(\sum_{p=0}^{N} a_p (1-x^2) T_p(x)\right)'' = \sum_{p=0}^{N} b_p T_p(x), \tag{2.38}$$

and

$$\sum_{p=0}^{N} a_p (1-x^2) T_p(x) = \sum_{p=0}^{N+2} d_p T_p(x), \tag{2.39}$$

where

$$b_p = (p+1)(p+2)\,a_p + \frac{1}{c_p} \sum_{\substack{k=p+2 \\ k+p=even}}^{N} (6k)\,a_k,$$

and

$$d_p = -\frac{c_{p-2}}{4} a_{p-2} + \frac{4 - c_{p-1} - c_p}{4} a_p - \frac{1}{4} a_{p+2}.\blacktriangle$$

In the last equalities, if the index p does not belong to the set $\{0, 1, 2, ..., N\}$, the coefficient a_p is considered zero.

5. [103] Let $v^N(x) := \sum_{p=0}^{N} a_p T_p(x)$; then

$$-\left(1-x^2\right)\left(v^N(x)\right)'' = \sum_{p=0}^{N} \sigma_p T_p(x), \tag{2.40}$$

and

$$\left(1-x^2\right)\left(v^N(x)\right)' = \sum_{p=0}^{N+1} \tau_p T_p(x), \tag{2.41}$$

where

$$\sigma_p = p(p-1)a_p - \frac{1}{c_p} \sum_{\substack{k=p+2 \\ k+p=even}}^{N} (2k)\, a_k,$$

and

$\tau_p = \dfrac{p-1}{2}a_{p-1} + \dfrac{p+1}{2}a_{p+1},\ p = 0,1,2,...,N+1;\ a_{-1} = a_{N+1} = a_{N+2} = 0.$▲

6. [187] Solve by various spectral methods the problem

$$\begin{cases} u''(x) + 400 \cdot u(x) = -400 \cdot \cos^2(\pi x) - 2\pi^2 \cos(2\pi x), \ 0 < x < 1, \\ u(0) = u(1) = 0. \end{cases}$$

Hint The problem is a difficult one due to the presence of rapidly growing solutions of the corresponding homogeneous equation. First, rewrite the problem on $[-1 \ \ 1]$ using the change of variables $x \in [-1 \ \ 1] \Leftrightarrow y \in [a \ \ b],\ y = \frac{1}{2}\left[(b-a)x + (a+b)\right]$. The solution is depicted in Fig. 2.8.▲

7. [98] Solve by spectral collocation methods the following problem corresponding to Bessel equation, namely

$$\begin{cases} u''(x) + \frac{1}{x} \cdot u'(x) + \frac{x^2 - \upsilon^2}{x^2} \cdot u(x) = 0, \ 0 < x < 1, \\ u(0) = 0, \ \ u(600) = 1 \end{cases}$$

for $\upsilon = 100$. *Hint* The difficulty of this problem is due to the fact that the two linearly independent solutions of the above equation are $J_\upsilon(x)$ and $Y_\upsilon(x)$, i.e., Bessel functions of the first and second kinds, respectively. It is well known that $J_\upsilon(x)$ behaves like x^υ as $x \to 0$ and $Y_\upsilon(x)$ behaves like $x^{-\upsilon}$. Most numerical algorithms have serious trouble in finding the decaying solution. In addition, this is a very large-scale computation, since the interval $[0 \ \ 600]$ contains almost 100 wavelengths of the solution of the problem. More than that, the coefficients of the equation are singular at the ends of its interval of definition. In spite of all these inconveniences the spectral methods perform fairly well.▲

8. Solve by Chebyshev tau, Chebyshev Galerkin and Chebyshev collocation methods the boundary value problem ([39], P. 74-105)

$$\begin{cases} -u'' + tu = \left(-t^2 + 2t + 1\right)e^t + t^2 - t, \ 0 < t < 1, \\ u(0) = u(1) = 0. \end{cases}$$

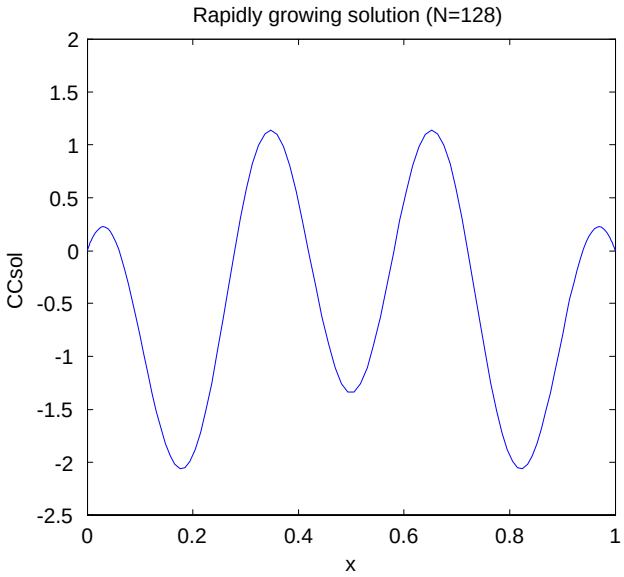

Figure 2.8: Rapidly growing solution

Hint The (CC) solution is depicted in Fig. 2.9 and agrees with the exact solution with an error equal to $3.3307e-015$. It is advisable to solve this problem by classical finite element method with linear elements and compare the errors. ▲

9. Observe the Runge's phenomenon for the *Runge's function*

$$f(x) := 1/\left(1+x^2\right),$$

by comparing the errors in the Lagrangian polynomial of interpolation and Chebyshev interpolation of this function. *Hint* (see for instance Quarteroni and Saleri, [170], P. 80).▲

10. Solve comparatively, by pseudospectral and Chebyshev Galerkin methods, the following boundary value problems ([62]):

$$\begin{cases} u^{(iv)} = -2e^x + 3u, \ 0 < x < 1, \\ u(0) = 1, \ u(1) = e, \\ u'(0) = 1, \ u'(1) = e, \end{cases}$$

with the exact solution $u(x) = e^x$, and the nonlinear problem

$$\begin{cases} u^{(iv)} = 6e^{-4u} - \frac{12}{(1+x)^4}, \ 0 < x < 1, \\ u(0) = 0, \ u(1) = \ln 2, \\ u'(0) = 1, \ u'(1) = 0.5, \end{cases}$$

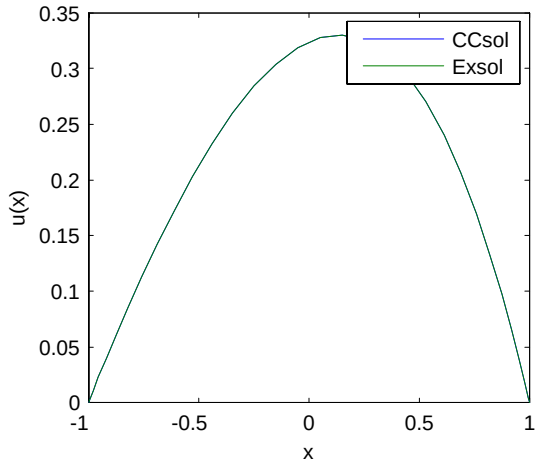

Figure 2.9: The (CC) solution to a linear two-point boundary value problem

which has the exact solution given by $u(x) = \ln(1+x)$. *Hint* First, rewrite the problems on the interval $[-1\ 1]$ and then reformulate the problems into standard form, i.e., homogenize the boundary conditions. These conversion techniques are quite elementary, and many of them are mentioned, for instance, in the paper of Ascher and Russell [7].▲

11. Solve the following homogeneous Dirichlet problem (see [209]):

$$\begin{cases} \varepsilon u'' - u = -1,\ 0 < x < 1,\ 0 < \varepsilon << 1, \\ u(0) = u(1) = 0, \end{cases}$$

whose solution has a boundary layer in the neighborhoods of the points $x = 0$ and $x = 1$. *Hint* The exact solution is given by

$$u(x) = \frac{2\sinh(x/2\varepsilon)\sinh((1-x)/2\varepsilon)}{\cosh(1/2\varepsilon)}.$$

▲

Chapter 3

Spectral methods for p. d. e.

3.1 Parabolic problems

Roughly speaking, any numerical method for the general *linear parabolic problem*

$$u_t = Lu,$$

where $u(x,\cdot) \in H,\ L : H \to H$, is a *linear spatial operator,* i.e.,

$$Lu := (p(x)\,u_x)_x - \sigma(x)\,u(x,\mathrm{t}) + h(x),\ p(x) > 0,\ \sigma(x) \geq 0,\ -1 < x < 1,$$

and H is a Hilbert space, consists of *three steps.*

The *first* one is to choose a finite dimensional subspace of H, say X_N, and the *second* is to choose a projection operator $Q_N : H \to X_N$. Consequently, the spatial discretization to our problem reads

$$\partial u^N/\partial t = Q_N L u^N, \quad u^N \in X_N. \tag{3.1}$$

There are three ways which have been used previously in order to choose the operator P_N, namely *Galerkin, tau and collocation.*

The last step is to solve the finite dimensional (N dimensional) system of differential equations (3.1).

To be more specific let us consider the $1D$ *initial-boundary value problem* (i. b. v. p.)

$$(2P) \quad \begin{cases} u_t = S(x)\,u_{xx}, \quad x \in (-1,1) \\ u(\pm 1,t) = 0,\ t > 0, \\ u(x,0) = u_0(x), \end{cases} \tag{3.2}$$

where $S(x) > \delta > 0$.

In (CG) method we can choose

$$\begin{aligned} \phi_n &= T_n - T_0, \quad n\ even; \\ \phi_n &= T_n - T_1, \quad n\ odd, \end{aligned} \tag{3.3}$$

and consequently the approximate solution $u^N(x,t) := \sum_{j=2}^{N} a_j(t)\,\phi_j(x)$ satisfies the boundary conditions for any $t > 0$. The system (3.1) becomes

$$\left(\frac{\partial u^N}{\partial t} - Lu^N, \phi_n\right)_{0,\omega} = 0, \quad n = 2, 3, \ldots, N,$$

and has the explicit form

$$\int_{-1}^{1} [\frac{\partial u^N}{\partial t} - S(x)\frac{\partial^2 u^N}{\partial x^2}] \frac{\phi_n}{\sqrt{1-x^2}} dx = 0, \quad n = 2, 3, \ldots, N.$$

It is readily seen that for non-constant $S(x)$ it is difficult, or at least time consuming, to solve this system for coefficients $\{a_N\}$.

In (CT) method, we set

$$u^N(x,t) := \sum_{j=0}^{N+2} a_j(t)\,T_j(x),$$

and require

$$\left(\frac{\partial u^N}{\partial t} - Lu^N, T_n\right)_{0,\omega} = 0, \quad n = 0, 1, 2, 3, \ldots, N.$$

Explicitly, the system (3.1) means the system of differential equations

$$\int_{-1}^{1} [\frac{\partial u^N}{\partial t} - S(x)\frac{\partial^2 u^N}{\partial x^2}] \frac{T_n}{\sqrt{1-x^2}} dx = 0, \quad n = 2, 3, \ldots, N,$$

together with the boundary conditions (restrictions)

$$\begin{aligned} \sum_{j=0}^{N+2} a_j T_j(1) &= \sum_{j=0}^{N+2} a_j = 0, \\ \sum_{j=0}^{N+2} a_j T_j(-1) &= \sum_{j=0}^{N+2} (-1)^j a_j = 0. \end{aligned}$$

Unfortunately, we face at least the same complications for getting the coefficients as we had for the Galerkin method.

A word of caution: *in both forms of u^N above, the dependence on t (of a_j) is not explicit in each and every line!*

In (CC) method, we set

$$u^N(x,t) := \sum_{j=0}^{N} a_j(t)\,\phi_j(x),$$

with ϕ_j defined in (3.3), and demand

$$\frac{\partial u^N}{\partial t} - S(x)\frac{\partial^2 u^N}{\partial x^2} = 0, \quad at\ x = x_j, \quad j = 1, 2, \ldots, N-1, \tag{3.4}$$

for some nodes x_j. If these nodes x_j are chosen to be Chebyshev-Gauss-Lobatto nodes, (1.34), i.e., $\cos(\pi j/N)$, $j = 0, 1, 2, \ldots, N-1, N$, so that the boundary values are included, there is an efficient way to solve (3.4), by taking into account the orthogonality of trigonometric functions. Explicitly, if we set (we "hide" again the dependence on t)

$$u^N(x) := \sum_{n=0}^{N} a_n \phi_n(x),$$

then

$$a_n = \frac{2}{N c_n} \sum_{k=0}^{N} \frac{1}{c_k} u^N(x_k) \cos\frac{\pi n k}{N},\ c_0 = c_N = 2,\ c_k = 1, 1 \le k \le N-1.$$

We know that

$$\frac{\partial^2 u^N(x_j)}{\partial x^2} = \sum_{n=0}^{N} b_n T_n(x_j),$$

where the coefficients b_n may be found from

$$c_n b_n = \sum_{\substack{p=n+2 \\ p+n\ even}}^{N} p\left(p^2 - n^2\right) a_p.$$

We then go back to the physical space and solve the *system of* $N-1$ *ordinary differential equations:*

$$\begin{aligned} \frac{\partial u^N}{\partial t}(x_j) - S(x_j)\frac{\partial^2 u^N}{\partial x^2}(x_j) &= 0, \quad j = 1, 2, \ldots, N-1, \\ u^N(x_0) &= u^N(x_N) = 0. \end{aligned} \tag{3.5}$$

This procedure is very efficient and may be generalized without any difficulty to nonlinear problems. We will make use exclusively of this procedure.

However, in practice we would use the Chebyshev polynomials to interpolate u spatially and then to evaluate the spatial derivatives at the desired points x_j. Finally, the solution of the system of ordinary differential equations (3.5) would be advanced in time, starting with initial data $u(x, 0) = u_0(x)$, $u_0(\pm 1) = 0$. We use some *finite difference scheme* for the original differential equation to find the time derivatives at points x_j in physical space.

Remark 47 *J. Butcher in his monograph [24], and more recently in the survey paper [25], provides a detailed perspective of Runge-Kutta methods as well as*

multistep methods used to solve different types of systems of differential equations including the stiff case. L. Shampine in [173] and [174] discusses the design of software for such systems that is easy to use for simple problems but still capable of solving complicated problems.

Remark 48 *For a more general problem, i.e., the* ***nonlinear*** *i. b. v. p.*

$$\begin{cases} u_t = (a(x)u_x)_x + f(x,u), \ 0 < x < L, \ 0 < t < s, \\ u(x,0) = u_0(x), \ 0 \le x \le L, \\ \alpha_0 u(0,t) - (1-\alpha_0)u_x(0,t) = \beta_0, \ 0 < t < s, \\ \alpha_1 u(L,t) - (1-\alpha_1)u_x(L,t) = \beta_1, \ 0 < t < s, \end{cases}$$

where $a(x) > 0$ is sufficiently smooth on $[0,L]$ and also $f : [0,L] \times \mathbb{R} \to \mathbb{R}$ is a smooth function with $0 \le \alpha_i \le 1$, $i = 1,2$ H. Matano shows in [137] that any solution that neither blows up in a finite time nor grows up as $t \to \infty$ should converge to some equilibrium solution as $t \to \infty$. We try to approximate numerically exclusively this type of solutions.

Example 49 *Let us consider the i. b. v. p. for heat equation*

$$\begin{cases} u_t = 3u_{xx}, \ 0 < x < L, \ t > 0, \\ u(0,t) = u(L,t) = 0, \ t > 0, \\ u(x,0) = L\left(1 - \cos\frac{2\pi x}{L}\right), \ 0 \le x \le L, \end{cases} \tag{3.6}$$

with the close solution

$$u(x,t) = \sum_{n=1}^{\infty} \frac{-16L}{(2n-1)\pi\left[(2n-1)^2 - 4\right]} \sin\frac{(2n-1)\pi x}{L} \exp\frac{-3(2n-1)^2\pi^2 t}{L^2}.$$

First, we use the linear map $y = \frac{L}{2}(x+1)$ to re-write the problem in the range $[-1,1]$ and then the MATLAB code Heat2 in order to solve the attached (3.5) using Runge-Kutta scheme (`ode45` MATLAB code).

The output of this code is presented in Fig. 3.1.

Remark 50 *We observe in the second sub-picture some wiggles (oscillations of order $O\left(10^{-7}\right)$) of the solution at final moment of time, near the endpoints. They illustrate the aliasing or Gibbs-phenomenon. If we had used a genuine spectral method (tau or Galerkin) and the boundary conditions had been periodic, the error would have decayed at an exponential rate.*

Example 51 *Let us consider a nonlinear i. b. v. p. which contains Burgers' equation with artificial viscosity $\varepsilon \neq 0$. It reads*

$$\begin{cases} \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \varepsilon\frac{\partial^2 u}{\partial x^2}, \ |x| < 1, \ t > 0, \\ u(\pm 1, t) = 0, \\ u(x,0) = -\sin\pi x, \end{cases} \tag{3.7}$$

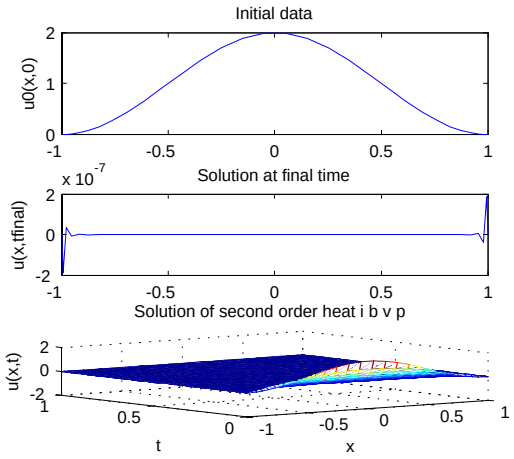

Figure 3.1: The solution to heat initial boundary value problem

and is also carefully analyzed in the paper of [12]. Its stationary counterpart is considered in [132]. The form of boundary conditions, i.e., Dirichlet homogenous, suggest to solve the problem by Chebyshev collocation method. In order to advance in time we used Runge-Kutta method (`ode45` MATLAB code). The output of this MATLAB code is depicted in Fig. 3.2. The solution curves become steeper and steeper as the time proceeds. They confirm the results from the paper [12].

Remark 52 *If we consider the Burgers equation on the entire real line and require that the solution goes to zero as $x \to \pm\infty$, the Hermite collocation method becomes suitable. The solution which starts with the initial data $u_0(x) := 0.5 \ \sec h(x)$ is available in Fig. 3.3. In spite of the large N used it is affected by the numerical noise of the method.*

The last example in this section refers to a reaction-convection-diffusion equation.

Example 53 *(Fischer's problem [113], [168]) Consider the reaction-convection-diffusion problem*

$$\begin{cases} u_t + au_x = bu_{xx} + u(1-u), \ 0 < x < 1, \ 0 < t < \infty, \\ u(0,t) = u(1,t) = 1, \ 0 < t < \infty, \\ u(x,0) = \phi(x), \ 0 < x < 1. \end{cases} \tag{3.8}$$

Higham and Owren show in [113] that if the initial data $\phi(x)$ belongs to C^1 and satisfies either $0 \le \phi(x) \le 1$ for $0 \le x \le 1$, or $\phi(x) \ge 1$ for all $0 \le x \le 1$, then the solution $u(x,t)$ is bounded for all $t > 0$ and converges pointwise to the

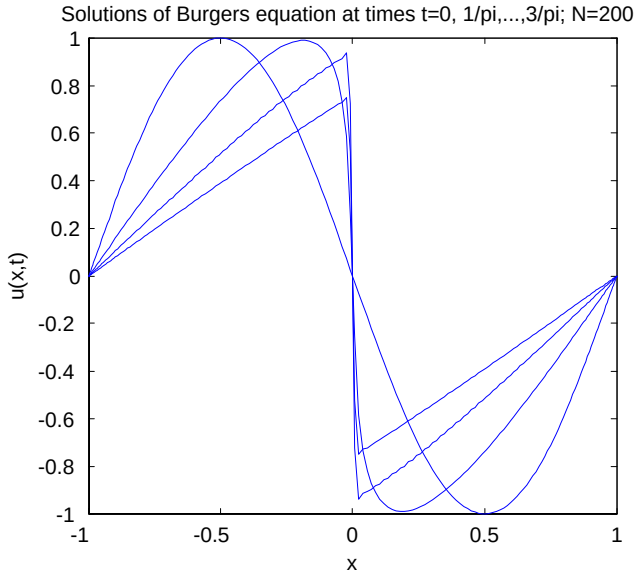

Figure 3.2: The solution to Burgers' problem

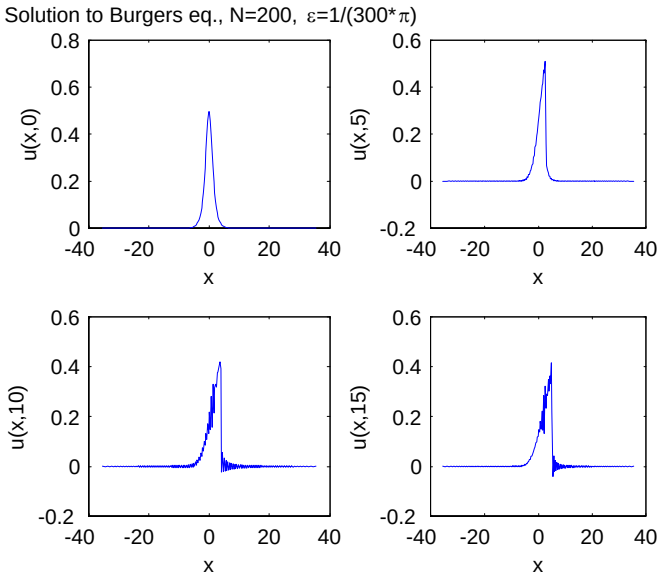

Figure 3.3: The Hermite collocation solution to Burgers' equation

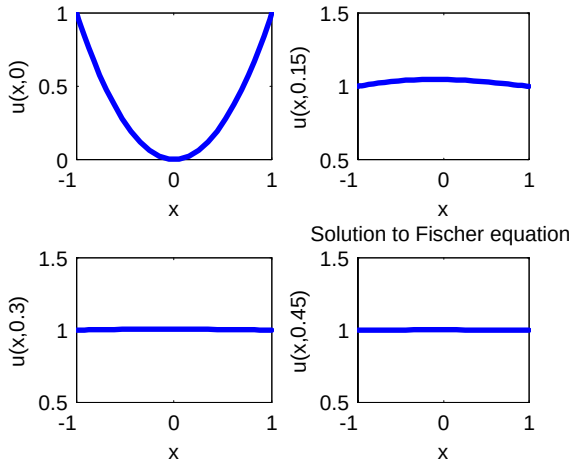

Figure 3.4: The solution to Fischer's equation on a bounded interval

steady state $u(x,t) = 1$ *as* $t \to \infty$. *Using a Chebyshev pseudospectral method in order to discretize the spatial operator and the Runge-Kutta method (MATLAB* `ode45`*) in order to advance in time, we verify extremely easy this theoretical result. The Fig. 3.4 contains the proof and shows the rapidity of convergence.*

3.2 Conservative p. d. e.

In this section we shall consider three very important types of parabolic nonlinear equations, namely the Schrödinger equation, the Ginzburg-Landau equation and the Körteveg- de Vries equations. They all can be seen as Hamiltonian systems. The numerical techniques for Hamiltonian ordinary differential equations can be found in a large number of articles as well as in some monographs, but the partial differential equations case has not been completely developed.

The *nonlinear Schrödinger equation* (NLS for short) has important applications in nonlinear optics, deep water waves and also plasma physics (see for instance [56] p. 285). It reads

$$i \cdot u_t + u_{xx} + |u|^2 u = 0, \tag{3.9}$$

and is supplemented with various boundary conditions. When the boundary conditions mean a bounded solution for large x, i.e., $|u(x,t)| \to \infty$ as $|x| \to \infty$ and the initial data decay with respect to x (see [202])

$$u(x,0) = \sec h(x/2) \exp(ix),$$

the solution we obtain is shown in the Fig. 3.5. The above boundary conditions

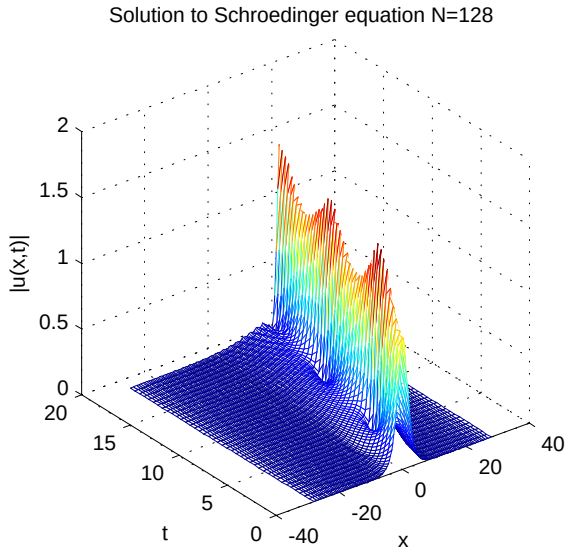

Figure 3.5: The solution to Schrödinger equation

imposed the use of Hermite differentiation matrices in the collocation method.

The equation (3.9) along with periodic boundary conditions

$$u\left(x+D\right)=u\left(x\right),\ D>0,$$

is an infinite dimensional integrable Hamiltonian system with the Hamiltonian

$$H\left(u,u^{*}\right)=\int_{0}^{D}\left(\left|u_{x}\right|^{2}-\frac{1}{2}\left|u\right|^{4}\right)dx.$$

The equation bears its name because it corresponds to the quantum Schrödinger equation with $|u|^2$ as the potential.

The complex *Ginzburg-Landau equation* (GL for short) reads

$$u_{t}=u-\left(1+iR\right)\left|u\right|^{2}u+\left(1+ib\right)u_{xx},\quad 0<x<l, \tag{3.10}$$

where the parameters R, b and l are real and the field $u\left(x,t\right)$ has complex values. Several types of boundary conditions are attached to this equation. There is a huge literature gathered around this equation. We refer only to the papers [23], [123], [200] and the monograph of Drazin [56], P. 286.

This is an amplitude equation governing the time evolution of the most unstable mode of perturbation around an equilibrium state. We also find the 3.10 equation as a model of superconductivity. In both cases the unknown variable

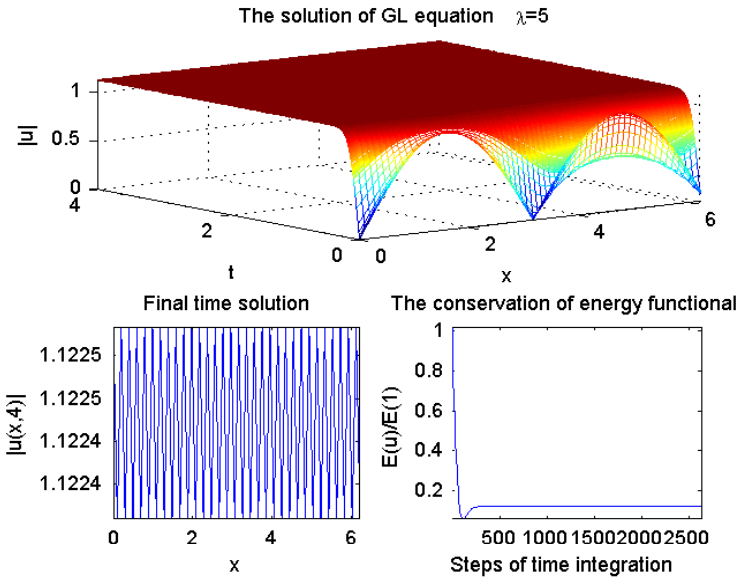

Figure 3.6: The solution to Ginzburg-Landau equation

of the equation denotes a complex-valued order parameter and it characterizes a macroscopic physical state. In this work we consider the following GL equation with periodic boundary conditions ([123]):

$$\begin{cases} \psi_t = \psi_{xx} + \lambda\left(1 - |\psi|^2\right)\psi, \ \lambda > 0, \\ \psi(x,t) = \psi\left(x + 2\pi, t\right), \ x \in \mathbb{R}. \end{cases} \tag{3.11}$$

It is easily seen that this equation gives a gradient flow of the *energy functional*

$$E\left(\psi\right) := \int_0^{2\pi} \left\{ \frac{1}{2}\left|\psi_x\right|^2 + \frac{\lambda}{4}\left(1 - |\psi|^2\right)^2 \right\} dx. \tag{3.12}$$

Thus any solution to 3.11 converges to a set of equilibrium solutions, which are given by solving the problem

$$\begin{cases} \psi_{xx} + \lambda\left(1 - |\psi|^2\right)\psi = 0, \ \lambda > 0, \\ \psi(x,t) = \psi\left(x + 2\pi, t\right), \ x \in \mathbb{R}. \end{cases} \tag{3.13}$$

The periodicity of boundary conditions lets us solve the problem 3.11 by a *Fourier pseudospectral method.* The solution corresponding to initial data $\psi(x,0) := \sin\left(x\right)$ and $N = 64$ is depicted in Fig. 3.6, a), b). The variation of the functional (3.12) can be observed in the same figure at the point c). It is

important to notice that at $t = 4$ the solution oscillates around the null solution with a precision of order $O\left(10^{-3}\right)$.

The third important example is the so called *Korteweg-de Vries equation,* (KdV for short) which reads

$$u_t + u \cdot u_x + u_{xxx} = 0. \tag{3.14}$$

It is a nonlinear partial differential equation arising in the study of a number of different physical systems, e.g., water waves, plasma physics, anharmonic lattices, and elastic rods. *MathSciNet* already lists more than 1000 articles on this subject. R. Miura in his paper [144] gives very interesting historical remarks on this equation. Thus, Korteweg and de Vries in 1857 obtained the model equation

$$\frac{\partial \eta}{\partial t} = \frac{3}{2}\sqrt{\frac{g}{l}}\frac{\partial}{\partial x}\left(\frac{1}{2}\eta^2 + \frac{2}{3}\alpha\eta + \frac{1}{3}\sigma\frac{\partial^2 \eta}{\partial x^2}\right), \tag{3.15}$$

in one space dimension and time, where η is the surface elevation above the equilibrium level l, α is a small arbitrary constant related to the uniform motion of liquid, g is the gravitational constant, and $\sigma = l^3/3 - Tl/\rho g$ with surface capillary tension T and density ρ. From the original equation (3.15), the transformations

$$t' := \frac{1}{2}\sqrt{\frac{g}{l\sigma}}t,\ x' := -\frac{x}{\sigma},\ u := -\frac{1}{2}\eta - \frac{1}{3}\alpha,$$

give us

$$u_t - 6u \cdot u_x + u_{xxx} = 0,$$

where we have dropped the primes. Moreover, this equation is Galilean invariant, i.e., is unchanged by the transformation

$$t' := t,\ x' := x - ct,$$

where c is some constant. This corresponds to going to a steady moving reference frame with a velocity c. We solve numerically this equation in the form

$$u_t = V'(u)_x + \nu u_{xxx},\quad V(u) := \frac{\alpha}{3}u^3 + \frac{\rho}{2}u^2, \tag{3.16}$$

on the spatial domain $[-1,\ 1]$, with *periodic boundary conditions* and initial condition $u(x, 0) = \cos(\pi x)$. We used

$$\alpha = -\frac{3}{8},\ \rho = -\frac{1}{10} \text{ and } \nu = -\frac{2}{3} \times 10^{-3}.$$

These values were used in [119], Ch. 14 and [9]. Their important feature is that $|\nu| << 1$. We solved the equation (3.16) using the Fourier pseudospectral method in order to discretize the spatial operators and the Runge-Kutta method of order 4 to progress in time. Our results are depicted in Fig. 3.7. The initial data breaks up into *a train of solitons* which repeatedly interact.

For the same values of α and ρ but $\nu = -10^{-6}$ a shock develops well before t reaches the value 0.5. This behavior is shown in Fig. 3.8. This behavior is analogue with that of Burgers' equation (see Fig. 3.2).

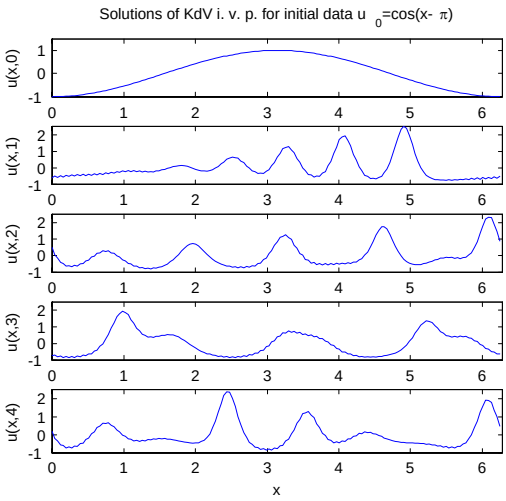

Figure 3.7: Numerical solution for KdV equation by Fourier pseudospectral method, N=160

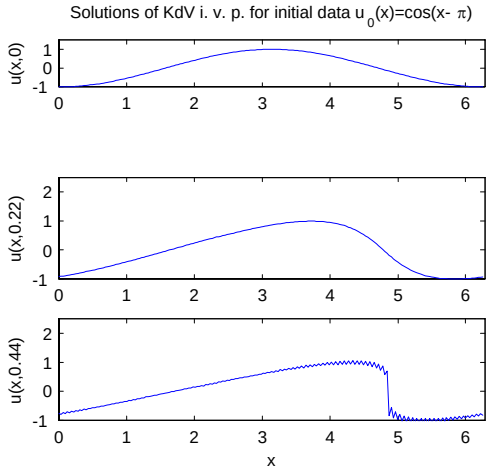

Figure 3.8: Shock like solution to KdV equation

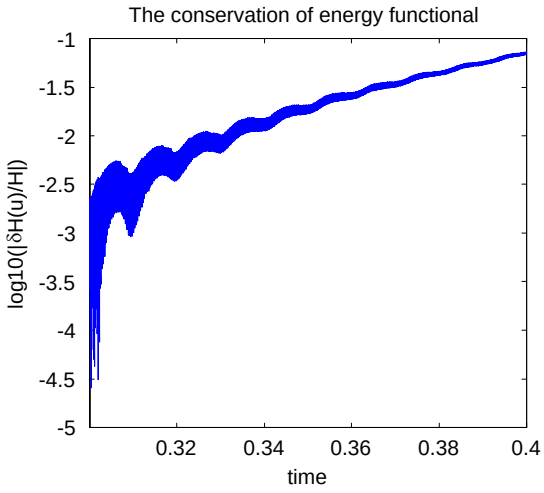

Figure 3.9: The conservation of the Hamiltonian of KdV equation

Remark 54 *The KdV equation in its Hamiltonian form reads*

$$u_t = \frac{\partial}{\partial x}\frac{\delta H}{\delta u}, \text{ with } H := \int \left(V(u) - \frac{\nu}{2}(u_x)^2 \right) dx,$$

where $\delta()/\delta u$ represents the Frechet derivative. It is also fairly important to know to what extent a numerical scheme, in this case RK for time evolution and Fourier pseudospectral, conserve the Hamiltonian. In Fig. 3.9 the evolution of numerical Hamiltonian corresponding to (3.16) is shown (log10(relative error in H) vs. time). It is not surprising that the Hamiltonian is not constant but this means that the dynamics of the differential system and that of the attached discrete system are no longer equivalent as time proceeds.

Remark 55 *It was shown (see for instance P D Lax [129] and [128]) that the KdV equation (3.14) possesses an infinite sequence of conservation laws of the form*

$$F_n = \int P_n dx,$$

where P_n is a polynomial in u and its derivatives up to order $n-1$. Three of them are classical:

$$\begin{aligned} F_0(u) &= \int 3u dx, \\ F_1(u) &= \int \frac{1}{2}u^2 dx, \\ F_2(u) &= \int \left(\frac{1}{6}u^3 - \frac{1}{2}u_x^2 \right) dx. \end{aligned}$$

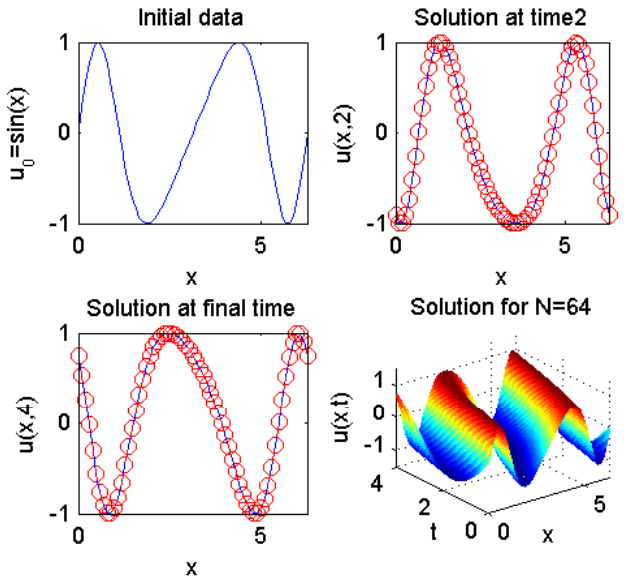

Figure 3.10: The solution to a first order hyperbolic problem, o (CS) solution and - exact solution

3.3 Hyperbolic problems

First, we shall consider the first order scalar equation

$$u_t + a(x)\, u_x = 0,\ t > 0, \quad 0 \le x \le 2\pi, \tag{3.17}$$

where $a(x) > 0$ is assumed periodic on $(0, 2\pi)$ and $u(x,t)$ satisfies periodic boundary conditions. Gottlieb and Turkel in [91] solve this problem by a *Fourier collocation (pseudospectral) method* coupled with *leap-frog* or second order RK method and show the stability of this algorithm.

We solved this problem for $a(x) := 1/(2+\cos(x))$ when the exact solution is $u(x,t) = \sin(2x + \sin(x) - t)$. For the time interval $[0,\ 2.5]$ the absolute value of the error at the half time is $1.3565e-006$ and it increases at the value $7.7932e-004$ at the final moment. The solution carried out is shown in Fig. 3.10

Second, D. Gottlieb solves in [92] a quite similar problem, namely

$$\begin{cases} u_t + x \cdot u_x = 0,\ t > 0,\ |x| \le 1, \\ u(x,0) = f(x),\ |x| \le 1, \\ u(\pm 1, t) = 0,\ t > 0, \end{cases} \tag{3.18}$$

and shows that the *Chebyshev collocation (pseudospectral) method* is stable. We solved this problem with initial data $f(x) := \left(x^2-1\right)^3$, as did Tal-Ezer in [191] and the numeric result are reported in Fig. 3.11.

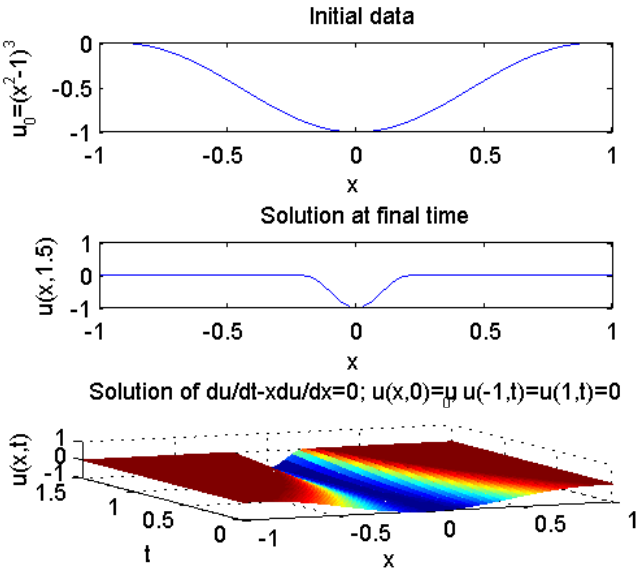

Figure 3.11: The solution to a particular hyperbolic problem

Our numerical results seem to be superior to those reported in [191].

Let us consider the second order wave equation together with Dirichlet boundary conditions, i.e.,

$$\begin{cases} u_{tt} = u_{xx}, \ (x,t) \in (-1,1) \times (0,.T] \\ u(\pm 1, t) = 0, \ t \in (0,.T] \\ u(x,0) = u_0(x), \ -1 \le x \le 1, \\ u_t(x,0) = v_0(x), \ -1 \le x \le 1. \end{cases} \tag{3.19}$$

Introducing a new variable $v := u_t$ the problem reduces to a system of differential equations in the variable t for the unknown vector $w := [u; v]$. The spatial derivative is discretized by Chebyshev pseudospectral derivative and the homogeneous Dirichlet boundary conditions are introduced, as usual, deleting the first and the last rows and columns. The solution corresponding to the initial data $u(x,0) = \sin(\pi x)$ and $u_t(x,0) = \cos(\pi x)$ is drawn in Fig. 3.12.

Remark 56 *The examples above illustrate the fact that the type of boundary conditions implies the choice of interpolation polynomials in spectral collocation method.*

Let us consider, at the end of this section, a nonlinear hyperbolic equation defined on the entire real axis, i.e., the so called *sine-Gordon equation.* It reads

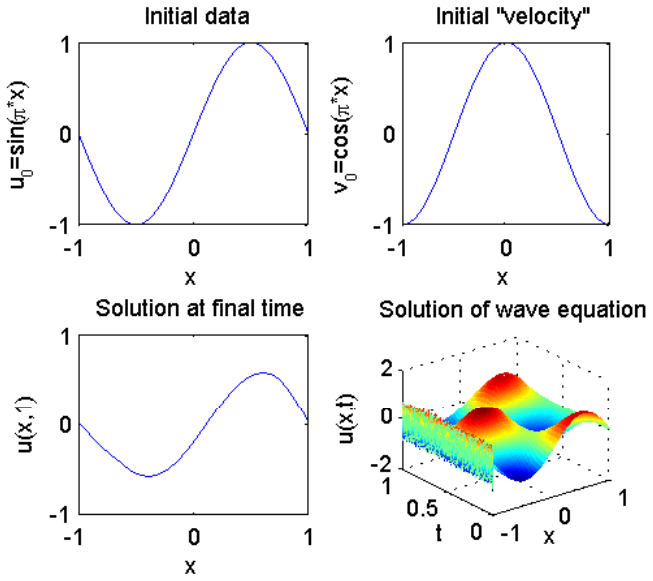

Figure 3.12: The solution to the wave equation

(see for nonlinear Klein-Gordon equations [44])

$$\begin{cases} u_{tt} = u_{xx} - \sin(u), \ (x,t) \in \mathbb{R} \times (0, .T] \\ |u(x,t)| \to 0, \ |x| \to \infty, \\ u(x,0) = u_0(x), \ u_t(x,0) = v_0(x), \ x \in \mathbb{R}, \end{cases} \tag{3.20}$$

and is related to the KdV and cubic NLS equations in the sense that all these equations admit *soliton* solutions. The equation describes nonlinear waves in elastic media, and it also has applications in relativistic field theory. Being completely integrable, this equation is solvable, at least in principle by the method of inverse scattering. In practice, however, this method of solution is curbersome to execute when arbitrary initial data $u(x,0)$ and $u_t(x,0)$ are prescribed.

However, the *solitons* were discovered by Zabuski and Kruskal in 1965. They found that solitary wave solutions had behavior similar to the superposition principle, despite the fact that waves themselves were highly nonlinear! They dubbed such waves *solitons*.

As the differential equation is defined on the entire real axis, we discretized the term u_{xx} by second order Hermite pseudo differential operator (matrix) of dimension N, i.e., $D_N^{(2),H}$. This way, the initial value problem for a partial differential equation (3.20) becomes an initial value problem for a system of

ordinary differential equations, namely

$$\begin{cases} U_{tt}^{N}(t) = D_{N}^{(2),H} \cdot U^{N}(t), \ t \in (0,T], \\ U^{N}(0) = u_0, \quad U_{t}^{N}(0) = v_0, \end{cases} \tag{3.21}$$

where $U^N(t)$ is a vector valued map defined as

$$U^{N}(t) := (u(x_1,t) \ u(x_2,t) ... u(x_N,t))^{T},$$

corresponding to a prescribed mesh $\{x_i\}_{i=1,...,N}$ applied to the spatial interval of integration.

In fact, the equation in (3.20) is a *nonlinear wave equation* (NLW for short) which has also a Hamiltonian structure, in the sense that we can rewrite it as (see for instance [79])

$$\begin{cases} u_t = v, \\ v_t = u_{xx} - V'(u), \quad V(u) = -\cos(u). \end{cases}$$

The Hamiltonian has the expression

$$H = \int \left(\frac{1}{2}(u_t)^2 + \frac{1}{2}(u_x)^2 + V(u) \right) dx. \tag{3.22}$$

The numerical integration of a large class of Hamiltonian partial differential equations by standard *symplectic integrators* is discussed in the paper of McLachlan [139] as well as in a series of papers of Omelyan, Mryglod and Folk [151], [152] and [153]. A brief survey of the theory and performance of symplectic integrators is also available in the paper of D. Markiewicz, [136].

Symplectic integrators, or *geometric numerical integrators*, have a remarkable capacity for capturing the long-time dynamics of Hamiltonian systems correctly and easily. Consequently, we used the following *symplectic and symmetric integrators* to advance in time in (3.20), or in other words, to solve the system (3.21):

1. implicit Runge-Kutta methods; the so called Gauss methods (Hairer & Hairer, [101], p.15);
2. partitioned multistep methods ([101], p.16);
3. composed Störmer/Verlet method (leap-frog method, [101], p.35);

In order to check the accuracy of the numerical integrator, we consider the problem (3.20) with the known solution (the so-called "breather" solution)

$$u(x,t) = 4\tan^{-1}\left(\frac{\sin\left(t/\sqrt{2}\right)}{\cosh\left(x/\sqrt{2}\right)} \right),$$

and extract the corresponding initial conditions

$$u(x,0) = 0, \quad u_t(x,0) = 2\sqrt{2}\sec h\left(x/\sqrt{2}\right).$$

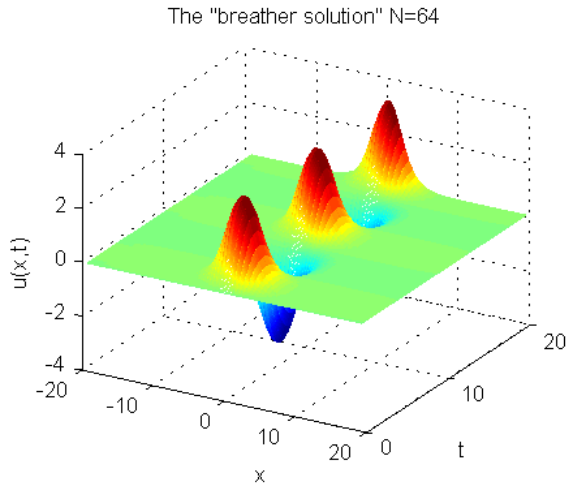

Figure 3.13: The "breather" solution to sine-Gordon equation

The Hermite pseudospectral-Störmer/Verlet solution of the problem is shown in Fig. 3.13. The error with respect to the above exact solution was $3.4265e{-}09$ for $N = 128$ irrespective of the used symplectic method. When the classical Runge-Kutta method of order 4 (`ode45` code in MATLAB) was used the situation was worse, i.e., the error attained only $1.7178e-05$! For larger N in the spectral approximation, i.e., $N = 200$ and using a fourth order compose Störmer/Verlet the error was much better, i.e., it attained the value $2.0305e-011$. We have to mention that in all the above computations the *scaling factor* was $b = 0.545$.

Remark 57 *In order to discretize the spatial operator u_{xx} in (3.20) as well as in wave equation we used alternatively the the Fourier pseudodifferential operator $D_N^{(2),F}$ and the sinc pseudodifferential operator $D_N^{(2),s}$ but the numerical results were inferior. For instance, in case of Fourier method the error increased to the value* 0.0084 *irrespective of time integration scheme.*

Remark 58 *The conservation of the Hamiltonian (3.22), when a fourth order Störmer/Verlet method was used to advance in time, is depicted in Fig. 3.14. This behavior is specific to symplectic methods and is better when compared with the non symplectic methods (see comparatively Fig. 3.9).*

Remark 59 *Gottlieb and Hesthaven, in a recent paper [95], review the current state of Fourier and Chebyshev collocation methods for the solution of hyperbolic problems with the emphasis on basic questions of accuracy and stability of numerical approximations.*

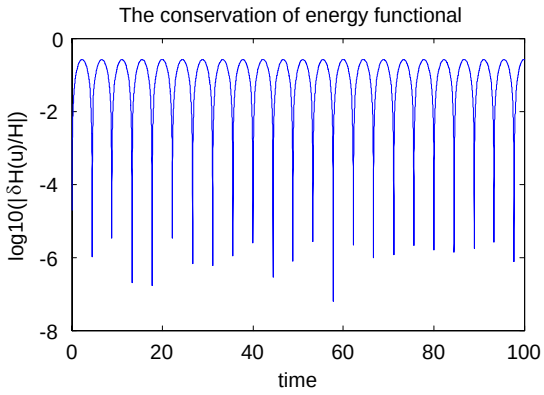

Figure 3.14: The variation of numerical sine-Gordon Hamiltonian

3.4 Problems

1. Solve the nonlinear parabolic problem [46]

$$\begin{array}{c} u_t - \left(u_x^2\right)_x = f(x,t), \ (x,t) \in (0,1) \times (0,.24] \\ u(x,0) = \sin(x), \quad x \in (0,1), \end{array}$$

with $f(x,t)$ and natural Neumann boundary conditions chosen so that the solution is $u(x,t) = e^t \sin(x)$. ▲

2. Solve the initial-boundary value problem for heat equation of order four (clamped boundary conditions)

$$\begin{cases} u_t = u_{xxxx}, \ -1 < x < 1, \ t > 0, \\ u(-1,t) = u(1,t) = 0, \ t > 0, \\ u'(-1,t) = u'(1,t) = 0, \ t > 0, \\ u(x,0) = 1 - \cos \pi (x+1), \ -1 \leq x \leq 1, \end{cases}$$

using a pseudospectral method in order to discretize the spatial differential operator and Runge-Kutta of order 4 (ode45, MATLAB code) in order to progress in time. *Answer* The solution is depicted in Fig. 3.15.▲

3. Solve numerically by Fourier pseudospectral method the (3.17) problem with $a(x) := 1$. *Hint* The analytic solution is $u(x,t) = \sin(x-t)$.▲

4. Solve numerically by Chebyshev pseudospectral method the (3.18) problem with $f(x) := \exp(1/(x^2-1))$. ▲

5. Solve numerically by Chebyshev pseudospectral method the (3.19) problem with Neumann boundary conditions instead of Dirichlet boundary conditions, i.e.,

$$u_x(\pm 1, t) = 0, \ t \in (0, .T],$$

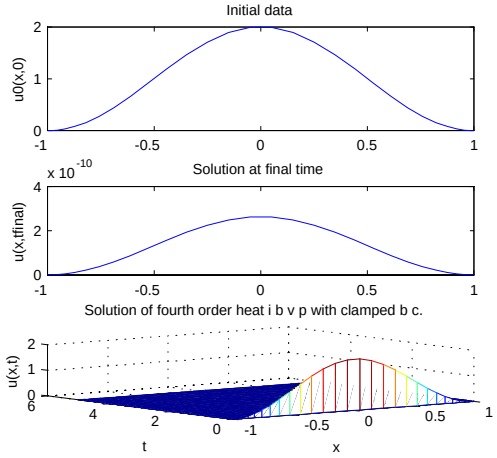

Figure 3.15: The solution to the fourth order heat equation

and the initial data $u(x,0) = \cos(\pi x)$ and $u_t(x,0) = \sin(\pi x)$. *Answer* The solution is drawn in Fig. 3.16. ▲

6. Solve the Schrödinger equation (3.9) with the $|u(x,t)| \to \infty$ as $|x| \to \infty$ boundary conditions and "shock" data initial conditions, i.e.,

$$u(x,0) := A \cdot \exp(-i\mu|x|), \ A \in \mathbb{R}, \ \mu > 0.$$

Hint For $A = \mu = 1$, the solution is represented in Fig. 3.17.▲

7. Solve the Schrödinger equation (3.9) with the initial data

$$u(x,0) := \frac{1}{2}\left(1 + \varepsilon \cos\left(\frac{\sqrt{2}}{4}x\right)\right), \varepsilon = 0.1,$$

which means a perturbation of plane wave solution (see [13], P. 3). The solution is depicted in Fig 3.18. ▲

8. Solve the wave equation, i.e., $u_{tt} = u_{xx}$, $(x,t) \in \mathbb{R} \times (0,.T]$ by a Hermite pseudospectral-Störmer/Verlet method starting with the initial data

$$u(x,0) = 0, \ u_t(x,0) = 2\sqrt{2}\sec h(x/\sqrt{2}).$$

Hint The solution is depicted in Fig. 3.19.▲

9. Solve by Fourier pseudospectral method the KdV equation (3.16) with periodic boundary conditions $u(-L) = u(L)$, $L > 0$. Use the following values of the parameters: $\alpha = -6$, $\rho = 0$, and $\nu = -1$. For the initial data take $u_0(x) := 6\sec h^2 x$. *Hint* The *soliton* solution is shown in Fig. 3.20.▲

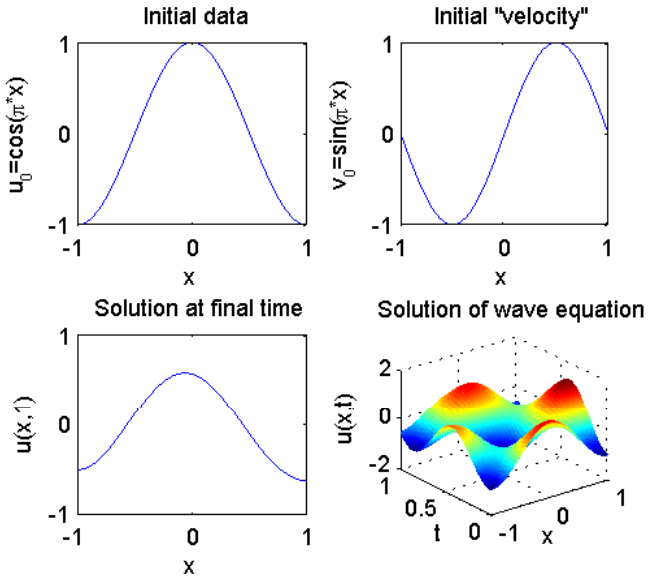

Figure 3.16: The solution to wave equation with Neumann boundary conditions

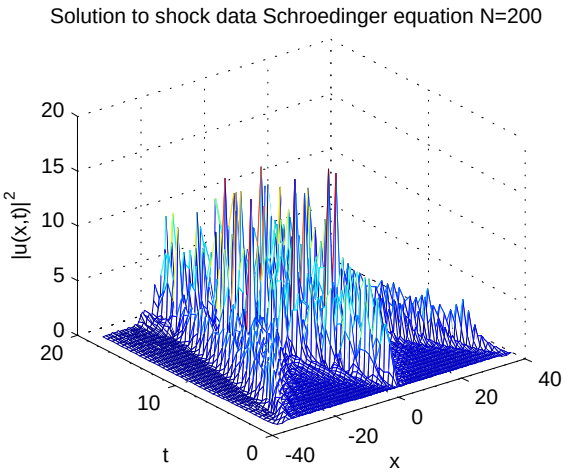

Figure 3.17: The solution to "shock" data Schrödinger equation

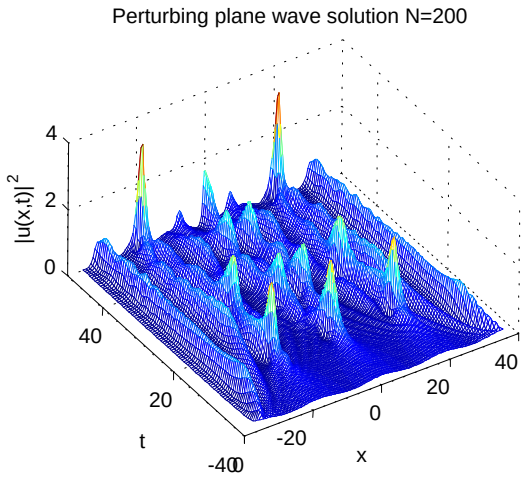

Figure 3.18: Perturbation of plane wave solution to Schrödinger equation

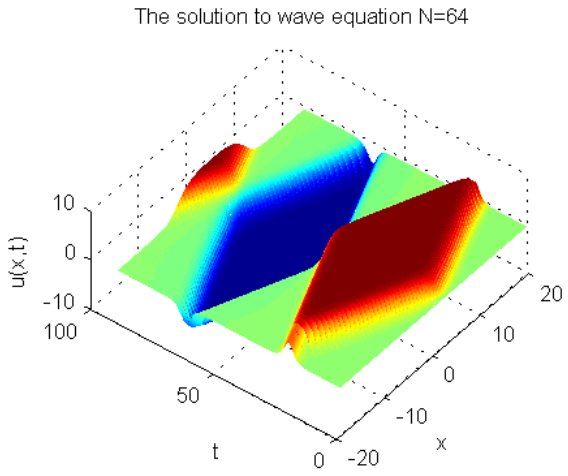

Figure 3.19: The solution to wave equation on the real line

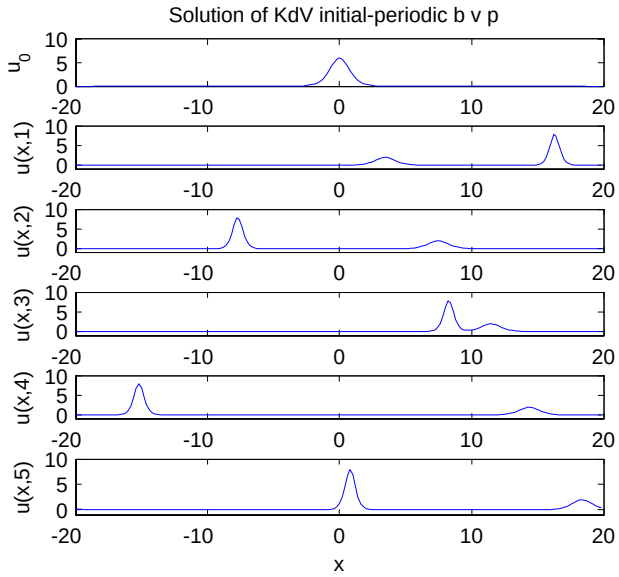

Figure 3.20: The soliton solution for KdV equation

10. Solve the *Fischer's* initial-boundary value problem on the real line(see for instance [143])

$$\begin{cases} u_t = u_{xx} + u\,(1-u)\,,\ t>0 \\ u(x,0) = \sin x, \\ u\,(x,t) \to 0,\ x \to \pm\infty,\ t>0. \end{cases}$$

Hint For a large positive L transform the interval $[-L\ L]$ into $[0\ 2\pi]$ and apply the Fourier pseudospectral method in order to discretize the spatial operator and Runge Kutta method to advance in time. The corresponding solution is displayed in Fig. 3.21. This solution blows up for $t > 1$, i.e., becomes unbounded. Compare this result with the numerical solution to Fischer's problem on a bounded interval, i.e., problem (3.8).▲

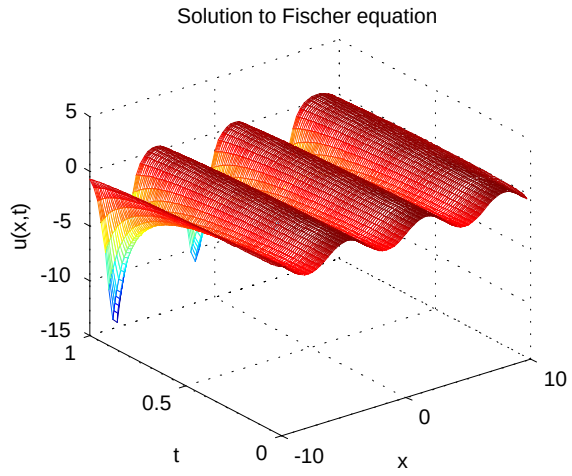

Figure 3.21: Solution to Fischer equation on the real line, $L = 10$, $N = 64$

Chapter 4

Efficient implementation

The aim of this part is to consider some results concerning the efficient implementation of some spectral methods. The chapter refers mainly to Chebyshev-tau and Chebyshev-Galerkin methods, i.e., to those methods which search the solutions in the *transformed space or spectral space.* In other words, we obtain the coefficients of the truncated series which represents the numerical solution, instead of the values of that in some points, as it is the case for collocation methods.

4.1 Second order Dirichlet problems for o. d. e.

Let us consider the *second order linear and homogeneous Dirichlet problem*

$$\begin{cases} Lu(x) = f(x), & x \in (-1,1), \\ u(\pm 1) = 0, \end{cases} \tag{4.1}$$

where the differential operator L is defined as

$$L() := \frac{d^2()}{dx^2} + \widetilde{L}(),$$

and $\widetilde{L}()$ is a first order linear differential operator.

First of all, we have to remark that a *second order non-homogeneous Dirichlet problem*

$$\begin{cases} Lu(x) = f(x), & x \in (-1,1), \\ u(-1) = a, \quad u(1) = b, \end{cases}$$

can be transformed in a homogenous one with the substitution

$$v(x) := u(x) - p(x),$$

where

$$p(x) := \frac{b-a}{2}x + \frac{b+a}{2}.$$

Consequently, we take into account exclusively the problem (4.1) and we focus our attention on the choice of the bases of our finite dimensional spaces such that the linear algebraic system obtained through discretization exhibit "good" properties such as well conditioning, spartness, low non-normality, etc.

Spectral methods involve representing the solution to a problem in terms of a truncated series of *smooth global functions.* In other methods such as finite elements and finite differences, the underlying expansion involves *local representations (interpolator)* such as piecewise polynomials. In practice, this means that the accuracy of spectral methods is superior. For instance, the rates of convergence associated with problems with smooth solutions are $O\left(\exp\left(-cN\right)\right)$ or $O\left(\exp\left(-c\sqrt{N}\right)\right)$, where N is the number of degrees of freedom in the expansion. In contrast, finite differences and finite elements yield rates of convergence that are only algebraic in N, typically $O\left(N^{-2}\right)$ or at most $O\left(N^{-4}\right)$. However, there is a price to be paid for using spectral methods instead of finite differences or finite elements, because:

- sparse matrices are replaced by full matrices;
- symmetric (normal) matrices are replaced by non-symmetric (non-normal) matrices;
- matrices with condition number $O(N^2)$ are replaced by worse conditioned matrices with condition number $O\left(N^4\right)$.

More than that, stability restrictions may become more severe and computer implementation, particularly for problems formulated on irregular domains, may not be fairly straightforward. Nevertheless, provided the solution is smooth enough, the rapid convergence of the spectral methods often compensates these shortcomings.

The issue of non-normality of matrices associated with the spectral approximations will be addressed in the second part of the work. Now we consider the conditioning of these matrices.

In his paper [103] Heinrichs derived an improved type of spectral methods with an $O\left(N^2\right)$ condition number. The main idea was to employ as trial functions polynomials fulfilling the homogeneous boundary conditions. Thus, he introduced the basis

$$X_N := \left\{\left(1-x^2\right)T_k\left(x\right),\ k=0,1,2,...,N\right\}, \tag{4.2}$$

and searched a solution of the form

$$u^N\left(x\right) = \sum_{k=0}^{N} a_k\left(1-x^2\right)T_k\left(x\right).$$

From the formulas (2.38) and (2.39) it becomes obvious that the matrices $B \in \mathbb{R}^{(N+1)\times(N+1)}$ and $D \in \mathbb{R}^{(N+1)\times(N+1)}$ with

$$b = B \cdot a, \tag{4.3}$$

and

$$d = D \cdot a, \tag{4.4}$$

where $a = (a_0\ a_1 ... a_N)^T$, $b = (b_0\ b_1 ... b_N)^T$, $d = (d_0\ d_1 ... d_N)^T$ have special structures (remember that b_k.and d_k are respectively the coefficients of $\left(u^N(x)\right)''$ and $u^N(x)$ with respect to system T_K). Thus, B is a positive upper triangular matrix with eigenvalues $(p+1)(p+2)$, $p = 0, 1, ..., N$ and hence an $O\left(N^2\right)$ condition number is expected. The system (4.3) can be solved in about $O(N)$ arithmetic operations. The matrix D is a band one with semibandwidth equals 3. Unfortunately, it is not symmetric but very close to, the first three rows being responsible for this inconvenient.

The (CT) method for 1D Helmholtz equation (2.19) with X_N using Heinrichs' basis (4.2) and Chebyshev polynomials as test functions reads

$$\left(B + \lambda^2 D\right) \cdot a = f, \tag{4.5}$$

where as usual f contains Chebyshev coefficients of function $f(x)$.

The strong form of (CC) method for the same problem reads

$$\left(\sum_{k=0}^{N} b_k T_k\right)(x_j) = f(x_j), \ j = 0, 1, 2, ..., N, \tag{4.6}$$

where the x_j are (CGaussL) nodes $x_j = \cos(j\pi/N)$.

Another attempt was made to improve the properties of the matrix of the system (2.20) in the classical (unmodified) (CT) method. Thus, if the solution has the form

$$u^N = \sum_{k=0}^{N} \widehat{u}_k T_k,$$

and its derivatives are denoted

$$\left(u^N\right)^{(\nu)} = \sum_{k=0}^{N} \widehat{u}_k^{(\nu)} T_k,$$

the next general result can be proved by simple algebraic manipulations.

Lemma 60 *For the coefficients* $\widehat{u}_k$, $\widehat{u}_k^{(1)}$, $\widehat{u}_k^{(2)}$ *hold*

$$\begin{aligned}
(i)\ & c_k \widehat{u}_k^{(1)} - \widehat{u}_{k+2}^{(2)} &&= 2(k+1)\widehat{u}_{k+1},\ k = 0, 1, \ldots, N-1, \\
(ii)\ & \frac{c_{k-2}\widehat{u}_{k-2}^{(1)}}{4k(k-1)} - \frac{\widehat{u}_k^{(1)}}{2(k^2-1)} + \frac{\widehat{u}_{k+2}^{(1)}}{4k(k+1)} &&= \frac{\widehat{u}_{k-1} - \widehat{u}_{k+1}}{4k},\ k = 0, 1, \ldots, N-2, \\
(iii)\ & c_k \widehat{u}_k^{(2)} - \widehat{u}_{k+2}^{(2)} &&= 2(k+1)\widehat{u}_{k+1}^{(1)},\ k = 0, 1, \ldots, N-2, \\
(iv)\ & \frac{c_{k-2}\widehat{u}_{k-2}^{(2)}}{4k(k-1)} - \frac{\widehat{u}_k^{(2)}}{2(k^2-1)} + \frac{\widehat{u}_{k+2}^{(2)}}{4k(k+1)} &&= \widehat{u}_k,\ k = 0, 1, \ldots, N-2.
\end{aligned}$$

To obtain a quasi-triangular form for the matrix of the (2.20) system, Dennis and Quartapelle [47] suggest to successively substrate the fourth row from the second, the fifth from the third and so on, and then each and every equation to be respectively divided by $2(k+1)$ $k=0,1,2,...,N-2$. Only the first two rows, corresponding to the boundary conditions, remain fully populated and consequently the matrix is improved. For the problem (2.19) the system can be solved directly. However, the condition number of the matrices generated by this procedure remains $O(N^4)$.

A variant of (CG) method, the so-called *Petrov-Galerkin method,* leads to banded matrices with condition number of order $O(N^2)$. The trial and test spaces are respectively

$$\begin{array}{c} X_N := span\left\{\Psi_k|\ \Psi_k(x)=\left(1-x^2\right)T_k(x),\ k=0,1,2,...,N\right\} \\ Y_N := span\left\{\Phi_k|\ \Phi_k(x)=T_k(x)-T_{k+2}(x),\ k=0,1,2,...,N\right\}. \end{array} \tag{4.7}$$

They both satisfy Dirichlet boundary conditions. To apply this method to problem (4.1) the following lemma is very handy.

Lemma 61

$$a)\ \left(\Phi_k,\Psi_j''\right)_\omega=\frac{\pi}{2}a_{kj},\quad a_{kj}=\begin{cases} c_k(k+1)(k+2),\ j=k, \\ k(k+1),\ j=k+2, \\ 0,\ otherwise \end{cases} \tag{4.8}$$

$$b)\ \left(\Phi_k,\Psi_j'\right)_\omega=\frac{\pi}{2}b_{kj},\quad b_{kj}=\begin{cases} -c_{k-1}\frac{k+1}{2},\ j=k-1, \\ -\frac{c_k+3}{2},\ \ j=k+1, \\ \frac{k+1}{2},\ \ j=k+3, \\ 0,\ otherwise \end{cases} \tag{4.9}$$

$$b)\ \left(\Phi_k,\Psi_j\right)_\omega=\frac{\pi}{2}c_{kj},\quad c_{kj}=\begin{cases} -\frac{c_{k-2}}{4},\ j=k-2, \\ \frac{-c_{k-1}+2c_k+2}{4},\ j=k, \\ -\frac{3}{4},\ \ j=k+2, \\ \frac{1}{4},\ j=k+4, \\ 0,\ otherwise. \end{cases} \tag{4.10}$$

4.2 Third and fourth order Dirichlet problems for o. d. e.

Let us consider first the third order boundary value problem, namely

$$\begin{cases} u'''=f(x),\ -1<x<1, \\ u(\pm1)=u'(-1)=0. \end{cases} \tag{4.11}$$

For the spectral approximation we modified the strong Chebyshev collocation scheme following an idea of W. Heinrichs from [107]. More exactly, we solved the discrete problem

$$\begin{cases} find\quad u^N\in\mathcal{P}_N\quad such\ that \\ \left(u^N(x_i)\right)'''=f(x_i),\quad x_i\in(-1,1),\ i=1,2,\ldots,N-2, \\ u^N(\pm1)=u^N(-1)=0, \end{cases} \tag{4.12}$$

where the abscissas x_i are the Chebyshev Gauss-Lobatto nodes defined by (1.34), i.e.,

$$x_i := \cos \frac{i\pi}{N-1},\ i = 0, 1, 2, ..., N-1.$$

The interpolator $u^N(x)$ is defined by

$$u^N(x) := \sum_{j=1}^{N-2} u(x_j) \frac{1+x}{1+x_j} l_j(x), \tag{4.13}$$

where $l_j(x) \in \mathcal{P}_{N-1}$, $j = 0, 1, 2, ..., N-1$ is the Lagrangian basis relative to the above nodes x_j, $j = 0, 1, 2, ..., N-1$. Since $l_j(\pm 1) = 0$, $j = 1, 2, ..., N-2$ the Dirichlet boundary conditions $u(\pm 1) = 0$, are automatically satisfied. The factor $(1+x)$ in front of $l_j(x)$ yields the Neumann boundary condition in $x = -1$. Hence $u^N(x)$ implicitly contains the boundary conditions. This is the major difference between the approximation (4.13) and the classical Chebyshev collocation representation (2.22) which does not contain the boundary conditions.

Now, the third derivative is given by

$$\left(u^N(x)\right)''' := \sum_{j=1}^{N-2} \frac{u(x_j)}{1+x_j} \left[3l_j''(x) + (1+x_j) l_j'''(x)\right],\ i = 1, 2, ..., N-2.$$

The implementation of this derivative was carried out using the MATLAB code `poldif.m` from [204]. Eventually, we solved the linear algebraic system of $N-2$ equations

$$\left(u^N(x_i)\right)''' = f(x_i), \quad x_i \in (-1, 1),\ i = 1, 2, \ldots, N-2,$$

for the $N-2$ unknowns $u(x_j)$, $j = 1, 2, ..., N-2$. When the right hand side term was such that the exact solution is $u(x) := \left(1-x^2\right)(1-x)\exp(-10x)$, the solution was found within an error of order $O\left(10^{-10}\right)$. Corresponding to $N = 64$ it is depicted in Fig. 4.1 along with the exact solution.

Remark 62 *As the solution $u(x)$ of the above problem is an analytic function and the error committed by the spectral method decays to zero at an* ***exponential rate,*** *we state that the method works with an exponential accuracy (see the paper of E. Tadmor [189] about the exponential accuracy of Fourier and Chebyshev differencing methods). This should be compared with the* ***polynomial decay rate*** *obtained by finite difference or finite elements methods.*

Let us consider the fourth order boundary value problem, namely

$$\begin{cases} u^{(iv)} + u = f(x),\ -1 < x < 1, \\ u(\pm 1) = u'(\pm 1) = 0, \end{cases} \tag{4.14}$$

with $f(x)$ such that the solution is $u(x) = \left(1-x^2\right)^2 \exp(x)$. In the (CG) method we have considered the solution was approximated by

$$u^N := \sum_{k=0}^{N} a_k \left(1-x^2\right)^2 T_k(x),$$

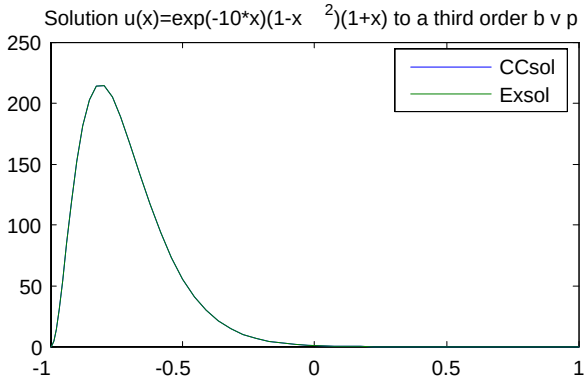

Figure 4.1: A Chebyshev collocation solution for a third order b. v. p.

and the trial functions have been $\left(1-x^2\right)^2 T_k(x)$. All these functions satisfy the boundary conditions. The algebra of this method is much more tedious than the algebra of the (CS) method. The only positive computational aspect is the fact that the matrix of the left hand side of the algebraic system for unknown coefficients a_k, is quasi symmetric, i.e., the Henrici number equals $5.159131440655628e{-}001$ (see for the definition of this measure of non-normality 6.3).

Remark 63 *It is worth noting at the end of this chapter that Legendre-Galerkin method for linear elliptic problems leads to symmetric and simpler linear systems than Chebyshev counterpart. Unfortunately, its efficiency is limited by the lack of fast transformation between physical space and transformed space. Furthermore the Legendre-Gauss-Lobatto nodes (LGL) are not available in an explicit form and their evaluation for large N may introduce significant rounding errors. (cf. J Shen, [177]).*

4.3 Problems

1. Solve the *second order differential linear eigenvalue problem* which is a regular one that looks singular

$$\begin{cases} -\left(w(x)u'\right)' = \lambda w(x)u, \ (-1,1), \ w(x) := \left(1-x^2\right)^{-1/2}, \\ u(\pm 1) = 0, \end{cases} \tag{4.15}$$

by (CT), (CC) and (CG) methods. Estimate in each and every case the condition number of "stiffness" matrix A associated with the numerical scheme. *Hint* A weak formulation reads as follows:

$$\begin{cases} find\ u \in H^1_{\omega,0}(-1,1)\ such\ that \\ (u',v')_\omega = \lambda(u,v)_\omega, \ \forall v \in H^1_{\omega,0}(-1,1). \end{cases}$$

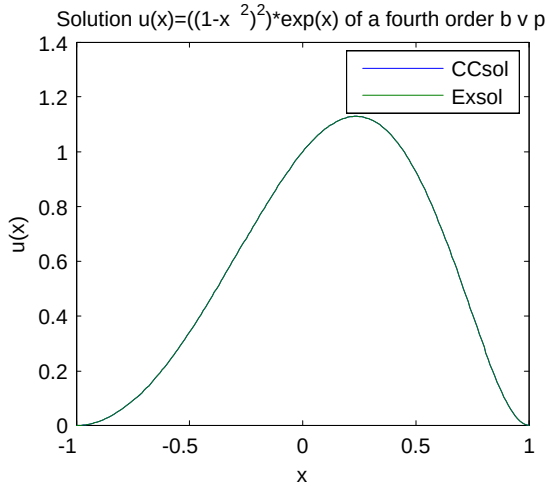

Figure 4.2: The (CG) solution to a fourth order problem, $N = 32$

Eventually, we get $\lambda_0 = 3.559279966,\ \lambda_9 = 258,8005854\ ,\ \lambda_{24} = 1572.635284$ results confirmed by J. D. Pryce in [167].

2. Solve by the same numerical methods the *second order differential linear eigenvalue problem* which is a regular one with oscillatory coefficients (see [167] and also [166])

$$\begin{cases} -u'' + q(x)\, u = \lambda u,\ x \in (-1,1)\,, \\ u\left(\pm\frac{\pi}{2}\right) = 0, \end{cases} \tag{4.16}$$

where the potential $q(x)$ is the so -called Coffey-Evans potential defined by $q(x) = -2\beta\cos 2x + \beta^2 \sin^2 2x,\ \beta = 20.;\ 30.;\ 50.$ Observe the dependence of eigenvalues on the parameter β.

3. Solve the singular eigenvalue problem

$$\begin{cases} -u'' + q(x)\, u = \lambda u,\ x \in (-1,1)\,, \\ u(\pm L) = 0, \end{cases} \tag{4.17}$$

where the potential $q(x)$ is the so-called *Morse potential* defined by $q(x) = \beta\exp(-2x) - 2\beta\exp(-x),\ \beta = 9$, and observe the dependence of the spectrum on L when $L \to \infty$.

4. Use a Chebyshev type method in combination with the change of variables

$$t := \frac{x-2}{x},$$

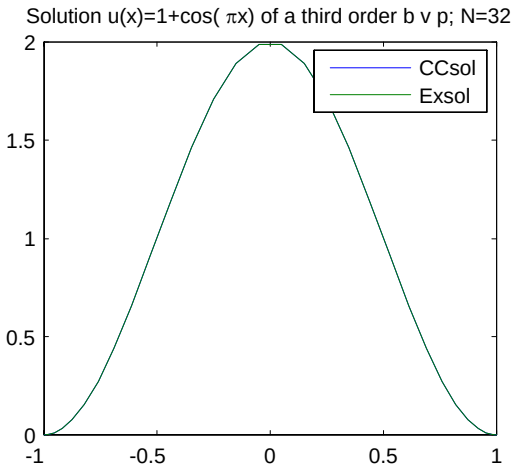

Figure 4.3: Another solution for the third order problem

in order to solve the *non-standard eigenvalue problem*

$$\begin{cases} -u'' + q(x)\,u = \lambda u, \ x \in [1, \infty), \\ \lambda u(1) + u'(1) = 0, \ u(x) \to 0, \ as \ x \to \infty, \end{cases}$$

where $q(x) = -\frac{1}{4x^2}$. *Hint* This problem has a continuous spectrum $(0, \infty)$ with a single eigenvalue $\lambda_0 = 0.0222989949$.

5. Solve the problem (4.11) when the exact solution is given by $u(x) := 1 + \cos(\pi x)$. In order to observe the spectral accuracy use the above modified (CC) method. *Hint* The rounding error accuracy is already reached for $N = 32$, i.e., $norm((CCsol - Exsol), inf) = 3.743672039036028e - 013$. The solution is represented in Fig. 4.3.▲

Chapter 5

Eigenvalue problems

In solving linear eigenvalue problems by a spectral method using (N+1) terms in the truncated spectral series, the lowest N/2 eigenvalues are usually accurate to within a few percent while the larger N/2 numerical eigenvalues differ from those of differential equation by such large amounts as to be useless.

J P Boyd's eigenvalue rule-of-thumb, [19], P. 132

As it is well known, the main trouble with numerical methods for differential eigenvalue problems consists in the appearance of *spurious eigenvalues.* This chapter is devoted to such problems. Standard as well as non-standard problems, i.e., eigenvalue problems with boundary conditions depending on spectral parameter, are in turn examined using spectral methods.

In a series of works, Golub and Wilkinson [87], Golub and Ortega [90] and more recently Golub and van der Vorst [89], it is observed that the computing of eigenvalues and vectors is essentially more complicated than solving linear systems. Research in this area of numerical linear algebra is very active, since there is a heavy demand for solving complicated problems associated with stability and perturbation analysis for practical applications. For standard problems, powerful tools are available, but there still remain many open problems.

J. P. Boyd in [18] makes several observations about the traps and snares in eigenvalue computations, but concludes that with a bit of care the pseudospectral method is a very robust and efficient method for solving differential eigenproblems, even with singular eigenmodes.

5.1 Standard eigenvalue problems

The advantages of spectral methods in solving differential eigenvalue problems were underlined for the first time by S. Orszag in his seminal paper [155]. There exist situations, for example in hydrodynamic stability, when the accuracy of

the numerical method is essential. In the paper quoted above the following *standard Orr-Sommerfeld problem* is analyzed:

$$(OS) \quad \begin{cases} \Phi^{(iv)} - 2\alpha^2\Phi'' + \alpha^4\Phi = i\alpha R\left[(U-\lambda)\left(\Phi'' - \alpha^2\Phi\right) - U''\Phi\right], \ x \in (-1,1), \\ \Phi(\pm 1) = \Phi'(\pm 1) = 0, \end{cases} \tag{5.1}$$

where $(\lambda, \Phi(x))$ is the *unknown eigenpair* and the parameters α, R and function $U(x)$ have physical signification, are arbitrary but fixed. The efficiency of the Chebyshev-tau method was overwhelming even for small order N of approximation.

The superiority of spectral approximation in approximating the spectrum of a differential operator was underlined also by Weideman and Trefethen in their paper [203]. They were concerned with the classical second order eigenvalue problem

$$\begin{cases} u''(x) = \lambda \cdot u(x), \quad x \in (-1,1), \\ u(\pm 1) = 0, \end{cases} \tag{5.2}$$

which has the eigenvalues in the closed form $\lambda_k = -\frac{k^2\pi^2}{4}$, $k = 0, 1, 2, \ldots$.

In order to emphasize the capabilities of spectral methods in solving differential eigenvalue problems we consider other four special problems:

a) the *clamped road* problem, i.e., the fourth order eigenvalue problem (see Funaro and Heinrichs [76])

$$\begin{cases} \Phi^{(iv)}(x) = \lambda \cdot \Phi(x), \quad x \in (-1,1), \\ \Phi(\pm 1) = \Phi'(\pm 1) = 0, \end{cases}$$

b) a *fourth-order eigenvalue problem* with a third derivative term

$$\begin{cases} \Phi^{(iv)}(x) + R \cdot \Phi'''(x) = \lambda \cdot \Phi''(x), \quad x \in (-1,1), \\ \Phi(\pm 1) = \Phi'(\pm 1) = 0, \end{cases} \tag{5.3}$$

where R is a real parameter and the eigencondition reads

$$\left(R^2 + 4\lambda\right)^{1/2}\left[1 - \frac{\cosh\left(\left(R^2+4\lambda\right)^{1/2}\right)}{\cosh(R)}\right] + \frac{2\lambda\sinh\left(\left(R^2+4\lambda\right)^{1/2}\right)}{\cosh(R)} = 0;$$

c) a *third order problem*, namely

$$\begin{cases} u'''(x) = \lambda u(x), \quad x \in (-1,1), \\ u(\pm 1) = 0, \ u'(-1) = 0, \end{cases} \tag{5.4}$$

with the eigencondition

$$e^{3\lambda^{1/3}} - 2\sin\left(\sqrt{3}\lambda^{1/3} + \frac{\pi}{6}\right) = 0,$$

which implies the following approximative values for eigenvalues

$$\lambda_k = -\left\{\left(k + \frac{1}{6}\right)\frac{\pi}{\sqrt{3}}\right\}^3, \quad k = 1, 2, \ldots .$$

d) and *the Orr-Sommerfeld problem* (5.1) with $U(x) := 1 - x^2$, i.e., corresponding to Poiseuille flow.

The first three problems were solved using modified Chebyshev collocation methods. The difference with respect to the classical (CC) method consists in the fact that each and every trial function satisfies the boundary conditions.

Consequently, for the first two problems we use as *trial functions* the product of the weight $\left(1-x^2\right)^2$ and the cardinal functions, i.e.,

$$\left(1-x^2\right)^2 l_j(x),\ j = 0, 1, 2, ..., N.$$

They satisfy all boundary conditions. The pseudospectral derivative matrices are, as usual, obtained from the paper of Weideman and Reddy, [204], p.479. It is worth noting at this moment that the cardinal functions are rewritten in the form (see [71], P. 15)

$$l_j(x) = \frac{(-1)^j}{c_j}\frac{1-x^2}{(N-1)^2}\frac{T'_{N-1}(x)}{x-x_j},$$

where $c_1 = c_N = 2$ and $c_2 = ... = c_{N-1} = 1$ and

$$x_j = \cos\left(\frac{(j-1)\pi}{N-1}\right),\ j = 1, 2, ..., N,$$

are the (1.34) nodes, i.e., the Chebyshev points of the second kind or, equivalently, the extreme points on $[-1,\ 1]$ of $T_{N-1}(x)$.

The first 14th eigenvalues of the problem a), for $N = 16$, are displayed in the Table 1.

N = 16
$1.71337968904269 \times 1.0e+007$
1.62401642422808
0.03856105241227
0.03597677558196
0.00711035649235
0.00603075735206
0.00305369477203 ↕
0.00193928004466
0.00108631975968
0.00055710074688
0.00024964874758
0.00009136018866
0.00002377210675
$0.00000312852439 \times 1.0e+007$

Table 1 The first eigenvalues for problem a)

The matrix associated with this problem is quasi fully populated as can be seen in the Fig. 5.1 where the locations of nonzero elements are plotted.

The first two exact eigenvalues for the second problem with $R = 0$, are -9.8696044 and -20.1907286. They are exactly reproduced by our computations. For such R the problem is self-adjoint and all the eigenvalues are real

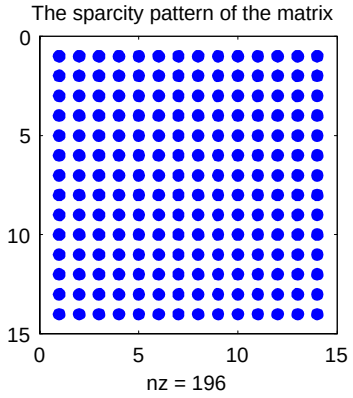

Figure 5.1: The sparcity pattern for the (CC) matrix

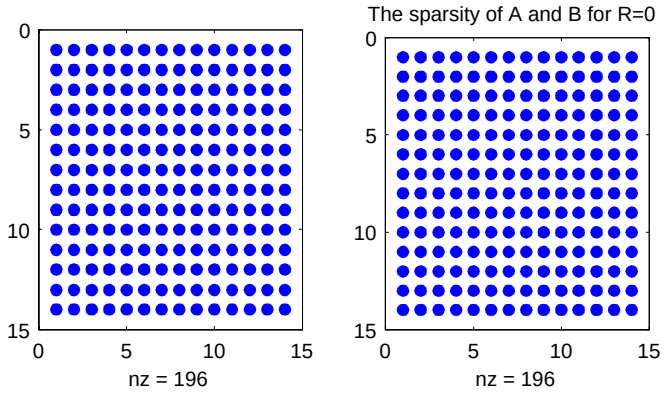

Figure 5.2: The sparsity pattern for matrices A and B; (CC) method

and less than zero. The eigenvalue problem is a generalized one this time, i.e., of the form

$$A \cdot x = \lambda B \cdot x,$$

and the sparsity patterns of A and B are depicted in Fig. 5.2. The matrices remain fully populated.When the R is non-zero the problem is not self-adjoint and the eigenvalues are no longer real, though the real part is negative. For $R = 4$, the exact eigenvalues are $-17.9129218 \pm 9.45840144$ and they are reproduced accurately by our computations. All eigenvalues of this problem are depicted in Fig. 5.3. It may also be shown that $\lambda = -R^2/4$ is not an eigenvalue.

Remark 64 *This problem was solved by Gardner, Trogdon and Douglas using a laborious modified Chebyshev-tau method in their paper [80]. They used the EISPACK driver RG on a Cray-2 supercomputer with single precision (64 bit)*

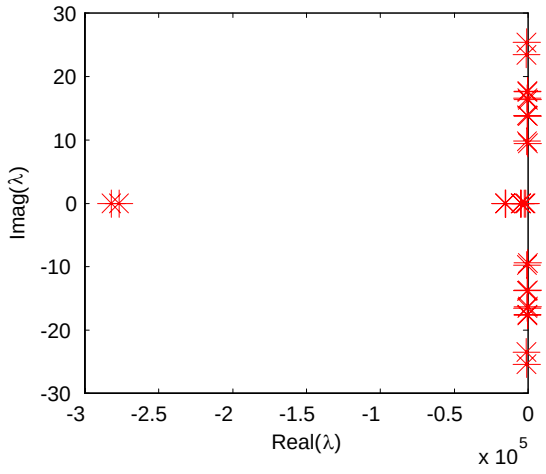

Figure 5.3: The set of eigenvalues when N=20 and R=4

arithmetic. No spurious eigenvalues are reported in their paper and the convergence is clearly indicated by magnitudes of the tau coefficients. Our numerical experiments were carried out on more modest computing equipment but in spite of this no spurious eigenvalues were observed. We used the EIG and EIGS codes from MATLAB. For $R = 0$ the problem is also solved in the paper of McFadden, Murray and Boisvert, [138], P. 229.

The problem c) is the eigenvalue problem attached to problem (4.11). The (CC) approach for that reads

$$\left\{ \begin{array}{c} find \quad u^N \in \mathcal{P}_N \quad such\ that \\ \left(u^N(x_i)\right)''' = \lambda u^N(x_i), \quad x_i \in (-1,1),\ i = 1,2,\ldots,N-2, \\ u^N(\pm 1) = u^N(-1) = 0. \end{array} \right.$$

Hence we have $N-2$ inner collocation and 3 boundary conditions. These are $N+1$ conditions which uniquely determine the spectral approximation u^N defined again by (4.13).

N = 64
$-2.524068539853232e+004$
$-2.081344964224966e+004$
$-1.696640042472849e+004$
$-1.362044995037454e+004$
$-1.362044995037454e+004$
$-8.308798671337761e+003$
$-6.270519503199299e+003$
$-4.596233100588465e+003$
$-3.250141182185935e+003$
$-2.196440035490058e+003$
$-1.399326817084188e+003$
$-8.229985429508488e+002$
$-4.316522518291442e+002$
$-1.894849798380508e+002$
$-6.069365841193836e+001$
$-9.482406935491648e+000$

Table 2 The first 16th eigenvalues of the problem c)

We have to notice that only the first 16th eigenvalues of this problem were computed with a reasonable accuracy. They are represented in Table 2 and suggest that the Boyd's rule-of-thumb is too optimistic.

The fourth problem, (5.1), was solved in [164] by a *modified Chebyshev Galerkin approach*. More specifically, as trial basis was used the following *Heinrich's basis* ([105])

$$\Psi_k(x) := \left(1-x^2\right)^2 T_k(x), \quad k = 0, 1, 2, ..., N$$

and as a test basis, the *Shen's basis* ([176])

$$\Phi_k(x) := T_k(x) - \frac{2(k+2)}{k+3} T_{k+2}(x) + \frac{k+1}{k+3} T_{k+4}(x), \quad k = 0, 1, 2, ..., N.$$

This approach leads to *banded matrices,* as results from the lemma below.

Lemma 65 *[163] For $\Psi_k(x)$ and $\Phi_k(x)$ defined as above we have $\left(\Phi_i, \Psi_j^{(k)}\right)_2 := \frac{2}{\pi} a_{ij}^k$, where*

$$a_{ij}^4 = \begin{cases} c_i (i+1)(i+2)(i+3)(i+4), & j = i, \\ -2i(i+1)(i+2)(i+4), & j = i+2, \\ i(i+1)^2(i+2), & j = i+4, \\ 0, & otherwise, \end{cases} \tag{5.5}$$

$$a_{ij}^2 = \begin{cases} \frac{c_{i-2}(i+1)(i+2)}{4}, & j = i-2, \\ -\frac{2i(i+3)+4c_i-3(c_{i-1}-d_{i-1})}{2}, & j = i, \\ \frac{3(i+2)(i+1)}{2} + \frac{\delta_{i0}}{2}, & j = i+4, \\ -(i+1)(i+2), & j = i+4, \\ \frac{(i+1)(i+2)}{4}, & j = i+6, \\ 0, & otherwise, \end{cases} \tag{5.6}$$

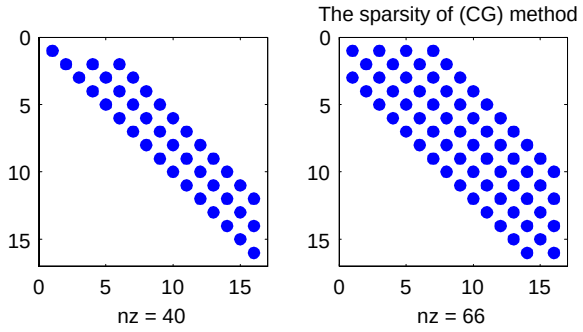

Figure 5.4: The sparsity of (CG) method

$$a_{ij} = \begin{cases} \frac{c_{i-4}}{16}, & j = i-4, \\ -\frac{c_{i-2}(3i+8)-3(c_{i-3}-d_{i-3})}{8(i+3)}, & j = i-2, \\ \frac{15i+35c_i-22(c_{i-1}-d_{i-1})+5(c_{i-2}-d_{i-2})}{16(i+3)}, & j = i, \\ -\frac{5i+6+4c_i-c_{i-1}+d_{i-1}}{4(i+3)}, & j = i+2, \\ \frac{15i+22+3c_i}{16(i+3)}, & j = i+4, \\ -\frac{3i+4}{8(i+3)}, & j = i+6, \\ \frac{i+1}{16(i+3)}, & j = i+8, \\ 0, & otherwise, \end{cases} \tag{5.7}$$

where $\delta_{m,n}$ is the Kronecker symbol, c_i are defined as in (1.2) and

$$d_i := \begin{cases} 0, & if \quad i < 0, \\ 1, & if \quad i \geq 0. \end{cases} .$$

With these bases the condition number of the fourth order differentiation is reduced to $O\left(N^4\right)$. No spurious eigenvalues were obtained and we achieved an accuracy up to eight digits. In fact, for $\alpha = 1.00,\ R = 10000,\ N = 48$ we obtained for the first eigenvalue the value $\lambda = 0.23752651907 + 0.00373967171i$ and the paper of Orszag [155] furnishes $\lambda = 0.23752648882 + 0.003739677062i$.

Remark 66 *The problem b) was also solved by a Chebyshev-Galerkin method with the Heinrich's basis and the Shen's basis. The matrices A and B are simpler this time. The matrix A is upper triangular and the matrix B is banded. For $R = 0$ and $N = 16$ their sparsity patterns can be seen in Fig. 5.4. To get these patterns we used the MATLAB command* **`spy`**.

5.2 Theoretical analysis of a model problem

The aim of this section is to carry out some estimations concerning the numerical results obtained when the problem

$$\begin{cases} u^{(iv)}(x) = \lambda \cdot u(x), \quad x \in (-1,1), \\ u(\pm 1) = u'(\pm 1) = 0. \end{cases} \tag{5.8}$$

is solved by Chebyshev-Galerkin method. We shall use some ideas from the paper of Weideman and Trefethen [203].

The method means to solve the problem

$$\begin{cases} find\ u^N \in X_N, \\ such\ that\ \left(\left(u^N\right)^{(iv)}, v\right)_{0,\omega} = \lambda \left(u^N, v\right)_{0,\omega}, \quad \forall v \in X_N \end{cases} \tag{5.9}$$

where $X_N := \mathcal{P}_{N+4} \cap H^4_\omega(-1,1) \cap H^2_\omega(-1,1)$. For the space X_N we consider two distinct bases, namely the *Heinrich's basis*

$$\left\{\Phi_k | \Phi_k(x) := \left(1-x^2\right)^2 T_k(x),\ k = 0,1,2,...,N\right\} \tag{5.10}$$

and the *Shen's basis*

$$\left\{\Psi_k | \Psi_k(x) := T_k(x) - \frac{2(k+2)}{k+3} T_{k+2}(x) + \frac{k+1}{k+3} T_{k+4}(x),\ k = 0,1,...,N\right\}. \tag{5.11}$$

Let us take Φ_k from (5.10) as shape (test) functions and Ψ_k from (5.11) as trial functions. It means that

$$u^N(x) := \sum_{j=0}^{N} a_k \Phi_k(x),$$

and the weak formulation (5.9) reads

$$\begin{cases} find\ u^N \in X_N, \\ such\ that\ \left(\left(u^N\right)^{(iv)} - \lambda u^N, \Psi_k\right)_{0,\omega} = 0, \quad k = 0,1,2,...,N. \end{cases}$$

Mathematically, this means an algebraic $N+1$ dimensional *generalized eigenvalue problem*

$$\widetilde{D^4} \cdot x = \lambda \widetilde{I}_{N+1} \cdot x,$$

where $\widetilde{I}_{N+1}$ is the matrix of order $N+1$,

$$\widetilde{I}_{N+1} := [(\Phi_k, \Psi_j)_{0,\omega}]_{k,j=0,1,...,N},$$

and $\widetilde{D}^4$ is the matrix, of the same order,.defined by

$$\widetilde{D}^4 := [\left(\Phi_k^{(iv)}, \Psi_j\right)_{0,\omega}]_{k,j=0,1,...,N}.$$

An upper bound for the numerical eigenvalues of the problem (5.9) is given by the following estimation (cf. I. S. Pop, P. 46)

$$\lim_{n\to\infty} \sup |\lambda| \leq 2.77 \cdot 10^{-4} \cdot N^8.$$

Remark 67 *The last estimation is very rough but the main computational difficulty related to this eigenvalue problem lies in the rapid growth in the difference between the orders of magnitude of the computed eigenvalues. Thus, for* N *around* 150 *results become too large to represent as conventional floating-point values (overflow).*

5.3 Non-standard eigenvalue problems

Let us consider the following differential eigenvalue problem where the eigenvalue parameter λ enters into boundary conditions:

$$\left\{ \begin{array}{c} \Phi^{(iv)} - 2\alpha^2\Phi'' + \alpha^4\Phi = i\alpha R\left[(U-\lambda)\left(\Phi'' - \alpha^2\Phi\right) - U''\Phi\right],\ x \in (0,1), \\ (i), (ii) \quad \Phi(1) = \Phi'(1) = 0, \\ (iii) \quad \left[\Phi''(0) + \alpha^2\Phi(0)\right](\lambda - U(0)) + \Phi(0)\,U''(0) = 0 \\ (iv) \quad U''(0)\,\Phi'''(0) + i\alpha\left[R(\lambda - U(0)) + 3i\alpha\right]U''(0)\,\Phi'(0) - \\ -i\alpha\left[2\cot(\beta) + \alpha^2 Ca + (\lambda - U(0))\,RU'(0)\right]\left[\Phi''(0) + \alpha^2\Phi(0)\right] = 0. \end{array} \right. \tag{5.12}$$

The problem comes from physicochemical hydrodynamics, with parameters α, β, τ, Ca and R ranging in specified intervals of real axis and the *basic flow* is

$$U(x) := \left(1 - x^2\right) + (1 - x)\,\tau.$$

We call this a *non-standard Orr-Sommerfeld eigenvalue problem* and observe that the differential operators are identical in (5.1) and (5.12). The difference appears at the last two boundary conditions, where the spectral parameter λ enters linearly. It is not easy to implement these boundary conditions in a collocating or a Galerkin method. In our paper [82] we introduced a *modified Chebyshev-tau method* in order to circumvent these difficulties.

For the *classical Chebyshev-tau formulation* of the problem (5.12) we transform it first (linearly) on the interval $(-1\ \ 1)$ and then make use of the following spaces and the corresponding bases

$$\begin{array}{rcl} X_N & : & = \mathcal{P}_{N+4} = span\left\{T_k,\ \ k = 0,1,2,...,N+4\right\}, \\ Y_N & : & = \mathcal{P}_N = span\left\{T_k,\ \ k = 0,1,2,...,N\right\}. \end{array} \tag{5.13}$$

This formulation reads

$$\left\{ \begin{array}{c} find \quad \Phi_N \in X_N, \\ such\ that \quad (L_t\Phi_N, \varphi)_{0,\omega} = 0, \quad \forall \varphi \in Y_N, \\ B.C.'s, \end{array} \right. \tag{5.14}$$

where L_t denotes the differential operator and by B. C. we mean an explicit impose of boundary conditions. Replacing φ by T_k, $k = 0, 1, 2, ..., N$ successively and taking $\Phi_N := \sum_{j=0}^{N+4} a_j T_j$ we get $N+1$ equations for the coefficients a_j, $j = 0, 1, 2, ..., N+4$ and the spectral parameter λ. The remaining four equations are given by the boundary conditions in ± 1. Hence we have obtained a *generalized eigenvalue problem* of the form $Aa = \lambda Ba$.

The discretization matrices generated by this method are badly conditioned and not sparse. This can be illustrated, for example, for the fourth order differentiation matrix corresponding to Φ_N, namely

$$\Phi_N^{(iv)} = \sum_{j=0}^{N+4} a_j^{(iv)} T_j,$$

where

$$a_j^{(iv)} = \frac{1}{c_j} \sum_{\substack{p=j+4, \\ p+j \ even}}^{N+4} p\left[p^2\left(p^2-4\right)^2 - 3j^2p^4 + 3j^4p^2 - j^2\left(j^2-4\right)^2\right] a_p. \quad (5.15)$$

Thus, even though this method has theoretically a spectral accuracy, it is strongly affected by round off errors. Moreover, for our eigenvalue problem it generates two spurious eigenvalues.

The difference between this classical method and the modified version consists in the spaces involved in the discretization process. We define the functions

$$\Theta_i(x) := \left(1-x^2\right) T_i(x), \ i \geq 0,$$

and approximate the solution Φ by $\Phi_N := \sum_{j=0}^{N+2} a_j \Theta_i$. Clearly,

$$\widetilde{X}_N := span\left\{\Theta_i, \ i = 0, 1, 2, ..., N+2\right\} = \left\{v \in \mathcal{P}_{N+4} | v(1) = v'(1) = 0\right\},$$

and therefore each and every $\Phi_N \in \widetilde{X}_N$ satisfies the boundary conditions in 1.

For the definition of test functions we use the functions below

$$\Psi_i^1(x) := \frac{2}{\pi} d_i^N \left[\frac{2i+3}{4(i+1)} T_i(x) - T_{i+1}(x) + \frac{2i+1}{4(i+1)} T_{i+2}(x)\right], \quad i \geq 0,$$

where $d_i^N = 1$ if $0 \leq i \leq N$ and $d_i^N = 0$ otherwise. Now let

$$\Psi_i^{k+1}(x) := \frac{1}{2i+k+2}\left(\Psi_i^k(x) + \Psi_{i+1}^k(x)\right), \ i \geq 0, \ k \geq 1.$$

The choice of the test function spaces is justified by the following lemma.

Lemma 68 *[163] For $k \geq 1$ we have*

$$\widetilde{Y}_N := span\left\{\Psi_i^k,\ i = 0, 1, 2, ..., N\right\} = \left\{v \in \mathcal{P}_{N+2} | v(1) = v'(1) = 0\right\}.$$

Proof. The case $k = 1$ is obvious. For $k > 1$ the mathematical induction can be applied easily. ■

With the two spaces defined above we can proceed with the discretization of the problem (5.12). The new formulation reads

$$\begin{cases} & find \quad \Phi_N \in \widetilde{X}_N, \\ such\ that & (L_t \Phi_N, \varphi)_{0,\omega} = 0, \quad \forall \varphi \in \widetilde{Y}_N, \\ & B.C.'s, \end{cases}$$

but now B. C. stands only for the two boundary conditions in -1, which still have to be imposed explicitly. For this discretization we take $\Psi_i^5,\ \ i = 0, 1, 2, ..., N$ as test functions and construct the corresponding discretization matrices. The choice of these test functions is justified by the following lemma.

Lemma 69 *[163] Let $\Theta_i(x)$ and Ψ_i^5 as above. For $i = 0, 1, 2, ..., N$ and $j \geq 0$ we have $\left(\Psi_i^5, \Theta_i^{(iv)}\right)_{0,\omega} = d_j^{N+2} a_{ij}$, where*

$$a_{ij} = \begin{cases} \frac{i+4}{4(2i+5)}, & j = i+2, \\ -\frac{(i+2)(i+4)}{(2i+5)(2i+7)} & j = i+3, \\ \frac{i+2}{4(2i+7)} & j = i+4, \\ 0 & otherwise. \end{cases}$$

Proof. These relations can be obtained from orthonormal relations for Chebyshev polynomials. The algebra is quite tedious but not difficult, so we do not reproduce it here. ■

Remark 70 *In this approach, the discretization matrix for the fourth order derivative is banded. Compared with the one in the classical approach, which, as revealed in (5.15), is only upper triangular and more difficult to compute, it is better conditioned and therefore the method features much more stability. Similar discretization matrices are obtained for lower order derivatives or other differential operators.*

Remark 71 *The bases adopted here are suited only for boundary conditions in (5.12). Analogous ideas can be exploited for other types of (homogeneous) boundary conditions.*

Remark 72 *The numerical results reported in our paper [82], which refers to the most unstable mode, for various values of the physical parameters were confirmed in the papers of Greenberg and Marleta [97], P. 1842, [96], P. 380, Malik, [135], P. 32, and Hill and Straughan, [114], P. 2124.*

5.4 Problems

1. Solve by suitable spectral methods the "Rossby waves" problems, namely

$$\begin{cases} u'' + \left\{ \frac{\beta - U''}{U - c} - \alpha^2 \right\} \cdot u = 0, \ y_1 < y < y_2, \\ u(y_1) = u(y_2) = 0, \end{cases}$$

where $\alpha,\ \beta$ will be taken as constants and the complex (phase velocity) $c = c_r + ic_i$ is to be found as an eigenvalue so that a mode is unstable if and only if $\alpha c_i > 0$. a) $U(y) = \left(\frac{y}{\pi}\right)^2,\ -\pi \leq y \leq \pi,\ \alpha^2 = .5,\ \beta = 1.$; b) $U(y) = \sin y,\ -\pi \leq y \leq \pi,\ \alpha^2 = 1.,\ \beta = 1.$; c) $U(y) = \frac{y}{\pi},\ -\pi \leq y \leq \pi,\ \alpha^2 = 1.,\ \beta = 1.$; d) $U(y) = -\exp(-y),\ 0 \leq y \leq \infty,\ \alpha^2 = 1.,\ \beta = 1.$;▲

2. [179] [180] A A Shkalikov considers the following eigenvalue problem

$$\begin{cases} -i\varepsilon^2 z'' + q(x) z = \lambda z, \ |x| \leq 1, \\ z(-1) = z(1) = 0, \end{cases}$$

in order to understand the behavior of the spectrum of the standard Orr-Sommerfeld eigenvalue problem (5.1) as the Reynolds number tends to infinity. From the physical point of view it means that the viscous fluid "tends" to be ideal. Assume that $\varepsilon^2 := (\alpha R)^{-1}$, $q(x)$ is the same as in (5.1) and is an analytic monotonous function. In these conditions, Shkalikov shows that the spectrum of this simplified problem lies in the closure of the semi-strip

$$\Pi = \{\lambda |\ \operatorname{Im}\lambda < 0,\ -1 < \operatorname{Re}\lambda < 1\}.$$

Verify numerically this result for $q(x) := \sin(\pi x/2)$, $q(x) := (x+1)^2$. For $q(x) := x$, our numerical results are depicted in Fig. 5.5.▲.

3. Prove that the boundary conditions in the problem (5.4) are chosen such that the eigenvalues λ are real and positive. *Hint* This can easily be shown by setting up the energy norm, i.e., $\|u\|^2 := \int_{-1}^{1} u^2(x)\,dx$, and proving that $\lambda \|u\|^2 \leq 0$. In fact, we can start with the equality $\int_{-1}^{1} u(x) \cdot u'''(x)\,dx = \lambda \int_{-1}^{1} u^2(x)\,dx$ and then integrate by parts. ▲

4. Show that all eigenvalues of the clamped road problem are real and positive. ▲

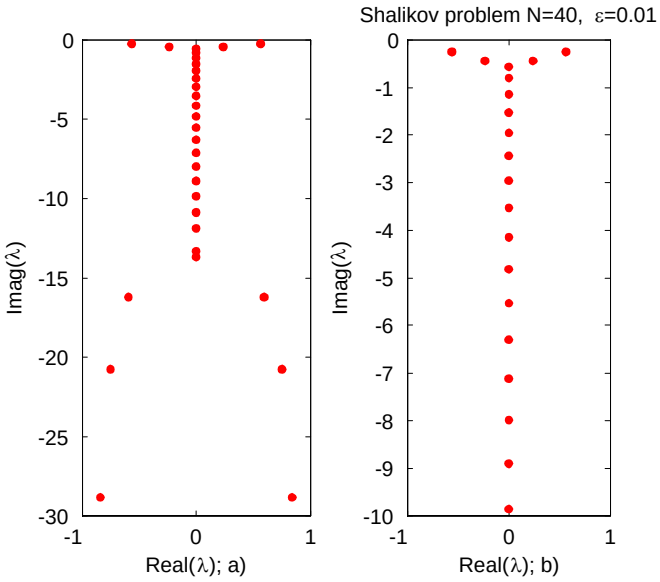

Figure 5.5: The spectrum for Shkalikov's problem. a) the first 30 largest imaginary parts; b) the first 20 largest imaginary parts;

Part II

Second Part

Chapter 6

Non-normality of spectral approximation

Eigenvalues and eigenvectors are an imperfect tool for analyzing non-normal matrices and operators, a tool that has often been abused. Physically, it is not always the eigenmodes that dominate what one has observed in a highly non-normal system. Mathematically, eigenanalysis is not always an efficient means to the end that really matters: understanding behavior.

L. N. Thefethen, Pseudospectra of matrices, 1992

The main drawback of Chebyshev spectral methods, tau, Galerkin as well as pseudospectral, consists in the fact that, due to the non-uniform weight associated with the Chebyshev polynomials, they produce *non-symmetric matrices* even for self-adjoint elliptic problems. Consequently, it is fairly important to have a measure of this anomaly. More generally, it means a lack of normality of Chebyshev approximation.

There are two major concepts with respect to the measure of the non-normality of square matrices. The first one is due to P. Henrici [108] (see also Eberlein [59], Elsner and Paardekooper [63] and Chaitin-Chatelin and Fraysse [36, p.160]) and it provides some *scalar measures of non-normality*. The second, more recently introduced, is that of *pseudospectrum* of a matrix and is systematically treated by L. N. Trefethen in a large series of papers from which we quote only [193], [194] and [195]. As it is well known, the non-normality of matrices and, consequently, of linear finite dimensional operators assumed, is responsible for a high spectral sensitivity. This sensitivity leads to a surprising and sometimes critical behavior of some numerical algorithms and procedures. We mean by such critical behavior the fact that the eigenvalues of a matrix A characterize the evolution of related quantities such as $\|\exp(tA)\|$ and $\|A^n\|$ only in the long run and not in the "transitory regime", i.e., for "small" t or n.

Unfortunately, this regime is of overwhelming importance, for example, in the behavior of dynamic systems with non-normal linear part. Such systems could undertake a catastrophic behavior or become chaotic.

In spite of this, the estimation of non-normality is not yet a routine matter among scientists and engineers who deal with such matrices. However, a realistic approach of some important numerical methods (such as spectral type methods) and problems (mainly those non-self-adjoint) imposes the quantification of the non-normality of the matrices involved. In this respect, we try to rewrite some of the results of P. Henrici [108, p.27] on scalar measures of non-normality and introduce a new one. Consequently, in the second section, after some preliminary remarks, we introduce an *euclidean relative departure from normality of an arbitrary matrix* and derive an upper universal bound for that. The bound implies two factors. The first one depends solely on the dimension of the matrix, and is of order $O\left(n^{\frac{3}{4}}\right)$, and the second depends mainly on the structure of the matrix.

Two remarks are in order at this point. First, the structure of the matrix is intimately related to the numerical method, or shortly, the structure means the numerical method. Second, the latter factor called *the non-normality ratio* is itself a measure of non-normality. For this one we provide an upper bound which is sharp and at the same time practical. It is crucial to observe that, for an arbitrary measure of non-normality, the problem of the existence of an upper bound over the set of non-normal matrices of a specific dimension remains an open one.

However, the non-normality ratio furnishes a scale on which each and every matrix can be measured and, eventually, several numerical methods applied to a specific problem can be compared to one another.

With the measure of non-normality established above, we introduce also a *relative distance* of a matrix A to the set of normal matrices and provide an upper bound for that.

In the third section, we write down the matrices corresponding to a Chebyshev-Galerkin method with different trial and test basis functions considered in some of our previous papers [163] and [164]. In the fourth section, we analyze some particular matrices which come from numerical analysis of second order and fourth order one-dimensional (1D for short) differential operator. More exactly, we take into account the above quoted (CG) method, a (CG) method introduced by J. Shen in [176] (CGS for short), and two variants of the classical Chebyshev-tau method. For typical choices of parameters in these problems the ratios of non-normality are worked out. They show the efficiency of this measure of non-normality in detecting this anomaly even when the pseudospectrum fails.

For the particular case of *complex Schroedinger operator,* which is considered as highly non-normal, we make use also of Chebyshev collocation method to discretize it and compare the non-normality of matrices involved with the non-normality of the three methods above.

6.1 A scalar measure of non-normality

In his seminal paper [108], P. Henrici introduced the following departure from normality of an arbitrary $n \times n$ matrix A with complex entries ($A \in C^{n\times n}$) :

$$\Delta_\nu(A) := \inf_{\underbrace{A = U(\Lambda + M)U^*}_{(Schur\ decomposition)}} \nu(M), \tag{6.1}$$

where U is unitary and Λ is diagonal and is made up of the eigenvalues of A. The symbol $*$ denotes as usual the conjugate transpose of a vector or a matrix, while ν stands for a norm of A (see for example [188, p.116] for the equivalence of various norms). The main result from the above quoted paper reads as follows:

$$\Delta_\varepsilon(A) \leq \left(\frac{n^3 - n}{12}\right)^{1/4} (\varepsilon(A^*A - AA^*)^{1/2}, \tag{6.2}$$

where ε stands for the Frobenius (*euclidian*) norm of A.

In order to quantify more rigorously the non-normality we introduce a new scalar measure with the next definition.

Definition 73 *For any $A \in C^{n\times n}$, $A \neq O$- the null matrix, we define an **euclidean relative departure from normality** of a matrix A with the equality*

$$\frac{dep_\varepsilon(A)}{\varepsilon(A)} := \frac{\Delta_\varepsilon(A)}{\varepsilon(A)},$$

and the map

$$H : C^{n\times n} \to \mathbf{R}_+, \qquad H(A) := \frac{\sqrt{\varepsilon(A^*A - AA^*)}}{\varepsilon(A)} \tag{6.3}$$

called the non-normality ratio.

Remark 74 *This map is itself a scalar measure of non-normality in the sense of definition from [63, p.108], i.e., $H(A) = 0$ iff A is normal $(A^*A - AA^* = 0)$.*

The main result is contained in the following theorem. It extends our previous result reported in [83].

Theorem 75 *For any $A \in \mathbb{C}^{n\times n}$ the non-normality ratio satisfies the inequality*

$$0 \leq H(A) \leq 2^{\frac{1}{4}}, \tag{6.4}$$

and the euclidean relative departure is bounded by

$$\frac{dep_\varepsilon(A)}{\varepsilon(A)} \leq \left(\frac{n^3 - n}{12}\right)^{1/4} H(A). \tag{6.5}$$

Moreover, for the eigenvalues λ_i of a non-normal matrix A we have the inequality

$$\left(1-\left(\frac{n^3-n}{12}\right)^{\frac{1}{2}}\left(H\left(A\right)\right)^2\right)\leq\frac{\sum_i\left|\lambda_i\right|^2}{\varepsilon^2\left(A\right)}\leq\left(1-\frac{1}{2}\left(H\left(A\right)\right)^4\right)^{\frac{1}{2}}. \tag{6.6}$$

Proof. The double inequality (6.4) is due to P. J. Eberlein [59, p.996]. In fact, she showed that

$$\varepsilon\left(A^*A-AA^*\right)\leq\sqrt{2}\varepsilon^2\left(A\right). \tag{6.7}$$

This inequality is sharp, i.e., the right hand side equality holds *iff* the matrix A has the special form

$$A=\alpha\left(vw^*\right) \tag{6.8}$$

where v and w are orthonormal (column) vectors and $\alpha\in\mathbb{C}$ is arbitrary and $\alpha\neq 0$. The left hand side equality holds, of course, *iff* the matrix is normal. The bound in (6.5) is a simple consequence of (6.2) and the definition of the map H. The left hand side and the right hand side inequalities in (6.6) are respectively direct consequences of inequalities from [108, p.28], see also [124, p.110] and [124, p.109]. ■

Corollary 76 *Incidentally, the inequality (6.6) shows that any matrix $A\neq O$ which satisfies (6.8) has the property that every eigenvalue equals zero. It is a very suggestive description of the most non-normal matrices.*

Remark 77 *For an arbitrary matrix, the previous bound (6.5) furnishes a fairly direct and practical estimation of its departure from normality. This bound holds together with that from [130, p.466]. Moreover, we have for the* ***relative distance*** *to the set $\mathcal{N}$ of normal matrices the estimation*

$$\frac{dist\left(A,\mathcal{N}\right)}{\varepsilon\left(A\right)}\leq\frac{dep_\varepsilon\left(A\right)}{\varepsilon\left(A\right)}\leq\left(\frac{n^3-n}{12}\right)^{1/4}H\left(A\right).$$

Remark 78 *The Bauer-Fike theorem (see for instance the monograph of P. G. Ciarlet [39] P. 59 or that of Stoer and Burlirsch [187] P. 388) bounds the pseudospectra in terms of the condition number of a matrix of eigenvectors (a scalar measure of non-normality in some sense!). In fact, in [39] the following theorem is proved:*

Theorem 79 (Bauer-Fike) *Let A be a diagonalizable matrix, P a matrix such that*

$$P^{-1}AP=diag\left(\lambda_i\right):=D,$$

and $\|\cdot\|$ a matrix norm satisfying

$$\left\|diag\left(d_i\right)\right\|=\max_i\left|d_i\right|,$$

for every diagonal matrix. Then, for every matrix δA and for each and every eigenvalue λ of the matrix $A+\delta A$ there exists at least one index i such that

$$|\lambda-\lambda_i| \leq cond\,(P)\,\|\delta A\|\,,$$

where $cond(P) = \|P\|\left\|P^{-1}\right\|$, *and the matrix norms* $\|\cdot\|_1, \|\cdot\|_2$, *and* $\|\cdot\|_\infty$ *all satisfy the hypothesis in the statement of the theorem.*

Due to the fact that **normal matrices** are diagonalizable by a unitary matrix in the above estimation we get $cond\,(P)=1$ for such matrices. However, the *condition number* introduced above is called *condition number for the matrix A relative to the eigenvalue problem.*

6.2 A C G method with different trial and test basis functions

Let us consider the second order 1D two-point boundary value problem

$$u''+\mu\cdot u'-\lambda\cdot u=f\,(x)\,,\quad u\,(-1)=u\,(1)=0, \tag{6.9}$$

and the fourth order 1D two-point boundary value problem

$$u^{(iv)}-\mu\cdot u''+\lambda\cdot u=f\,(x)\,,\quad u\,(\pm 1)=u'\,(\pm 1)=0. \tag{6.10}$$

The *classical Galerkin method* solves such problems using *the same basis in the trial (shape functions space or projection space) and test spaces.* In this respect we refer to the papers of J. Shen [175] and [176], where spectral Galerkin methods are analyzed in detail for the 1D-3D second and the fourth elliptic operators by using Legendre and respectively Chebyshev polynomials.

Anyway, it was observed that such Galerkin methods have important inconveniences. The most serious drawback seems to be the fact that due to the increased condition number of the matrices resulting in the discretization process, the computational round off errors deteriorate the expected theoretical (spectral) accuracy. Several attempts were made in order to circumvent these disadvantages. In some of our previous papers [82], [164] and [163], we considered methods involving different trial and test bases. We obtained better conditioned banded matrices and the elimination of spurious eigenvalues in non-standard Orr-Sommerfeld eigenvalue problems.

Thus, following the idea of W. Heinrichs ([105]) we search solutions to problem (6.9) in the form of the expansion

$$u^N\,(x)=\sum_{k=0}^{N} a_k\cdot w_k\,(x)\,, \tag{6.11}$$

where $w_k\,(x)=\left(1-x^2\right)T_k\,(x)$ and $T_k\,(x)$ is the k^{th} order Chebyshev polynomial. This basis satisfies apriori the above homogeneous boundary conditions. As test functions we make use of those from the paper of Shen ([176, p.3]), i.e.,

$$v_k(x) = d_k\left(T_k(x) - T_{k+2}(x)\right), \quad k = 0, 1, \ldots, N.$$

The Galerkin formulation for the problem (6.9) requires the following scalar products

$$\begin{aligned}
\left(w_p^{''}, v_k\right)_{0,\varpi} &= \frac{\pi}{2}\begin{cases} -c_p(p+1)(p+2), & p = k \\ d_k(p-1)(p-2), & p = k+2 \\ 0, & otherwise \end{cases}, \\
\left(w_p^{'}, v_k\right)_{0,\varpi} &= \frac{\pi}{4}\begin{cases} -c_{k-1}(k+1), & p = k-1 \\ (3d_k - c_k)(k+1), & p = k+1 \\ -d_k(k+1), & p = k+3 \\ 0, & otherwise \end{cases}, \\
(w_p, v_k)_{0,\varpi} &= \frac{\pi}{8}\begin{cases} -c_{k-2}, & p = k-1 \\ 2c_k + 2d_k - c_{k-1}, & p = k \\ -(c_k + 2d_k), & p = k+2 \\ d_k, & p = k+4 \\ 0, & otherwise \end{cases},
\end{aligned}$$

$p, k = 0, 1, 2, \ldots, N$, where $\varpi(x) = \left(1 - x^2\right)^{\frac{-1}{2}}$ is the Chebyshev weight, $(\cdot,\cdot)_{0,\varpi}$ stands for the scalar product $(w, v)_{0,\varpi} = \int_{-1}^{1} w \cdot v \cdot \varpi dx$ in the weighted space $L^2_\varpi(-1, 1)$ and the coefficients c_k and d_k are defined as usual by

$$c_k = \begin{cases} 0, & k < 0 \\ 2, & k = 0 \\ 1, & k > 0 \end{cases}, \qquad d_k = \begin{cases} 0, & k < 0 \\ 1, & k \geq 0 \end{cases}.$$

As far as we know, they were for the first time reported in the work of I. S. Pop [165, p.29]. Consequently, the (CG) method for (6.9) reads as follows

$$\left(w_p^{''}, v_k\right)_{0,\varpi} + \mu \cdot \left(w_p^{'}, v_k\right)_{0,\varpi} - \lambda \cdot (w_p, v_k)_{0,\varpi} = (f, v_k)_{0,\varpi}, \quad p, k = 0, 1, 2, \ldots, N. \tag{6.12}$$

For the fourth order problem (6.10) we use the following expansion of the solution

$$u^N = \sum_{k=0}^{N} a_k \cdot \psi_k,$$

where the *trial (shape) functions* $\psi_k(x) = \left(1 - x^2\right)^2 T_k(x)$ satisfy the boundary conditions in (6.10). As *test functions* we make use of the same test functions as those used by Shen [176, p.7], namely

$$\varphi_k(x) = T_k(x) - \frac{2(k+2)}{k+3} T_{k+2}(x) + \frac{k+1}{k+3} T_{k+4}(x).$$

In this situation, the (CG) method requires the next three scalar products $\left(\varphi_i, \psi_j^{(iv)}\right)_{0,\omega}$, $\left(\varphi_i, \psi_j''\right)_{0,\omega}$, and $\left(\varphi_i, \psi_j\right)_{0,\omega}$, for $i, j = 0, 1, 2, \ldots, N$. They are available in the relations (5.5), (5.6) and (5.7) respectively. They generate matrices which are analyzed with respect to their condition number in [165, p.32].

With these scalar products, the (CG) method for (6.10) reads

$$\left(\varphi_i, \psi_j^{(iv)}\right)_{0,\omega} - \mu \cdot \left(\varphi_i, \psi_j''\right)_{0,\omega} + \lambda \cdot \left(\varphi_i, \psi_j\right)_{0,\omega} = \left(f, v_j\right)_{0,\varpi}, \quad i, j = 0, 1, 2, \ldots, N. \tag{6.13}$$

This means an algebraic system for the unknowns a_k, $k = 0, 1, 2, ..., N$.

It is well known that whenever $\mu \geq 0$, $\lambda > 0$ the problem (6.10) has a unique solution in $H_{\varpi}^{s}(-1,1) \cap H_{0,\varpi}^{2}(-1,1)$ for $s \geq 2$. The spectral methods approximate this solution much more exactly than classical finite differences and finite elements methods.

6.3 Numerical experiments

If you find yourself computing eigenvalues of non-normal matrices, try perturbing the entries by a few percent and see what happens! If the effect on the eigenvalues is negligible, it is probably safe to forget about non-normality. If the effect is considerable, the time has come to be more careful.

L. N Trefethen, Pseudospectra of matrices, 1992

6.3.1 Second order problems

First, we consider the Helmholtz problem with a typical choice of λ, namely:

$$u'' - \lambda \cdot u = 0, \;\; u(-1) = u(1) = 0, \;\; \lambda = N^2.$$

We take into account the discretization matrices (the so-called stiffness matrices) provided by the use of two spectral methods, namely Chebyshev-tau and Chebyshev-Galerkin. The non-normality ratios of the stiffness matrices involved are displayed in Table 3. Thus, the first two rows in this Table contain the non-normality ratios corresponding to the standard (unmodified) Chebyshev-tau method analyzed by Gottlieb and Orszag [90, p.119] and respectively to the improved (more numerically stable) quasi-tridiagonal method from the same monograph [90, p.120].

We refer also to the well known monograph of Canuto, Hussaini, Quarteroni and Zang [33, p.129], for the mathematics of the spectral methods as well as for the technique of improving the algebraic system furnished by Chebyshev-tau method.

The third row displays the non-normality ratios of the matrices involved in the Chebyshev-Galerkin method proposed by J. Shen [176, p.4], for short CGS

method, and the fourth row displays the non-normality ratios of the matrices furnished by the left hand side of (6.12), (see also [82], [163] and [164]). The superiority of the second method is evident.

	N=8	N=64	N=128	N=256	N=512
CT-standard	1.0040	0.9851	0.9847	0.9846	0.9845
CT-improved	0.8071	0.8201	0.8202	0.8202	0.8202
CGS	0.1779	0.0618	0.0432	0.0303	0.0214
CG	0.4345	0.1641	0.1170	0.0831	0.0589
CT-Laplace	1.0584	1.0579	1.0583	1.0585	1.0586

Table 3 The non-normality ratios for stiffness matrix associated to Helmholtz problem

The second order operator

$$r\frac{d}{dr}(r\frac{du}{dr}) - \lambda \cdot u, \qquad \lambda = N^2,$$

with homogeneous boundary conditions is discretized using only the Chebyshev-tau method. It has obvious applications to the Laplace's equation on the unit disk. A detailed description of the entries of the corresponding matrices is available in [33, p.132]. The corresponding non-normality ratios are displayed in the fifth row of the first Table. They confirm the superiority of the Chebyshev-Galerkin method. This superiority comes from the fact that the boundary conditions are incorporated into the test and trial functions and do not perturb the structure of the matrix of the method. At the same time, we have to observe that, for both equations considered, Chebyshev-tau method leads to extremely non-normal matrices $\left(2^{\frac{1}{4}} = 1.1892\right)$.

The two variants of CG methods are again compared in the following two Tables. The problem (6.9) is solved with a typical choice of the parameters λ and μ, namely $\lambda = N^2$ and $\mu = \pm N$.

	N=8	N=64	N=128	N=256	N=512
CGS	0.2449	0.0803	0.0558	0.0391	0.0275
CG	0.4448	0.1562	0.1109	0.0786	0.0556

Table 4 The non-normality ratios corresponding to mu=N

	N=8	N=64	N=128	N=256	N=512
CGS	0.4011	0.1303	0.0906	0.0634	0.0446
CG	0.3638	0.1647	0.1187	0.0847	0.0602

Table 5 Non-normality ratios corresponding to mu=-N

Remark 80 *For the above mentioned choices of parameters λ and μ, the terms in Galerkin formulation (6.12) are all of them of the same order, namely $O\left(N^2\right)$. In our report [85] we have considered the situations $\mu = 0.0$, $\lambda = 0.1$, and $\mu = O(N)$, $\lambda = 0.1$ for the same values of the cut-off parameter N. The qualitative behavior of the non-normality ratios of the associated matrices does not change.*

6.3.2 Fourth order problems

In the next three tables are reported the results obtained for these problems. For (CT) method , details on entries of associated matrices could be found, for example, in [165, p.30]. The matrices associated with (CG) method are furnished by the left hand side of (6.13) and those associated with CGS method are available in [176, p.7].

	N=64	N=128	N=512
CT	1.0032	0.9995	0.9988
CG	0.4014	0.2964	0.1536
CGS	0.1665	0.1212	0.0620

Table 6 The non-normality ratios for $\mu = 0.1, \lambda = 0.1$

	N=64	N=128	N=512
CT	1.0032	0.9999	0.9988
CG	0.1728	0.1190	0.1411
CGS	3.4282e-04	4.5645e-04	8.7946e-04

Table 7 The non-normality ratios for $\mu = 0.1$, $\lambda = 256^4$;

	N=64	N=128	N=512	N=1024
CG	0.2597	0.2097	0.1450	0.1077
CGS	0.0404	0.0384	0.0534	0.0423

Table 8 The non-normality ratios for $\mu = 256^2$, $\lambda = 0.1$

As a direct relation between scalar measures of non-normality and pseudospectra is still an open problem, the pseudospectra of matrices corresponding to second column ($N = 128$) of Table 7 are depicted and compared in the next three figures.

However, it is apparent from Figures 6.1 and 6.2 that the "amplitude" of variations (contours) of pseudospectra of the associated matrices decreases at the same time with the non-normal ratio.

This "amplitude" even vanishes for quasi normal matrices, i.e., those corresponding to the non-normality ratio

$$H(CGS) = 4.5645e - 04.$$

The pseudospectrum of such matrices, depicted in Figure 6.3, reduces to a set of points very close to the real axis.

Remark 81 *In all the three figures 6.1, 6.2 and 6.3 the large dots are the eigenvalues. The matrices are non-normal, and their pseudospectra are accordingly much bigger than the $\varepsilon-$neighborhoods about their spectra.*

Remark 82 *It is quite surprising that for both Chebyshev-Galerkin methods considered, the non-normality ratio of the stiffness matrices decreases, as dimension N increases, thus improving their normality (see Tables 1-6). Small fluctuations are observed in Tables 5 and 6 for (CGS) method but they do not affect the above conclusions.*

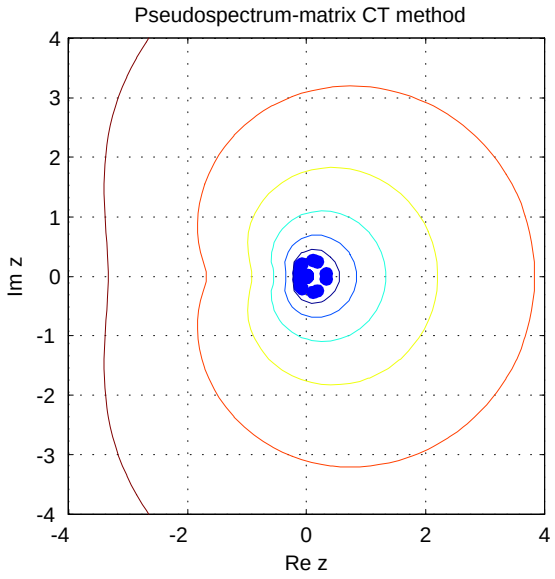

Figure 6.1: The pseudospectrum, (CT) method, $N = 128$, $\lambda = 256^4$, $\mu = 0$

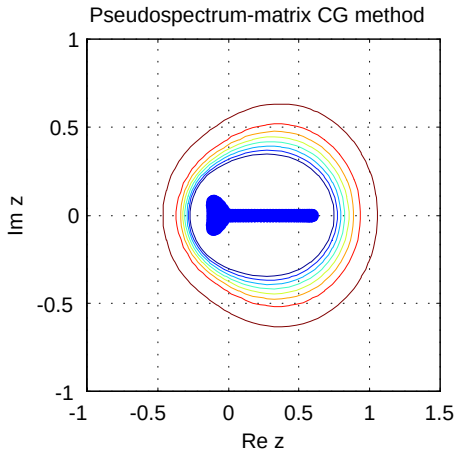

Figure 6.2: The pseudospectrum for (CG) method $N = 128$, $\lambda = 256^4$

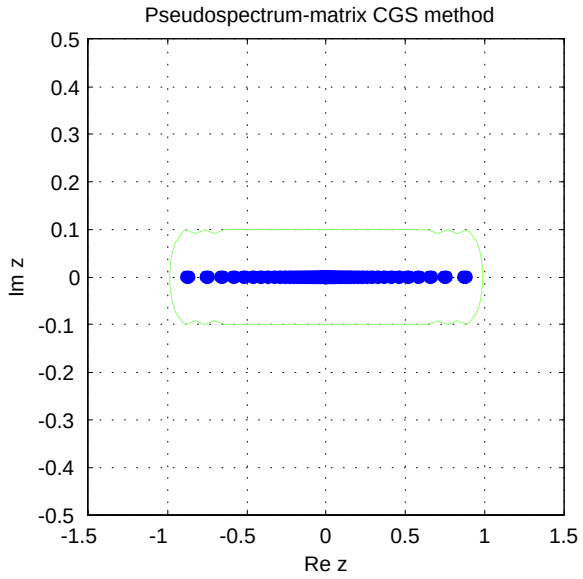

Figure 6.3: The pseudospectrum, (CGS) method, $N = 128$, $\lambda = 256^4$, $\mu = 0$

In the case of tau method the non-normality is mainly caused by the perturbations introduced with direct enforcing of the boundary conditions. For this method the non-normality of the matrices involved is high and quite constant.

However, with respect to normality, the (CG) methods are far better than (CT) methods.

We also have to remark the superiority of the (CGS) method considered by Shen as compared to our (CG) method. It is worth noting that this method produces quite normal matrices (see Table 7).

As a final remark, it is important to underline that whenever the pseudospectrum fails to observe the non-normality, as it is clear from Figure 6.3, our scalar measure remains an indicator for that.

6.3.3 Complex Schrödinger operators

In his paper [194] L. N. Trefethen considers the following Schrödinger operator

$$Au\,(x) := u'' + \left(cx^2 - dx^4\right) u, \quad c = 3 + i3, \quad d = \frac{1}{16}. \tag{6.14}$$

It is discretized by a Chebyshev collocation method on a finite interval $[-L, L]$ with boundary conditions $u\,(\pm 1) = 0$. First, the interval $[-L, L]$ is approximated by the set of (CGaussL) points

$$x_j := L \cos\left(\frac{j\pi}{N+1}\right), \quad j = 0, 1, 2, \ldots, N+1,$$

and the operator A is then approximated on this grid by an $N \times N$ matrix A_N defined by the following prescription. For any $N-$vector v, $A_N v$ is the $N-$vector obtained by two steps:

- let $p(x)$ the unique polynomial of degree $\leq N+1$ with $p(\pm L) = 0$ and $p(x_j) = v_j$ for $j = 1, 2, \ldots, N$;
- for $j = 1, 2, \ldots, N$, $\quad (A_N v)_j = p''(x_j) + \left(cx_j^2 - dx_j^4\right) p(x_j)$.

At the same time with matrix A we consider a "conditioned" (equivalent) form of that, namely $B := WAW^{-1}$, where the *weight matrix* W has the structure

$$W := diag(w_1, w_2, \ldots, w_N),$$

and the weights are

$$w_j := \frac{\pi\sqrt{L^2 - x_j^2}}{2(N+1)}.$$

The fact that the matrix B is more normal than A is apparent from the Table 9.

N	H(A_N)	H(B_N)
64	0.3745	0.1262
140	0.3946	0.1329
160	0.3949	0.1330
200	0.3949	0.1330

Table 9 The non-normality ratios

The non-normality of these matrices is compared with that of discretization matrices corresponding to (CT), (CG) and (CGS) methods in the Table 10.

	N=64	N=256	N=512	N=1024
CT	0.9912	0.9916	0.9844	0.9844
CG	0.2236	0.1161	0.0826	0.0586
CGS	0.1247	0.0637	0.0452	0.0320

Table 10 The non-normality ratios for (CT), (CG) and (CGS) methods.

Again, the (CGS) method is the best. The pseudospectrum of the (CT) discretization matrix for $N = 16$ is displayed in Fig. 6.4. Another aspect of non-normality is the fact that almost all eigenvalues gather around the origin.

Remark 83 ***The normality of collocations methods*** *In their paper [196], Trefethen and Trummer, analyzed the spectra of the Fourier, Chebyshev and Legendre pseudodifferential matrices with respect to the fact that they determine the allowable time step in an explicit time integration of parabolic and hyperbolic partial differential equations. They observe that these eigenvalues are extraordinarily sensitive to rounding errors and other perturbations. Our numerical experiments thoroughly confirm this assertion. More than that, we observe that the Fourier pseudodifferential matrix is a* ***circulant matrix*** *and consequently is a normal one. Our numerical experiments confirmed that their Henrici number*

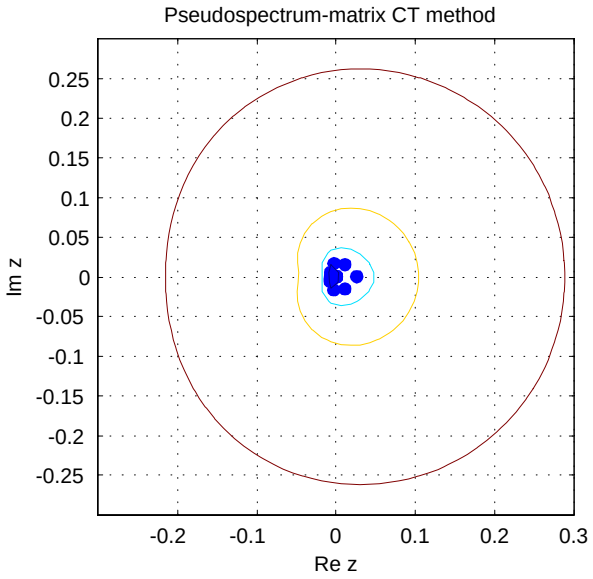

Figure 6.4: The pseudospectrum for the (CT) method

equals zero. The eigenvalues of such a matrix (denoted $D_N^{(1),F}$*) are all complex. The pseudospectrum and the log of the norm of its resolvent are displayed in Fig. 6.5. On the other hand, the Henrici number of Chebyshev pseudodifferential matrices is much larger, i.e.,* $HA(D_N^{(1),C}) = 0.8928$ *and is quite independent of* N*. The pseudospectrum and the log of the norm of its resolvent for* $N = 16$ *are displayed in Fig. 6.6. Unfortunately, these results do not confirm those of Trefethen and Trummer, [196], p.1011, but the norm of the resolvent underlines the sensitivity of eigenvalues to the rounding off errors. For Hermite pseudodifferential matrices we obtained Henrici numbers of order* 0.2589.

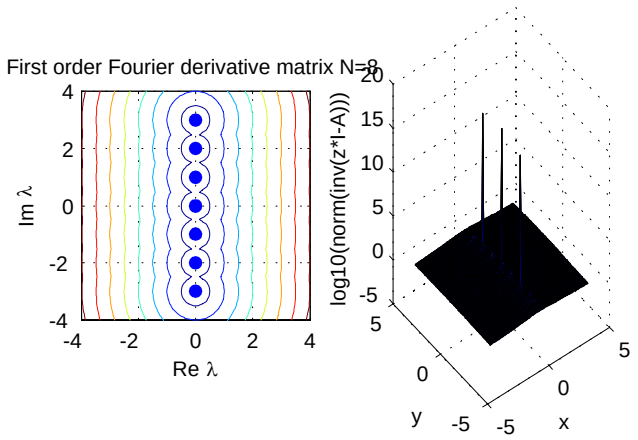

Figure 6.5: The pseudospectrum and the norm of the resolvent for $D_8^{(1),F}$

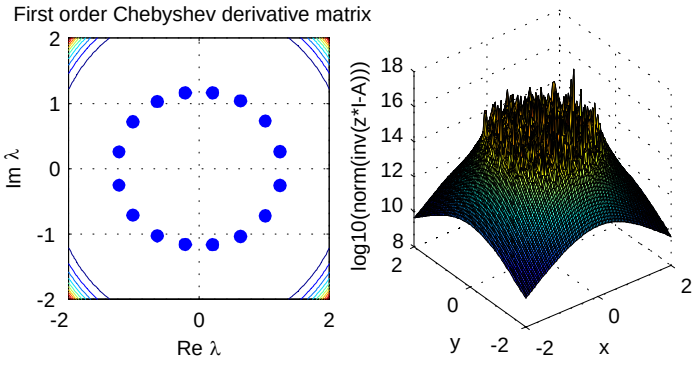

Figure 6.6: The pseudospectrum and the norm of the resolvent for $D_{16}^{(1),C}$. The large dots are the eigenvalues.

Chapter 7

Concluding remarks

"Spectral methods are like Swiss watch. They work beautifully, but a little dust in the gear stops them entirely."
Philip L. Roe, quoted by J P Boyd, SIAM Rev., 46(2004)

Two main conclusions were drawn after all the numerical experiments were carried out.

The first one refers to the fact that the Chebyshev collocation method (pseudospectral), in spite of its non-normality, remains the most feasible and the most implementable spectral method. It succeeded in the linear and, most important, nonlinear problems of elliptic, parabolic and hyperbolic types.

The second one underlines the fact that the symplectic methods, implicit as well as explicit, in conjunction with Chebyshev collocation methods, perform much better than the conventional methods in long time integration of Hamiltonian partial differential equations. They reproduce accurately the structure of the phase space and produce Hamiltonians with narrow oscillations. This way, they do not alter the nature of Hamiltonian systems.

Chapter 8

Appendix

8.1 Lagrangian and Hermite interpolation

Let $f : [a,b] \to \mathbb{R}$ be a sufficiently smooth function and $a \leq x_0 \leq x_1 \leq x_2 \leq ... \leq x_N \leq b$ a partition of the interval $[a,b]$.

The well known *Lagrangian interpolation formula* is

$$f(x) = \sum_{k=0}^{N} l_k(x) f(x_k) + \frac{1}{(N+1)!} \Pi(x) f^{(N+1)}(\xi), \tag{8.1}$$

where $l_k(x)$ are *Lagrangian basis polynomials,* called also *cardinal functions,* defined by

$$l_k(x) = \Pi_{\substack{j=0 \\ k \neq j}}^{N} \left(\frac{x - x_j}{x_k - x_j} \right), \quad k = 0, 1, 2, ..., N.$$

We observe that the set of interpolating polynomials $\{l_k(x)\}_{k=0}^{N}$ satisfies

$$l_k(x_j) = \delta_{kj}.$$

The *Hermite interpolation formula* matches not only $f(x)$ but also $f'(x)$ at the same points. It reads

$$f(x) = \sum_{k=0}^{N} h_k(x) f(x_k) + \sum_{k=0}^{N} h_k^*(x) f'(x_k) + \frac{1}{(2N+2)!} f^{(2N+2)}(\xi) \Pi^2(x), \tag{8.2}$$

where

$$h_k(x) = \{1 - 2l_k'(x_k)(x - x_k)\} l_k^2(x), \; h_k^*(x) = (x - x_k) l_k^2(x),$$

and in both formulas, Lagrangian and Hermite, ξ is in the range bounded by extreme values of x and x_k. They produce, respectively, Lagrangian quadrature formulas and Gauss quadrature formulas, whose errors are well established in classical monographs of numerical analysis (see for instance [43]).

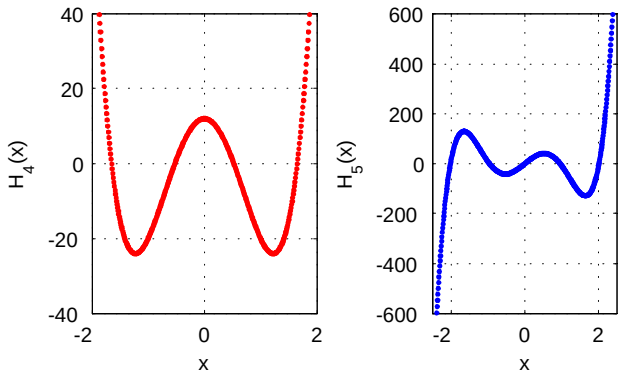

Figure 8.1: The 4th and the 5th order Hermite polynomials

However, we employed a more direct approach of Hermite polynomials. The Hermite polynomials $H_n(x)$, its first and second derivatives at $x \in \mathbb{R}$ satisfy the following recurrence relations (see for a detailed introduction D. Funaro [73])

$$\begin{cases} H_0(x) = 1, \\ H_1(x) = 2x, \\ H_n(x) = 2xH_{n-1}(x) + 2(n-1)H_{n-2}(x), \ n \geq 2, \end{cases}$$

$$H_n'(x) = 2nH_{n-1}(x), \quad n \geq 1, \quad \textit{and} \quad H_n''(x) = 4n(n-1)H_{n-2}(x), \ n \geq 2.$$

The roots of the Hermite polynomial of degree N, indexed in ascending order, $x_1, ..., x_N$ satisfy the asymptotic estimation $-x_1 = x_N = O\left(\sqrt{N}\right)$ as $N \to \infty$. This is visible in Fig. (8.1). Each and every *spectral collocation method* for solving differential equations is based on *weighted interpolants* of the form

$$f(x) \cong p_{N-1}(x) = \sum_{j=1}^{N} \frac{\alpha(x)}{\alpha(x_j)} \phi_j(x) f_j, \tag{8.3}$$

where $\{x_j\}_{j=1,2,...,N}$ is a set of *distinct interpolation nodes,* $\alpha(x)$ is a *weight function,* $f_i := f(x_i)$ and the set of *interpolating functions* $\{\phi_j(x)\}_{j=1,2,...,N}$ satisfies

$$\phi_j(x_k) = \delta_{jk}.$$

It is clear that this formula includes, as some particular cases, the formulas (8.1) and (8.2). A list of commonly used nodes, weights, and interpolating functions are tabulated in the paper of Weideman and Reddy [204]. These include the Chebyshev, Hermite, and Laguerre expansions as well as two well known nonpolynomial cases, namely trigonometric (Fourier) and *sinc* interpolants.

Associated with an interpolant such as (8.3) is the concept of a *collocation derivative operator or pseudospectral derivative.* This operator is generated by

taking l derivatives of (8.3) and evaluating the result at the nodes $\{x_j\}_{j=1,2,...,N}$:

$$f^{(l)}(x_k) \cong \sum_{j=1}^{N} \frac{d^l}{dx^l} \left[\frac{\alpha(x)}{\alpha(x_j)} \phi_j(x) \right]_{x=x_k} f_j, \ k = 1, 2, ..., N.$$

The derivative operator may be represented by a matrix $D^{(l)}$ (to be more explicit, sometimes we add an upper index for the type of expansion, i.e., C, H, L, F, s and a lower index for the order N of the approximation), the differentiation matrix, with entries

$$D_{k,j}^{(l)} := \frac{d^l}{dx^l} \left[\frac{\alpha(x)}{\alpha(x_j)} \phi_j(x) \right]_{x=x_k}. \tag{8.4}$$

The numerical differential process may be carried out as the matrix-vector product

$$\mathbf{f}^{(l)} = D^{(l)} \cdot \mathbf{f},$$

where $\mathbf{f}$ (respective $\mathbf{f}^{(l)}$) is the vector of function values (respective approximate derivative values) at the nodes $\{x_j\}_{j=1,2,...,N}$.

With respect to Hermite case, we notice that

$$p_{N-1}(x) = \sum_{j=1}^{N} \frac{e^{-x^2/2}}{e^{-x_j^2/2}} \cdot \frac{H_N(x)}{H'_N(x_j)(x - x_j)} \cdot f_j,$$

which means that the *weight function* $\alpha(x)$ is

$$\alpha(x) := e^{-x^2/2},$$

i.e., a *Gaussian type function.*

For a differentiation formula $D_N^{(l),H}$ we refer to Funaro [73] and for its implementation in MATLAB environment to Weideman and Reddy [204]. We also remark that

$$D_N^{(l),H} \neq \left(D_N^{(1),H} \right)^l,$$

and that the accuracy (error analysis) of Hermite interpolation formulas in general, is not analyzed.

Remark 84 *When a physical problem is posed on an infinite domain, i.e., the real line, a variety of spectral methods have been developed using the Hermite polynomials as a natural choice of basis functions because of their close connection to the physics. With respect to the implementation of Hermite collocation method we observe that the real line* $(-\infty, \infty)$ *can be mapped to itself by change of variables*

$$x := b \cdot \widetilde{x},$$

where b *is any positive real number called scaling factor. Consequently, due to the chain rule, the first-derivative matrix corresponding to* $b = 1$ *should be*

multiplied by b, the second-derivative matrix by b^2, etc. At the same time, the nodes are rescaled to x_k/b. It means that the Hermite differentiation process is exact for functions of the form

$$e^{-b^2x^2/2}p(x),$$

where $p(x)$ is any polynomial of degree $N-1$ or less. The freedom offered by the parameter b can be exploited to optimize the Hermite differentiation process; see Tang [192] and also P. Anhaouy [162], P. 87.

Remark 85 *F. Stenger in [185] and [186] provides a fairly complete summary of numerical methods based on sinc functions or Whittaker Cardinal functions. Sinc approximation excels for problems whose solutions have singularities, or infinite domains, or boundary layers. The sinc function is defined by*

$$\sin c(x) := \frac{\sin(\pi x)}{\pi x},$$

and a sinc expansion of f reads

$$C(f,h)(x) = \sum_{k\in\mathbb{Z}} f(kh)\sin c\left(\frac{x}{h}-k\right),\ x\in\mathbb{R}.$$

The function $C(f,h)(x)$ provides an incredible accurate approximation on $\mathbb{R}$ to functions that are analytic and uniformly bounded, i.e.,

$$\sup_{x\in\mathbb{R}}|f(x)-C(f,h)(x)| = O\left(e^{-\pi/h}\right),\ h\to 0.$$

8.2 Sobolev spaces

8.2.1 The Spaces $\mathrm{C}^m\left(\overline{\Omega}\right),\ m\geq 0$

Let $\Omega := (a,b)^d,\ d=1,2,3$ and let's denote by $\overline{\Omega}$ the closure of Ω ,i.e., the closed poly-interval $[a,b]^d$. For every multi-index $\alpha=(\alpha_1,...,\alpha_d)$ of non-negative integers, set $|\alpha| := \alpha_1+...+\alpha_d$ and $D^a v = \partial^{|\alpha|}v/\partial x_1^{\alpha_1}...\partial x_d^{\alpha_d}$.

We denote by $\mathrm{C}^m\left(\overline{\Omega}\right)$ the vector space of the functions $v:\overline{\Omega}\to\mathbb{R}$ such that for each multi-index α with $0\leq|\alpha|\leq m$, $D^a v$ exists and is continuous in $\overline{\Omega}$. Since a continuous function on a closed bounded poly-interval is bounded there, one can set

$$\|u\| := \sup_{0\leq|\alpha|\leq m}\sup_{x\in\overline{\Omega}}|D^a v(x)|.$$

It means a norm for which $\mathrm{C}^m\left(\overline{\Omega}\right)$ is a Banach space.

A function belongs to $\mathrm{C}^\infty\left(\overline{\Omega}\right)$ *iff* it belongs to $\mathrm{C}^m\left(\overline{\Omega}\right)$ for all $m>0$.

8.2.2 The Lebesgue Integral and Spaces $\mathbf{L}^p(a,b)$, $1 \leq p \leq \infty$.

A complete introduction to the Lebesgue integral can be found in many text books (see, for instance, Adams, R.A.,[3]). We assume known the notions of Lebesgue measure, measurable sets, measurable functions, and Lebesgue integral. Since two integrable functions which differ on a set of zero measure have the same integral, they can be identified from the point of view of the Lebesgue integration theory, i.e., they belong to the same equivalence class. This identification is always presumed here an in the sequel.

Let (a,b) be a bounded interval of $\mathbb{R}$ and let $1 \leq p < +\infty$. We denote by $L^p(a,b)$ the space

$$L^p(a,b) := \{u : (a,b) \to \mathbb{R}; u \text{ measurable}, \int_a^b |u(x)|^p \, dx < +\infty\}$$

Endowed with the norm

$$\|u\|_{L^p(a,b)} := \left(\int_a^b |u(x)|^p \, dx \right)^{1/p},$$

it is a Banach space.

If $p = \infty$, $L^\infty(a,b)$ is the space of measurable functions $u : (a,b) \to \mathbb{R}$ such that $|u(x)|$ is bounded outside a set of measure zero. If M denotes the smallest real number such that $|u(x)| \leq M$ outside the set of measure zero we define the norm of this space by

$$\|u\|_{L^\infty(a,b)} := ess \sup_{x \in (a,b)} |u(x)| = M.$$

$L^\infty(a,b)$ is again a Banach space.

The index $p = 2$ is of special interest because $L^2(a,b)$ is not only a Banach space but a Hilbert space with the scalar product

$$(u,v)_{L^2(a,b)} := \int_a^b u(x) \, v(x) \, dx.$$

The previous definitions can be extended in a straightforward way to more than one space dimension and we get the spaces $L^p(\Omega)$ $1 \leq p \leq \infty$.

8.2.3 Infinite Differentiable Functions and Distributions

We denote by $\mathcal{D}(\Omega)$, Ω bounded in $\mathbb{R}^d$, $d = 1, 2, 3$, the vector space of all infinitely differentiable functions $\phi : \Omega \to \mathbb{R}$, for which there exists a closed set $K \subset \Omega$ such that $\phi \equiv 0$ outside K. In other words, the function ϕ has **compact support** in Ω. We say that a sequence of functions $\phi_n \in \mathcal{D}(\Omega)$ **converges** in $\mathcal{D}(\Omega)$ to a function $\phi \in \mathcal{D}(\Omega)$ as $n \to \infty$ if there exists a common set $K \subset \Omega$ such that all the ϕ_n vanish outside K and $D^\alpha \phi_n \to D^a \phi$ uniformly on K as $n \to \infty$, for all non-negative multi-indices α.

a) Distributions

Let $T : \mathcal{D}(\Omega) \to \mathbb{R}$ be a linear mapping (form). We shall denote the value of T on the element $\phi \in \mathcal{D}(\Omega)$ by $\langle T, \phi \rangle$. T is said to be **continuous** if, for each sequence $\phi_n \in \mathcal{D}(\Omega)$ which converges in $\mathcal{D}(\Omega)$ to a function $\phi \in \mathcal{D}(\Omega)$ as $n \to \infty$, one has

$$\langle T, \phi_n \rangle \to \langle T, \phi \rangle .$$

A **distribution** is **a linear continuous form on** $\mathcal{D}(\Omega)$. The set of all the distributions on Ω is a vector space denoted by $\mathcal{D}'(\Omega)$.

Example 86 (i) *Each integrable function $f \in L^p(a,b)$ can be identified with a distribution T_f defined by*

$$\langle T_f, \phi \rangle := \int_a^b f(x)\,\phi(x)\,dx, \ \forall \phi \in \mathcal{D}(\Omega).$$

(ii) Let $x_0 \in (a,b)$. The linear form on $\mathcal{D}(\Omega)$

$$\langle \delta_{x_0}, \phi \rangle := \phi(x_0), \ for\ all\ \phi \in \mathcal{D}(\Omega),$$

is a distribution, which is commonly (but improperly!) called the "Dirac function".

We notice that if T_1 and T_2 are two distributions, then they are called "equal in the sense of distributions" if

$$\langle T_1, \phi \rangle = \langle T_2, \phi \rangle, \ for\ all\ \phi \in \mathcal{D}(\Omega).$$

b) The Derivative of Distributions

Let α be a non-negative multi-index and $T \in \mathcal{D}'(\Omega)$ an arbitrary distribution. The distribution $D^\alpha T$ defined by

$$\langle D^\alpha T, \phi \rangle := (-1)^{|\alpha|} \langle T, D^\alpha \phi \rangle, \ for\ all\ \phi \in \mathcal{D}(\Omega),$$

is called **the $\alpha-$ distributional derivative of** T.

Remark 87 *It is extremely important to observe that each function $u(x)$ from $L^p(\Omega)$, $1 \le p < \infty$* ***is infinitely differentiable in the sense of distributions,*** *and the following Green's (or integration by parts) formula holds*

$$\langle D^\alpha u, \phi \rangle := (-1)^{|\alpha|} \int_\Omega u(x)\, D^\alpha \phi(x)\, dx, \ \forall\ \phi \in \mathcal{D}(\Omega).$$

At the same time, if a function is continuously differentiable (in the classical sense), it is of course differentiable in the sense of distributions. The converse statement, generally speaking, is not true!

In general, *the distributional derivative of an integrable function can be an integrable function or merely a distribution.* We say that the α -distributional derivative of an integrable function $u \in L^1(\Omega)$ is an integrable function if there exists $g \in L^1(\Omega)$ such that

$$\langle D^\alpha u, \phi \rangle = \int_\Omega g(x)\,\phi(x)\,dx, \ \textit{for all } \phi \in \mathcal{D}(\Omega).$$

Example 88 *Consider the function*

$$u(x) = \frac{1}{2}|x|, \ x \in (-1,1).$$

Note that u is not classically differentiable at the origin. The first derivative in the distributional sense is represented by the step function

$$v(x) = \begin{cases} 1/2, & x > 0, \\ -1/2, & x < 0, \end{cases}$$

which is an integrable function. Please verify these statements! Hint:

$$\begin{aligned} \langle Du, \phi \rangle &= (-1)\int_{-1}^{1} u\phi' dx = \\ &= -\frac{1}{2}\left\{-x\phi|_{-1}^{0} + \int_{-1}^{0}\phi dx + x\phi|_{0}^{1} - \int_{0}^{1}\phi dx\right\} = \\ &= \frac{1}{2}\left\{\int_{-1}^{0}(-\phi)\,dx + \int_{0}^{1}\phi dx\right\} \\ &= \int_{-1}^{1} v(x)\,\phi(x)\,dx. \end{aligned}$$

Example 89 *Consider the function $v(x)$ now defined. Note that classical derivative is zero at all points $x \neq 0$. The first derivative of v in the sense of distributions is the Dirac distribution δ_0 at the origin. This distribution can not be represented as a function. More exactly, we have*

$$\begin{aligned} \langle Dv, \phi \rangle &= (-1)\frac{1}{2}\left\{-\int_{-1}^{0}\phi' dx + \int_{0}^{1}\phi' dx\right\} = \\ &= \phi(0) = \langle \delta_0, \phi \rangle. \end{aligned}$$

Functions having a certain number of distributional derivatives which can be represented by integrable functions play a fundamental role in the modern theory of partial differential equations. The spaces of these functions are named **Sobolev spaces.**

8.2.4 Sobolev Spaces and Sobolev Norms

We introduce hereafter some relevant Hilbert spaces, which occur in the numerical analysis of boundary value problems (see also [38] and [33]). They are spaces of square integrable functions, which possess a certain number of derivatives in the sense of distributions, representable as square integrable functions.

a) The Spaces $\mathbf{H}^m(a,b)$ and $\mathbf{H}^m(\Omega)$, $m \geq 0$

Let (a,b) be a bounded interval of the real line, and let $m \geq 0$ be an integer. We define $H^m(a,b)$ to be the vector space of the following functions

$$H^m(a,b) := \left\{ v \in L^2(a,b);\ for\ 0 \leq k \leq m, \atop \frac{d^k v}{dx^k} \in L^2(a,b) \right\}.$$

$H^m(a,b)$, endowed with the scalar product

$$(u,v)_{H^m(a,b)} := \sum_{k=0}^{m} \int_a^b \frac{d^k v}{dx^k}(x) \frac{d^k u}{dx^k}(x)\, dx,$$

becomes a Hilbert space. The associated norm is

$$\|v\|_{H^m(a,b)} = \left(\sum_{k=0}^{m} \int_a^b \left\| \frac{d^k v}{dx^k} \right\|^2_{L^2(a,b)} \right)^{\frac{1}{2}}.$$

The Sobolev spaces $H^m(a,b)$ form a hierarchy of Hilbert spaces in the sense that

$$...H^{m+1}(a,b) \subset H^m(a,b) \subset ...H^0(a,b) = L^2(a,b),$$

each inclusion being continuous. Clearly, if a function u has m classical continuous derivatives in $[a,b]$, then u belongs to $H^m(a,b)$; in other words, $C^m[a,b] \subset H^m(a,b)$ with continuous inclusion. Conversely, if u belongs to $H^m(a,b)$ for $m \geq 1$, then u has $m-1$ classical continuous derivatives in $[a,b]$, i.e., $H^m(a,b) \subset C^{m-1}[a,b]$ with continuous inclusion. This is an example of so-called "Sobolev Imbedding theorems".

As a matter of fact, $H^m(a,b)$ can be equivalently defined as

$$H^m(a,b) := \left\{ v \in C^{m-1}[a,b];\ \frac{d}{dx} v^{(m-1)} \in L^2[a,b] \right\},$$

where the last derivative is considered in the sense of distributions.

Functions in $H^m(a,b)$ can be approximated arbitrarily well by infinitely differentiable functions in $[a,b]$, in the distance induced by the norm of $H^m(a,b)$. In other words,

$$C^\infty[a,b] \text{ is dense in } H^m(a,b).$$

Set now $\Omega := (a,b)^d$, $d = 2,3$, which means $\underbrace{(a,b) \times (a,b)}_{d\ times}$. Given a multi-index $\alpha := (\alpha_1, ..., \alpha_d)$ of non-negative integers, we set $|\alpha| := \alpha_1 + ... + \alpha_d$ and

$$D^\alpha v := \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} ... \partial x_d^{\alpha_d}}.$$

The previous definition of Sobolev spaces can be extended to higher space dimensions as follows. We define

$$\begin{aligned} H^m(\Omega) \quad : \quad &= \{v \in L^2(\Omega);\ \textit{for each non-negative} \\ &\quad \textit{multi-index } \alpha\ |\alpha| \le m, \\ &\quad \textit{distributional derivative } D^\alpha v \in L^2(\Omega)\}, \end{aligned}$$

the **scalar product**

$$(u,v)_m := \sum_{|\alpha|\le m} \int_\Omega D^\alpha u(x)\, D^\alpha v(x)\, dx,$$

and the induced **norm**

$$\|v\|_{H^m(\Omega)} := \left(\sum_{|\alpha|\le m} \|D^\alpha v(x)\|^2_{L^2(\Omega)} \right)^{1/2}.$$

In this case, we have a weaker Sobolev inclusion, namely

$$H^m(\Omega) \subset C^{m-2}\left(\overline{\Omega}\right),\ m \ge 2.$$

On the other hand, as in the 1D case,

$$C^\infty\left(\overline{\Omega}\right) \text{ is dense in } H^m(\Omega).$$

b) The Spaces $\mathbf{H}^1_0(a,b)$ and $\mathbf{H}^1_0(\Omega)$

Dirichlet conditions are among the simplest and most common boundary conditions to be associated with a differential operator. Therefore, the subspaces of Sobolev spaces H^m spanned by the functions satisfying homogeneous Dirichlet boundary conditions play a fundamental role.

Since the functions of $H^1(a,b)$ are continuous up to the boundary, by the Sobolev Imbedding Theorems, it is meaningful to introduce the following subspace of $H^1(a,b)$, namely

$$H^1_0(a,b) := \{v \in H^1(a,b);\ v(a) = v(b) = 0\}.$$

This is a Hilbert space with respect to the same scalar product of $H^1(a,b)$. It is often preferable to endow $H^1_0(a,b)$ with a different, although equivalent, scalar product. This is defined as

$$[u,v] := \int_a^b \frac{du}{dx}\frac{dv}{dx} dx.$$

By *Poincaré inequality*, (1.57) this is indeed a scalar product on $H^1_0(a,b)$ and the associated norm reads

$$\|v\|_{H^1_0(a,b)} = \left(\int_a^b \left|\frac{du}{dx}\right|^2 dx \right)^{1/2}.$$

This norm is equivalent to the $H^1(a,b)$ −norm, in the sense that there exists a constant $C > 0$ such that

$$C\|v\|_{H^1(a,b)} \leq \|v\|_{H_0^1(a,b)} \leq \|v\|_{H^1(a,b)}.$$

Again, this follows from the Poincaré inequality.

The functions of $H_0^1(a,b)$ can be approximated arbitrarily well in this norm not only by infinitely differentiable functions on $[a,b]$ but also by infinitely differentiable functions which vanish identically in a neighborhood of $x = a$ and $x = b$. In other words,

$$\mathcal{D}(a,b) \text{ is dense in } H_0^1(a,b).$$

We now turn to more space dimensions. If again $\Omega := (a,b)^d$, $d = 2,3$, the functions of $H^1(\Omega)$ need not be continuous on the closure of Ω. Thus, their pointwise values on the boundary $\partial\Omega$ of Ω need not be defined. However, it is possible to extend the trace operator $v \to v|_{\partial\Omega}$ (classically defined for functions from $C\left(\overline{\Omega}\right)$ so as to be a linear continuous mapping between $H^1(\Omega)$ and $L^2(\partial\Omega)$ (see [3], for the rigorous definition of the trace of functions from $H^1(\Omega)$). With that in mind, it is meaningful to define $H_0^1(\Omega)$ as the space

$$H_0^1(\Omega) := \left\{v \in H^1(\Omega);\ v|_{\partial\Omega} = 0\right\}.$$

This is a Hilbert space for the scalar product of $H^1(\Omega)$, or for the scalar product

$$[u,v] := \int_\Omega \nabla u \cdot \nabla v dx.$$

The associated norm is defined by

$$\|u\|_{H_0^1(\Omega)} = \left(\int_\Omega |\nabla u|^2\right)^{1/2}.$$

and is equivalent to the $H^1(\Omega)$ −norm, by the Poincaré inequality.

Concerning the approximation of functions of $H_0^1(\Omega)$ by infinitely differentiable functions, the following result holds

$$\mathcal{D}(\Omega) \text{ is dense in } H_0^1(\Omega).$$

8.2.5 The Weighted Spaces

Let $\omega(x)$ be a *weight function* on the interval $[-1,1]$, i.e., a continuous, strictly positive and integrable function on $(-1,1)$. For $1 \leq p < \infty$ we denote by $L_\omega^p(-1,1)$ the Banach space of measurable functions $u : (a,b) \to \mathbb{R}$ such that $\int_{-1}^1 |u(x)|^p \omega(x)\, dx < \infty$. Consequently, this space can be endowed with the norm

$$\|u\|_{L_\omega^p(-1,1)} := \left(\int_{-1}^1 |u(x)|^p \omega(x)\, dx\right)^{1/p}.$$

For $p=\infty$ we set $L_{\omega}^{\infty}(-1,1)=L^{\infty}(-1,1)$ and for $p=2$ the space becomes a Hilbert one with the scalar product

$$(u,v)_{L_{\omega}^{2}(-1,1)}:=\int_{-1}^{1}u(x)\,v(x)\,\omega(x)\,dx.$$

In the definition of a Sobolev space, one can require that the function as well as its distributional derivative be square integrable with respect to a weight function. This is the most natural background in dealing with spectral methods. Thus, in a perfect analogous way with the above analysis we can construct the spaces $L_{\omega}^{2}(\Omega)$, $H_{\omega}^{m}(-1,1)$, $H_{\omega}^{m}(\Omega)$, $H_{\omega,0}^{1}(-1,1)$ and $H_{\omega,0}^{1}(\Omega)$.

Remark 90 *In order to simplify the writing, corresponding to the most important case $p=2$, we adopt the following short notations for the norms and scalar products in the respective spaces*

$$\begin{gathered}\|\cdot\|,\ (\cdot,\cdot),\ \|\cdot\|_{\omega},\ (\cdot,\cdot)_{\omega},\ L^{2},\ L_{\omega}^{2};\\ \|\cdot\|_{m},\ (\cdot,\cdot)_{m},\ \|\cdot\|_{m,\omega},\ (\cdot,\cdot)_{m,\omega},\ H^{m},\ H_{\omega}^{m};\\ \|\cdot\|_{1,0},\ (\cdot,\cdot)_{1,0},\ \|\cdot\|_{1,\omega,0},,\ (\cdot,\cdot)_{1,\omega,0},\ H_{0}^{1},\ H_{\omega,0}^{1}.\end{gathered}$$

The domains on which these norms and scalar products are considered result from context.

8.3 MATLAB codes

In the Appendix B of their monograph [33], Canuto, Hussaini, Quarteroni and Zang present a listing of self-contained FORTRAN routines for computing Fourier and Chebyshev collocation derivatives. They are geared towards applications in multidimensional problems. D. Funaro in [73], using the same environment, offers a series of basic algorithms in order to allow a smooth start in the development of more extensive codes.

L. N. Trefethen introduces in his monograph [197] some fairly useful MATLAB codes which implement spectral methods. Weideman and Reddy present in their paper [204] a software suite which consists of 17 MATLAB functions for solving differential problems by the spectral collocation methods. These functions enable the user to generate spectral differential matrices based on Chebyshev, Fourier, Hermite, Laguerre and sinc interpolants. We used these functions in the implementation of our collocation methods.

However, in this Appendix we introduce also some MATLAB codes which implement Chebyshev tau and Chebyshev Galerkin methods. Here are the codes.

1) The MATLAB code `Chebyshev_tau.m`

```
% Chebyshev-tau method u''+u=f, u(-1)=u(1)=0, f(x)=x^2+x;
clear all, close all
N=129; % The order of approximation
```

```
k=-1.; % the parameter lambda
A=zeros(N+1,N+1); % the entries of matrix A
A(N+1,:)=1;
i=1:(N+1);A(N,:)=(-1).^(i+1);A(1,1)=-k;
for p=3:2:N+1
A(1,p)=(1/2)*(p-1)^3;
end
for i=2:N-1
A(i,i)=-k;
for p=(i+2):2:N+1
A(i,p)=(p-1)*(((p-1)^2)-((i-1)^2));
end
end\vspace{0.3cm}
% The matrix CI transforms from Chebyshev space to physical space
\vspace{0.3cm}
for j=1:N+1
for k=1:N+1
CI(j,k)=cos(pi*(j-1)*(k-1)/N);
end
end
F=zeros(N+1,1); F(1)=F(1)+1./2.; F(2)=F(2)+1.; F(3)=F(3)+1./2.;
X=A\F; % Solve the algebraic system AX=F
```

% Transform the Chebyshev tau solution from SPECTRAL SPACE to PHYSICAL SPACE

```
U=CI*X;
m=1:N+1;XG=cos(pi*(m-1)./N); %(CGaussL) nodes (\QTSN{ref}{CGL})
% THE closed SOLUTION
Sol=XG.^2+XG-2+(1/cos(1.)).*cos(XG)-(1/sin(1.)).*sin(XG);
Diff=U-Sol';
Maxerror=max(abs(Diff))
fsize=10;
plot(XG,U,XG,Sol), xlabel('x','Fontsize',fsize)
ylabel('u(x)','Fontsize',fsize)
title('u"+u=x^2+x, u(-1)=u(1)=0, N=128','Fontsize',fsize)
legend('Chebyshev-tau Sol','Exact Sol')
% Obtain Fig. 2.1
```

2) The MATLAB code `LargeScale.m`

```
% Solve the t p b v p u''+(k^2+3)u=5sin(kx), u(-1)=-sink; u(+1)=sink
% The second order operator u'' is discretized by Chebyshev-collocation
% L. Greengard, V. Rokhlin, On the Numerical Solution of Two-Point B V P,
```

```
% Communications on Pure and Applied Mathematics, Vol XLIV, 419-452(1991)
% The differentiation matrices from Weideman and Reddy
clear all, close all
pi=4.*atan(1.);
N = input(' Dimension of the differentiation matrix: N = ? ');
g=[1 0 0;1 0 0]; % B. C. for CC method
[x,D2t]=cheb2bc(N,g); % Differentiation matrices
kapa=200;
A=D2t+(kapa^2+5)*eye(size(D2t));
F=5*sin(x.*kapa)-x.*(kapa^2+5)*sin(kapa);
v=A\F; % Solve the linear algebraic system
x=flipud(x);v=flipud(v);
x=[-1;x;1];v=[0;v;0];
CCsol=v+x.*sin(kapa);
Exsol=sin(x.*kapa);
MaxDifference=max(abs(CCsol-Exsol)) % The accuracy of computation
plot(x,CCsol,x,Exsol), legend('CCsol','Exsol')% Plot and compare solutions
axis([-1 1 -1 1])
xlabel('x','FontSize',10)
ylabel('CCsol / EXsol','FontSize',10)
% Obtain the Figure 2.3
```

3) The MATLAB code `SingPerturb.m`

```
% Solve the t p b v p eps*u''-u'=.5, u(-1)=0; u(+1)=0
% The second order operator u'' is discretized by Chebyshev-collocation method
% L. Greengard, V. Rokhlin, On the Numerical Solution of Two-Point B V P,
% Communications on Pure and Applied Mathematics, Vol XLIV, 419-452(1991)
% The differentiation matrices from Weideman and Reddy
clear all, close all
N = input(' Dimension of the differentiation matrix: N = ? ');
g=[1 0 0;1 0 0]; % B. C. for CC method
% Differentiation matrices (first and second orders) with enforced
% boundary conditions
[x,D2t,D1t]=cheb2bc(N,g);
x=flipud(x);y=[-1;x;1];
eps=[.1 .01 .001 .0001 1.e-05];
for i=1:5
A=eps(i)*D2t-D1t; F=.5*ones(size(x));
v=A\F;
v=[0;v;0];v=flipud(v);
CCsol(:,i)=v+y.*.5+1.5;
ac=.5+1/(exp(-2/eps(i))-1);bc=ac-.5;
EXsol=ac+1.5-bc*exp((y-1)/eps(i));
format long
eps(i)
```

```
MaxDifference=max(abs(CCsol(:,i)-EXsol))
clear v ac bc EXsol MaxDifference
end
fsize=10;
plot(y,CCsol(:,1),'-r',y,CCsol(:,2),'.g',y,CCsol(:,3),'+b',y,CCsol(:,4),'-c'...
,y,CCsol(:,5),'-m')
title(' A singularly perturbed problem (N=1024)','FontSize',fsize)
legend([repmat('\epsilon=',5,1), num2str(eps')],2)
xlabel('x','FontSize',10)
ylabel('u(x,\epsilon)','FontSize',10)
hold off
% Obtain the Figure 2.4
```

4) The MATLAB code `Therm.m`

```
% Solve the t p b v p u''+u^3=0, u(-1)=u(+1)=0
% The second order operator u'' discretized by Chebyshev-collocation method
%
clear all, close all
N=64; pi=4.*atan(1.); % Order of approximation
g=[1 0 0;1 0 0]; % B. C. for CC method
[x,D2t,D1t]=cheb2bc(N,g); % Differentiation matrices
u0=(1-x.^2).^2; % Make a starting guess at the solution
lambda=5;
% Solve the nonlinear algebraic system
options=optimset('Display','off','LevenbergMarquardt','on')
[u,Fval,exitflag]=fsolve(@therm,u0,options,D2t,lambda) % Call optimizer
sol=[0;u;0]; x=[1;x;-1];fsize=10;
sum(sum(Fval.*Fval))
plot(x,sol)
xlabel('x','FontSize',fsize)
ylabel('u(x)','FontSize',fsize)
title('The solution to average temperature in a r-d process','FontSize',fsize)
% Obtain the Figure 2.5
The function therm is available in the next routine:

function [F,J]=therm(x,D2t,lambda)
%
x1=ones(size(x));
F=D2t*x+lambda.*diag(x1)*(x.^3) ; % objective function
```

5) The MATLAB code `Heat2.m`

```
% MATLAB code Heat2
% Solve ibvp for heat equation du/dt=S(x)u''(x,t)
```

```
% Initial data u(x,0)=u0, boundary conditions u(-1)=u(1)=0;
clear, clf
N=32; pi=4.*atan(1.);    % Order of approximation
g=[1 0 0;1 0 0];         % B. C. for CC method
tfinal=1.;               % Final time of integration
[x,D2t]=cheb2bc(N,g);    % Differentiation matrix (second order) with enforced b. c.
% Use the routine cheb2bc from Weideman & Reddy
u=zeros(size(x));
u0=1-cos(pi.*(x+1));
subplot(3,1,1); plot(x,u0),ylabel('u0(x,0)'), title('Initial data')
tspan=[0:tfinal/100:tfinal]; % Step size
options=odeset('RelTol',1e-04,'AbsTol',1e-6);
[t,u]=ode45('parab',tspan,u0,options,D2t);  % Runge-Kutta solver
x=[1;x;-1];                          % The spatial grid
% [m,n]=size(u)
ufinal=[0,u(m,:),0];
subplot(3,1,2); plot(x,ufinal),ylabel('u(x,tfinal)'), title('Solution at final time')
u1=zeros(m,1);
u=[u1,u,u1];
subplot(3,1,3); mesh(x,t,u), xlabel('x'), ylabel('t'), zlabel('u(x,t)'),
title('Solution of heat initial-boundary value problem')

function dudt=parab(t,w,flag,D2t)
% Function to compute the RHS of heat 1D eq.
dudt=zeros(size(w));                % Preallocate column vector dudt
dudt=D2t*w;
dudt=12.*dudt;
% Obtain Figure 3.1
```

6) The MATLAB code `Burgersibvp.m`

```
% A MATLAB code for Burgers equation using Hermite collocation
% Solving initial boundary value problem for Burgers equation
% du/dt=-(theta/2)*(u^2)'-(1-theta)*uu')+eps*u''
% Initial data u(x,0)=u0, boundary conditions u=0 at the end of real line
% !!
clear all; close all;
N=200; pi=4.*atan(1.); % Order of approximation
b=.545; % B. C. for HermiteC method
eps=1/(100*pi); % Difusion coefficient
[x,D]=herdif(N,2,b); % Differentiation matrices
D2t=D(:,:,2); D1t=D(:,:,1); fsize=10;
u0=0.5*sech(x);
subplot(2,2,1),plot(x,u0)
xlabel('x','FontSize',fsize)
ylabel(['u(x,',num2str(0),')'],'FontSize',fsize)
```

```
title('Solution to Burgers eq., N=200,\epsilon=1/(300*\pi)','FontSize',fsize)
% Shampine, L. F. and M. W. Reichelt, "The MATLAB ODE Suite,"
% SIAM Journal on Scientific Computing, Vol. 18, 1997, pp 1-22.
options=odeset('RelTol',1e-04,'AbsTol',1e-6);
for i=1:3
t0=(i-1)*5; tf=i*5;
tspan=[t0:tf/200:tf]; % Step size
[t,u]=ode113('Burgers',tspan,u0,options,eps,D2t,D1t);
[m,n]=size(u);
ufinal=u(m,:);i1=i+1;
subplot(2,2,i1),plot(x,ufinal)
xlabel('x','FontSize',fsize)
ylabel(['u(x,',num2str(tf),')'],'FontSize',fsize)
u0=u(m,:);
clear u ufinal
end
xlabel('x','FontSize',fsize)
ylabel(['u(x,',num2str(tf),')'],'FontSize',fsize)
% Obtain Figure 3.3



function du=Burgers(t,w,flag,eps,D2t,D1t)
% Function to compute the RHS of Burgers eq.
% the convective term has the conservative form (uu/2)'
theta=2/3;
du=zeros(size(w)); % Preallocate column vector du
convterm=zeros(size(w));
convterm=(theta*D1t*(w.*w)/2+(1-theta)*w.*(D1t*w));
du=-convterm+eps*(D2t*w);
```

7) The MATLAB code `FourthEq.m`

```
% Solve a standard fourth order b. v. problem by a
% Chebyshev Galerkin (CG) method
clear all; close all
N = input(' Dimension of the differentiation matrix: N = ? ');
c=ones(N,1);c(1)=2;pi=4*atan(1);
m=1:N;a40=c'.*m.*(m+1).*(m+2).*(m+3); size(a40) % Get the matrix A4
m=1:N-2;a42=-2*(m-1).*m.*(m+1).*(m+3);
m=1:N-4;a44=(m-1).*(m.^2).*(m+1);
A4=diag(a40)+diag(a42,2)+diag(a44,4); A4=pi*A4/2;
for k=5:N                                          % Get the matrix A0
a0_4(k-4)=c(k-4)/16;
end
for k=3:N
```

```
a0_2(k-2)=-c(k-2)*(3*(k-1)+8)/(8*(k+2));
end
a0_2(2)=a0_2(2)+1/16;
m=1:N;a00=(15*(m-1)+35*c')./(16.*(m+2));
a00(2)=a00(2)-22/(16*(2+2));a00(3)=a00(3)+5/(16*(3+2));
for m=1:N-2
a02(m)=-(5*(m-1)+6+4*c(m))/(4*(m+2));
end
a02(2)=a02(2)-1/(4*(2+2));
for m=1:N-4
a04(m)=(15*(m-1)+22+3*c(m))/(16*(m+2));
end
m=1:N-6;a06=-(3*(m-1)+4)./(8*(m+2));
m=1:N-8;a08=m./(16*(m+2));
A0=diag(a0_4,-4)+diag(a0_2,-2)+diag(a00)+diag(a02,2)+diag(a04,4)+...
diag(a06,6)+diag(a08,8); A0=pi*A0/2;
% Get the matrix A in the linear algebraic system A*U=F
A=A4+A0;
% Get the right hand side F in the linear system A*U=F
for n=1:N
Rhsf=@(x)2*(exp(x)).*(1+40*x+34*(x.^2)+8*(x.^3)+(x.^4)).*((1-(x.^2)).^(-1/2)).*...
(cos((n-1)*acos(x))-(2*(n+1)/(n+2))*cos((n+1)*acos(x))+...
(n/(n+2))*cos((n+3)*acos(x)));
Q(n)=quadl(Rhsf,-1,1);
end
tol=1e-09;format long e
maxit=105;y=-1:.01:1;
CGcoef=A\(Q')
format long e
Exsol=((1-y.^2).^2).*exp(y); % The close solution
CGsol=zeros(size(y));
for n=1:N
CGsol=CGsol+((1-y.^2).^2).*cos((n-1)*acos(y))*CGcoef(n);
end
MaxDifference=max(abs(CGsol-Exsol)) % The accuracy of computation
plot(y,CGsol,y,Exsol), legend('CCsol','Exsol') % Plot and compare solutions
xlabel('x','FontSize',10)
ylabel('u(x)','FontSize',10)
title('Solution u(x)=((1-x^2)^2)*exp(x) of a fourth order b v p')
condeig(A)
B=A'*A-A*A';s=norm(A,1);
Henrici=sqrt(norm(B,'fro'))/norm(A,'fro') % the non-normality ratio
% Obtain Figure 4.2
```

Remark 91 *We can now compare the implementation of spectral collocation methods and spectral Galerkin methods. It is clear that whenever reliable codes*

for the differentiation matrices are well established, the collocation methods lead to more compact, more simple and more flexible codes than Galerkin methods.

8) The MATLAB code `Sine_Gordon.m`

```
% This script file solves the sine-Gordon equation
% u_tt=u_xx-sin u on the real line using one of the following
% differentiation matrices: (1) Hermite, (2) sinc, or (3) Fourier.
% The solution is displayed as a mesh plot.
clear all; close all
method = input(' Which method: (1) Hermite, (2) sinc, (3) Fourier? ')
N = input(' Order of differentiation matrix: N = ? ')
tfinal = input(' Final time: t = ? ')
if method == 1
b = input(' Scaling parameter for Hermite method: b = ? ');
[x,D] = herdif(N,2,b); % Compute Hermite differentiation matrices
D = D(:,:,2); % Extract second derivative
D1 = D(:,:,1); size(x)
elseif method == 2
h = input(' Step-size for sinc method: h = ? ');
[x,D] = sincdif(N,2,h); % Compute sinc differentiation matrices
D = D(:,:,2); % Extract second derivative
D1 = D(:,:,1);
elseif method == 3
L = input(' Half-period for Fourier method: L = ? ');
[x,D] = fourdif(N,2); % Compute Fourier second derivative
x = L*(x-pi)/pi; % Rescale [0, 2pi] to [-L,L]
D = (pi/L)^2*D;
[x,D1] = fourdif(N,1);
end
Nsteps=1.e+04;
% Integrate in time by a symplectic method
%[T,P,Q]=gni_irk2('SinG1',[],[],[],x,D,tfinal);
%[T,P,Q]=gni_lmm2('SinG1',[],[],[],x,D,tfinal);
[T,P,Q]=gni_comp('SinG1',[],[],[],x,D,tfinal,Nsteps);
% Obtain Figure 3.13
subplot(1,2,1),mesh(x,T,P); view(30,30); % Generate a mesh plot of u
M=length(T);
for k=1:M
for j=1:N
ExSol(k,j)=4*atan(sin(T(k)/sqrt(2))/cosh(x(j)/sqrt(2)));
end
end
Error=max(max(abs(ExSol-P)))
[m,n]=size(P)
fsize=10;
xlabel('x','FontSize',fsize)
```

```
ylabel('t','FontSize',fsize)
zlabel('u(x,t)','FontSize',fsize)
title(['The "breather solution" N=',num2str(N)],'FontSize',fsize)
for k=1:m
qd=D1*(Q(k,:))';
pqd=fourint(qd,x);
Tu=(pqd.^2)/2+((P(k,:)').^2)/2;
Vu=-cos(Q(k,:)');
H(k)=trapz(x,Vu)+trapz(x,Tu);
end
H=(H-H(1))/H(1);
subplot(1,2,2),plot(T,log10(abs(H)))
% Obtain Figure 3.14
plot(T,log10(abs(H)))
xlabel('time','FontSize',fsize)
ylabel('log10(|\deltaH(u)/H|)','FontSize',fsize)
title('The conservation of energy functional','FontSize',fsize)
```

9) The MATLAB code `KdVivpFourier.m`

```
% Solve KdV eq by Fourier spectral method with
% periodic boundary conditions
clear all; close all
N=160; tfinal=4*pi; pi=4*atan(1.);t0=0;
[x,D3]=fourdif(N,3);[x,D1]=fourdif(N,1);
u0=zeros(size(x));
u0=cos(x-pi);a=-3/8;ro=-.1; niu=(-2/3)*1.e-03;
fsize=10; figure(1)
subplot(5,1,1); plot(x,u0);ylabel(['u(x,',num2str(t0),')'],'FontSize',fsize);
axis([0 2*pi -1 1.5])
title('Solutions of KdV i. v. p. for initial data u_0(x)=cos(x-\pi)','FontSize',fsize)
options=odeset('RelTol',1e-03,'AbsTol',1e-4);
for i=1:4
tf=t0+.10;
[t,u]=ode45(@KdVFourier,[t0 tf],u0,options,a,ro,niu,D3,D1);
[m,n]=size(u);i1=i+1;
subplot(5,1,i1); plot(x,u(m,:));ylabel(['u(x,',num2str(tf),')'],'FontSize',fsize);
axis([0 2*pi -1 2.5])
u0=u(m,:);
t0=tf;
end
xlabel('x','FontSize',fsize);
% Obtain Figure 3.7

for k=1:m
ud=D1*(u(k,:))';
```

```
pu=fourint(u(k,:),x);
pud=fourint(ud,x);
Tu=-niu*(pud.^2)/2;
Vu=a*(pu.^3)/3+ro*(pu.^2)/2;
H(k)=trapz(x,Vu)+trapz(x,Tu);
end
H=(H-H(1))/H(1);
figure(2)
plot(t,log10(abs(H)))
xlabel('time','FontSize',fsize)
ylabel('log10(|\deltaH(u)/H|)','FontSize',fsize)
title('The conservation of energy functional','FontSize',fsize)
% Obtain Figure 3.9

function du=KdVFourier(t,w,a,ro,niu,D3t,D1t)
% Function to compute the RHS of KdV eq.
% by Fourier spectral methods
pi=4*atan(1.);
theta=0.;
du=zeros(size(w)); % Preallocate column vector du
convterm=zeros(size(w));
convterm=D1t*(pi*ro*w)+a*pi*theta*D1t*(w.*w)+2*a*pi*(1-theta)*w.*(D1t*w);
du=convterm+niu*(pi^3)*(D3t*w);
```

10) The MATLAB code `Fischer_Ibvp.m`

```
%  This script file solves the Fischer equation
%  u_t=u_xx+u(1-u) on the real line using one of the following
%  differentiation matrices: (1) Hermite, (2) sinc, or (3) Fourier.
clear all; close all
method = input(' Which method: (1) Hermite, (2) sinc, (3) Fourier? ')
N = input(' Order of differentiation matrix: N = ? ')
tfinal = input(' Final time: tf = ? ')
if method == 1
b = input(' Scaling parameter for Hermite method: b = ? ');
[x,D] = herdif(N,2,b);              % Hermite differentiation matrices
    D = D(:,:,2);                   % Extract second derivativ
elseif method == 2
h = input(' Step-size for sinc method: h = ? ');
[x,D] = sincdif(N,2,h);             % Compute sinc differentiation matrices
    D = D(:,:,2);                   % Extract second derivative
elseif method == 3
L = input(' Half-period for Fourier method: L = ? ');
[x,D] = fourdif(N,2);               % Compute Fourier second derivative
    x = L*(x-pi)/pi;                % Rescale [0, 2pi] to [-L,L]
    D = (pi/L)^2*D;
```

```
end
u0=sin(x);Nsteps=1.e+02;t0=0;
options=odeset('RelTol',1e-03,'AbsTol',1e-3);
tspan=[t0 tfinal];
[t,u]=ode45(@Fischer,tspan,u0,options,D);
mesh(x,t,u);                          % Generate a mesh plot of solution u
xlabel('x','FontSize',10);ylabel('t','FontSize',10);
zlabel('u(x,t)','FontSize',10);
title('Solution to Fischer equation','FontSize',10);
% Obtain Figure 3.21

 function dw = Fischer(t,u,D)
%  The function dw = Fischer(t,w,D) computes the right-hand side
%  of the Fischer equation with the aid of an NxN differentiation matrix
%  D.
N=length(u);
dw=zeros(N,1);
dw = D*u+u.*(1-u);
```

Bibliography

[1] Ablowitz, M.J., Herbst, B.M., Schober, C., *On the Numerical Solution of the Sine-Gordon Equation, I Integrable Discretizations and Homoclinic Manifolds,* J. Comput. Phys., 126(1996), 299-314

[2] Ache, G.A., Cores, D., *Note on the Two-Point Boundary Value Numerical Solution of the Orr-Sommerfeld Stability Equation,* J. Comput. Phys., 116(1995), 180-183

[3] Adams, R. A., *Sobolev Spaces,* Academic Press, New York-San Francisco-London, 1975

[4] Andrew, A. L., *Centrosymmetric Matrices, SIAM Rev.,* 40(1998), 697-698

[5] Andrews, L. C., *Special Functions of Mathematics for Engineers,* International Edition 1992, McGraw-Hill, Inc. 1992

[6] Antohe, V., Gladwell, I., *Performance of two methods for solving separable Hamiltonian systems,* J. Comput. Appl. Math., 125(2000), 83-92

[7] Ascher, U. M., Russel, R. D., *Reformulation of Boundary Value Problems into "Standard" Form,* SIAM Rev. 23(1981), 238-254

[8] Ascher, U. M., Mattheij, R, M. M., Russel, R. D., *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations,* Prentice Hall, Englewood Cliffs, 1988

[9] Ascher, U. M., McLachlan, R. I., *Multisymplectic box schemes and the Korteweg-de Vries equation,* Preprint submitted to Elsevier Sciences, 31 July, 2003

[10] Atkinson, K.E., *Elementary Numerical Analysis,* John Wiley &Sons, 1985

[11] Babuska, I., Aziz, K., *Survey Lectures on the Mathematical Foundation of the Finite Element Method,* in The Mathematical Foundations of the Finite Element Method with Application to Partial Differential Equations, Ed. A. K. Aziz, Academic Press, London/New York 1972, pp. 3-359

[12] Basdevant, C., Deville, M., Haldenwang, P., Lacroix, J. M., Ouazzani, J., Peyret, R., Orlandi, P., Patera, A.T., *Spectral and Finite Solutions of the Burgers Equation,* Computers and Fluids, 14(1986) 23-41

[13] Berland, H., Islas, A. L., Schoder, C., *Conservation of phase properties using exponential integrators on the cubic Schrödinger equation,* Preprint Norvegian University of Sciences and Technology, Trondheim, Norway, 2006

[14] Bernardi, C., Maday, Y., *Approximations Spectrales de Problems aux Limites Elliptiques,* Springer Verlag, Paris, 1992

[15] Birkhoff, G., Rota, G.-C., *Ordinary Differential Equations,* John Wiley & Sons, Second Edition, 1969

[16] Bjoerstad, P.E., Tjoestheim, B.P., *Efficient algorithms for solving a fourth-equation with the spectral-Galerkin method,* SIAM J. Sci. Stat. Comput. 18(1997), 621-632

[17] Boyd, J. P., *Numerical Computations of a Nearly Singular Nonlinear Equation: Weakly Nonlocal Bound States of Solitons for the Fifth-Order Korteweg-de Vries Equation,* J. Comput. Phys. 124(1996), 55-70

[18] Boyd, J. P., *Traps and Snares in Eigenvalue Calculations with Application to Pseudospectral Computations of Ocean Tides in a Basin Bounded by Meridians,* J. Comput. Phys., 126(1996), 11-20

[19] Boyd, J. P., *Chebyshev and Fourier Spectral Methods,* Second Edition, DOVER Publications, Inc., 2000

[20] Boyd, J. P., *Trouble with Gegenbauer reconstruction for defeating Gibbs' phenomenon: Runge phenomenon in the diagonal limit of Gegenbauer polynomial approximations,* J. Comput. Phys., 204(2005), 253-264

[21] Boyd, J. P., *A Chebyshev/rational Chebyshev spectral method for the Helmholtz equation in a sector on the surface of a sphere: defeating corner singularities,* J. Comput. Phys., 206(2005), 302-310

[22] Breuer, K.S., Everson, R.M., *On the Errors Incurred Calculating Derivatives Using Chebyshev Polynomials,* J. Comput. Phys., 99(1992), 56-67

[23] Brusch, L., Torcini, A., van Hecke, M., Zimmermann, M. G., Bär, M., *Modulated amplitude waves and defect formation in the one-dimensional complex Ginzburg-Landau equation,* Physica D, 160(2001) 127-148

[24] Butcher, J. C., *The Numerical Analysis of Ordinary Differential Equations-Runge-Kutta and General Linear Methods,* John Wiley &Sons, 1987

[25] Butcher, J. C., *Numerical methods for ordinary differential equations in the 20th century,* J. Comput. Appl. Math., 125(2000), 1-29

[26] Butzer, P., Jongmans, F., *P. L. Chebyshev (1821-1894) A Guide to his Life and Work,* J. Approx. Theory, 96(1999) 111-138

[27] Cabos, Ch., *A preconditioning of the tau operator for ordinary differential equations,* ZAMM 74(1994) 521-532

[28] Canuto, C., *Boundary Conditions in Chebyshev and Legendre Methods,* SIAM J. Numer. Anal., 23(1986) 815-831

[29] Canuto, C., *Spectral Methods and a Maximum Principle,* Math. Comput., 51(1988), 615-629

[30] Canuto, C., Quarteroni, A., *Approximation results for orthogonal polynomials in Sobolev spaces,* Math. Comp. 38(1982), 67-86

[31] Canuto, C., Quarteroni, A., *Variational Methods in the Theoretical Analysis of Spectral Approximations,* in Spectral Methods for Partial Differential Equations, Ed. by R. G. Voigt, D. Gottlieb, M. Y. Hussaini, SIAM-CBMS, P. 55-78

[32] Canuto, C., Quarteroni, A., *Spectral and pseudo-spectral methods for parabolic problems with non-periodic boundary conditions,* Calcolo 18(1981), 197-218

[33] Canuto, C, Hussaini, M.Y., Quarteroni, A., Zang, T.A., *Spectral Methods in Fluid Dynamics,* Springer Verlag, 1987

[34] Carpenter, M. H., Gottlieb, D., *Spectral Methods on Arbitrary Grids,* J. Comput. Phys., 129(1996), 74-86

[35] Coron, J-M, Crepeau, E, *Exact boundary controllability of a nonlinear KdV equation with critical lengths,* J. Eur. Math. Soc., 6(2004) 367-398

[36] Chaitin-Chatelin, F. and V. Fraysse, *Lectures on Finite Precision Computation,* SIAM Philadelphia, 1996

[37] Chan, T. F., Kerkhoven, T., *Fourier Methods with Extended Stability Intervals for the Korteweg-De Vries Equation,* SIAM J. Numer. Anal., 22(1985), 441-454

[38] Ciarlet, P.G., *The Finite Element Method for Elliptic Equations,* North-Holland Publishing Company, 1978

[39] Ciarlet, P.G., *Introduction to Numerical Linear Algebra and Optimization,* Cambridge Texts in Applied Mathematics, CUP, 1989

[40] Cooley, J. W., Tukey, J. W., *An Algorithm for the Machine Calculation of Complex Fourier Series,* Math. Comput., 19(1965) 297-301

[41] Cooley, J. W., Lewis, P. A. W., Welch, P. D., *The Fast Fourier Transform Algorithm: Programming Considerations in the Calculation of Sine, Cosine, and Laplace Transforms,* J. Sound. Vib., 12(1970) 315-337

[42] Davis, P.J., *Interpolation and Approximation,* Blaisdell Pub. Co., New-York, 1963

[43] Davis, P.J., Rabinowitz, P., *Methods of Numerical Integration,* New York: Academic Press, 1975

[44] Deeba, E., Khuri, S. A., *A decomposition Method for Solving the Nonlinear Klein-Gordon Equation,* J. Comput. Phys., 124(1996), 442-448

[45] Deeba, E., Khuri, S. A., Xie, S., *An Algorithm for Solving Boundary Value Problems,* J. Comput. Phys., 159(2000), 125-138

[46] Dendy, J. E. Jr., *Galerkin's Method for some Highly Nonlinear Problems,* SIAM J Numer. Anal., 14(1997), 327-347

[47] Denis, S.C.R., Quartapelle, L., *Spectral Algorithms for Vector Elliptic Equations in a Spherical Gap,* J. Comput. Phys., 61(1985), 218-241

[48] Deuflhard, P., Hohmann, A., *Numerical Analysis in Modern Scientific Computing; An Introduction,* Springer Verlag, 2003

[49] Doha, E.H., Bhrawy, A.H., *Efficient spectral-Galerkin algorithms for direct solution for second-order differential equations using Jacobi polynomials,* Numer. Algor., 42(2006), 137-164

[50] Doha, E.H., Bhrawy, A.H., *Efficient spectral-Galerkin algorithms for direct solution for fourth-order differential equations using Jacobi polynomials,* Appl. Numer. Math., 58(2008), 1224-1244

[51] Doha, E.H., Abd-Elhameed, W. M., Bhrawy, A.H., *Efficient spectral ultraspherical-Galerkin algorithms for the direct solution of 2n th-order linear differential equations,* Appl. Math. Modell., 33(2009), 1982-1996

[52] Doha, E.H., Abd-Elhameed, W. M., *Efficient spectral ultraspherical-dual-Petrov-Galerkin algorithms for the direct solution of (2n+1) th-order linear differential equations,* Math. Comput. Simul., 79(2009), 3221-3242

[53] Doha, E.H., Bhrawy, A.H., Abd-Elhameed, W. M., *Jacobi spectral Galerkin method for elliptic Neumann problems,* Numer. Algor., 50(2009), 67-91

[54] Don, W. S., Gottlieb, D., *The Chebyshev-Legendre Method: Implementing Legendre Methods on Chebyshev Points,* SIAM J. Numer. Anal., 31(1994), 1519-1534

[55] Dongarra, J.J., Straughan, B., Walker, D.W., *Chebyshev tau- QZ algorithm for calculating spectra of hydrodynamic stability problems,* Appl. Numer. Math. 22(1996), 399-434

[56] Drazin, P.G., *Nonlinear Systems,* Cambridge University Press, 1992

[57] Drazin, P.G., Beaumont, D.N., Coaker, S.A., *On Rossby waves modified by basic shear, and barotropic instability,* J. Fluid Mech. 124(1982), 439-456

[58] Driscoll, T. A., Fornberg, B., *A Block Pseudospectral Method for Maxwell's Equations,* J. Comput. Phys., 140(1998), 47-65

[59] Eberlein, P. J., *On measures of non-normality for matrices,* Amer. Math. Monthly 72(1965), 995-996

[60] Elbarbary, E. M. E., Ei-Sayed, S. M., *Higher order pseudospectral differentiation matrices,* Appl. Numer. Math., 55(2005), 425-438

[61] El-Daou, M.K., Ortiz, E.L., Samara, H., *A Unified Approach to the Tau Method and Chebyshev Series Expansion Techniques,* Computers Math. Applic. 25(1993), 73-82

[62] El-gamel, M., *A comparison between the Sinc-Galerkin and the modified decomposition methods for solving two-point boundary-value problems,* J. Comput. Phys., 223(2007), 369-383

[63] L. Elsner, M. H. C. Paardekooper, *On Measure of Nonnormality of Matrices*, Linear Algebra Appl. 92:107-124 (1897)

[64] Engquist, B., Osher, S., *One-Sided Difference Approximations for Nonlinear Conservation Laws,* Math. Comp. 36(1981) 321-351

[65] Fatone, L., Funaro, D., Yoon, G. J., *A convergence analysis for the superconsistent Chebyshev method,* Appl. Numer. Math., 58(2008), 88-100

[66] Fishelov, D., *The Spectrum and the Stability of the Chebyshev Collocation Operator for Transonic Flow,* Math. Comput. 51(1988), 559-579

[67] Fornberg, B., *On the Instability of Leap-Frog and Crank-Nicolson Approximations of a Nonlinear Partial Differential Equation,* Math. Comput. 27(1973) 45-57

[68] Fornberg, B., *Generation of Finite Difference Formulas on Arbitrary Spaced Grids,* Math. Comput., 51(1988), 699-706

[69] Fornberg, B., Sloan, D., M., *A review of pseudospectral methods for solving partial differential equations,* Acta Numerica, 1994, 203-267

[70] Fornberg, B., Driscoll, T.A., *A Fast Spectral Algorithm for Nonlinear Wave Equations with Linear Dispersion,* J. Comput. Phys. 155(1999), 456-467

[71] Fox, L., Parker, I.B., *Chebyshev Polynomials in Numerical Analysis,* Oxford Mathematical Handbooks, O U P, 1968

[72] Funaro, D., *A Preconditioning Matrix for the Chebyshev Differencing Operator,* SIAM J. Numer. Anal., 24(1987) 1024-1031

[73] Funaro, D., *FORTRAN routines for spectral methods,* http://cdm.unimo.it/home/matematica/funaro.daniele/fman.pdf

[74] Funaro, D., *A Variational Formulation for the Chebyshev Pseudospectral Approximation of Neumann Problems,* SIAM J. Numer. Anal. 27(1990), 695-703

[75] Funaro, D., *A New Scheme for the Approximation of Advection-Diffusion Equations by Collocation,* SIAM J. Numer. Anal., 30(1993), 1664-1676

[76] Funaro, D., Heinrichs,W., *Some results about the pseudospectral approximation of one-dimensional fourth-order problems,* Numer. Math. 58(1990), 399-418

[77] Funaro, D., Kavian, O., *Approximation of Some Diffusion Evolution Equations in Unbounded Domains by Hermite Functions,* Math. Comput. 57(1991), 597-619

[78] Garcia-Archilla, B., *A Spectral Method for the Equal Width Equation,* J. Comput. Phys., 125(1996), 395-402

[79] Gardner, C. S. *Korteweg-de Vries Equation and Generalizations. IV. The Korteweg-de Vrie Equation as a Hamiltonian System,* J. Math. Phys., 12(1971), 1548-1561

[80] Gardner, D. R., Trogdon, S. A., Douglass, R. D., *A Modified Tau Spectral Method That Eliminates Spurious Eigenvalues,* J. Comput. Phys., 80(1989)137-167

[81] Gheorghiu, C.I., Pop, S.I.,*On the Chebyshev-tau approximation for some singularly perturbed two-point boundary value problems,* Rev. Roum. Anal. Numer. Theor. Approx., 24(1995), 117-124, Zbl M 960.44077

[82] Gheorghiu, C.I., Pop, S.I., *A Modified Chebyshev-tau Method for a Hydrodynamic Stability Problem,* Proceedings of I C A O R, vol. II, pp.119-126, Cluj-Napoca, 1997 MR 98g:41002

[83] Gheorghiu, C.I., *A Constructive Introduction to Finite Elements Method,* Qvo Vadis, Cluj-Napoca, 1997

[84] Gheorghiu, C. I., *On the Scalar Measure of Non-Normality of Matrices; Dimension vs. Structure,* General Mathematics, U. L. B. Sibiu, 11:21-32 (2003)

[85] Gheorghiu, C. I., *On the spectral Characterization of some Chebyshev-Type Methods; Dimension vs. Structure,* Studia Univ. "Babes-Bolyai", Mathematica, L(2005) 61-66

[86] Gheorghiu, C. I., Trif, D., *The numerical approximation to positive solution of some reaction-diffusion problems,* PU. M. A., 11(2000) 243-253

[87] Golub, G. H., Wilkinson, J. H., *Ill-Conditioned Eigensystems and the Computation of the Jordan Canonical Form,* SIAM Review, 18(1976) 578-619

[88] Golub, G. H., Ortega, J. M., *Scientific Computing and Differential Equations-An Introduction to Numerical Methods,* Academic Press, 1992

[89] Golub, G. H., van der Vorst, H. A., *Eigenvalue computation in the 20th century,* J. Comput. Appl. Math., 123(2000), 35-65

[90] Gottlieb, D., Orszag, S.A., *Numerical Analysis of Spectral Methods: Theory and Applications,* SIAM, 1977

[91] Gottlieb, D., Turkel, E., *On Time Discretization for Spectral Methods,* Studies in Appl. Math., 63(1980), 67-86

[92] Gottlieb, D., *The Stability of Pseudospectral-Chebyshev Methods,* Math. Comp., 36(1981), 107-118

[93] Gottlieb, D., Hussaini, M.Y., Orszag, S.A., *Theory and Applications of Spectral Methods,* in Spectral Methods for Partial Differential Equations, Ed. by R.G. Voigt, D. Gottlieb, M.Y. Hussaini, SIAM-CBMS, P. 1-54, 1984

[94] Gottlieb, D., Orszag, S.A., Turkel, E., *Stability of Pseudospectral and Finite-Difference Methods for Variable Coefficient Problem,* Math. Comp. 37(1981), 293-305

[95] Gottlieb, D., Hesthaven, J. S., *Spectral methods for hyperbolic problems,* J. Comput. Appl. Math., 128(2001), 83-131

[96] Greenberg, L., Marleta, M., *Numerical methods for higher order Sturm-Liouville problems,* J. Comput. Appl. Math., 125(2000) 367-383

[97] Greenberg, L., Marleta, M., *Numerical Solution of Non-Self-Adjoint Sturm-Liouville Problems and Related Systems,* SIAM J. Numer. Anal., 38(2001), 1800-1845

[98] Greengard, L., Rokhlin, V., *On the Numerical Solution of Two-Point Boundary Value Problems,* Comm. Pure Appl. Math. XLIV(1991), 419-452

[99] Guo, B.-y, Wang, Z.-q, Wan, Z.-s, Chu, D., *Second order Jacobi approximation with applications to fourth-order differential equations,* Appl. Numer. Math., 55(2005), 480-502

[100] Hamming, R. W., *Introduction to Applied Numerical Analysis,* International Student Edition, McGraw-Hill, Inc. 1971

[101] Hairer, E., Hairer, M., *GniCodes-Matlab Programs for Geometric Numerical Integration,* http://www.unige.ch/math/folks/hairer

[102] Heinrichs, W., *Spectral Methods with Sparse Matrices,* Numer. Math., 56(1989), 25-41

[103] Heinrichs, W., *Improved Condition Number for Spectral Methods,* 53(1998), Math. Comp., 103-119

[104] Heinrichs, W., *Stabilization Techniques for Spectral Methods,* J. Sci. Comput., 6(1991), 1-19

[105] Heinrichs, W., *A Stabilized Treatment of the Biharmonic Operator with Spectral Methods,* SIAM J. Sci. Stat. Comput., 12(1991), 1162-1172

[106] Heinrichs, W., *Strong Convergence Estimates for Pseudospectral Methods,* Appl. Math., (1992), 401-417

[107] Heinrichs, W., *Spectral Approximation of Third-Order Problems,* J. Scientific Computing, 14(1999), 275-289

[108] Henrici,P., *Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices,* Numer. Math. 4, 24-40(1962)

[109] Henrici, P., *Discrete Variable Methods in Ordinary Differential Equations,* John Wiley & Sons, Inc., New York-London, 1962

[110] Henrici, P., *Essentials of Numerical Analysis with Pocket Calculator Demonstrations,* John Wiley & Sons, Inc., New York, 1982

[111] Hiegemann, M., Strauss, K., *On a Chebyshev matrix operator method for ordinary differential equations with non-constant coefficients,* Acta Mech., 105(1994), 227-232

[112] Hiegemann, M., *Chebyshev matrix operator method for the solution of integrated forms of linear ordinary differential equations,* Acta Mech. 122(1997), 231-242

[113] Higham, D. J., Owren, B., *Nonnormality Effects in a Discretised Nonlinear Reaction-Convection-Diffusion Equation,* J. Comput. Phys., 124(1996), 309-323

[114] Hill, A. A., Straughan, B., *Linear and non-linear stability thresholds for thermal convection in a box,* Math. Meth. Appl. Sci. 29(2006), 2123-2132

[115] Holden, H., Karlsen, K. H., Risebro, N. H., *Operator Splitting Methods for Generalized Korteweg-De Vries Equations,* J. Comput. Phys. 153(1999), 203-222

[116] Huang, W., Sloan, D.M., *The pseudospectral method for third-order differential equations,* SIAM J. Numer. Anal. 29(1992), 1626-1647

[117] Huang, W., Sloan, D.M., *The pseudospectral method for solving differential eigenvalue problems,* J. Comput. Phys. 111(1994), 399-409

[118] Hunt, B. R., Lipsman, R. L., Rosenberg, J. M., Coombes, K. R., Osborn, J. E., Stuck, G. J., *A guide to MATLAB for Beginners and Experienced Users,* Cambridge University Press, 2001

[119] Iserles, A., *A First Course in the Numerical Analysis of Differential Equations,* Cambridge University Press, 1996

[120] Ismail, M. S., *Numerical solution of coupled nonlinear Schrödinger equation by Galerkin method,* Math. Comput. Simul. 78(2008), 532-547

[121] Johnson, C., *Numerical solutions of partial differential equations by the finite element method,* Cambridge University Press, 1987

[122] Jung, J-H, Shizgal, B. D., *On the numerical convergence with the inverse polynomial reconstruction method for the resolution of the Gibbs phenomenon,* J. Comput. Phys., 224(2007), 477-488

[123] Kosugi, S., Morita, Y., Yotsutani, S., *Complete bifurcation diagram of the Ginzburg-Landau equation with periodic boundary conditions,* Communications on Pure and Applied Analysis, 4(2005) 665-682

[124] R. Kress, H. L. de Vries, R. Wegmann, *On Nonnormal Matrices,* Linear Algebra Appl. 8, 109-120(1974)

[125] Kreiss, H-O., Oliger, J., *Stability of the Fourier Method,* SIAM J. Numerical Analysis, 16(1979), 421-433

[126] Lanczos, C., *Applied Analysis,* Prentice Hall Inc., Englewood Cliffs, N. J., 1956

[127] Lax, P.D., Milgram, A.N., *Parabolic equations,* Annals of Math. Studies, No. 33(1954) Princeton Univ. Press

[128] Lax, P.D., *Integrals of Nonlinear Equations of Evolution and Solitary Waves,* Communications on Pure and Applied Mathematics, XXI(1968), 467-490

[129] Lax, P.D., *Almost periodic solutions of the KdV equation,* SIAM Rev., 18(1976) 351-375

[130] Lee, S. L., *A Practical Upper Bound for Departure from Normality,* SIAM J. Matrix Anal. Appl., 16, 462-468, April 1995

[131] Lindsay, K.A., Odgen, R.R., *A Practical Implementation of Spectral Methods Resistant to the Generation of Spurious Eigenvalues,* Intl. J. Numer. Fluids 15(1992), 1277-1294

[132] Maday, Y., Quarteroni, A., *Legendre and Chebyshev Spectral Approximations of Burgers' Equation,* Numer. Math. 37(1981) 321-332

[133] Maday, Y., *Analysis of Spectral Projectors in One-Dimensional Domains,* Math. Comp. 55(1990), 537-562

[134] Maday, Y., Metivet, B., *Chebyshev Spectral Approximation of Navier-Stokes Equations in a Two Dimensional Domain,* Model Math. Anal. Numer. (M^2AN)21(1987), 93-123

[135] Malik, S.V., *Stability of the interfacial flows,* Technical Report, University of the West of England, Bristol-England, 2003

[136] Markiewicz, D., *Survey on Symplectic Integrators,* Preprint Univ. California at Berkeley, Spring 1999

[137] Matano, H., *Convergence of solutions of one-dimensional semilinear parabolic equations,* J. Math. Kyoto Univ. (JMKYAZ) 18-2(1978), 221-227

[138] McFaden, G.B., Murray, B.T., Boisvert, R. F., *Elimination of Spurious Eigenvalues in the Chebyshev Tau Spectral Methods,* J. Comput. Phys., 91(1990) 228-239

[139] McLachlan, R., *Symplectic integration of Hamiltonian wave equations,* Numer. Math., 66(1994), 465-492

[140] Mead, J. L., Renaut, R. A., *Optimal Runge-Kutta Methods for First Order Pseudospectral Operators,* J. Comput. Phys., 152(1999), 404-419

[141] Melenk, J.M., Kirchner, N.P., Schwab, C., *Spectral Galerkin Discretization for Hydrodynamic Stability Problems,* Computing 65(2000), 97-118

[142] Miele, A., Aggarwal, A. K., Tietze, J. L., *Solution of Two-Point Boundary-Value Problems with Jacobian Matrix Characterized by Large Positive Eigenvalues,* J. Comput. Phys., 15(1974), 117-133

[143] Mitchell, A.R., Murray, B.A., Sleeman, B.D., *Numerical Solution of Hamiltonian Systems in Reaction-Diffusion by Symplectic Difference Schemes,* J. Comput. Phys., 95(1991), 339-358

[144] Miura, R., *The Korteweg-De Vries Equation: A Survey of Results,* SIAM Rev., 18(1976), 412-459

[145] Monro, D. M., *Interpolation by Fast Fourier and Chebyshev Transforms,* Int. J. for Num. Met. in Engineering, 14(1979) 1679-1692

[146] Moser, J., *Recent developments in the theory of Hamiltonian systems,* SIAM Rev., 28(1986), 459-485

[147] Murty, V. N., *Best Approximation with Chebyshev Polynomials,* SIAM J. Numer. Anal., 8(1971), 717-721

[148] Necas, J., *Sur une methode pour resoudre les equations aux derivee partielles du type elliptiques, voisine de la variationelle,* Ann. Sc. Norm. Sup. Pisa 16(1962), 305-326

[149] Nield, D. A., *Odd-Even Factorization Results for Eigenvalue Problems,* SIAM Rev., 36(1994), 649-651

[150] Nikolsky, S.M., *A Course of Mathematical Analysis,* Mir Publishers Moscow, 1981.

[151] Omelyan, I. P., Mryglod, I. M., Folk, R., *Molecular dynamics simulations of spin and pure liquids with preservation of all the conservation laws,* Phys. Rev. E 64, 016105(2001)

[152] Omelyan, I. P., Mryglod, I. M., Folk, R., *Construction of high-order force-gradient algorithms for integration of motion in classical and quantum systems,* Phys. Rev. E 66, 026701(2002)

[153] Omelyan, I. P., Mryglod, I. M., Folk, R., *Optimized Verlet-like algorithms for molecular dynamics simulations,* Phys. Rev. E 65, 056706(2002)

[154] O'Neil, P.V., *Advanced Engineering Mathematics,* International Student Edition, Chapman & Hall, 3rd. edition, 1991

[155] Orszag, S., *Accurate solutions of the Orr-Sommerfeld stability equation,* J. Fluid Mech., 50(1971), 689-703

[156] Orszag, S., *Comparison of Pseudospectral and Spectral Approximations,* Studies in Appl. Math. 51(1971), 253-259

[157] Ortiz, E.L., *The Tau method,* SIAM J. Numer. Anal., 6(1969), 480-492

[158] Ortiz, E.L., Samara, H., *An Operational Approach to the Tau Method for the Numerical Solution of Non-Linear Differential Equations,* Computing, 27(1981), 15-25

[159] OuldKaber, S.M., *A Legendre Pseudospectral Viscosity Method,* J. Comput. Phys., 128(1996), 165-180

[160] Parter, S. V., Rothman, E. E., *Preconditioning Spectral Collocation Approximations to Elliptic Problems,* SIAM J. Numer. Anal. 32(1995) 333-385

[161] Petrila, T., Trif, D., *Basics of Fluid Mechanics and Introduction to Computational Fluid Dynamics,* Kluwer Academic Publishers, Boston/Dordrecht/London, 2004

[162] Pichmony Anhaouy, *Fourier Spectral Methods for Solving the Korteweg-de Vries Equation,* Thesis, Simon Fraser University, 2000

[163] Pop, I.S., *Numerical Approximation of Differential Equations by Spectral Methods,* Technical report, "Babes-Bolyai" University, Cluj-Napoca, 1995 (in Romanian)

[164] Pop, I.S., Gheorghiu, C.I., *A Chebyshev-Galerkin Method for Fourth Order Problems,* Proceedings of I C A O R, vol. II, pp.217-220, 1997 MR 98g:41002

[165] Pop, I.S., *A stabilized approach for the Chebyshev-tau method,* Studia Univ. "Babes-Bolyai", Mathematica, 42(1997), 67-79

[166] Pruess, S., Fulton, C. T., *Mathematical Software for Sturm-Liouville Problems,* A C M Trans. Math. Softw. 19(1993), 360-376

[167] Pryce, J. D., *A Test Package for Sturm-Liouville Solvers,* A C M Trans. on Math. Software, 25(1999), 21-57

[168] Qiu, Y., Sloan, D. M., *Numerical Solution of Fischer's Equation Using a Moving Mesh Method,* J. Comput. Phys., 146(1998), 726-746

[169] Quarteroni, A., Valli, A., *Numerical Approximation of Partial Differential Equations,* Springer Verlag, Berlin/Heidelberg, 1994

[170] Quarteroni, A., Saleri, F., *Scientific Computing with MATLAB and Octave,* Second Ed., Springer, 2006

[171] Raltson, A., Rabinowitz, Ph., *A First Course in Numerical Analysis,* McGraw Hill, 1978

[172] Roos, H.G., Pfeiffer, E., *A Convergence Result for the Tau Method,* Computing 42(1989), 81-84

[173] Shampine, L. F., Reichelt, M. W., *The MATLAB O D E Suite,* SIAM J. Sci. Comput., 18(1997), 1-22

[174] Shampine, L. F., *Design of software for ODE,* J. Comput. Appl. Math., 205(2007),

[175] Shen, J., *Efficient spectral-Galerkin method I, Direct solvers of second and fourth equations using Legendre polynomials,* SIAM J. Sci. Stat. Comput., 15(1994), 1489-1505

[176] Shen, J., *Efficient Spectral-Galerkin Method II. Direct Solvers of Second and Fourth Order Equations by Using Chebyshev Polynomials,* SIAM J. Sci. Comput., 16(1995), 74-87

[177] Shen, J., *Efficient Chebyshev-Legendre Galerkin methods for elliptic problems,* Proceedings of ICOSAHOM'95, Houston J. Math. (1996) 233-239

[178] Shen, J., Temam, R., *Nonlinear Galerkin Method Using Chebyshev and Legendre Polynomials I. The One-Dimensional Case,* SIAM J. Numer. Anal. 32(1995) 215-234

[179] Shkalikov, A.A., *Spectral portrait of the Orr-Sommerfeld operator with large Reynolds numbers,* arXiv:math-ph/0304030v1, 22Apr2003

[180] Shkalikov, A.A., *Spectral portrait and the resolvent growth of a model problem associated with the Orr-Sommerfeld equation,* arXiv:math.FA/0306342v1, 24Jun2003

[181] Simmons, G. F., *Differential Equations with Applications and Historical Notes,* McGraw-Hill Book Company, 1972

[182] Sloan, D.M., *On the norms of inverses of pseudospectral differential matrices,* SIAM J. Numer. Anal., 42(2004), 30-48

[183] Solomonoff, A., Turkel, E., *Global Properties of Pseudospectral Methods,* J. Comput. Phys., 81(1989) 239-276

[184] Solomonoff, A., *A Fast Algorithm for Spectral Differentiation,* J. Comput. Phys., 98(1992), 174-177

[185] Stenger, F., *Numerical Methods Based on Whittaker Cardinal, or Sinc Functions,* SIAM Rev. 23(1981), 165-224

[186] Stenger, F., *Summary of Sinc numerical methods,* J. Comput. Appl. Math., 121(2000), 379-420

[187] Stoer, J., Bulirsch, R., *Introduction to Numerical Analysis,* Springer Verlag, New York, Heidelberg, Berlin, 1980

[188] B. J. Stone, *Best possible ratios of certain matrix norms,* Numer. Math. 4,114-116(1962)

[189] Tadmor, E., *The Exponential Accuracy of Fourier and Chebyshev Differencing Methods,* SIAM J. Numer. Anal., 23(1986), 1-10

[190] Tadmor, E., *Stability Analysis of Finite-Difference, Pseudospectral and Fourier-Galerkin Approximations for Time-Dependent Problems,* SIAM Rev., 29(1987), 525-555

[191] Tal-Ezer, H., *A Pseudospectral Legendre Method for Hyperbolic Equations with an Improved Stability Condition,* J. Comput. Phys., 67(1986), 175-172

[192] Tang, T., *The Hermite spectral method for Gaussian-type functions,* SIAM J. Sci. Comput. 14(1993), 594-606

[193] Trefethen, L. N., *Pseudospectra of linear operators,* SIAM Review, 39(1997)

[194] Trefethen, L. N., *Computation of Pseudospectra,* Acta Numerica, 247-295(1999)

[195] Trefethen, L. N., Reichel, L., *Eigenvalues and Pseudoeigenvalues of Toeplitz Matrices,* Linear Algebra Appl., 162-164:153-158(1992)

[196] Trefethen, L. N., Trummer, M. R., *An Instability Phenomenon in Spectral Methods,* SIAM J. Numer. Anal. 24(1987), 1008-1023.

[197] Trefethen, L. N., *Spectral Methods in MATLAB,* SIAM, Philadelphia, PA, 2000

[198] Trefethen, L. N., Embree, M., *Spectra and Pseudospectra; The Behavior of Nonnormal Matrices,* Princeton University Press, Princeton and Oxford, 2005

[199] Tretter, Ch., *A Linearization for a Class of $\lambda-$Nonlinear Boundary Eigenvalue Problems*, J. Mathematical Analysis and Applications, 247(2000), 331-355

[200] van Saarloos, W., *The Complex Ginzburg-Landau equation for beginners,* in Spatio-temporal Patterns in Nonequilibrium Complex Systems, eds. P. E. Cladis and P. Palffy-Muhoray, Santa Fe Institute, Studies in the Science of Complexity, proceedings XXI, Addison-Wesley, Reading, 1994

[201] Varga, R. S., *On Higher Order Stable Implicit Methods for Solving Parabolic Partial Differential Equations*, J. Math. and Phys., XL (1961), 220-231

[202] Venakides, S., *Focusing Nonlinear Schroedinger equation: Rigorous Semiclassical Asymptotics,* http://www.iacm.forth.gr/anogia05/Docs/venakides_material.pdf

[203] Weideman, J.A.C., Trefethen, L.N., *The Eigenvalues of Second -Order Spectral Differentiations Matrices,* SIAM J. Numer. Anal. 25(1988), 1279-1298

[204] Weideman, J.A.C., Reddy, S. C., *A MATLAB Differentiation Matrix Suite,* ACM Trans. on Math. Software, 26(2000), 465-519

[205] Weinan, E., *Convergence of spectral methods for Burger's equation,* SIAM J. Numer. Anal., 29(1992), 1520-1541

[206] Welfert, B. D., *Generation of pseudospectral differentiation matrices,* SIAM J. Numer. Anal., 34(1997), 1640-1657

[207] Wright, T. G., Trefethen, L. N., *Large-Scale Computation of Pseudospectra Using ARPACK and EIGS,* SIAM J. Sci. Comput., 23(2001), 591-605

[208] Wright, T. G., *http://www.comlab.ox.ac.uk/oucl/work/tom.wright/psgui/*, University of Oxford, 2000

[209] Wu, X., Kong, W., Li, C., *Sinc collocation method with boundary treatment for two-point boundary value problem,* J. Comput. Appl. Math., 196(2006), 229-240

[210] Zebib, A., *A Chebyshev Method for the Solution of Boundary Value Problems,* J. Comput. Phys., 53(1984), 443-455

[211] Zebib, A., *Removal of Spurious Modes Encountered in Solving Stability Problems by Spectral Methods,* J. Comput. Phys., 70(1987), 521-525

Index