



Sinc-Galerkin solution of second-order hyperbolic problems in multiple space dimensions
by Kelly Marie McArthur

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in
Mathematics

Montana State University

© Copyright by Kelly Marie McArthur (1987)

Abstract:

A fully Galerkin method in space and time is developed for the second-order hyperbolic problem in one, two, and three space dimensions. Using sinc basis functions and the sinc quadrature rule, the discrete system arising from the orthogonalization of the residual is easily assembled. (Two equivalent matrix formulations of the systems are given. One lends itself to scalar computation while the other is more natural in a vector computing environment. In fact, it is shown that passing from one to the other is simply a notational change. In either setting the move from one to two or three space dimensions does not significantly affect the ease of implementation. Intermediate diagonalization of each matrix representing the discretization of a second partial leads to the diagonalization of the overall system.

The method was tested on an extensive class of problems. Numerical results indicate the method has an exponential convergence rate for analytic and singular problems. Moreover that is independent of the spatial dimension.

SINC-GALERKIN SOLUTION OF SECOND-ORDER HYPERBOLIC
PROBLEMS IN MULTIPLE SPACE DIMENSIONS

by

Kelly Marie McArthur

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Mathematics

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 1987

D 378
M1195
Cop. 2

ii

APPROVAL

of a thesis submitted by

Kelly Marie McArthur

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

May 20, 1987
Date

Kenneth L. Bowers
Chairperson,
Graduate Committee

Approved for the Major Department

May 20, 1987
Date

K. F. Lisk
Head, Mathematical
Sciences Department

Approved for the College of Graduate Studies

5-20-87
Date

MB Maher
Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this thesis is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this thesis should be referred to University Microfilms International, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute copies of the dissertation in and from microfilm and the right to reproduce and distribute by abstract in any format."

Signature

Kelly Marie McArthur

Date

May 21, 1987

ACKNOWLEDGMENTS

The author thankfully acknowledges the superb typing of Ms. Rene' Tritz and the invaluable computer assistance of Eric Greenwade. Thanks also go to Professor Norman Eggert for his careful reading of this work. In addition, the author recognizes Professor Frank Stenger for the revival and extension of sinc function theory. His clear exposes and willingness to discuss new problems have led to yet another generation of sinc function students. Finally, the author thanks Professor Kenneth L. Bowers and Professor John Lund, whose guidance is responsible for this work. They possess the rare ability to teach one to teach oneself.

TABLE OF CONTENTS

	Page
1. INTRODUCTION.....	1
2. THE SINC-GALERKIN SOLUTION OF ORDINARY DIFFERENTIAL EQUATIONS.....	8
Interpolation on $(-\infty, \infty)$	9
Quadrature on $(-\infty, \infty)$	18
Interpolation and Quadrature on Alternative Arcs.....	22
Sinc-Galerkin Method.....	26
3. THE SINC-GALERKIN METHOD FOR THE WAVE EQUATION.....	35
4. SOLUTION OF THE DISCRETE SINC-GALERKIN SYSTEM...	56
Classical Methods.....	56
Solution of the Discrete System in One Spatial Variable.....	61
Solution of the Systems in Two and Three Spatial Variables.....	65
5. NUMERICAL EXAMPLES OF THE SINC-GALERKIN METHOD..	71
Exponentially Damped Sine Waves.....	78
Singular Problems.....	82
Problems Dictating Noncentered Sums in All Indices.....	86
REFERENCES CITED.....	91

LIST OF TABLES

Table		Page
1.	Machine Storage Information for the Matrices $B^{(j)}$ and $B_D^{(j)}$	58
2.	Numerical Results for Example 5.1.....	79
3.	Numerical Results for Example 5.2.....	80
4.	Numerical Results for Example 5.3.....	81
5.	Numerical Results for the Damped Sine Wave.....	81
6.	Numerical Results for Example 5.4.....	82
7.	Numerical Results for Example 5.5.....	84
8.	Numerical Results for Example 5.6.....	85
9.	Numerical Results for the Singular Problems....	85
10.	Numerical Results for Example 5.7.....	86
11.	Numerical Results for Example 5.8.....	88
12.	Numerical Results for Example 5.9.....	89
13.	Numerical Results for Problems Dictating Noncentered Sums in All Indices.....	89

LIST OF FIGURES

Figure		Page
1.	The Domain of Dependence.....	3
2.	The Numerical Domain of Dependence for the Explicit, Centered Finite Difference Method.....	4
3.	$S(0,1)(x) = \text{sinc}(x)$, $x \in \mathbb{R}$	10
4.	The Domain D_S	14
5.	The Conformal Map χ	23
6.	The Regions D_E , D_W , and D_S	38
7.	The Basis Functions $S_0(x)$ and $S_0^*(t)$	39

ABSTRACT

A fully Galerkin method in space and time is developed for the second-order hyperbolic problem in one, two, and three space dimensions. Using sinc basis functions and the sinc quadrature rule, the discrete system arising from the orthogonalization of the residual is easily assembled. Two equivalent matrix formulations of the systems are given. One lends itself to scalar computation while the other is more natural in a vector computing environment. In fact, it is shown that passing from one to the other is simply a notational change. In either setting the move from one to two or three space dimensions does not significantly affect the ease of implementation. Intermediate diagonalization of each matrix representing the discretization of a second partial leads to the diagonalization of the overall system.

The method was tested on an extensive class of problems. Numerical results indicate the method has an exponential convergence rate for analytic and singular problems. Moreover that is independent of the spatial dimension.

CHAPTER 1

INTRODUCTION

The general second-order, linear partial differential equation

$$(1.1) \quad Au_{xx} + Bu_{xy} + Cu_{yy} + Du_x + Eu_y + Fu = G,$$

where A, B, C, D, E, F , and G are functions of x and y only, is classified at the point (x,y) via the discriminate $(B^2 - 4AC)(x,y)$. That is, when $(B^2 - 4AC)(x,y)$ is positive, zero, or negative the equation at (x,y) is hyperbolic, parabolic, or elliptic, respectively. The present chapter reviews classical results for a purely hyperbolic, constant coefficient problem corresponding to $A = -1, C = 1$, and $B = D = E = F = 0$. Including boundary and initial conditions, the specific partial differential equation examined is

$$(1.2) \quad \begin{aligned} u_{tt}(x,t) - u_{xx}(x,t) &= G(x,t) & (x,t) &\in (0,1) \times (0,\infty) \\ u(0,t) &= u(1,t) = 0 & t &\geq 0 \\ u(x,0) &= f(x) & 0 \leq x &\leq 1 \\ u_t(x,0) &= g(x) & 0 \leq x &\leq 1. \end{aligned}$$

This problem is referred to as a one-dimensional wave equation since it models the displacement of a vibrating string with initial displacement f , initial velocity g , and subject to an external force G .

The equation in (1.2) is representative of one of the two hyperbolic canonical forms due to the absence of the term u_{xt} . In contrast, the distinguishing feature of the alternative canonical form is that its only second-order term is the mixed partial u_{xt} . The change of variables $\xi = x + t$ and $\eta = x - t$ transforms the equation in (1.2) to

$$(1.3) \quad -4u_{\xi\eta} = G((\xi + \eta)/2, (\xi - \eta)/2) \\ (\xi, \eta) \in (0, \infty) \times (-\infty, 1)$$

It is well-known that the correct change of variables is found by solving two ordinary differential equations which depend on the discriminate. For a complete discussion see Farlow [1]. By integrating (1.3), restoring the variables x and t , and applying the initial and boundary conditions, d'Alembert explicitly solved (1.2) [2]. His result of 1747 is

$$(1.4) \quad u(x, t) = \frac{1}{2} \left\{ f(x + t) + f(x - t) + \int_{x-t}^{x+t} g(\alpha) d\alpha \right. \\ \left. + \int_0^t \int_{x-(t-\beta)}^{x+(t-\beta)} G(\alpha, \beta) d\alpha d\beta \right\}$$

where f , g , and G are extended as odd periodic functions when necessary.

The solution (1.4) is the sum of homogeneous and particular solutions, $u^H(x, t)$ and $u^P(x, t)$, respectively, where

$$(1.5) \quad u^H(x, t) = \frac{1}{2} \left\{ f(x + t) + f(x - t) + \int_{x-t}^{x+t} g(\alpha) d\alpha \right\}$$

and

$$(1.6) \quad u^P(x, t) = \frac{1}{2} \int_0^t \int_{x-(t-\beta)}^{x+(t-\beta)} G(\alpha, \beta) d\alpha d\beta.$$

From the form of the homogeneous solution, it is apparent that u^H at a fixed point (x_0, t_0) depends only on the initial conditions over the interval $[x_0 - t_0, x_0 + t_0]$ on the x -axis. As a result $[x_0 - t_0, x_0 + t_0]$ is called the interval of dependence. The particular solution may be used to find the domain of dependence shown by the shaded region in Figure 1 below. The curves representing the change of variables ξ and η are called characteristics of (1.2). Notice that the characteristics cut the interval of dependence out of the x -axis and define the domain of dependence.

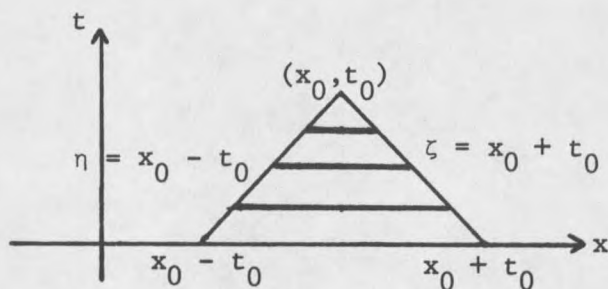


Figure 1. The Domain of Dependence.

The significance of the domain of dependence is the constraint it imposes on schemes used to numerically solve the one-dimensional wave equation (1.2). This constraint is

called the Courant-Friedrichs-Lewy (or CFL) condition [3]. A scheme which provides a ready illustration of the condition is the centered finite difference method defined explicitly by the difference equation

$$(1.7) \quad U_{i,j} = m^2(U_{i+1,j-1} + U_{i-1,j-1}) + 2(1 - m^2)U_{i,j-1} - U_{i,j-2} + (\Delta t)^2 G(x_i, t_{j-1}) .$$

Here $U_{i,j}$ approximates u at $(x_i, t_j) = (i\Delta x, j\Delta t)$ and $m = \Delta t/\Delta x$ is the ratio of the stepsizes. A detailed derivation of the time-marching scheme (1.7) is included in Ames [4]. Stepping back in time determines the finite difference gridpoints which influence $U_{i,j}$ (see Figure 2 below). These gridpoints blanket the numerical domain of dependence outlined in Figure 2. The CFL criteria states that a necessary condition for the convergence of (1.7) (as Δt and $\Delta x \rightarrow 0$) is that the numerical domain of dependence must contain the analytic domain of dependence. This requires $m \leq 1$. The computational impact of restricting $m \leq 1$ is briefly discussed in Chapter 4.

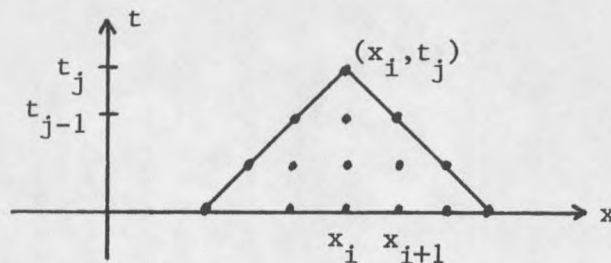


Figure 2. The Numerical Domain of Dependence for the Explicit, Centered Finite Difference Method.

The method of the present work satisfies the CFL condition trivially. This technique, called the Sinc-Galerkin method, builds an approximate solution for (1.2) valid on the entire domain. The approximate is a truncated generalized Fourier expansion of basis elements which are tensor products of sinc functions composed with suitable conformal maps. The support of each basis element is $(0,1) \times (0,\infty)$ hence, the method may be termed a spectral method and its numerical domain of dependence is identically the domain of the partial differential equation.

To ease the description of the Sinc-Galerkin method applied to (1.2), its counterpart for ordinary differential equations is derived in Chapter 2. Here the basis elements are single sinc functions composed with conformal maps. The pertinent sinc function properties needed to construct an approximate solution are reviewed. Further, it is shown that when $2N + 1$ basis functions are used to define this approximate, the optimal exponential order of convergence, $O(e^{-K\sqrt{N}})$, $K > 0$, occurs even in the presence of singularities. Stenger [5] discusses a more general setting than that considered in Chapter 2. The chapter closes with the formulation of the discrete linear system whose solution specifies the approximate. This system is symmetric, a result that Lund [6] has shown depends on the correct choice of weighted inner product.

Chapter 3 extends the Sinc-Galerkin method to (1.2) (with $f = g = 0$) and then further to the analogous wave equations in two and three space dimensions. At this writing, the most common procedure for solving partial differential equations with a semi-infinite time interval is a Galerkin discretization of the spatial domain with time dependent coefficients. The result is a system of ordinary differential equations usually solved via difference techniques. Botha and Pinder [3] develop Galerkin discretizations of space using finite element basis functions. In contrast, Gottlieb and Orszag [7] use globally defined spatial basis elements. They show that in space many of these spectral methods exhibit an exponential convergence rate. However, because Gottlieb and Orszag consider only finite difference techniques for the temporal domain, the time solution has finite-order accuracy. They acknowledge the incompatibility of the error statement in time versus space with the following remark:

No efficient, infinite-order accurate time-differencing methods for variable coefficient problems are yet known. The current state-of-the-art of time-integration techniques for spectral methods is far from satisfactory on both theoretical and practical grounds...[7].

The point of view taken here differs from the two sources just cited by carrying the Galerkin discretization into the time domain. Chapter 5 reports numerical results which attest to the success of this notion.

Besides developing the Sinc-Galerkin method, Chapter 3 introduces notation to facilitate the description of the resulting discrete systems. These systems are posed in two algebraically equivalent matrix forms. The choice of form to use depends somewhat on available computing facilities. Chapter 4 discusses this topic along with algorithms for the solution of the linear systems in either form.

The nine examples included in Chapter 5 are broken into groups of three. Each group highlights a feature of the Sinc-Galerkin method for problems in one, two, and three space dimensions. For instance, the first three examples have analytic solutions while the second three have combinations of both algebraic and logarithmic singularities. The numerical results show that the rate of convergence is not affected by this singular behavior. The last three examples show the dramatic reduction in the size of the discrete system solved when care is exercised in parameter selections. Finally, each group indicates that the asymptotic error $O(e^{-K\sqrt{N}})$, $K > 0$, is attained independent of the dimension of the wave equation.

CHAPTER 2

THE SINC-GALERKIN SOLUTION OF ORDINARY
DIFFERENTIAL EQUATIONS

The goal of this chapter is to derive the discrete Sinc-Galerkin system necessary to build an approximate to the solution of

$$\begin{aligned} (2.1) \quad Lf(x) &\equiv f''(x) + v(x)f(x) = \sigma(x) \quad a < x < b \\ f(a) &= f(b) = 0 \end{aligned}$$

valid on the interval (a,b) . A symmetric matrix formulation for the system can be posed and is, in fact, easy to solve numerically. The resulting approximate solution converges to the true solution on (a,b) at the rate $O(e^{-K\sqrt{N}})$ where K is a positive constant and $2N + 1$ basis functions are used to build the approximate. Further, the convergence rate is maintained in the presence of singularities (the solution has an unbounded derivative) on the boundary. To prove these statements a background in general sinc function theory is necessary. In particular, the foundation for the error analysis of the Sinc-Galerkin method is the error associated with the truncated sinc quadrature rule.

Interpolation on $(-\infty, \infty)$

Numerical sinc function methods are rooted in E.T. Whittaker's [8] work concerning interpolation of a function at the integers. Rather than using well-known expansions based on polynomials, Whittaker sought an expansion whose properties are far more distinguished when applied in the proper setting. The foundation of the series is the sinc function

$$(2.2) \quad \text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}, \quad x \in \mathbb{R}$$

shown in Figure 3 below. The resulting formal cardinal series for a function f is

$$(2.3) \quad \sum_{k=-\infty}^{\infty} f(k) \text{sinc}(x - k)$$

To generalize (2.2) and (2.3) to handle interpolation on any evenly spaced grid define for $h > 0$

$$(2.4) \quad S(k, h)(x) = \text{sinc}\left(\frac{x - kh}{h}\right)$$

and denote the Whittaker cardinal function by

$$(2.5) \quad C(f, h)(x) = \sum_{k=-\infty}^{\infty} f(kh) S(k, h)(x)$$

whenever this series converges. In engineering literature (2.5) is often called a band-limited series.

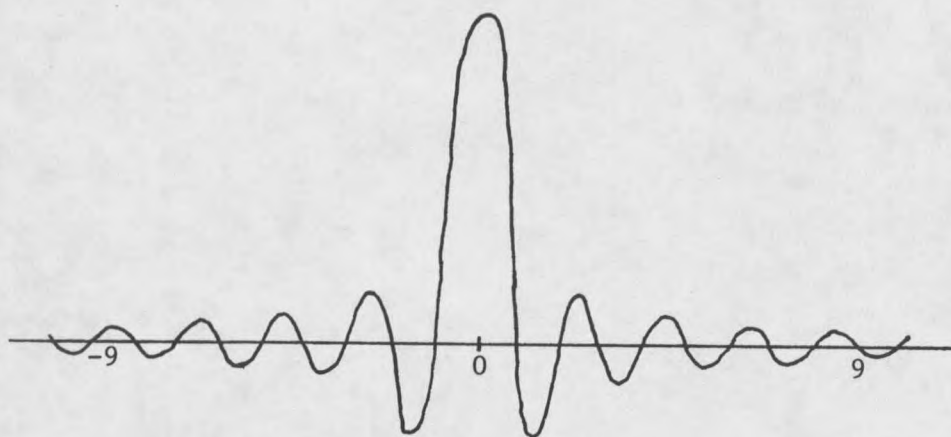


Figure 3: $S(0,1)(x) = \text{sinc}(x)$, $x \in \mathbb{R}$.

With regard to using (2.5) as an approximation tool, two important classes of functions must be identified. The first is the very restricted class for which (2.5) is exact. The second is the class of functions f for which the difference between f and $C(f,h)$ is small. J.M. Whittaker [9] and McNamee, et.al. [10] accomplish this identification task by displaying a natural link between the Whittaker series and aspects of Fourier series and integrals.

The Fourier transform for a function g is

$$(2.6) \quad \hat{g}(x) = \int_{-\infty}^{\infty} g(t)e^{ixt}dt .$$

A fundamental result of Fourier analysis is that if $g \in L^2(\mathbb{R})$ then $\hat{g} \in L^2(\mathbb{R})$ and g is recovered from $\hat{g}$ by the Fourier inversion integral

$$(2.7) \quad g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{g}(x) e^{-ixt} dx.$$

For a select set of functions, the Paley-Wiener Theorem shows that the inverse transform has compact support. Their result is

Theorem (2.8): If $g \in L^2(\mathbb{R})$, entire, and there exist positive constants A and C so that $|g(w)| \leq C \exp\{A|w|\}$ where $w \in \mathbb{C}$; then

$$(2.9) \quad g(w) = \frac{1}{2\pi} \int_{-A}^A \hat{g}(x) e^{-ixw} dx.$$

Showing that the sinc function satisfies the hypotheses of Theorem (2.8) with $A = \pi$ and $C = 1$ is straightforward. An elementary calculation gives

$$(2.10) \quad \text{sinc}(w) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ixw} dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi_{(-\pi, \pi)} e^{-ixw} dx;$$

hence, an immediate consequence of Theorem (2.8) is

$$(2.11) \quad \widehat{\text{sinc}}(x) = \int_{-\infty}^{\infty} \text{sinc}(t) e^{ixt} dt = \chi_{(-\pi, \pi)}(x).$$

To accommodate the translated sinc function appearing in (2.4), a change of variables in (2.10) gives

$$(2.12) \quad \widehat{S(k, h)}(x) = \int_{-\infty}^{\infty} \text{sinc}\left(\frac{t - \xi}{h}\right) e^{ixt} dt$$

$$= h e^{ix\xi} \chi_{(-\pi/h, \pi/h)}(x)$$

for fixed real ξ .

The support of the characteristic function in (2.12) prompts the definition of a class of functions, called the Paley-Wiener class, which is naturally associated with the Whittaker series.

Definition (2.13): Let $B(h)$ be the set of functions g such that $g \in L^2(\mathbb{R})$, g is entire, and $|g(w)| \leq C \exp\{\pi|w|/h\}$ where $w \in \mathbb{C}$.

Before the Whittaker function can be discussed, one result is vital.

Theorem (2.14): If $g \in B(h)$ then $g(w) = \frac{1}{h} \int_{-\infty}^{\infty} g(t) \operatorname{sinc}\left(\frac{t-w}{h}\right) dt$.

A proof of Theorem (2.14) is found in [10]. For completeness the converse of Theorem (2.14) is

Theorem (2.15): If $g \in L^2(\mathbb{R})$ then $k(w) = \frac{1}{h} \int_{-\infty}^{\infty} g(t) \operatorname{sinc}\left(\frac{t-w}{h}\right) dt$ is in $B(h)$.

Proof: Using Parseval's Theorem with the inner product

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(t) \overline{g(t)} dt \quad \text{establishes the growth estimate}$$

for k . Application of Morera's Theorem proves entirety and the Cauchy-Schwarz inequality yields $k \in L^2(\mathbb{R})$.

The significance of all the preceeding work is in the subsequent elegant theorem.

Theorem (2.16): If $g \in B(h)$ then $g(w) = \sum_{k=-\infty}^{\infty} \alpha_k S(k,h)(w)$

where $\alpha_k = \frac{1}{h} \int_{-\infty}^{\infty} g(t) \operatorname{sinc}\left(\frac{t - kh}{h}\right) dt = g(kh).$

Proof: The Paley-Wiener Theorem, the identity

$$(2.17) \quad e^{-ixw} = \frac{h}{\pi} \sin\left(\frac{\pi w}{h}\right) \sum_{k=-\infty}^{\infty} (-1)^k \frac{e^{ikhx}}{w+kh}$$

and the uniform convergence theorem justify the ensuing steps:

$$\begin{aligned} g(w) &= \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} \hat{g}(x) e^{-ixw} dx \\ &= \frac{h}{2\pi^2} \sin\left(\frac{\pi w}{h}\right) \int_{-\pi/h}^{\pi/h} \hat{g}(x) \left\{ \sum_{k=-\infty}^{\infty} (-1)^k \frac{e^{ikhx}}{w+kh} \right\} dx \\ &= \frac{h}{\pi} \sum_{k=-\infty}^{\infty} \frac{(-1)^k \sin(\pi w/h)}{w + kh} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{g}(x) e^{ikhx} dx \right\} \\ &= \frac{h}{\pi} \sum_{k=-\infty}^{\infty} \frac{(-1)^k \sin(\pi w/h)}{w + kh} g(-kh) \\ &= \sum_{k=-\infty}^{\infty} g(kh) \operatorname{sinc}\left(\frac{w - kh}{h}\right). \end{aligned}$$

Hence for $f \in B(h)$, $f(w) = C(f,h)(w)$ for all $w \in \mathbb{C}$. Note that this is a stronger result than originally sought. The initial quest was for a class of functions such that the Whittaker series was exact on $\mathbb{R}$. An even stronger statement is derived from Theorem (2.16) and the identity

$$(2.18) \quad \frac{1}{h} \int_{-\infty}^{\infty} S(k,h)(t) S(\lambda,h)(t) dt = \delta_{\lambda k}^{(0)} = \begin{cases} 1, & \text{if } \lambda = k \\ 0, & \text{if } \lambda \neq k \end{cases};$$

that is,

Theorem (2.19): The set $\left\{ \frac{1}{\sqrt{h}} S(k,h) \right\}_{k=-\infty}^{\infty}$ is a complete orthonormal set in $B(h)$ [10].

Unfortunately the set $B(h)$ is extremely restrictive and some relaxation is necessary if (2.5) is used as a practical interpolatory tool. McNamee, et. al. [10] identifies a set, here called $B^p(D_S)$, where $C(f,h)$ is not exact but its approximation to f is very good. In particular, the domain of analyticity for $B^p(D_S)$ is D_S .

Definition (2.20): $D_S = \{z: z = x + iy, |y| < d \in \mathbb{R}, d > 0\}$

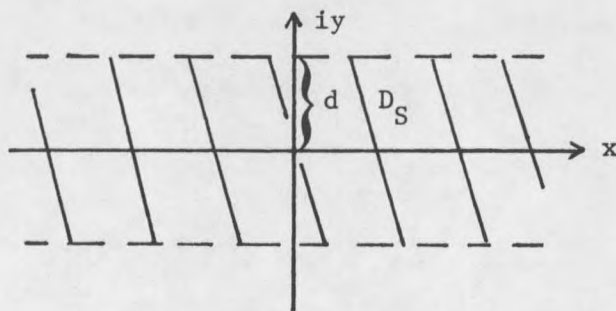


Figure 4: The Domain D_S .

The class $B^p(D_S)$ is specified by Definition (2.21).

Definition (2.21): $B^p(D_S)$ is the set of functions f such that

(2.22) f is analytic in D_S ,

(2.23) $\int_{-d}^d |f(t+iy)| dy = O(|t|^a)$ as $t \rightarrow \pm\infty$ where $0 \leq a < 1$,

and for $p = 1$ or 2

$$(2.24) \quad N^p(f, D_S) = \lim_{y \rightarrow d} \left[\left\{ \int_{-\infty}^{\infty} |f(t+iy)|^p dt \right\}^{1/p} + \left\{ \int_{-\infty}^{\infty} |f(t-iy)|^p dt \right\}^{1/p} \right] < \infty.$$

The exact form of the error is given by Theorem (2.25).

Theorem (2.25): If $f \in B^p(D_S)$ then $\epsilon(f)(x) = f(x) - C(f, h)(x)$ where

$$(2.26) \quad \epsilon(f)(x) = \frac{\sin(\pi x/h)}{2\pi i} \int_{-\infty}^{\infty} \left\{ \frac{f(t-id)}{(t-x-id)\sin[\pi(t-id)/h]} - \frac{f(t+id)}{(t-x+id)\sin[\pi(t+id)/h]} \right\} dt.$$

Moreover, if $f \in B^1(D_S) \equiv B(D_S)$ then

$$(2.27) \quad \|\epsilon(f)\|_{\infty} \leq \frac{N^1(f, D_S)}{2\pi d \sinh(\pi d/h)} \equiv \frac{N(f, D_S)}{2\pi d \sinh(\pi d/h)}$$

while if $f \in B^2(D_S)$ then

$$(2.28) \quad \|\varepsilon(f)\|_{\infty} \leq \frac{N^2(f, D_S)}{2\sqrt{\pi d} \sinh(\pi d/h)}$$

and

$$(2.29) \quad \|\varepsilon(f)\|_2 \leq \frac{N^2(f, D_S)}{\sinh(\pi d/h)}.$$

For a proof see Stenger [11]. Worth noting, is that the error statement of Theorem (2.25) is valid only on the real line. Theorem (2.16), recall, applies to the complex plane. However, the original goal was approximating on the real line and Theorem (2.25) certainly satisfies that.

Of far greater interest is the order statement derived from (2.27), (2.28), and (2.29). As $h \rightarrow 0$ $\sinh(\pi d/h) \rightarrow \infty$; hence, $\|\varepsilon(f)\|_{\infty, 2} \rightarrow 0$ independent of whether $f \in B^1(D_S)$ or $B^2(D_S)$. Moreover, the rate of convergence is governed by $\sinh(\pi d/h)$; i.e., $1/\sinh(\pi d/h) = O(\exp(-\pi d/h))$ as $h \rightarrow 0$.

Although the exponential convergence rate is attractive, to be of practical importance it must be maintained when (2.5) is truncated. Denote the truncated Whittaker series for a function f by

$$(2.30) \quad C_{M,N}(f,h)(x) = \sum_{k=-M}^N f(kh) \operatorname{sinc}\left(\frac{x - kh}{h}\right)$$

where it is assumed $C(f,h)(x)$ converges.

Theorem (2.31): If $f \in B^p(D_S)$ for $p = 1$ or 2 , $d > 0$ and there exist positive constants α and β such that

$$(2.32) \quad |f(x)| \leq L \begin{cases} e^{\alpha x} & \text{if } x \in (-\infty, 0) \\ e^{-\beta x} & \text{if } x \in [0, \infty) \end{cases}$$

then choosing

$$(2.33) \quad N = \left\lceil \frac{\alpha}{\beta} M + 1 \right\rceil$$

and

$$(2.34) \quad h = (\pi d / (\alpha M))^{\frac{1}{2}}$$

gives

$$(2.35) \quad \|f - C_{M,N}(f,h)\|_{\infty} \leq C M^{\frac{1}{2}} e^{-(\pi d \alpha M)^{\frac{1}{2}}}$$

where C is a constant dependent on f .

Proof: From Theorem (2.25) there exists a constant L_1 such that $|f(x) - C(f,h)(x)| \leq L_1 e^{-\pi d/h}$ for all $x \in \mathbb{R}$. Using the triangle inequality and (2.32),

$$\begin{aligned} |f(x) - C_{M,N}(f,h)(x)| &\leq L_1 e^{-\pi d/h} + \sum_{k=M+1}^{\infty} |f(-kh)| + \sum_{k=N+1}^{\infty} |f(kh)| \\ &\leq L_1 e^{-\pi d/h} + L \left\{ \sum_{k=M+1}^{\infty} e^{-\alpha kh} + \sum_{k=N+1}^{\infty} e^{-\beta kh} \right\} \\ &\leq L_1 e^{-\pi d/h} + L \left\{ \frac{e^{-\alpha Mh}}{\alpha h} + \frac{e^{-\beta Nh}}{\beta h} \right\} \end{aligned}$$

Now if N and h are defined by (2.33) and (2.34), then

$$\begin{aligned} |f(x) - C_{M,N}(f,h)(x)| &\leq L_1 e^{-(\pi d \alpha M)^{\frac{1}{2}}} + L \left(\frac{1}{\alpha} + \frac{1}{\beta} \right) \sqrt{\frac{\alpha M}{\pi d}} e^{-(\pi d \alpha M)^{\frac{1}{2}}} \\ &= C M^{\frac{1}{2}} e^{-(\pi d \alpha M)^{\frac{1}{2}}} \end{aligned}$$

The choice of N and h is dictated by balancing asymptotic errors. The truncation errors for the lower and upper sums

are of orders $O(e^{-\alpha Mh})$ and $O(e^{-\beta N h})$, respectively. Equating these order statements gives $N = \alpha M/\beta$. Using (2.33) is a convenience to guarantee N is an integer. The choice for h is deduced by equating the truncation error order $O(e^{-\alpha Mh})$ and the order of the interpolation error $O(e^{-\pi d/h})$. Hence, truncating sums for computational feasibility need not be at the expense of the exponential convergence rate. This rate carries over during the development of the sinc quadrature rule.

Quadrature on $(-\infty, \infty)$

The interpolation results just reviewed are the groundwork for the derivation of the sinc quadrature rule on the real line. As with interpolation, the quadrature formula involves infinite sums; hence, a primary task is deducing the error caused by truncation. This leads to a numerically practical means of estimating an integral over the real line. The use of conformal maps generalizes the sinc quadrature to alternative curves.

A few preliminaries are necessary before stating the quadrature theorem analogous to Theorem (2.25).

$$\text{Lemma (2.36): } \frac{1}{h} \int_{-\infty}^{\infty} \text{sinc}\left(\frac{x - kh}{h}\right) dx = 1, \quad h > 0, k \in \mathbb{Z}.$$

$$\text{Lemma (2.37): } \int_{-\infty}^{\infty} \frac{e^{iax}}{x - z_0} dx = 2\pi i \begin{cases} e^{iaz_0}, & \text{Im}(z_0) > 0 \\ 0, & \text{Im}(z_0) < 0 \end{cases}.$$

Lemmas (2.36) and (2.37) are proved using standard contour integration.

Lemma (2.38): If $f \in L^1(\mathbb{R})$ then $\int_{-\infty}^{\infty} C(f, h)(x) dx = h \sum_{k=-\infty}^{\infty} f(kh)$.

where $h > 0$ and $k \in \mathbb{Z}$.

Proof: By the integral test $\sum_{k=-\infty}^{\infty} |f(kh)|$ converges.

Since $|f(kh) \operatorname{sinc}\left(\frac{x - kh}{h}\right)| \leq |f(kh)|$, $\sum_{k=-\infty}^{\infty} f(kh) \operatorname{sinc}\left(\frac{x - kh}{h}\right)$

converges absolutely and, therefore, by the Weierstrass M-test, uniformly. Thus

$$\begin{aligned} \int_{-\infty}^{\infty} C(f, h)(x) dx &= \int_{-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f(kh) \operatorname{sinc}\left(\frac{x - kh}{h}\right) dx \\ &= \sum_{k=-\infty}^{\infty} f(kh) \int_{-\infty}^{\infty} \operatorname{sinc}\left(\frac{x - kh}{h}\right) dx \\ &= h \sum_{k=-\infty}^{\infty} f(kh) . \end{aligned}$$

These three lemmas along with Theorem (2.25) provide the background to prove the following.

Theorem (2.39): If $f \in B(D_S) \equiv B^1(D_S)$ then

$$(2.40) \quad \int_{-\infty}^{\infty} f(x) dx = h \sum_{k=-\infty}^{\infty} f(kh) + \eta(f)$$

where

$$(2.41) \quad \eta(f) \equiv \int_{-\infty}^{\infty} \varepsilon(f)(x) dx$$

and

$$(2.42) \quad |\eta(f)| \leq \frac{e^{-\pi d/h} N(f, D_S)}{2 \sinh(\pi d/h)}.$$

Here $\varepsilon(f)(x)$ and $N(f, D_S)$ are defined in (2.26) and (2.24) (with $p = 1$), respectively.

$$\begin{aligned} \text{Proof: } \int_{-\infty}^{\infty} \varepsilon(f)(x) dx &= \int_{-\infty}^{\infty} f(x) dx - \int_{-\infty}^{\infty} C(f, h)(x) dx \\ &= \int_{-\infty}^{\infty} f(x) dx - h \sum_{k=-\infty}^{\infty} f(kh) \end{aligned}$$

From the left hand side, using (2.26)

$$\begin{aligned} \eta(f) &\equiv \int_{-\infty}^{\infty} \varepsilon(f)(x) dx \\ &= \int_{-\infty}^{\infty} \frac{\sin(\pi x/h)}{2\pi i} \left\{ \int_{-\infty}^{\infty} \frac{f(t-id) dt}{(t-id-x) \sin(\pi(t-id)/h)} \right. \\ &\quad \left. - \int_{-\infty}^{\infty} \frac{f(t+id) dt}{(t+id-x) \sin(\pi(t+id)/h)} \right\} dx \end{aligned}$$

$$= \frac{e^{-\pi d/h}}{2i} \int_{-\infty}^{\infty} \left\{ \frac{f(t+id)e^{i\pi t/h}}{\sin(\pi(t+id)/h)} - \frac{f(t-id)e^{-i\pi t/h}}{\sin(\pi(t-id)/h)} \right\} dt$$

where Fubini's Theorem and (2.37) with $\alpha = \pi/h$ and

$z_0 = t \pm id$ give the last equality. Bounding the last expression leads to (2.42).

The result of Theorem (2.39) is that sinc quadrature is the familiar trapezoidal rule on the real line. However, the rate of convergence for (2.40) is $O(e^{-2\pi d/h})$ rather than the usual $O(h^2)$ that is associated with the trapezoidal rule when f'' is bounded. Sinc function properties lead to the restriction $f \in B(D_S)$ which, in turn, accounts for the vastly accelerated convergence rate.

As with interpolation, numerical practicality calls for truncating the sinc quadrature, nee' trapezoidal, series. Define the truncated trapezoidal series

$$(2.43) \quad T_{M,N}(f,h) \equiv h \sum_{k=-M}^N f(kh), \quad h > 0.$$

Analogous to Theorem (2.31) is Theorem (2.44).

Theorem (2.44): If $f \in B(D_S)$, $d > 0$ and f satisfies (2.32) for some positive constants α and β , then choosing N and h as in (2.33) and (2.34) gives

$$(2.45) \quad \left| \int_{-\infty}^{\infty} f(x)dx - T_{M,N}(f,h) \right|$$

$$\leq K_1 e^{-2\pi d/h} + \frac{K}{\alpha} e^{-\alpha Mh} + \frac{K}{\beta} e^{-\beta Nh}$$

$$\leq C e^{-(2\pi d\alpha M)^{1/2}}$$

where K_1 , K , and C are constants.

The proof follows in a like fashion to that of Theorem (2.31); for greater detail see Stenger [12], [13] and Lund [6]. Once again selections for N and h are motivated by asymptotically balancing errors. Finally, note the increased rate of convergence for quadrature, $O(e^{-(2\pi d\alpha M)^{1/2}})$, versus interpolation, $O(e^{-(\pi d\alpha M)^{1/2}})$.

Interpolation and Quadrature on Alternative Arcs

The preceding results hold for $x \in (-\infty, \infty)$. To be useful in the setting of building a Sinc-Galerkin approximate for the solution of a differential equation, the results must be generalized to alternative intervals. Stenger [5] shows that conformal maps provide the means of extension. His discussion is briefly outlined here with the goal in mind the statement of the quadrature rule.

Let D be a simply connected domain and D_S be as in Definition (2.20). Given $a, b \in \partial D$ such that $a \neq b$, there is a conformal map $\chi: D \rightarrow D_S$ (see Figure 5) satisfying

$$(2.46) \quad \psi(R) \equiv \chi^{-1}(R) = \gamma$$

and

$$(2.47) \quad \lim_{\substack{z \rightarrow a \\ z \in \gamma}} \chi(z) = -\infty, \quad \lim_{\substack{z \rightarrow b \\ z \in \gamma}} \chi(z) = +\infty.$$

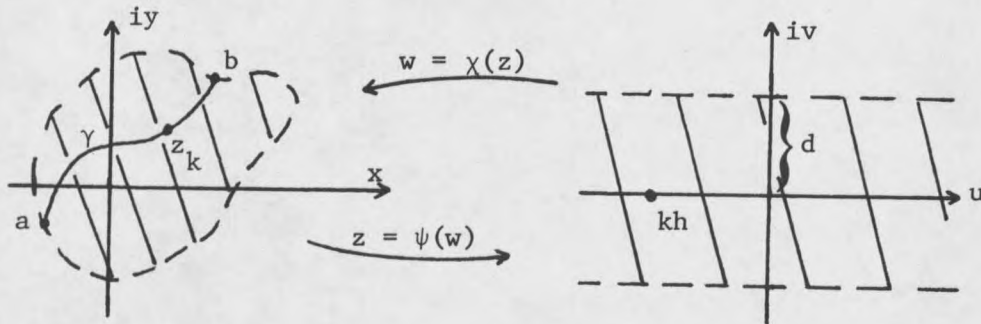


Figure 5: The Conformal Map χ .

With respect to D , the following definition is similar to (2.21).

Definition (2.48): Let $B(D)$ be the class of functions such that f is holomorphic on D ;

$$(2.49) \quad \int_{\gamma(t+L)} |f(w)dw| = O(|t|^a) \text{ as } t \rightarrow \infty$$

where $a \in [0,1)$ and $L = \{iy: |y| < d\}$; and

$$(2.50) \quad N(f,D) \equiv \liminf_{\mathcal{C} \subset D \rightarrow \partial D} \int_{\mathcal{C}} |f(w)dw| < \infty$$

where $\mathcal{C}$ is a simple closed curve in D .

Rather than developing interpolation and quadrature rules for D from scratch, two theorems provide a natural link to the past work.

Theorem (2.51): If ψ is a conformal map of D_S onto the simply connected domain D and if $f \in B(D)$, then $F \in B(D_S)$ where

$$(2.52) \quad F(w) \equiv f(\psi(w))\psi'(w) .$$

Also

Theorem (2.53): If $\chi: D \rightarrow D_S$ is a conformal, one-to-one and onto mapping then $\psi = \chi^{-1}$ is conformal.

Stenger [13] remarks on Theorem (2.51) while Theorem (2.53) is a standard complex analysis result. Finally a revised version of the exponential decay property (2.32) is needed.

Definition (2.54): Let $f \in B(D)$, χ be a conformal map satisfying the conditions of Theorem (2.53), and $\tau \in \gamma = \psi(R) = \chi^{-1}(R)$. Then f/χ' is said to decay exponentially with respect to χ if there exist positive constants K , α and β such that

$$(2.55) \quad \left| \frac{f(\tau)}{\chi'(\tau)} \right| \leq K \begin{cases} \exp(-\alpha|\chi(\tau)|) , & \tau \in \gamma_L \\ \exp(-\beta|\chi(\tau)|) , & \tau \in \gamma_R \end{cases}$$

where

$$(2.56) \quad \gamma_L = \{z: z \in \psi((-\infty, 0))\}$$

$$\gamma_R = \{z: z \in \psi([0, \infty))\}.$$

Theorem (2.51) and Definition (2.54) suggest that the conformal map is incorporated into the previous theorems and

definitions quite easily. The interpolation and quadrature rules that follow support this notion.

Theorem (2.57): Let $f \in B(D)$ and $\chi: D \rightarrow D_S$ satisfy the hypotheses of Theorem (2.53). Further, suppose f/χ' decays exponentially with respect to χ . If $z \in \gamma = \psi(R) = \chi^{-1}(R)$ and $z_k = \psi(kh)$, $k \in \mathbb{Z}$, $h > 0$, then selecting N and h as in (2.33) and (2.34), respectively, gives

$$\begin{aligned}
 (2.58) \quad & \left| \frac{f(z)}{\chi'(z)} - \sum_{k=-M}^N \frac{f(z_k)}{\chi'(z_k)} \operatorname{sinc}\left(\frac{\chi(z) - kh}{h}\right) \right| \\
 &= \left| \frac{f(z)}{\chi'(z)} - C_{M,N}\left(\frac{f}{\chi'}, h\right) \circ \chi(z) \right| \\
 &\leq K e^{-(\pi d \alpha M)^{\frac{1}{2}}}
 \end{aligned}$$

where K is a constant independent of z .

Note that to interpolate f rather than f/χ' , simply let $F = f\chi'$ and substitute into (2.58) assuming F satisfies the hypotheses of Theorem (2.57). Equation (2.58) shows that the rate of convergence for sinc interpolation on a curve γ remains $O(e^{-(\pi d \alpha M)^{\frac{1}{2}}})$. Similarly, the rate of convergence for the sinc quadrature rule is unchanged.

Theorem (2.59):

- (a) If $f \in B(D)$ and χ , ψ , z , and z_k are as in Theorem (2.57), then for h sufficiently small

$$(2.60) \quad \left| \int_{\gamma} f(z) dz - h \sum_{k=-\infty}^{\infty} \frac{f(z_k)}{\chi'(z_k)} \right| \leq \frac{N(f, D)}{1-e^{-d}} e^{-2\pi d/h}$$

$$\equiv K_2 e^{-2\pi d/h}.$$

(b) Further, if f/χ' decays exponentially with respect to χ then

$$(2.61) \quad \left| \int_{\gamma} f(z) dz - h \sum_{k=-M}^N \frac{f(z_k)}{\chi'(z_k)} \right| \leq K_2 e^{-2\pi d/h} + \frac{K}{\alpha} e^{-\alpha Mh}$$

$$+ \frac{K}{\beta} e^{-\beta Nh}.$$

(c) Finally, if N and h satisfy (2.33) and (2.34), respectively, then for some constant C depending on f

$$(2.62) \quad K_2 e^{-2\pi d/h} + \frac{K}{\alpha} e^{-\alpha Mh} + \frac{K}{\beta} e^{-\beta Nh} \leq C e^{-(2\pi d\alpha M)^{1/2}}.$$

Stenger [12] provides proofs of Theorems (2.57) and (2.59) for centered sums, that is, $M = N$. Lund [6] expands on these to handle $M \neq N$. Both of their proofs illustrate the rationale for the use of conformal maps; namely, under suitable assumptions the maps are a means of transferring a standard set of results from D_S to various domains D . Most significant, is that conformality preserves the error.

Sinc-Galerkin Method

The previous discussion is the background necessary to approach the numerical solution of differential equations

via the Sinc-Galerkin method. For the purposes of this work, it is sufficient to examine the method applied to a specific class of differential equations. Hence, consider the second-order, self-adjoint boundary value problem

$$(2.63) \quad \begin{aligned} Lf(x) &= f''(x) + v(x)f(x) = \sigma(x), \quad a < x < b \\ f(a) &= f(b) = 0 \end{aligned}$$

Let D be a simply connected domain with

$$(2.64) \quad \gamma = \{(x, y) : a < x < b, y = 0\} \subset D$$

and

$$(2.65) \quad \{(a, 0), (b, 0)\} \subset \partial D.$$

With regard to D , the Sinc-Galerkin method to approximate the solution of (2.63) is summarized as follows. Define the set of basis functions $\{S_i\}_{i=-M}^N$ by

$$(2.66) \quad S_i(x) = S(i, h) \circ \chi(x), \quad h > 0$$

where $\chi: D \rightarrow D_S$ is described in (2.46), (2.47), and Theorem (2.53). See Figure 5. Next define the approximate solution $\hat{f}_m$ by

$$(2.67) \quad \hat{f}_m(x) = \sum_{i=-M}^N f_i S_i(x), \quad m = M + N + 1.$$

To determine the unknown coefficients, f_i , orthogonalize the residual with respect to the basis elements; that is,

$$(2.68) \quad \begin{aligned} 0 &= (L\hat{f}_m - \sigma, S_p) \\ &= (\hat{f}_m'', S_p) + (v\hat{f}_m - \sigma, S_p), \quad -M \leq p \leq N. \end{aligned}$$

Here, the inner product $(\cdot, \cdot)$ is

$$(2.69) \quad (u, v) = \int_a^b u(x)v(x)w(x)dx$$

where w is a weight function yet to be specified and the quadrature rule used to evaluate the inner product is

(2.61). The resulting discrete linear system is solved for f_i , $-M \leq i \leq N$.

A general review of Galerkin methods is found in Botha and Pinder [3]. Their discussion includes several criteria by which to judge the quality of a method. These include: (i) the ease of constructing the basis elements; (ii) the ease of evaluating the inner product; (iii) the ease of solving the system; and (iv) the accuracy of the method. With regard to the Sinc-Galerkin method, the first is demonstrated by (2.66) while the second is shown to depend on the sinc basis elements and sinc quadrature rule. For the third, it is shown that an adroit choice of weight function leads to a system which is easily solved. Lastly, the accuracy follows directly from Theorem (2.59).

Stenger [5] thoroughly discusses $w = 1/X'$ and Lund [6] introduces $w = 1/(X')^{1/2}$. Stenger's choice of weight function handles regular singular problems but yields a nonsymmetric linear system. Lund's weight function results in a symmetric linear system while possibly restricting the class of functions to which the method applies. The development of the discrete system for a general weight motivates the

choice $w = 1/(\chi')^{1/2}$ as well as addresses the questions raised in the previous paragraph.

Continuing with (2.68), for $-M \leq p \leq N$ integrate the inner product (f'', S_p) by parts twice to get

$$\begin{aligned}
 0 &= \int_a^b f''(x) S_p(x) w(x) dx \\
 &\quad + \int_a^b [v(x)f(x) - \sigma(x)] S_p(x) w(x) dx \\
 &= BT + \int_a^b f(x) \{ [S_p(x)w(x)]'' + v(x)S_p(x)w(x) \} dx \\
 &\quad - \int_a^b \sigma(x) S_p(x) w(x) dx \\
 (2.70) \quad &= BT + \int_a^b f(x) \left\{ \left[\frac{d^2}{d\chi^2} S(p, h) \circ \chi(x) \right] (\chi'(x))^2 w(x) \right. \\
 &\quad + \left[\frac{d}{d\chi} S(p, h) \circ \chi(x) \right] (\chi''(x)w(x) + 2\chi'(x)w'(x)) \\
 &\quad + S(p, h) \circ \chi(x) (w''(x) + v(x)w(x)) \left. \right\} dx \\
 &\quad - \int_a^b \sigma(x) S(p, h) \circ \chi(x) w(x) dx
 \end{aligned}$$

where BT is the boundary term

$$(2.71) \quad BT = \left\{ f'(x) S_p(x) w(x) - f(x) [S_p(x) w(x)]' \right\} \Big|_a^b .$$

For now, BT is assumed to vanish. The exact assumptions

governing $BT = 0$ are discussed in Chapter 3 when specific conformal maps are used. To apply the sinc quadrature (2.61) to (2.70) several suppositions are necessary. First, $f[S_p w]$ satisfies the hypotheses of Theorem (2.59). Second, $f v S_p w$ and $\sigma S_p w$ are in $B(D)$. It is unnecessary for $f v S_p w / \chi'$ and $\sigma S_p w / \chi'$ to decay exponentially with respect to χ . This is because for $g \in B(D)$, the quadrature rule (2.60) for g integrated against the sinc yields point evaluation. That is,

$$(2.72) \quad \int_a^b g(x) S_p(x) dx = g(ph) + O(e^{-\pi d/h})$$

To proceed, the following three identities are useful.

$$(2.73) \quad S(p, h) \circ \chi(x) \Big|_{x=x_i} = \delta_{pi}^{(0)} = \begin{cases} 1, & p = i \\ 0, & p \neq i \end{cases}$$

$$(2.74) \quad h \left[\frac{d}{d\chi} S(p, h) \circ \chi(x) \right] \Big|_{x=x_i} = \delta_{pi}^{(1)} = \begin{cases} 0, & p = i \\ \frac{(-1)^{i-p}}{i-p}, & p \neq i \end{cases}$$

and

$$(2.75) \quad h^2 \left[\frac{d^2}{d\chi^2} S(p, h) \circ \chi(x) \right] \Big|_{x=x_i} = \delta_{pi}^{(2)} = \begin{cases} -\pi^2/3, & p = i \\ \frac{-2(-1)^{i-p}}{(i-p)^2}, & p \neq i \end{cases}$$

Applying the quadrature rule results to (2.70) yields

$$\begin{aligned}
(2.76) \quad 0 = & \sum_{i=-M}^N \left\{ f(x_i) \left[\frac{1}{h^2} \delta_{pi}^{(2)} \chi'(x_i) w(x_i) \right. \right. \\
& + \frac{1}{h} \delta_{pi}^{(1)} \left(\frac{\chi''(x_i)}{\chi'(x_i)} w(x_i) + 2w'(x_i) \right) \\
& \left. \left. + \delta_{pi}^{(0)} \left(\frac{w''(x_i)}{\chi'(x_i)} + \frac{v(x_i)w(x_i)}{\chi'(x_i)} \right) \right] \right\} \\
& - \sum_{i=-M}^N \delta_{pi}^{(0)} \frac{\sigma(x_i)w(x_i)}{\chi'(x_i)} + o(e^{-(\pi d \alpha M)^{1/2}}).
\end{aligned}$$

An equivalent matrix formulation is

$$\begin{aligned}
(2.77) \quad & \left\{ \frac{1}{h^2} I^{(2)} D(\chi'w) + \frac{1}{h} I^{(1)} D \left[\frac{\chi''w}{\chi'} + 2w' \right] \right. \\
& \left. + D \left[\frac{w'' + vw}{\chi'} \right] \right\} \vec{f} = D \left(\frac{\sigma w}{\chi'} \right) \vec{f}
\end{aligned}$$

where

$$(2.78) \quad \vec{f} = [f_{-M}, f_{-M+1}, \dots, f_{N-1}, f_N]^T,$$

$$(2.79) \quad D(g) = \begin{bmatrix} g(x_{-M}) & & & \\ & g(x_{-M+1}) & & \\ & & \ddots & \\ & & & g(x_N) \end{bmatrix}_{m \times m}$$

$$\begin{aligned}
 I^{(2)} &= \left(\delta_{pi}^{(2)} \right) \\
 (2.80) \quad &= \begin{bmatrix} \frac{-\pi^2}{3} & \frac{2}{1^2} & \frac{-2}{2^2} & \dots & \frac{2(-1)^m}{(m-1)^2} \\ \frac{2}{1^2} & \frac{-\pi^2}{3} & \frac{2}{1^2} & \dots & \frac{2(-1)^{m-1}}{(m-2)^2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{2(-1)^m}{(m-1)^2} & \frac{2(-1)^{m-1}}{(m-2)^2} & \dots & \dots & \frac{-\pi^2}{3} \end{bmatrix}_{m \times m}
 \end{aligned}$$

and

$$\begin{aligned}
 I^{(1)} &= \left(\delta_{pi}^{(1)} \right) \\
 (2.81) \quad &= \begin{bmatrix} 0 & -1 & \frac{1}{2} & \dots & \frac{(-1)^{m-1}}{m-1} \\ 1 & 0 & -1 & \dots & \frac{(-1)^{m-2}}{m-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{(-1)^{m-1}}{m-1} & -\frac{(-1)^{m-2}}{m-2} & \dots & \dots & 0 \end{bmatrix}
 \end{aligned}$$

For an arbitrary weight function it is unclear how to solve (2.77). In particular, the skew-symmetric nature of $I^{(1)}$ causes difficulties. It can be argued that as $h \rightarrow 0$ $I^{(2)}$ is the dominant matrix in the system. This is somewhat satisfying in that $I^{(2)}$ is a symmetric, negative definite

Toeplitz matrix whose spectra lies in the interval $(-\pi^2, 0)$.

Lund's [6] notion is to consider

$$(2.82) \quad \frac{\chi''}{\chi'} w + 2w' = 0$$

or equivalently

$$(2.83) \quad \frac{\chi''}{\chi'} = \frac{-2w'}{w}$$

Integrating both sides of (2.83) yields $w = 1/(\chi')^{1/2}$ and the system now simplifies to

$$(2.84) \quad \left\{ \frac{1}{h^2} I^{(2)} D((\chi')^{1/2}) + D \left[\left(\left(\frac{1}{(\chi')^{1/2}} \right)'' + \frac{\nu}{(\chi')^{1/2}} \right) \frac{1}{\chi'} \right] \right\} \vec{f} \\ = D[\sigma/(\chi')^{3/2}] \vec{f}$$

Multiplying through by $D(\chi')$ and factoring gives

$$(2.85) \quad A\vec{F} = D(\sigma/(\chi')^{1/2}) \vec{f}$$

where

$$(2.86) \quad A = D(\chi') \left\{ \frac{1}{h^2} I^{(2)} + D \left[\left(\left(\frac{1}{(\chi')^{1/2}} \right)'' + \frac{\nu}{(\chi')^{1/2}} \right) \frac{1}{(\chi')^{3/2}} \right] \right\} D(\chi')$$

and

$$(2.87) \quad \vec{F} = D(1/(\chi')^{1/2}) \vec{f}$$

The matrix A is a real symmetric, negative definite matrix; hence, there exists an orthogonal matrix Q such that

$$(2.88) \quad A = Q\Lambda Q^T$$

where Λ is the diagonal matrix of the eigenvalues of A .

Solving for $\vec{F}$,

$$(2.89) \quad \vec{F} = Q\Lambda^{-1}Q^T D(\sigma/(\chi')^{1/2}) \vec{f}$$

and, in turn, $\vec{f}$ is recovered via

$$(2.90) \quad \vec{f} = D((\chi')^{\frac{1}{2}})\vec{F} .$$

In review, a symmetric coefficient matrix for the self-adjoint problem (2.1) is attained by selecting $w = 1/(\chi')^{\frac{1}{2}}$. The class of singular problems that $w = 1/(\chi')^{\frac{1}{2}}$ handles appears to be somewhat more restricted than if $w = 1/\chi'$ [6]. However, this statement is dependent on the method of proof in [6]. When $w = 1/(\chi')^{\frac{1}{2}}$ is known to apply to a singular problem, the rate of convergence remains $O(\exp(-K\sqrt{N}))$ where $2N + 1$ basis functions are used.

CHAPTER 3

THE SINC-GALERKIN METHOD FOR THE WAVE EQUATION

Various numerical methods for solving general second-order hyperbolic problems can be found in the literature [3]. Usually a scheme is first developed for the classic vibrating string problem

$$\begin{aligned}
 (3.1) \quad & Lu(x,t) = u_{tt}(x,t) - u_{xx}(x,t) = f(x,t) \\
 & (x,t) \in (0,1) \times (0,\infty) \\
 & u(0,t) = u(1,t) = 0 \quad t \in [0,\infty) \\
 & u(x,0) = u_t(x,0) = 0 \quad x \in [0,1] .
 \end{aligned}$$

Once this model problem is investigated, several generalizing routes are possible. Among them are adaptations to handle nonconstant coefficients, nonlinearities, or higher space dimensions. With regard to the Sinc-Galerkin method, the present work examines higher space dimensions. In this chapter the discrete linear Sinc-Galerkin system for (3.1) is derived in some detail and from it a natural extension to the linear systems for two and three space dimensions is exhibited. The actual solution of the systems is postponed until Chapter 4.

A common procedure for solving (3.1) is a Galerkin discretization of the spatial domain with time dependent coefficients. This gives a system of ordinary differential

equations typically solved by difference techniques (O.D.E. solvers) [14]. Drawbacks of this method include the necessity to artificially truncate the time domain and the fact that the numerical solution is valid only on a finite time grid. In contrast, the technique of this work implements a Galerkin scheme in time as well as space. The success of the method is largely due to the choice of basis elements; these basis elements are tensor products of sinc functions composed with suitable conformal maps for the intervals $(0,1)$ and $(0,\infty)$. The conformal map for $(0,\infty)$ is the means for building an approximate solution to (3.1) valid on the infinite time domain.

The structure of the discrete Sinc-Galerkin system is computationally efficient for at least two reasons. As a result of the identities (2.73) - (2.75) and the sinc quadrature rule (2.61), the coefficients of the unknowns and the constant terms are easily evaluated. Note that this property is independent of the weight function. A property which is intimately connected to the selection of the weight is the symmetry of the system. In the present chapter it is shown that a generalization of $w = 1/(x')^{1/2}$ leads to a symmetric linear system. Two matrix formulations for the discrete system are given. The more convenient formulation to implement is dependent on the computational environment, i.e. the available computer hardware.

To determine the Sinc-Galerkin approximate for the solution of (3.1) the fundamental steps used in Chapter 2 remain unchanged. First, select a set of basis functions, then using these basis functions define an approximate solution. The unknown coefficients appearing in the approximate are determined by orthogonalizing the residual with respect to the basis elements. The orthogonalization is defined with a weight function and evaluated by means of the quadrature rule (2.61).

To construct basis functions on the region $\{(x,t): 0 < x < 1, 0 < t < \infty\}$, basis functions on the intervals $(0,1)$ and $(0,\infty)$ are built and then their tensor products are formed. So to begin, define the conformal maps

$$(3.2) \quad \phi(z) = \ln\left(\frac{z}{1-z}\right)$$

and

$$(3.3) \quad \uparrow(w) = \ln(w)$$

The map ϕ carries the eye-shaped region

$$(3.4) \quad D_E = \left\{ z = x + iy: \left| \arg\left(\frac{z}{1-z}\right) \right| < d \leq \pi/2 \right\}$$

onto the infinite strip D_S given by Definition (2.20).

Similarly, the map $\uparrow$ carries the infinite wedge

$$(3.5) \quad D_W = \{w = t + is: |\arg(w)| < d \leq \pi/2\}$$

onto D_S . These regions are depicted in Figure 6.

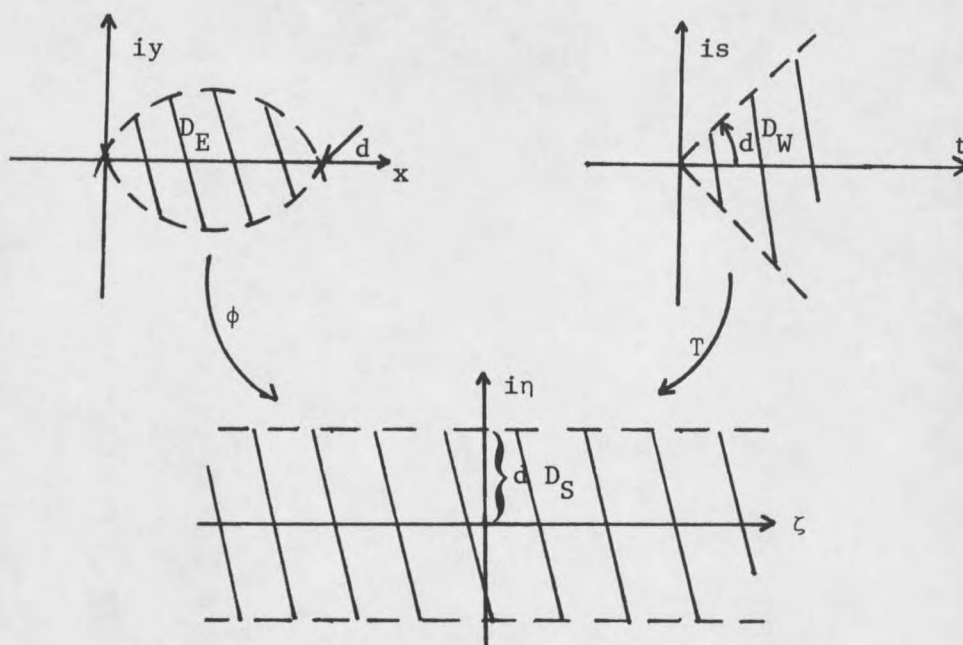


Figure 6: The Regions D_E , D_W , and D_S .

The compositions

$$(3.6) \quad S_i(x) = S(i, h_x) \circ \phi(x)$$

and

$$(3.7) \quad S_j^*(t) = S(j, h_t) \circ \tau(t)$$

define the basis elements for $(0,1)$ and $(0,\infty)$, respectively (see Figure 7 below). The "mesh sizes" h_x and h_t represent the mesh sizes in D_S for the uniform grids $\{kh_x\}$, $-\infty < k < \infty$, and $\{ph_t\}$, $-\infty < p < \infty$. The sinc grid-points $x_k \in (0,1)$ in D_E and $t_p \in (0,\infty)$ in D_W are the inverse images of the equispaced grids; that is,

$$(3.8) \quad x_k = \phi^{-1}(kh_x) = e^{kh_x} / (1 + e^{kh_x})$$

and

$$(3.9) \quad t_p = \tau^{-1}(ph_t) = e^{ph_t}.$$

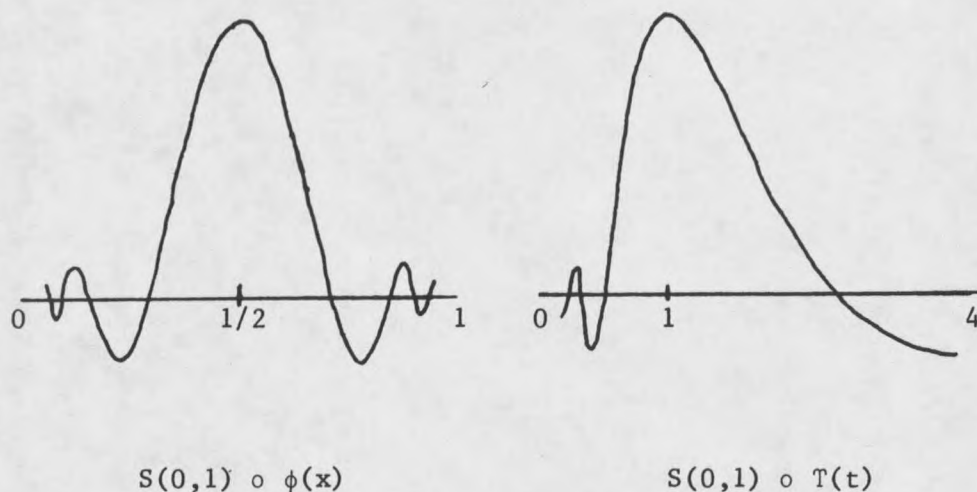


Figure 7: The Basis Functions $S_0(x)$ and $S_0^*(t)$.

The fully Galerkin approximate is now defined by

$$(3.10) \quad \hat{u}_{m_x, m_t}(x, t) = \sum_{i=-M_x}^{N_x} \sum_{j=-M_t}^{N_t} u_{ij} S_i(x) S_j^*(t),$$

$$m_x = M_x + N_x + 1$$

$$m_t = M_t + N_t + 1$$

The inner product used to orthogonalize the residual is

$$(3.11) \quad (u, v) = \int_0^\infty \int_0^1 u(x, t) v(x, t) \frac{dx dt}{[\phi'(x) \tau'(t)]^{1/2}}$$

and may be viewed as the double integral analogue of (2.69) with $w = 1/(\chi')^{1/2}$. Direct development of the discrete Sinc-Galerkin system for (3.1) is obtained via

$$(3.12) \quad 0 = (L\hat{u}_{m_x, m_t} - f, S_k S_p^*) , \quad \begin{array}{l} -M_x \leq k \leq N_x \\ -M_t \leq p \leq N_t \end{array}$$

However, this procedure obscures the important parameter selections needed to implement the method. The development of the Sinc-Galerkin systems for the one-dimensional problems

$$(3.13) \quad \begin{array}{ll} u''(t) = r(t) & 0 < t < \infty \\ u(0) = u'(0) = 0 \end{array}$$

and

$$(3.14) \quad \begin{array}{ll} u''(x) = s(x) & 0 < x < 1 \\ u(0) = u(1) = 0 \end{array}$$

yields all of the sinc matrices for the two-dimensional problem while clearly disclosing the parameter selections just mentioned.

Beginning with (3.13), perform two integrations by parts in the orthogonalization of the residual (see (2.70) for detail). Once this is done, apply the quadrature rule (2.60) in Theorem (2.59) with x replaced by τ to get

$$(3.15) \quad \begin{aligned} 0 &= \int_0^\infty [u''(t) - r(t)] \{S(p, h_t) \circ \tau(t) (\tau'(t))^{-\frac{1}{2}}\} dt \\ &= \int_0^\infty u(t) \{S(p, h_t) \circ \tau(t) (\tau'(t))^{-\frac{1}{2}}\}'' dt \\ &\quad - \int_0^\infty r(t) \{S(p, h_t) \circ \tau(t) (\tau'(t))^{-\frac{1}{2}}\} dt \end{aligned}$$

$$\begin{aligned}
&= h_t \sum_{j=-\infty}^{\infty} u(t_j) \left(\frac{1}{h_t^2} \delta_{pj}^{(2)} - \frac{1}{4} \delta_{pj}^{(0)} \right) (\tau'(t_j))^{\frac{1}{2}} \\
&- h_t \frac{r(t_p)}{(\tau'(t_p))^{3/2}} + I_u + I_r.
\end{aligned}$$

The second equality in (3.15) assumes the boundary term

$$\begin{aligned}
(3.16) \quad &\{S(p, h_t) \circ \tau(t)(\tau'(t))^{-\frac{1}{2}} u'(t) \\
&- [S(p, h_t) \circ \tau(t)(\tau'(t))^{-\frac{1}{2}}]' u(t)\} \Big|_0^{\infty}
\end{aligned}$$

vanishes. Using the definitions of τ (3.3) and $S(p, h_t) \circ \tau$, this is true as long as

$$(3.17) \quad \lim_{t \rightarrow \infty} u'(t) \sqrt{t} / \lambda n(t) = 0,$$

$$(3.18) \quad \lim_{t \rightarrow 0^+} u(t) / \sqrt{t} = 0,$$

and

$$(3.19) \quad \lim_{t \rightarrow \infty} u(t) / (\sqrt{t} \cdot \lambda n(t)) = 0.$$

With regard to the last equality in (3.15), the nodes t_p are defined by (3.9) and the identity

$$(3.20) \quad (\tau'(t))^{-3/2} [(\tau'(t))^{-\frac{1}{2}}]'' = -1/4$$

is utilized. The integrals I_u and I_r are explicitly defined in [15] and represent the exact error terms in (2.60). From Theorem (2.59), I_u vanishes with order $O(e^{-\pi d/h})$ if

$$(3.21) \quad u(z) [S(p, h_t) \circ \tau(z)(\tau'(z))^{-\frac{1}{2}}]'' \in B(D_W)$$

while the same statement holds for I_r if

$$(3.22) \quad r(t) \{S(p, h_t) \circ \tau(t)(\tau'(t))^{-\frac{1}{2}}\} \in B(D_W).$$

To truncate the infinite sum note that

$u(t)\{S(p, h_t) \circ \tau(t)(\tau'(t))^{-1/2}\}$ behaves like $u(t)(\tau'(t))^{3/2}$ near $t = 0$ and $t = \infty$. Hence, condition (2.55) simplifies to

$$(3.23) \quad |u(t)(\tau'(t))^{1/2}| \leq K \begin{cases} t^\gamma, & t \in (0, 1) \\ t^{-\delta}, & t \in [1, \infty) \end{cases}$$

for positive constants K , γ , and δ . Since (3.23) implies (3.18) and (3.19) the only additional assumption necessary for (3.16) to vanish is (3.17). Applying the truncated quadrature rule (2.61) to (3.15) yields

$$(3.24) \quad \begin{aligned} 0 &= h_t \sum_{j=-M_t}^{N_t} u(t_j) \left(\frac{1}{h_t^2} \delta_{pj}^{(2)} - \frac{1}{4} \delta_{pj}^{(0)} \right) (\tau'(t_j))^{1/2} \\ &\quad - h_t \frac{r(t_p)}{(\tau'(t_p))^{3/2}} + O(\exp(-\pi d/h_t)) \\ &\quad + O(\exp(-\gamma M_t h_t)) + O(\exp(-\delta N_t h_t)) \end{aligned}$$

If h_t and N_t are chosen by

$$(3.25) \quad h_t = (\pi d / (\gamma M_t))^{1/2}$$

and

$$(3.26) \quad N_t = \left\lceil \frac{\gamma}{\delta} M_t + 1 \right\rceil$$

then the errors are asymptotically balanced and the final linear system is

$$(3.27) \quad \begin{aligned} \sum_{j=-M_t}^{N_t} \left(\frac{1}{h_t^2} \delta_{pj}^{(2)} - \frac{1}{4} \delta_{pj}^{(0)} \right) (\tau'(t_j))^{1/2} u(t_j) \\ - r(t_p) / (\tau'(t_p))^{3/2} = O(\exp(-(\pi d \gamma M_t)^{1/2})). \end{aligned}$$

Therefore, if

$$(3.28) \quad \hat{u}_{m_t}(t) = \sum_{j=-M_t}^{N_t} u_j S(j, h_t) \circ \Upsilon(t), \quad m_t = M_t + N_t + 1$$

is an assumed approximate solution of (3.13) then since

$\hat{u}_{m_t}(t_p) = u_p$, the discrete Sinc-Galerkin system for (3.13)

is defined by the left-hand side of (3.27) with $u(t_j)$ replaced by u_j . When $p = -M_t, \dots, N_t$ the matrix formulation for (3.27) follows from (2.84) and reads

$$(3.29) \quad A_t D((\Upsilon')^{1/2}) \vec{u} = D((\Upsilon')^{-3/2} r) \vec{1}$$

where

$$(3.30) \quad A_t = \frac{1}{h_t^2} I^{(2)} - \frac{1}{4} I$$

and the remaining terms are defined by (2.78) and (2.79).

A_t is a real symmetric, negative definite matrix. To maintain these properties in the discrete system (3.29) the change of variables

$$(3.31) \quad \vec{v} = D((\Upsilon')^{-1/2}) \vec{u}$$

leads to

$$(3.32) \quad A_t \vec{v} = D((\Upsilon')^{-1/2} r) \vec{1}$$

where the matrix

$$(3.33) \quad A_t = D(\Upsilon') A_t D(\Upsilon')$$

The solution $\vec{u}$ is found via the procedure outlined in (2.88) through (2.90).

Before considering (3.14) note that the selection h_t in (3.25) asymptotically balances the error terms $\exp(-\pi d/h_t)$ and $\exp(-\gamma M_t h_t)$ while the selection N_t in (3.26) balances

the asymptotic errors $\exp(-\delta N_t h_t)$ and $\exp(-\gamma M_t h_t)$. The error term $\exp(-\delta N_t h_t)$ arises as a consequence of the inequality (3.23) which assumes $u(t)$ decays algebraically at infinity. For many problems the solution decays exponentially as

$$(3.34) \quad |u(t)(\uparrow'(t))^{\frac{1}{2}}| \leq K e^{-\delta t}, \quad t \in [1, \infty).$$

In this case, Lund [6] shows that the choice

$$(3.35) \quad N_t = \left\lceil \frac{1}{h_t} \ln \frac{\gamma}{\delta} M_t h_t + 1 \right\rceil$$

significantly reduces the size of the discrete systems solved with no loss of accuracy.

The preceding discussion applied to (3.14) follows a parallel development. The map $\uparrow$ is replaced by the map ϕ of (3.2) (since ϕ is compatible with the interval $(0,1)$) and h_x is substituted for h_t . Again orthogonalizing the residual and integrating by parts twice renders an equation similar to (3.15) with a boundary term on $(0,1)$ analogous to (3.16). To guarantee the boundary term vanishes it is assumed that

$$(3.36) \quad \lim_{x \rightarrow 0^+} \frac{u'(x)\sqrt{x}}{\ln(x)} = \lim_{x \rightarrow 1^-} \frac{u'(x)\sqrt{1-x}}{\ln(1-x)} = 0.$$

Example 5.4 in Chapter 5 illustrates a case where u' is unbounded as $x \rightarrow 0^+$ yet (3.36) is satisfied. Further discussion is included in that example. The exponential decay condition (2.55) simplifies to

$$(3.37) \quad |u(x)(\phi'(x))^{\frac{1}{2}}| \leq L \quad \begin{cases} x^\alpha, & x \in (0, \frac{1}{2}) \\ (1-x)^\beta, & x \in [\frac{1}{2}, 1) \end{cases}$$

for positive constants L , α , and β where f and χ are replaced by $u(\phi')^{3/2}$ and ϕ , respectively. Hence the selections

$$(3.38) \quad h_x = (\pi d / (\alpha M_x))^{1/2}$$

and

$$(3.39) \quad N_x = \left\lceil \frac{\alpha}{\beta} M_x + 1 \right\rceil$$

result in the balanced asymptotic error rate $O(\exp(-(\pi d \alpha M_x)^{1/2}))$.

Therefore, if

$$(3.40) \quad \hat{u}_{m_x} = \sum_{i=-M_x}^{M_t} u_i S(i, h_x) \circ \phi(x), \quad m_x = M_x + N_x + 1$$

is an assumed approximate solution of (3.14) then the discrete Galerkin system for the $\{u_i\}$ is given by

$$(3.41) \quad A_x \vec{w} = D((\phi')^{-1/2} s) \vec{1}$$

where the matrix

$$(3.42) \quad A_x = D(\phi') A_t D(\phi')$$

and A_x is the same as A_t in (3.30) with h_t replaced by h_x .

As with A_t , A_x is symmetric and negative definite. Further, when $M_x = N_x$, $D(\phi')$ is centrosymmetric. Thus since A_x is Toeplitz, A_x is centrosymmetric. Finally, the vector $\vec{w}$ is related to the vector of unknowns $\vec{u}$ by

$$(3.43) \quad \vec{w} = D((\phi')^{-1/2}) \vec{u}.$$

The separate one-dimensional problems serve to identify the matrices A_x and A_t as well as disclosing the parameter selections. Returning to the two-dimensional hyperbolic

problem (3.1) and its approximate solution (3.10), the unknown coefficients $\{u_{ij}\}$ are found by orthogonalizing the residual. One possible formulation for the resulting discrete Sinc-Galerkin system is

$$(3.44) \quad A_x V - V A_t = G$$

where the matrices A_t and A_x are identified in (3.33) and (3.42). The $m_x \times m_t$ matrices V and G are defined by

$$(3.45) \quad V = D((\phi')^{-\frac{1}{2}}) U D((\tau')^{-\frac{1}{2}})$$

and

$$(3.46) \quad G = D((\phi')^{-\frac{1}{2}}) F D((\tau')^{-\frac{1}{2}})$$

where U and F are the $m_x \times m_t$ matrices which consist of the coefficients $\{u_{ij}\}$ for (3.10) and $f(x_i, t_j)$, respectively. The form of (3.44) has two easily discerned motivations. One is that representing the unknowns as a matrix is more naturally suited to the orthogonalization procedure and subsequent use of the sinc quadrature rule. The second motivation is that transforming the unknown matrix U by (3.45) introduces the symmetric coefficient matrices A_x and A_t . As a direct consequence of this symmetry, the system (3.44) is easy to solve numerically. Chapter 4 discusses the solution technique at some length.

A second alternative for posing the discrete system occurs when the unknowns are arrayed as a vector. In some sense, representing $\{u_{ij}\}$ as a vector versus a matrix is purely a notational matter; however, the computational aspects of storing coefficient matrices and numerically

solving the discrete system can vary greatly depending on whether $\{u_{ij}\}$ is regarded as a vector or matrix.

Two background terms provide the notational machinery to rewrite the discrete system posed as in (3.44) as a system defined with $\{u_{ij}\}$ as a vector.

Definition (3.47): Let A be an $m \times n$ matrix and B be a $p \times q$ matrix. The Kronecker or tensor product of A and B is the $mp \times nq$ matrix

$$A \otimes B \equiv \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix}.$$

The second term, concatenation, loosely refers to representing a finite ordered array as a vector. Eventually concatenation is defined for an array with n -subscripts. For now, a precise definition when there are one or two subscripts is sufficient. If $b = (b_i)$, $1 \leq i \leq m$, then the concatenation of b is denoted

$$(3.48) \quad \text{co}(b) \equiv [b_1, b_2, \dots, b_m]^T.$$

Hence, for a column vector $\vec{c}$, $\text{co}(\vec{c}) = \vec{c}$ while for a row vector $\vec{r}$, $\text{co}(\vec{r}) = \vec{r}^T$. When $B = (b_{ij})$, $1 \leq i \leq m$ and $1 \leq j \leq n$, then $\text{co}(B)$ is the $mn \times 1$ vector

$$(3.49) \quad \text{co}(B) = \begin{bmatrix} \text{co}(b_{i1}) \\ \text{co}(b_{i2}) \\ \vdots \\ \text{co}(b_{in}) \end{bmatrix}$$

which may be regarded as stacking the columns of a matrix. Davis [16] includes a more thorough but slightly different discussion of concatenation and the Kronecker product. In particular, Davis defines concatenation as stacking the rows of a matrix rather than its columns.

Besides the notational apparatus just introduced, the following theorems ease the transition to a system whose unknowns are represented as a vector.

Theorem (3.50): If A and B are matrices of identical dimension, α and β are scalars, then

$$(3.51) \quad \text{co}(\alpha A + \beta B) = \alpha \text{co}(A) + \beta \text{co}(B) .$$

Theorem (3.52): Let A , X , and B be matrices whose dimensions are compatible with the product AXB . Then

$$(3.53) \quad \text{co}(AXB) = (B^T \otimes A) \text{co}(X) .$$

Proof: Beginning with AXB , its ij -th element is

$$(AXB)_{ij} = \sum_{k=1}^m a_{ik} \sum_{\lambda=1}^n x_{k\lambda} b_{\lambda j}$$

$$= \sum_{\lambda=1}^n b_{\lambda j} \sum_{k=1}^m a_{ik} x_{k\lambda}$$

The last line is precisely $(B^T \otimes A) \text{co}(X)$.

To apply Theorems (3.50) and (3.52), a more convenient form of (3.44) is

$$(3.54) \quad A_x V I_{m_t} - I_{m_x} V A_t = G$$

where I_q ($q = m_t, m_x$) is the $q \times q$ identity matrix. Concatenating (3.54) and using the symmetry of A_t , the discrete system admits the equivalent representation

$$(3.55) \quad (I_{m_t} \otimes A_x - A_t \otimes I_{m_x}) \text{co}(V) = \text{co}(G)$$

which is referred to as the Kronecker sum form. The coefficient matrix

$$(3.56) \quad B^{(2)} = I_{m_t} \otimes A_x - A_t \otimes I_{m_x} = (B_{ij}) \quad , \quad -M_t \leq i, j \leq N_t$$

has an easily discerned block form. The square blocks B_{ij} have dimension $m_x \times m_x$ and are given by

$$(3.57) \quad B_{ij} = \delta_{ij}^{(0)} A_x - (A_t)_{ij} I_{m_x} \quad , \quad -M_t \leq i, j \leq N_t$$

where $(A_t)_{ij}$ is the ij -th element of the matrix A_t . The vectors $\text{co}(V)$ and $\text{co}(G)$ are related to U and F by

$$(3.58) \quad \begin{aligned} \text{co}(V) &= \text{co}(D((\phi')^{-\frac{1}{2}}) U D((\tau')^{-\frac{1}{2}})) \\ &= \{D((\tau')^{-\frac{1}{2}}) \otimes D((\phi')^{-\frac{1}{2}})\} \text{co}(U) \end{aligned}$$

and

$$(3.59) \quad \text{co}(G) = \{D((\tau')^{-\frac{1}{2}}) \otimes D((\phi')^{-\frac{1}{2}})\} \text{co}(F)$$

From (3.58), it is evident that (3.55) is the matrix

formulation which arises via a natural ordering of the sinc gridpoints from left to right, bottom to top where (x_i, t_j) follows (x_k, t_p) if $t_j > t_p$ or if $t_j = t_p$ and $x_i > x_k$.

The previous discussion may be generalized to the second order hyperbolic problems in two and three space dimensions. Explicitly, the problems considered are

$$\begin{aligned} Lu(x,y,t) &= u_{tt} - \nabla^2 u = f(x,y,t) \\ (x,y,t) &\in (0,1)^2 \times (0,\infty) \end{aligned}$$

$$\begin{aligned} (3.60) \quad u|_{\partial(0,1)^2} &= 0 & t \in [0,\infty) \\ u|_{t=0} &= u_t|_{t=0} = 0 & (x,y) \in [0,1]^2 \end{aligned}$$

and

$$\begin{aligned} Lu(x,y,z,t) &= u_{tt} - \nabla^2 u = f(x,y,z,t) \\ (x,y,z,t) &\in (0,1)^3 \times (0,\infty) \end{aligned}$$

$$\begin{aligned} (3.61) \quad u|_{\partial(0,1)^3} &= 0 & t \in [0,\infty) \\ u|_{t=0} &= u_t|_{t=0} = 0 & (x,y,z) \in [0,1]^3 \end{aligned}$$

where $(0,1)^n$ and $[0,1]^n$ ($n = 2,3$), respectively, are the open and closed n -dimensional unit cubes. The spatial variables are all on the same interval as a matter of ease rather than necessity. In particular, this simplifies the construction of basis elements on each spatial interval because the map ϕ (3.2) is used repetitively. Hence the Sinc-Galerkin approximate solutions to (3.60) and (3.61), respectively, are

$$(3.62) \quad \hat{u}_{m_x, m_y, m_t}(x, y, t) = \sum_{i=-M_x}^{N_x} \sum_{j=-M_y}^{N_y} \sum_{\lambda=-M_t}^{N_t} u_{ij\lambda} S_{ij\lambda}(x, y, t)$$

and

$$(3.63) \quad \hat{u}_{m_x, m_y, m_z, m_t}(x, y, z, t) = \sum_{i=-M_x}^{N_x} \sum_{j=-M_y}^{N_y} \sum_{k=-M_z}^{N_z} \sum_{\lambda=-M_t}^{N_t} u_{ijk\lambda} S_{ijk\lambda}(x, y, z, t)$$

where

$$(3.64) \quad S_{ij\lambda}(x, y, t) = S_i(x) S_j(y) S_{\lambda}^*(t) ,$$

$$(3.65) \quad S_{ijk\lambda}(x, y, z, t) = S_i(x) S_j(y) S_k(z) S_{\lambda}^*(t) ,$$

$$(3.66) \quad m_q = M_q + N_q + 1 \quad q = x, y, z, t ,$$

and S_p ($p = i, j, k$) and S_{λ}^* are given by (3.6) and (3.7). As usual, the unknown coefficients $\{u_{ij\lambda}\}$ or $\{u_{ijk\lambda}\}$ are found by orthogonalizing the residual using the weight function

$$(3.67) \quad w(x, y, t) = 1/(\phi'(x)\phi'(y)\psi'(t))^{\frac{1}{2}}$$

for the three-dimensional problem and

$$(3.68) \quad w(x, y, z, t) = 1/(\phi'(x)\phi'(y)\phi'(z)\psi'(t))^{\frac{1}{2}}$$

for the four-dimensional problem.

Analogous to the two-dimensional case, the parameter selections are deduced from separate one-dimensional problems. The form of the one-dimensional differential equations in y and z are assumed to be like that of problem (3.14). The choices for the y parameters are

$$(3.69) \quad h_y = (\pi d / (5M_y))^{\frac{1}{2}}$$

and

$$(3.70) \quad N_Y = \left\| \frac{\zeta}{\eta} M_Y + 1 \right\|$$

based upon the assumption

$$(3.71) \quad |u(y)(\phi'(y))^{\frac{1}{2}}| \leq L \quad \begin{cases} y^{\zeta} & y \in (0, \frac{1}{2}) \\ (1-y)^{\eta} & y \in [\frac{1}{2}, 1) \end{cases}$$

where L , ζ , and η are positive constants. With respect to the z parameters, an assumption like (3.71) with ζ replaced by μ and η replaced by ν gives

$$(3.72) \quad h_Z = (\pi d / (\mu M_Z))^{\frac{1}{2}}$$

and

$$(3.73) \quad N_Z = \left\| \frac{\mu}{\nu} M_Z + 1 \right\|$$

To list the linear systems for $\{u_{ij\lambda}\}$ and $\{u_{ijk\lambda}\}$ additional notational devices are necessary. For $U = (u_{ij})$, $1 \leq i \leq m$ and $1 \leq j \leq n$, U is represented as a matrix. To view $U^{(3)} = (u_{ij\lambda})$, $1 \leq i \leq m$, $1 \leq j \leq n$, and $1 \leq \lambda \leq p$, as a matrix define

$$(3.74) \quad \begin{aligned} \text{mat}(U^{(3)}) &= \text{mat}((u_{ij\lambda})) \\ &= [\text{co}(u_{ij1}), \text{co}(u_{ij2}), \dots, \text{co}(u_{ijp})] \end{aligned}$$

Conveniently, concatenating the array $U^{(3)}$ is equivalent to concatenating the matrix; i.e.,

$$(3.75) \quad \begin{aligned} \text{co}(U^{(3)}) &= \text{co}((u_{ij\lambda})) \\ &= \begin{bmatrix} \text{co}(u_{ij1}) \\ \text{co}(u_{ij2}) \\ \vdots \\ \text{co}(u_{ijp}) \end{bmatrix} \\ &= \text{co}(\text{mat}(U^{(3)})) \end{aligned}$$

A recursive definition suffices to generalize concatenation for the n -subscripted array $U^{(n)} = (u_{i_1 i_2 \dots i_n})$,

$1 \leq i_j \leq m_j, 1 \leq j \leq n$, as follows:

$$\begin{aligned}
 \text{co}(U^{(n)}) &= \text{co}((u_{i_1 i_2 \dots i_n})) \\
 (3.76) \quad &\equiv \begin{bmatrix} \text{co}(u_{i_1 i_2 \dots i_{(n-1)} 1}) \\ \text{co}(u_{i_1 i_2 \dots i_{(n-1)} 2}) \\ \vdots \\ \text{co}(u_{i_1 i_2 \dots i_{(n-1)} m_n}) \end{bmatrix}
 \end{aligned}$$

In turn, using (3.76) a recursive formula for the matrix representation is easily given by

$$\begin{aligned}
 \text{mat}(U^{(n)}) &= \text{mat}((u_{i_1 i_2 \dots i_n})) \\
 (3.77) \quad &\equiv [\text{co}(u_{i_1 i_2 \dots i_{(n-1)} 1}), \text{co}(u_{i_1 i_2 \dots i_{(n-1)} 2}), \\
 &\quad \dots, \text{co}(u_{i_1 i_2 \dots i_{(n-1)} m_n})]
 \end{aligned}$$

Hence, going from $\text{mat}(U^{(n)})$ to $\text{co}(U^{(n)})$ involves unraveling the last index of $U^{(n)}$. These devices provide a convenient means to compactly write the systems for unknown coefficients when the spatial dimension is greater than or equal to two.

In two spatial variables, the system analogous to (3.44) is

$$(3.78) \quad \{I_{m_Y} \otimes A_X + A_Y \otimes I_{m_X}\} \text{mat}(V^{(3)}) - \text{mat}(V^{(3)}) A_t = \text{mat}(G^{(3)})$$

where for $-M_x \leq i \leq N_x$, $-M_y \leq j \leq N_y$ and $-M_t \leq \lambda \leq N_t$

$$(3.79) \quad \begin{aligned} \text{mat}(V^{(3)}) &= \text{mat}((v_{ij\lambda})) \\ &= D_{yx} \text{mat}(U^{(3)}) D((\tau')^{-\frac{1}{2}}) \end{aligned}$$

$$(3.80) \quad D_{yx} = D((\phi'(y))^{-\frac{1}{2}}) \otimes D((\phi'(x))^{-\frac{1}{2}})$$

and A_y is the same as A_x in (3.42) with h_x replaced by h_y . The matrix $\text{mat}(G^{(3)})$ is defined just like $\text{mat}(V^{(3)})$ with $U^{(3)}$ replaced by $F = (f(x_i, y_j, t_\lambda))$ where the point (x_i, y_j, t_λ) is a sinc gridpoint. Adding a third variable yields

$$(3.81) \quad \begin{aligned} &\{I_{m_z m_y} \otimes A_x + I_{m_z} \otimes A_y \otimes I_{m_x} + A_z \otimes I_{m_y m_x}\} \text{mat}(V^{(4)}) \\ &- \text{mat}(V^{(4)}) A_t = \text{mat}(G^{(4)}) \end{aligned}$$

where

$$(3.82) \quad I_{pq} = I_p \otimes I_q$$

and for $-M_x \leq i \leq N_x$, $-M_y \leq j \leq N_y$, $-M_z \leq k \leq N_z$ and $-M_t \leq \lambda \leq N_t$

$$(3.83) \quad \begin{aligned} \text{mat}(V^{(4)}) &= \text{mat}((v_{ijk\lambda})) \\ &= D_{zyx} \text{mat}(U^{(4)}) D((\tau')^{-\frac{1}{2}}) \end{aligned}$$

$$(3.84) \quad D_{zyx} = D((\phi'(z))^{-\frac{1}{2}}) \otimes D((\phi'(y))^{-\frac{1}{2}}) \otimes D((\phi'(x))^{-\frac{1}{2}})$$

The matrix A_z is A_x with h_x replaced by h_z , and

$$(3.85) \quad \text{mat}(G^{(4)}) = D_{zyx} \text{mat}(F) D((\tau')^{-\frac{1}{2}})$$

where $F = (f(x_i, y_j, z_k, t_\lambda))$. Note that in (3.78) and (3.81) the coefficient matrices which multiply $\text{mat}(V^{(j)})$, $j = 3, 4$, on the left represent the discretization of the Laplacian operators in (3.60) and (3.61), respectively. Alternative definitions of $\text{mat}(V^{(n)})$ can break up the discretized Laplacian in n -dimensions.

The Kronecker sum form for the discrete system in two spatial variables is derived by concatenating (3.78). That is,

$$\begin{aligned}
 \text{co}(G^{(3)}) &= \text{co}(\text{mat}(G^{(3)})) \\
 &= \text{co} \left[\{I_{m_y} \otimes A_x + A_y \otimes I_{m_x}\} \text{mat}(V^{(3)}) I_{m_t} \right. \\
 &\quad \left. - I_{m_y m_x} \text{mat}(V^{(3)}) A_t \right] \\
 &= I_{m_t} \otimes \{I_{m_y} \otimes A_x + A_y \otimes I_{m_x}\} \text{co}(\text{mat}(V^{(3)})) \\
 (3.86) \quad &\quad - A_t^T \otimes I_{m_y m_x} \text{co}(\text{mat}(V^{(3)})) \\
 &= \left[I_{m_t} \otimes \{I_{m_y} \otimes A_x + A_y \otimes I_{m_x}\} \right. \\
 &\quad \left. - A_t \otimes I_{m_y m_x} \right] \text{co}(V^{(3)})
 \end{aligned}$$

Similarly, the vectorized version of (3.81) is

$$\begin{aligned}
 (3.87) \quad &\left[I_{m_t} \otimes \{I_{m_z m_y} \otimes A_x + I_{m_z} \otimes A_y \otimes I_{m_x} + A_z \otimes I_{m_y m_x}\} \right. \\
 &\quad \left. - A_t \otimes I_{m_z m_y m_x} \right] \text{co}(V^{(4)}) = \text{co}(G^{(4)})
 \end{aligned}$$

By definition of concatenation, the grid is swept in a natural ordering: first across the x-domain, then the y, followed by the z (for (3.87)), and lastly through the t-domain. A closer examination of this structure as well as the solution is developed in the next chapter.

CHAPTER 4

SOLUTION OF THE DISCRETE SINC-GALERKIN SYSTEM

Classical Methods

As mentioned in Chapters 1 and 3, typical Galerkin schemes for time dependent partial differential equations use a trial function which is a truncated expansion of basis functions defined over the spatial domain and having time dependent coefficients. For instance, the form of a trial function for (3.1) is

$$(4.1) \quad \hat{u}(x,t) = \sum_{i=1}^N c_i(t) \theta_i(x)$$

where $\{\theta_i\}_{i=1}^{\infty}$ is complete in some function space containing the true solution $u(x, \cdot)$. Orthogonalizing the residual with respect to θ_j , $1 \leq j \leq N$, leads to a system of ordinary differential equations in time. The literature abounds with algorithms to solve this type of system on a discrete, truncated time grid [14], [17].

In contrast, the Sinc-Galerkin method defines the basis functions on the entire space-time domain. The coefficients of the trial function are now constants, hence a system of ordinary differential equations is exchanged for a system of linear equations. Linear systems also arise when classic

schemes such as finite differences or finite elements are applied to (3.1) on a truncated time domain. Several well-established routines exist for the numerical solution of these systems. In light of developed techniques such as finite differences or more common Galerkin methods, a numerically feasible scheme which solves the discrete systems (derived in Chapter 3) is essential if the Sinc-Galerkin method is to represent a viable alternative for approximating solutions of (3.1), (3.60), and (3.61).

Conventional algorithms typically solve linear systems written in the form $B\vec{x} = \vec{b}$ where $\vec{x}$ is a vector of unknowns and B is called the coefficient matrix. In the present setting the discrete Sinc-Galerkin systems (3.55), (3.86), and (3.87) have the coefficient matrices

$$(4.2) \quad B^{(2)} = I_{m_t} \otimes A_x - A_t \otimes I_{m_x},$$

$$(4.3) \quad B^{(3)} = I_{m_t} \otimes \{I_{m_y} \otimes A_x + A_y \otimes I_{m_x}\} - A_t \otimes I_{m_y m_x}$$

and

$$(4.4) \quad B^{(4)} = I_{m_t} \otimes \{I_{m_z m_y} \otimes A_x + I_{m_z} \otimes A_y \otimes I_{m_x} \\ + A_z \otimes I_{m_y m_x}\} - A_t \otimes I_{m_z m_y m_x},$$

respectively. Table 1 lists the dimensions of $B^{(j)}$, $j = 2, 3, 4$, and suggests a first consideration for a feasible scheme -- machine storage. Methods which require full storage mode are often impractical in the setting of systems arising from the numerical solution of partial differential equations. This is certainly true here. For example, if a

Table 1. Machine Storage Information for the Matrices $B^{(j)}$ and $B_D^{(j)}$.

Equation	(3.55)	(3.86)	(3.87)
j	2	3	4
Dimension of $B^{(j)}$	$m_x m_t \times m_x m_t$	$m_x m_y m_t \times m_x m_y m_t$	$m_x m_y m_z m_t \times m_x m_y m_z m_t$
Nonzero elements of $B^{(j)}$	$m_x m_t (m_x + m_t - 1)$	$m_x m_y m_t (m_x + m_y + m_z - 2)$	$m_x m_y m_z m_t (m_x + m_y + m_z + m_t - 3)$
Nonzero elements of $B_D^{(j)}$	$m_x m_t m_t$	$m_x m_y m_t m_t$	$m_x m_y m_z m_t m_t$
Unknowns	$m_x m_t$	$m_x m_y m_t$	$m_x m_y m_z m_t$

coarse grid with $m_x = m_y = m_z = m_t = 10$ is used $B^{(4)}$ has order 10^4 ; i.e., 10^8 elements must be stored of which 370,000 are nonzero. The problem persists when $B^{(2)}$ has a fine grid like $m_x = m_t = 50$. In this case $B^{(2)}$ is 2500×2500 ; however, of the more than six million elements represented only 247,500 are nonzero. Moreover, with regard to position the regularity of the nonzero elements suggests the possibility of structure dependent algorithms with accompanying storage savings.

To put $B^{(2)}$ into perspective, consider the coefficient matrix for the usual centered finite difference scheme. Assume the order of the Sinc-Galerkin method is $O(e^{-\sqrt{m_x}})$ and that the true solution of (3.1) decays in time like e^{-t} . When $m_x = 50$, finite differences call for an approximate stepsize $h = .02$ to have an expected error of $O(h^2) = O(e^{-\sqrt{m_x}})$. Further, the time domain should not be truncated short of $t = 8$. This translates to 50 gridpoints in the spatial domain and 400 steps in time. Iterating requires the solution, at each time step, of a tridiagonal system with 148 nonzero entries. For purposes of comparison to $B^{(2)}$, these systems can be combined into one large system with 99,100 nonzero elements. Therefore, based on a count of nonzero elements finite differences is a more viable option than the present method. However, as demonstrated in Example 5.4, the $O(e^{-\sqrt{m_x}})$ convergence for the Sinc-Galerkin method is maintained when the solution u of (3.1) is

singular. Finite differences have no direct analogue maintaining $O(h^2)$ convergence in the presence of singularities.

The matrices occurring due to finite difference or finite element solutions of partial differential equations are often solved via iterative methods precisely because the methods can be coded to take advantage of well-structured, sparse coefficient matrices [4]. Given the structure of the discrete Sinc-Galerkin systems (3.55), (3.86), and (3.87) methods such as Jacobi, Gauss-Seidel, and SOR may, with some modification, be practical but they remain unexplored. A foreseeable modification is that greater advantage must be taken of the structure of $B^{(j)}$, $j = 2, 3, 4$, than just the location of nonzero elements. Storing only nonzero elements of $B^{(4)}$ is troublesome (see Table 1). Finally, with respect to convergence the success of an iterative method depends on the spectrum of the iteration matrix. In this direction, a detailed analysis of the spectrum of $B^{(j)}$, $j = 2, 3, 4$, would be useful. Whereas the author has done extensive numerical work on the spectrum of $B^{(2)}$, the analytic argument providing the proof of Conjecture (4.11) below remains elusive.

This chapter details two numerically viable methods which solve the discrete Sinc-Galerkin systems derived in Chapter 3. Both are direct methods that take advantage of symmetry and block structure to reduce machine storage as

well as ease implementation. The choice of technique is somewhat dependent on available computing facilities. For example, one algorithm is a modified block Gauss-Jordan elimination routine. Although it is possible to implement this method on a scalar machine, it is inherently suited to vector machine architecture. The other algorithm is well-suited to a scalar machine, in part, because it requires less machine storage than the block routine. Despite needing less storage, the scalar machine available to the author, a Honeywell Level 66, remains too small to implement the method in three space dimensions when

$m_x = m_y = m_z = m_t = 8$. In this instance a supercomputer is a necessity due to storage capabilities.

Solution of the Discrete System in One Spatial Variable

The discrete Sinc-Galerkin system (3.44) may be solved as follows. Since A_x and A_t are symmetric there are orthogonal matrices Q and P so that

$$(4.5) \quad Q^T A_x Q = \Lambda_x$$

and

$$(4.6) \quad P^T A_t P = \Lambda_t$$

where Λ_x , Λ_t are diagonal matrices containing the

eigenvalues $\{(\lambda_x)_i\}_{i=-M_x}^{N_x}$ and $\{(\lambda_t)_j\}_{j=-M_t}^{N_t}$ of A_x and A_t ,

respectively. If the change of variables

$$(4.7) \quad Z = Q^T V P$$

is made in (3.44) then the equation takes the form

$$(4.8) \quad \Lambda_x Z - Z \Lambda_t = H$$

where

$$(4.9) \quad H = Q^T G P.$$

The solution of (4.8), in component form, is given by

$$(4.10) \quad [(\lambda_x)_i - (\lambda_t)_j] z_{ij} = h_{ij}, \quad \begin{matrix} -M_x \leq i \leq N_x \\ -M_t \leq j \leq N_t \end{matrix}$$

where h_{ij} is the ij -th element of H . Once Z is determined from (4.10) the solution V of (3.44) is recovered via

$V = QZP^T$ and, in turn, using (3.45) gives U .

If $(\lambda_x)_i = (\lambda_t)_j$ for some indices i and j , the equation in (4.10) is inconsistent. In the array of examples listed in Chapter 5 this matching of eigenvalues never occurs. The author believes the following:

Conjecture (4.11): $\sigma(A_x) \cap \sigma(A_t) = \emptyset$ where $\sigma(A)$ denotes the spectrum of A .

This is connected, so the author believes, to the different nature of the spectrum of $Lu = -u$ on compact versus semi-infinite domains (see deBoor and Swartz [18]). While none of this has been proven, continuing analytic and numerical work supports the validity of the conjecture.

Verifying the consistency of the system posed as in (3.44) is equivalent to showing the matrix $B^{(2)}$ is nonsingular which, in turn, implies the existence of a unique solution for the linear equations (3.55). To

establish the connection a property of tensor products that Davis [16] states is useful; that is,

$$(4.12) \quad (A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

assuming each product is defined. Using (4.12) and transforming variables in (3.55) via

$$(4.13) \quad \text{co}(Z) = (P^T \otimes Q^T)\text{co}(V)$$

yields

$$(4.14) \quad \{I_{m_t} \otimes \Lambda_x - \Lambda_t \otimes I_{m_x}\}\text{co}(Z) = \text{co}(H)$$

where

$$(4.15) \quad \text{co}(H) = (P^T \otimes Q^T)\text{co}(G) .$$

Equation (4.14) shows that the eigenvalues of $B^{(2)}$ are given by all possible combinations of $(\lambda_x)_i - (\lambda_t)_j$, $-M_x \leq i \leq N_x$, $-M_t \leq j \leq N_t$. Equally significant is the fact that diagonalization of $B^{(2)}$ is accomplished by two intermediate diagonalizations of much smaller matrices. As a result, $B^{(2)}$ need never be stored. Indeed, if $m_x = m_t$, the largest array with respect to storage has $m_x m_t$ elements. This product represents the number of unknown coefficients in the trial function (3.10) (see Table 1). Finally, note that under the transformations (4.7) and (4.13) the algorithms solving (3.44) and (3.55), respectively, can be machine coded identically.

An alternative method of solution for (3.55) is described below. Consider the transformation

$$(4.16) \quad \begin{aligned} \text{co}(Z_D) &= \text{co}(Q^T V) \\ &= (I_{m_t} \otimes Q^T)\text{co}(V) . \end{aligned}$$

Rewriting (3.55) in terms of $\text{co}(Z_D)$ and multiplying through by $I_{m_t} \otimes Q^T$ yields

$$(4.17) \quad B_D^{(2)} \text{co}(Z_D) = \text{co}(H_D)$$

where

$$(4.18) \quad B_D^{(2)} = I_{m_t} \otimes \Lambda_x - A_t \otimes I_{m_x}$$

and

$$(4.19) \quad \text{co}(H_D) = \text{co}(Q^T G).$$

The matrix $B_D^{(2)}$ is a block matrix all of whose blocks are diagonal matrices. Explicitly, if the blocks are called $B_{D(ij)}$ then they are given by

$$(4.20) \quad B_{D(ij)} = \delta_{ij}^{(0)} \Lambda_x - (A_t)_{ij} I_{m_x} \quad -M_t \leq i, j \leq N_t$$

where $(A_t)_{ij}$ denotes the ij -th element of A_t . This improved structure has come at the minor expense of only one diagonalization of the $m_x \times m_x$ symmetric, negative definite matrix Λ_x . Moreover, machine storage for $B_D^{(2)}$ is considerably less than that needed to save the nonzero elements of $B^{(2)}$ (see Table 1). Block elimination techniques work extremely well on the system (4.17) as all matrix inversions and multiplications are performed on diagonal matrices. After the elimination procedure the solution is recovered from $\text{co}(V) = (I_{m_t} \otimes Q) \text{co}(Z_D) = \text{co}(QZ_D)$ from which U is found via $U = D((\Phi')^{\frac{1}{2}}) V D((\Upsilon')^{\frac{1}{2}})$.

Although block elimination techniques for $B_D^{(2)}$ may be performed on scalar machines, the structure of $B_D^{(2)}$ is

inherently suited to vectorized computation. A block Gauss-Jordan elimination routine solving (4.17) has been written in explicit vector FORTRAN and implemented on the CYBER 205 computer. Its performance was consistent with the results of the algorithm which diagonalizes both A_x and A_t .

Solution of the Systems in Two and Three Space Variables

The algorithms detailed in the previous section admit ready extensions for application to the discrete Sinc-Galerkin systems arising from (3.60) and (3.61). One extension is actually an assortment of methods utilizing block techniques. The second is based on diagonalizing A_x , A_y , A_t , and A_z (if it occurs in the system). The discussion begins with the second method.

A_y and A_z are symmetric hence there exist orthogonal matrices R and S such that

$$(4.21) \quad R^T A_y R = \Lambda_y$$

and

$$(4.22) \quad S^T A_z S = \Lambda_z$$

where Λ_y and Λ_z are diagonal matrices containing the eigenvalues of A_y and A_z , respectively. Equations (3.86) and (3.87) may be transformed using changes of variables analogous to (4.13) as follows. Let $Z^{(3)} = (z_{ij\lambda})$ and $Z^{(4)} = (z_{ijk\lambda})$, $-M_x \leq i \leq N_x$, $-M_y \leq j \leq N_y$, $-M_z \leq k \leq N_z$, $-M_t \leq \lambda \leq N_t$, be defined by

$$(4.23) \quad \text{co}(Z^{(3)}) = (P^T \otimes R^T \otimes Q^T) \text{co}(V^{(3)})$$

and

$$(4.24) \quad \text{co}(Z^{(4)}) = (P^T \otimes S^T \otimes R^T \otimes Q^T) \text{co}(V^{(4)})$$

where Q and P are given by (4.5) and (4.6), respectively.

Rewriting (3.86) in terms of $Z^{(3)}$ yields

$$(4.25) \quad \begin{aligned} & \left[I_{m_t} \otimes \{ I_{m_y} \otimes \Lambda_x + \Lambda_y \otimes I_{m_x} \} - \Lambda_t \otimes I_{m_y m_x} \right] \text{co}(Z^{(3)}) \\ & = (P^T \otimes R^T \otimes Q^T) \text{co}(G^{(3)}) \end{aligned}$$

while (3.87) in terms of $Z^{(4)}$ becomes

$$(4.26) \quad \begin{aligned} & \left[I_{m_t} \otimes \{ I_{m_z m_y} \otimes \Lambda_x + I_{m_z} \otimes \Lambda_y \otimes I_{m_x} + \Lambda_z \otimes I_{m_y m_x} \} \right. \\ & \quad \left. - \Lambda_t \otimes I_{m_z m_y m_x} \right] \text{co}(Z^{(4)}) \\ & = (P^T \otimes S^T \otimes R^T \otimes Q^T) \text{co}(G^{(4)}) \end{aligned}$$

If the arrays $H^{(3)} = (h_{ij\lambda})$ and $H^{(4)} = (h_{ijkl})$,

$-M_x \leq i \leq N_x$, $-M_y \leq j \leq N_y$, $-M_z \leq k \leq N_z$, $-M_t \leq \lambda \leq N_t$, are given by

$$(4.27) \quad \text{co}(H^{(3)}) = (P^T \otimes R^T \otimes Q^T) \text{co}(G^{(3)})$$

and

$$(4.28) \quad \text{co}(H^{(4)}) = (P^T \otimes S^T \otimes R^T \otimes Q^T) \text{co}(G^{(4)})$$

then the solutions of (4.25) and (4.26), respectively, are

$$(4.29) \quad [(\lambda_x)_i + (\lambda_y)_j - (\lambda_t)_\lambda] z_{ij\lambda} = h_{ij\lambda}$$

and

$$(4.30) \quad [(\lambda_x)_i + (\lambda_y)_j + (\lambda_z)_k - (\lambda_t)_\lambda] z_{ijk\lambda} = h_{ijk\lambda}.$$

As in (4.10), the possibility exists that (4.29) or (4.30) is inconsistent. Again this never happens for the examples discussed in Chapter 5. The author believes in the validity

of the analogue of Conjecture (4.11) for (4.29) and (4.30). That is, based on the array of examples tested, if the spatial domain is compact while the temporal is semi-infinite then, with respect to the sinc discretization the spectrum of the discretized Laplacian is disjoint from the spectrum of A_t . Assuming the conjecture is valid, the unknown coefficients $u_{ij\lambda}$ in the expansion (3.62) are recovered from (4.29) via (4.23) and (3.79). Similarly, (4.30) may be transformed using (4.24) followed by (3.83) to give $\{u_{ijk\lambda}\}$ appearing in the trial function (3.63).

One item remains to completely specify the algorithm. When solving the system as above the matrix-vector products which occur are highly structured. Taking advantage of that structure allows the products to be determined while retaining use of three and four subscripts. For instance, suppose $X = (x_{ijk})$, $1 \leq i \leq m$, $1 \leq j \leq n$, $1 \leq k \leq p$; $A = (a_{ij})$, $1 \leq i, j \leq m$; $B = (b_{ij})$, $1 \leq i, j \leq n$; and $C = (c_{ij})$, $1 \leq i, j \leq p$. The product $(C \otimes B \otimes A)co(X)$ is defined by

$$(4.31) \quad [(C \otimes B \otimes A)co(X)]_{ijk} = \sum_{t=1}^p c_{kt} \sum_{s=1}^n b_{js} \sum_{r=1}^m a_{ir} x_{rst}$$

where $[]_{ijk}$ indicates the ijk -th element. Similarly, if A , B , and C are as before; $X = (x_{ijk\lambda})$, $1 \leq i \leq m$, $1 \leq j \leq n$, $1 \leq k \leq p$, $1 \leq \lambda \leq q$; and $D = (d_{ij})$, $1 \leq i, j \leq q$; then $(D \otimes C \otimes B \otimes A)co(X)$ is given elementwise via

$$\begin{aligned}
 & [(D \otimes C \otimes B \otimes A) \text{co}(X)]_{ijkl} \\
 (4.32) \quad & = \sum_{v=1}^q d_{\lambda v} \sum_{t=1}^p c_{kt} \sum_{s=1}^n b_{js} \sum_{r=1}^m a_{ir} x_{rstv}
 \end{aligned}$$

Hence, the code needed to evaluate the vector $\text{co}(H^{(3)})$ or $\text{co}(H^{(4)})$ is merely a set of nested loops. Likewise, recovering the trial function coefficients from $Z^{(3)}$ and $Z^{(4)}$ is equally simple.

Note that the diagonalization of A_x , A_y , A_t , and A_z (when it occurs) is the definitive characteristic of the algorithm just described. Indeed, when these diagonalizations are carried out, the numerical solution of the linear systems posed as in (3.86) and (3.87) is indistinguishable from the numerical solution of (3.78) and (3.81), respectively. Further, the chief attributes of this method for the numerical solution of the discrete systems are a direct consequence of the diagonalizations. The first, implementation ease, is established by the preceding discussion. The second is the minimal machine storage needed. The maximum array size required to solve (3.86) is $m_x m_y m_t$ while for (3.87) it is $m_x m_y m_z m_t$ (see Table 1). These values represent the number of unknown coefficients occurring in the trial function and, as such, also give the minimum array size needed. Hence, with regard to machine storage, it is not possible to do better than this method.

The remaining class of algorithms is only briefly mentioned as it has yet to be implemented. The

distinguishing trait of the class is that at least one of the matrices A_x , A_y , A_t , or A_z (if it occurs) remains undiagonalized. Here A_t will not be diagonalized. As an example, the change of variables

$$(4.33) \quad \text{co}(Z_D^{(3)}) = (I_{m_t} \otimes R^T \otimes Q^T) \text{co}(V^{(3)})$$

transforms (3.86) to

$$(4.34) \quad B_D^{(3)} \text{co}(Z_D^{(3)}) = (I_{m_t} \otimes R^T \otimes Q^T) \text{co}(G^{(3)})$$

where

$$(4.35) \quad B_D^{(3)} = I_{m_t} \otimes \{I_{m_y} \otimes \Lambda_x + \Lambda_y \otimes I_{m_x}\} - A_t \otimes I_{m_y m_x}$$

$B_D^{(3)}$ is a block matrix all of whose blocks are diagonal matrices. Therefore, one appropriate choice of solution is a block Gauss-Jordan elimination routine. However, a complication arises due to the added space dimension. That is, either the block routine must be written to handle unknowns with three subscripts or the array of unknowns must be adapted to two subscripts. In the latter case, the block routines used to solve (4.17) could be applied although, they may not take advantage of all vectorization possible.

The dilemma is compounded for the system

$$(4.36) \quad B_D^{(4)} \text{co}(Z_D^{(4)}) = (I_{m_t} \otimes S^T \otimes R^T \otimes Q^T) \text{co}(G^{(4)})$$

where

$$(4.37) \quad B_D^{(4)} = I_{m_t} \otimes \{I_{m_z m_y} \otimes \Lambda_x + I_{m_z} \otimes \Lambda_y \otimes I_{m_x} \\ + \Lambda_z \otimes I_{m_y m_x}\} - A_t \otimes I_{m_z m_y m_x}$$

and

$$(4.38) \quad \text{co}(Z_D^{(4)}) = (I_{m_t} \otimes S^T \otimes R^T \otimes Q^T) \text{co}(V^{(4)})$$

Here the array of unknowns has four subscripts.

To further cloud the picture, the manner in which the transformed coefficient matrices are stored may depend on the transformation. In particular, if only A_x is diagonalized the resulting transformations of $B^{(3)}$ and $B^{(4)}$ require more storage in terms of nonzero elements than the matrices $B_D^{(3)}$ and $B_D^{(4)}$ (see Table 1).

Diagonalizing the matrices A_x , A_y , A_t , and possibly A_z appears the easier route in the setting of problems (3.1), (3.60), and (3.61) but the author believes that block techniques are more widely applicable. For instance, the inclusion in the partial differential equation of lower order terms and/or nonconstant coefficients may discourage diagonalization of the entire coefficient matrix.

CHAPTER 5

NUMERICAL EXAMPLES OF THE SINC-GALERKIN METHOD

The space-time Sinc-Galerkin method for (3.1), (3.60), and (3.61) was tested on a large class of problems whose known solutions exhibit differing behaviors. Choosing problems with known solutions allows for a more complete error evaluation. With regard to error, considerable numerical success is evident for all problems; hence, the nine examples reported herein are selected to highlight characteristics of the scheme. In addition the sample of nine breaks into three groups of three where within a group a given characteristic is illustrated for problems in one, two, and three space dimensions.

Examples 5.1 - 5.3 and 5.7 - 5.9 display a first trait; that is, optimal parameter selections for an expected convergence rate result in a discrete system of minimum size for the Sinc-Galerkin method. Moreover, symmetry of the discrete system is maintained for any choice of parameters (see Examples 5.7 - 5.9). A characteristic common to each of the nine examples is that the numerical solution (3.10), (3.62), or (3.63) (depending on space dimension) is global. In particular, by employing the conformal map $\hat{T}(t) = \Lambda_n(t)$ the approximate solution is valid on the infinite time

interval. Another significant property is exhibited by Examples 5.4 - 5.6 whose solutions are singular on their respective boundaries. With respect to implementation (that is, parameter selections), the Sinc-Galerkin method for these singular problems proceeds in the same fashion as for the analytic examples. Referring to the order statements following (5.7), the rate of convergence of the method is governed by the Lip α class to which the solution belongs--not the degree of singularity of the solution.

Perhaps the most distinguished feature of the Sinc-Galerkin method, indeed of spectral methods in general, is the potential exponential convergence rate [7]. For problem (2.1) this convergence is established analytically. Specifically, a direct consequence of Theorem (2.59) is that if $2N + 1$ basis functions are used to construct an approximate solution of (2.1) then the order $O(\exp(-K\sqrt{N}))$, $K > 0$, holds uniformly for both analytic and singular problems. With respect to problems (3.1), (3.60), and (3.61) and their respective approximates (3.10), (3.62), and (3.63), it is easy to establish the exponential convergence rate on the sinc grid. An analytic extension to uniform exponential convergence in multiple dimensions does not currently exist. Arguments based on analytic function theory (see Chapter 2 for the development in one variable) lead directly to problems in functions of several complex variables. Alternatively, arguments based on the

development as in Stenger [5] are complicated by the form of Green's functions in higher dimensions. The author believes that the exponential convergence rate of the Sinc-Galerkin method (indicated in the discussion following (5.12)) does hold uniformly in higher dimensions. Unfortunately, the development of analytic tools to verify this convergence has not kept pace with the development and numerical testing of the methods.

Since block techniques for the problems in two and three space dimensions have yet to be implemented, the results reported are for the systems solved via all possible diagonalizations. When feasible the code was run on a Honeywell Level 66 computer using a double precision ANS FORTRAN-77 compiler. For large systems arising from the discretization of (3.60) or (3.61) the array of unknowns exceeds the maximum array size allowed on the scalar machine mentioned. Hence, these systems were solved with the same code run on a CRAY XMP/48 using a single precision CFT.141 compiler. Single precision on the CRAY XMP/48 is equivalent to double precision on most scalar machines. Indeed, the CRAY XMP/48 reproduced the error results of smaller double precision Honeywell runs.

With regard to problems in one space dimension, the maximum absolute error between the numerical approximation, u_{iA} , and the true solution, $u(x_i, t_A)$, at the sinc gridpoints was determined and reported as

$$(5.1) \quad \|E_u^{(1)}(h)\| = \max_{i\lambda} |u_{i\lambda} - u(x_i, t_\lambda)|$$

where h is the stepsize associated with x . Similarly, for the problems in two and three space dimensions, the maximum errors are

$$(5.2) \quad \|E_u^{(2)}(h)\| = \max_{ij\lambda} |u_{ij\lambda} - u(x_i, y_j, t_\lambda)|$$

and

$$(5.3) \quad \|E_u^{(3)}(h)\| = \max_{ijk\lambda} |u_{ijk\lambda} - u(x_i, y_j, z_k, t_\lambda)|$$

respectively, where $u_{ij\lambda}$ and $u_{ijk\lambda}$ are numerical approximates of the true values $u(x_i, y_j, t_\lambda)$ and $u(x_i, y_j, z_k, t_\lambda)$. The notation .ddd- ν represents .ddd $\times 10^{-\nu}$.

To implement the method in one space dimension, assume the solution $u(x, t)$ of (3.1) satisfies

$$(5.4) \quad |u(x, t)| \leq Mx^{\alpha+\frac{1}{2}} (1-x)^{\beta+\frac{1}{2}} t^{\gamma+\frac{1}{2}} e^{-\delta t}$$

for some positive α , β , γ , and δ . As a consequence of (5.4) there exist constants K and L such that

$$(5.5) \quad |u(x, t)(\tau'(t))^{\frac{1}{2}}| \leq K \begin{cases} t^\gamma & , \quad t \in (0, 1) \\ e^{-\delta t} & , \quad t \in [1, \infty) \end{cases}$$

and

$$(5.6) \quad |u(x, t)(\phi'(x))^{\frac{1}{2}}| \leq L \begin{cases} x^\alpha & , \quad x \in (0, \frac{1}{2}) \\ (1-x)^\beta & , \quad x \in [\frac{1}{2}, 1) \end{cases}$$

hold uniformly for $(x, t) \in (0, 1) \times (0, \infty)$. These conditions taken in light of the one-dimensional problems (3.13) and (3.14), respectively, motivate the parameter selections for the approximate

$$(5.7) \quad \hat{u}_{m_x, m_t}(x, t) = \sum_{i=-M_x}^{N_x} \sum_{\lambda=-M_t}^{N_t} u_{i\lambda} S(i, h_x) \circ \phi(x) S(\lambda, h_t) \circ \tau(t),$$

$$m_x = M_x + N_x + 1$$

$$m_t = M_t + N_t + 1.$$

Recall that the asymptotic errors for the one-dimensional problem (3.13) are $O(\exp(-\pi d/h_t))$, $O(\exp(-\gamma M_t h_t))$, and $O(\exp(-\delta N_t h_t))$ (see (3.24)). These depend on the growth bound (3.23) which is analogous to (5.5). Similarly the asymptotic errors associated with the spatial problem (3.14) are $O(\exp(-\pi d/h_x))$, $O(\exp(-\alpha M_x h_x))$, and $O(\exp(-\beta N_x h_x))$. Once M_x is chosen, balancing the asymptotic errors for the one-dimensional problems with respect to $O(\exp(-\alpha M_x h_x))$ determines the following stepsizes and summation limits:

$$(5.8) \quad h_x = \pi / (2\alpha M_x)^{1/2}$$

where the angle d in Figure 6 is taken to be $\pi/2$,

$$(5.9) \quad N_x = \left\lceil \frac{\alpha}{\beta} M_x + 1 \right\rceil,$$

$$(5.10) \quad h_t = h_x,$$

and

$$(5.11) \quad M_t = \left\lceil \frac{\alpha}{\gamma} M_x + 1 \right\rceil.$$

Assuming the solution of (3.1) decays exponentially in time, a smaller value of N_t ,

$$(5.12) \quad N_t = \left\lceil \frac{1}{h_x} \ln \frac{\alpha}{\delta} M_x h_x + 1 \right\rceil,$$

may be chosen (see Lund [6]). Note that when $\frac{\alpha}{\beta} M_x$ or $\frac{\alpha}{\gamma} M_x$ is an integer these are the values used for N_x and M_t , respectively. The additional +1 appearing in (5.9), (5.11), and (5.12) guarantees that all the appropriate errors are at least $O(\exp(-\alpha M_x h_x))$.

Parameter selections for the approximate in two or three space dimensions are deduced via the same notion of balancing asymptotic errors. Hence, if the solution $u(x,y,t)$ of (3.60) satisfies

$$(5.13) \quad |u(x,y,t)| \leq K x^{\alpha+\frac{1}{2}} (1-x)^{\beta+\frac{1}{2}} y^{\zeta+\frac{1}{2}} (1-y)^{\eta+\frac{1}{2}} t^{\gamma+\frac{1}{2}} e^{-\delta t}$$

then the additional stepsize h_y and summation limits M_y and N_y for the approximate (3.62) are given by

$$(5.14) \quad h_y = h_x,$$

$$(5.15) \quad M_y = \left\lceil \frac{\alpha}{\zeta} M_x + 1 \right\rceil,$$

and

$$(5.16) \quad N_y = \left\lceil \frac{\alpha}{\eta} M_x + 1 \right\rceil.$$

Similarly for (3.61), assuming the solution $u(x,y,z,t)$ satisfies

$$(5.17) \quad |u(x,y,z,t)| \leq Kw(x,y,t) z^{\mu+\frac{1}{2}} (1-z)^{\nu+\frac{1}{2}}$$

where

$$(5.18) \quad w(x,y,t) = x^{\alpha+\frac{1}{2}} (1-x)^{\beta+\frac{1}{2}} y^{\zeta+\frac{1}{2}} (1-y)^{\eta+\frac{1}{2}} t^{\gamma+\frac{1}{2}} e^{-\delta t},$$

then the remaining parameters for the Sinc-Galerkin approximate (3.63) are

$$(5.19) \quad h_z = h_x ,$$

$$(5.20) \quad M_z = \left\| \frac{\alpha}{\mu} M_x + 1 \right\| ,$$

and

$$(5.21) \quad N_z = \left\| \frac{\alpha}{\nu} M_x + 1 \right\| .$$

For all examples in one or two space dimensions a sequence of runs with $M_x = 4, 8, 16$, and 32 is reported. In three space dimensions a run corresponding to $M_x = 32$ is not currently possible, even on the CRAY XMP/48, due to the considerable machine storage needed for the array of unknowns. With respect to all nine examples, whenever an error result is reported from a CRAY XMP/48 run a * appears to the left. This * also indicates that the run was too large to perform on the Honeywell machine available to the author.

In all cases reported, $M_t > N_t$ (see (5.12)), which yields a much smaller discrete system than the choice of N_t given by (3.26). The choice $M_t = N_t$ or that given by (3.26) results in larger matrices with no corresponding increase in accuracy. Worth noting is that as a consequence of the map $\Upsilon(t) = \Lambda n(t)$, the sinc gridpoints in time become quite large. Recalling (3.9), $t_{N_t} = e^{N_t h t}$. In all of the following results the selection $M_x = 16$, for example, leads to a choice of N_t that yields $t_{N_t} = e^{3(.7854)} = 10.55$. A

large number of iterations is typically required to approximate the solution for a t of similar magnitude when using time-marching schemes. Hence, in comparison, the Sinc-Galerkin method requires much smaller systems to attain the same accuracy as finite differences (see the discussion following Examples 5.7 - 5.9).

The parameter $\alpha = \frac{1}{2}$ is common to all nine examples hence the stepsize $h \equiv h_x = \pi/(2\alpha M_x)^{\frac{1}{2}}$ and the asymptotic error $O(\exp(-\alpha M_x h_x))$ are the same throughout this chapter. As a result; these values are not included in every table. Computationally the asymptotic rate (which for $d = \pi/2$ is $O(\exp[-\alpha M_x h_x]) = O(\exp[-\pi/(2\alpha M_x)^{\frac{1}{2}}])$ is consistently attained at the sinc gridpoints.

Exponentially Damped Sine Waves

Example 5.1: Damped sine wave in one-dimension.

$$\begin{aligned}Lv(x,t) &\equiv v_{tt}(x,t) - v_{xx}(x,t) \\&= \{\pi^2 + (2 - 4t + (1 + \pi^2)t^2)e^{-t}\}\sin(\pi x) \\v(0,t) &= v(1,t) = 0 \\v(x,0) &= \sin(\pi x), \quad v_t(x,0) = 0.\end{aligned}$$

This problem is transformed to the form (3.1) via

$u(x,t) = v(x,t) - \sin(\pi x)$, which yields

$$\begin{aligned}Lu(x,t) &= (2 - 4t + (1 + \pi^2)t^2)e^{-t}\sin(\pi x) \\u(0,t) &= u(1,t) = 0 \\u(x,0) &= u_t(x,0) = 0.\end{aligned}$$

The analytic solution is $u(x,t) = t^2 e^{-t} \sin(\pi x)$. The

parameters (see (5.4)) are $\alpha = \beta = 1/2$, $\gamma = 3/2$, and $\delta = 1$. Table 2 displays the maximum absolute error at the sinc gridpoints for the sequence $M_x = 4, 8, 16$, and 32 . Additionally a column with asymptotic error is given. Since $\alpha = \beta$, $M_x = N_x$. In this case the sum on i in (5.7) is referred to as a centered sum. The sum on k in (5.7) is a noncentered sum, as commonly happens when using (5.12).

Table 2. Numerical Results for Example 5.1.

M_x	N_x	M_t	N_t	h	$\ E_u^{(1)}(h)\ $	Asymptotic Error
4	4	2	1	1.57080	.378-1	.432-1
8	8	3	2	1.11072	.253-1	.118-1
16	16	6	3	.78540	.905-3	.188-2
32	32	11	4	.55536	.320-3	.138-3

Example 5.2: Damped sine wave in two-dimensions.

$$Lu(x,y,t) = (2 - 4t + (1 + 2\pi^2)t^2)e^{-t}\sin(\pi x)\sin(\pi y)$$

$$u|_{\partial(0,1)^2} = 0$$

$$u|_{t=0} = u_t|_{t=0} = 0,$$

where, recall, $(0,1)^n$ refers to the open n -dimensional unit cube. The analytic solution is

$$u(x,y,t) = t^2 e^{-t} \sin(\pi x) \sin(\pi y), \text{ hence the parameters are}$$

$\alpha = \beta = \zeta = \eta = 1/2$, $\gamma = 3/2$, and $\delta = 1$. Table 3 lists

$\|E_u^{(2)}(h)\|$, the maximum absolute error at the sinc

gridpoints. Note that the result corresponding to $M_x = 32$ was obtained on the CRAY XMP/48 as indicated by the *.

Table 3. Numerical Results for Example 5.2.

M_x	N_x	M_y	N_y	M_t	N_t	$\ E_u^{(2)}(h)\ $
4	4	4	4	2	1	.813-1
8	8	8	8	3	2	.760-1
16	16	16	16	6	3	.387-3
32	32	32	32	11	4	*.451-3

Example 5.3: Damped sine wave in three-dimensions.

$$Lu(x,y,z,t) = (2 - 4t + (1 + 3\pi^2)t^2)e^{-t}\sin(\pi x)\sin(\pi y)\sin(\pi z)$$

$$u|_{\partial(0,1)^3} = 0$$

$$u|_{t=0} = u_t|_{t=0} = 0$$

The true solution $u(x,y,z,t) = t^2e^{-t}\sin(\pi x)\sin(\pi y)\sin(\pi z)$ is the three-dimensional analogue of the previous two solutions. The results for this problem are given in Table 4.

Table 4. Numerical Results for Example 5.3.

M_x	N_x	M_y	N_y	M_z	N_z	M_t	N_t	$\ E_u^{(3)}(h)\ $
4	4	4	4	4	4	2	1	.325-0
8	8	8	8	8	8	3	2	*.123-1
16	16	16	16	16	16	6	3	*.451-3

With respect to Examples 5.1, 5.2, and 5.3, the maximum absolute value of the true solution is the same. Table 5 indicates that there is no consistent difference in the error results for the various space dimensions.

Table 5. Numerical Results for the Damped Sine Wave.

M_x	h	$\ E_u^{(1)}(h)\ $	$\ E_u^{(2)}(h)\ $	$\ E_u^{(3)}(h)\ $	Asymptotic Error
4	1.57080	.378-1	.813-1	.325-0	.432-1
8	1.11072	.253-1	.760-1	*.123-1	.118-1
16	.78540	.905-3	.387-3	*.451-3	.188-2
32	.55536	.320-3	*.451-3	----	.138-3

Singular Problems

Example 5.4: A singular problem in one space dimension.

$$Lu(x,t) = \left\{ (3 - 12t + 4t^2) x \ln(x) - \frac{4t^2}{x} \right\} \frac{e^{-t}}{4\sqrt{t}}$$

$$u(0,t) = u(1,t) = 0$$

$$u(x,0) = u_t(x,0) = 0$$

The true solution for this problem is

$u(x,t) = t^{3/2}e^{-t} x \ln(x)$. The solution is algebraically singular at $t = 0$ and logarithmically singular at $x = 0$. Although u_x is unbounded at $x = 0$, condition (3.36) is satisfied. Further, despite the different character of the singularities the method requires no modification with regard to implementation. For $\alpha = \beta = \frac{1}{2}$ and $\gamma = \delta = 1$, the stepsize is again $h = h_x = h_t = \pi/(M_x)^{1/2}$ and the asymptotic rate shown in Table 6 is achieved, despite the presence of singularities.

Table 6. Numerical Results for Example 5.4.

M_x	N_x	M_t	N_t	$\ E_u^{(1)}(h)\ $	Asymptotic Error
4	4	2	1	.321-2	.432-1
8	8	4	2	.405-2	.118-1
16	16	8	3	.852-3	.188-2
32	32	16	4	.767-4	.138-3

In fact, comparing Tables 6 and 2, this problem has slightly smaller errors with a minor increase in the number of gridpoints near $t = 0$ due to the algebraic singularity. The logarithmic singularity affects neither the performance of the method nor the system size.

Example 5.5: A singular problem in two space dimensions.

$$Lu(x,y,t) = \left\{ (3 - 12t + 4t^2) x \ln(x) y \ln(y) - 4t^2 \left[\frac{y}{x} \ln(y) + \frac{x}{y} \ln(x) \right] \right\} \frac{e^{-t}}{4\sqrt{t}}$$

$$u \Big|_{\partial(0,1)^2} = 0$$

$$u \Big|_{t=0} = u_t \Big|_{t=0} = 0$$

The true solution for this problem,

$u(x,t) = t^{3/2} e^{-t} x \ln(x) y \ln(y)$, exhibits the same algebraic singularity at $t = 0$ as the previous example. In addition, the solution is singular on the spatial boundary. If anything, the difficulties appear more severe here than in Example 5.4; however, the errors shown in Table 7 are slightly better than those in Table 6. The parameters used are $\alpha = \beta = \zeta = \eta = \frac{1}{2}$ and $\gamma = \delta = 1$.

Table 7. Numerical Results for Example 5.5.

M_x	N_x	M_y	N_y	M_t	N_t	$\ E_u^{(2)}(h)\ $
4	4	4	4	2	1	.221-2
8	8	8	8	4	2	.370-2
16	16	16	16	8	3	.235-3
32	32	32	32	16	4	*.530-4

Example 5.6: A singular problem in three space dimensions.

$$\begin{aligned}
 Lu(x,y,z,t) = & \left\{ (3 - 12t + 4t^2) x \ln(x) y \ln(y) z \ln(z) \right. \\
 & - 4t^2 \left[\frac{yz}{x} \ln(y) \ln(z) + \frac{xz}{y} \ln(x) \ln(z) \right. \\
 & \left. \left. + \frac{xy}{z} \ln(x) \ln(y) \right] \right\} \frac{e^{-t}}{4\sqrt{t}}
 \end{aligned}$$

$$u|_{\partial(0,1)^3} = 0$$

$$u|_{t=0} = u_t|_{t=0} = 0$$

The true solution is

$u(x,y,z,t) = t^{3/2}e^{-t} x \ln(x) y \ln(y) z \ln(z)$ and, with the addition of $\mu = \nu = \frac{1}{2}$, the parameters are identical to those in Example 5.5. Table 8 lists the results for the present example while Table 9 includes the errors for each of the singular problems.

Table 8. Numerical Results for Example 5.6.

M_x	N_x	M_y	N_y	M_z	N_z	M_t	N_t	$\ E_u^{(3)}(h)\ $
4	4	4	4	4	4	2	1	.354-2
8	8	8	8	8	8	4	2	*.286-3
16	16	16	16	16	16	8	3	*.405-4

Table 9. Numerical Results for the Singular Problems.

M_x	h	$\ E_u^{(1)}(h)\ $	$\ E_u^{(2)}(h)\ $	$\ E_u^{(3)}(h)\ $	Asymptotic Error
4	1.57080	.321-2	.221-2	.354-2	.432-1
8	1.11072	.405-2	.370-2	*.286-3	.118-1
16	.78540	.852-3	.235-3	*.405-4	.188-2
32	.55536	.767-4	*.530-4	----	.138-3

For runs corresponding to $M_x = 8$ and 16 the error associated with Example 5.6 is almost a full decimal place better than the singular examples in one and two space dimensions. Moreover, this improvement occurs despite the greater singularity of the problem. No such analogue exists when using finite differences.

Problems Dictating Noncentered Sums in All Indices

Example 5.7: Noncentered sums in one space dimension.

$$Lu(x,t) = \{(2 - 4t + t^2) x (1 - x)^3 - 6t^2(1 - x)(2x - 1)\}e^{-t}$$

$$u(0,t) = u(1,t) = 0$$

$$u(x,0) = u_t(x,0) = 0$$

The analytic solution of this problem is

$u(x,t) = t^2 e^{-t} x(1-x)^3$. The parameter selections $\alpha = 1/2$, $\beta = 5/2$, $\gamma = 3/2$, and $\delta = 1$ dictate noncentered sums in both space and time ($M_x \neq N_x$ and $M_t \neq N_t$) as shown with the errors in Table 10. Again $h \equiv h_x = h_t = \pi/(M_x)^{1/2}$.

Table 10. Numerical Results for Example 5.7.

M_x	N_x	M_t	N_t	$\ E_u^{(1)}(h)\ $	Asymptotic Error
4	1	2	1	.237-2	.432-1
8	2	3	2	.208-2	.118-1
16	4	6	3	.707-4	.188-2
32	7	11	4	.237-4	.138-3

Even though there are three-fourths fewer gridpoints in the region $\frac{1}{2} < x < 1$, the same asymptotic rate $\exp(-M_x h_x/2)$ is predicted. In comparison to Examples 5.1 and 5.4 the results are better even with the smaller system size. This

is because the convergence rate of the Sinc-Galerkin method is governed by the asymptotic behavior of the solution.

Example 5.8: Noncentered sums in two space dimensions.

$$\begin{aligned} Lu(x,y,t) = & 10\{(2 - 4t + t^2) x (1 - x)^3 y^2(1 - y) \\ & - 2t^2[3(1 - x)(2x - 1)y^2(1 - y) \\ & + x(1 - x)^3(1 - 3y)]\}e^{-t} \end{aligned}$$

$$u \Big|_{\partial(0,1)^2} = 0$$

$$u \Big|_{t=0} = u_t \Big|_{t=0} = 0$$

The solution of this problem,

$u(x,y,t) = 10t^2e^{-t}x(1 - x)^3 y^2(1 - y)$, yields the parameter selections $\alpha = 1/2$, $\beta = 5/2$, $\zeta = 3/2$, $\eta = 1/2$, $\gamma = 3/2$, and $\delta = 1$. As a result of these parameter selections the sums appearing in the approximate (3.62) are all noncentered. A more complete discussion of how small this system is in comparison to a finite difference solution follows the next example. Note that the multiplicative factor of 10 appearing on the right-hand side of the problem was chosen so that the solution was of the same order of magnitude as the previous example. Table 11 gives the results for the current example.

Table 11. Numerical Results for Example 5.8.

M_x	N_x	M_y	N_y	M_t	N_t	$\ E_u^{(2)}(h)\ $
4	1	2	4	2	1	.805-2
8	2	3	8	3	2	.489-2
16	4	6	16	6	3	.867-4
32	7	11	32	11	4	*.682-3

Example 5.9: Noncentered sums in three space dimensions.

$$\begin{aligned}
 Lu(x,y,z,t) = & 100\{(2 - 4t + t^2)x(1 - x)^3y^2(1 - y)z^3(1 - z)^2 \\
 & - 2t^2[3(1 - x)(2x - 1)y^2(1 - y)z^3(1 - z)^2 \\
 & + x(1 - x)^3(1 - 3y)z^3(1 - z)^2 \\
 & + x(1 - x)^3y^2(1 - y)(10z^3 - 12z^2 + 3z)]\}e^{-t}
 \end{aligned}$$

$$u \Big|_{\partial(0,1)^3} = 0$$

$$u \Big|_{t=0} = u_t \Big|_{t=0} = 0$$

The analytic solution of this problem is

$$u(x,y,z,t) = 100t^2e^{-t}x(1 - x)^3y^2(1 - y)z^3(1 - z)^2.$$

Analogous to the factor of 10 in Example 5.8, the factor of 100 is a magnitude adjustment. The parameters $\alpha = 1/2$, $\beta = 5/2$, $\zeta = 3/2$, $\eta = 1/2$, $\mu = 5/2$, $\nu = 3/2$, $\gamma = 3/2$, and $\delta = 1$ yield the summation parameters appearing in Table 12. The composite error results for Examples 5.7 - 5.9 are given in Table 13.

Table 12. Numerical Results for Example 5.9.

M_x	N_x	M_y	N_y	M_z	N_z	M_t	N_t	$\ E_u^{(3)}(h)\ $
4	1	2	4	1	2	2	1	.395-2
8	2	3	8	2	3	3	2	*.389-3
16	4	6	16	4	6	6	3	*.313-4

Table 13. Numerical Results for Problems Dictating Noncentered Sums in All Indices.

M_x	h	$\ E_u^{(1)}(h)\ $	$\ E_u^{(2)}(h)\ $	$\ E_u^{(3)}(h)\ $	Asymptotic Error
4	1.57080	.237-2	.805-2	.395-2	.432-1
8	1.11072	.208-2	.489-2	*.389-3	.118-1
16	.78540	.707-4	.867-4	*.313-4	.188-2
32	.55536	.237-4	*.682-3	----	.138-3

The last three examples illustrate how the parameter selections described by (5.8) - (5.12), (5.14) - (5.16), and (5.19) - (5.21) minimize the expenditure of computational effort for the Sinc-Galerkin method. This also increases the method's competitiveness with respect to alternatives like finite differences. As discussed in Chapter 4, one measure of competitiveness is the number of nonzero elements in the coefficient matrices. For instance, the asymptotic

error corresponding to runs using $M_x = 16$ is $.188 \times 10^{-2}$. Maintaining this error when using finite differences requires a stepsize $h_{FD} \approx .04$ and iteration of the time domain to at least $t = 6.3$. This translates to 25 gridpoints on any of the spatial intervals and 158 steps in time. Solving the finite difference scheme as one system yields a matrix with 674,200 nonzero elements for Example 5.8 and 21,595,000 for Example 5.9. Their corresponding Sinc-Galerkin matrices have 251,160 and 3,294,060 nonzero entries. This evidence supports Stenger's [13] statement that the computational efficiency of the Sinc-Galerkin method becomes more apparent in higher dimensions. Further, the author notes that the computational savings exhibited herein are, in large part, due to including the time domain in the Galerkin procedure.

REFERENCES CITED

1. Farlow, S.J. Partial Differential Equations for Scientists and Engineers, John Wiley and Sons, New York, 1982.
2. Weinberger, H.F. A First Course in Partial Differential Equations, John Wiley and Sons, New York, 1965.
3. Botha, J.F. and Pinder, G.F. Fundamental Concepts in the Numerical Solution of Differential Equations, John Wiley and Sons, 1983.
4. Ames, W.F. Numerical Methods for Partial Differential Equations, 2nd ed., Academic Press, New York, 1977.
5. Stenger, F. "A Sinc-Galerkin Method of Solution of Boundary Value Problems." Mathematics of Computation 33 (January 1979): 85-109.
6. Lund, J. "Symmetrization of the Sinc-Galerkin Method for Boundary Value Problems." Mathematics of Computation 47 (October 1986): 571-588.
7. Gottlieb, D. and Orszag, S.A. Numerical Analysis of Spectral Methods: Theory and Application, SIAM, Philadelphia, 1977.
8. Whittaker, E.T. "On the Functions Which Are Represented by the Expansions of the Interpolation Theory." Proc. Roy. Soc. Edinburgh 35 (1915): 181-194.
9. Whittaker, J.M. Interpolatory Function Theory, Cambridge University Press, London, 1935.
10. McNamee, J., Stenger, F., and Whitney, J.L. "Whittaker's Cardinal Function in Retrospect." Mathematics of Computation 25 (January 1971): 141-154.
11. Stenger, F. "The Approximate Solution of Convolution-Type Integral Equations." SIAM J. Math. Anal. 4 (August 1973): 536-555.

REFERENCES CITED--Continued

12. Stenger, F. "Integration Formulas via the Trapezoidal Rule." J. Inst. Maths. Applics. 12 (1973): 103-114.
13. Stenger, F. "Numerical Methods Based on Whittaker Cardinal, or Sinc Functions." SIAM Review 23 (April 1981): 165-223.
14. Gear, W.C. Numerical Initial Value Problems in Ordinary Differential Equations, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.
15. McArthur, K.M., Bowers, K.L., and Lund, J. "Numerical Implementation of the Sinc-Galerkin Method for Second-Order Hyperbolic Equations." To appear in Numerical Methods for Partial Differential Equations, (1987).
16. Davis, P.J. Circulant Matrices, John Wiley and Sons, New York, 1979.
17. Isaacson, E. and Keller, H.B. Analysis of Numerical Methods, John Wiley and Sons, New York, 1966.
18. deBoor, C. and Swartz, B. "Collocation Approximation to Eigenvalues of an Ordinary Differential Equation: Numerical Illustrations." Mathematics of Computation 36 (January 1981): 1-19.

MONTANA STATE UNIVERSITY LIBRARIES



3 1762 10021490 5