SPLINE-WAVELETS IN NUMERICAL SOLUTIONS OF DIFFERENTIAL EQUATIONS

BY

Bader Ahmed Al-Humaidi

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE
In

MATHEMATICAL SCIENCES

May 2001

UMI Number: 1406105

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by Al-Humaidi, Bader under the direction of his thesis advisor and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN MATHEMATICAL SCIENCES**.

Thesis Committee

Akca

Dr. Akca, Haydar

Dr. Attili, Basem

Zaman

Dr. Zaman, Fiazuddin

Dr. Al-Sabah, Walid
Department Chairman

Prof. Jannadi, Osama
Dean of Graduate Studies

20/5/2001
Date

Acknowledgments

All praise be to God, Allah the Almighty Lord, without whose mercy and clemency nothing would have been possible. I am deeply obliged to Doctor Haydar Akca, my thesis supervisor. Without him this work would not have been accomplished.

I wish to express my thanks and gratitude to KFUPM and thesis committee members: Dr. Haydar Akca, Dr. Fiazuddin Zaman, and special thanks to Dr. Basem Attili.

I also thank Dr. Adnan Al-Shakhs, Dr. Mohammed Al-Bar, Dr. Mohammed El-Gebeily, Mr. Hussain Al-Sofi, and Dr. Waleed Al-Sabah, chairman of Mathematical department, for their concern and encouragement throughout my thesis work.

I extend my thanks to my family and friends specially my parents, my wife and my daughter Fatimah.

THESIS ABSTRACT

Full Name of Student:	**Bader Ahmed Al-Humaidi**
TITLE OF STUDY:	**Spline Wavelets in Numerical Solutions of Differential Equations**
MAJOR FIELD:	**Mathematics**
DATE OF DEGREE:	**May 2001**

In this thesis, we will consider the numerical solution of two point boundary value problems. In particular a comparison between two methods will be presented. Namely, finite differences with Fourier basis and Wavelet-Galerkin methods. It will be shown that the latter method, Wavelet-Galerkin methods, is more efficient in terms of stability.

MASTER OF SCIENCE DEGREE

KING FAHD UNIVERSITY O FPETROLEUM & MINERALS

Dhahran, Saudi Arabia

May, 2001

خلاصة الرسالة

اسم الطالب: بدر أحمد الحميدي

عنوان الدراسة: طريقة المطوعات باستخدام المويجات لحل المعادلات التفاضلية عديداً

التخصص: رياضيات

تاريخ الشهادة: مايو ٢٠٠١م الموافق صفر ١٤٢٢هـ

في هذه الرسالة سوف نستعرض الحلول العددية لمسائل القيم الحدية ذات النقطتين. وبالتحديد، سوف نعقد مقارنة بين طريقتين ذات أسّس مختلفة. الأولى هي طريقة الفروق المحدودة باستخدام أسّس "فوريّر". والثانية هي طريقة "جالركن" باستخدام المويجات. هذه الدراسة ستوضح أن الطريقة الثانية هي أفضل ثباتاً.

درجة الماجستير في العلوم الرياضية

جامعة الملك فهد للبترول والمعادن

الظهران – المملكة العربية السعودية

التاريخ مايو ٢٠٠١

Contents

PREFACE

The rapid development of wavelets and spline functions is due primarily to their great usefulness in applications. The enormous literature published during the last decades, shows that the actual development of wavelets and spline theory has an essential influence on large areas of modern applied (numerical) mathematics, such as; data fitting, interpolation and approximation, numerical treatment of operator equations, control theory, probability and statistics, image and signal transform and so on.

The work is organized as follows. Chapter 1 contains relevant basic definitions and notations about spline functions. Chapter 2 will cover general introduction of Fourier and wavelet transforms. A brief overview of wavelet analysis, multiresolution analysis and highlighting the important properties of spline representation and construction of functions are presented in Chapter 3. Finally, In chapter 4, we briefly consider the application of wavelets to the numerical solution of boundary value problems. As a further research problem we consider a general second order boundary value problem. The existence as well as the convergence of the solution for this BVP is introduced using Spline functions.

Chapter 1

Spline Functions

1.1 Introduction

Spline functions are piecewise polynomials of degree n that are connected together (at points called knots) so as to have $n-1$ continuous derivatives. They were first considered from a mathematical point of view by Schoenberg [39], and became the object of rather intensive research in the late 1950s [1], [7] and [37-40].

In general, there are two main categories of problems in approximation theory. The first category consists of problems in which approximation of unknown function is sought, based on a series of data (often-measurable data) regarding this function. These problems are called fitting-data problems. The second category of problems arises within the mathematical modeling of various processes in nature. Since these models usually lead to solving certain equations, they are called operational equations problems. Examples include, but not limited to, boundary-value problems for ordinary and partial differential equations, integral equations and optimal control problems. In both categories a new function is sought, one that approximates an unknown function with given properties. Two steps are necessary to produce such functions, see [1], [7] and [40].

i. Choosing a class of suitable functions where the approximation is sought

ii. Choosing an approximation process (an algorithm) to produce the function most suitable for the problem.

The success of this approach depends heavily on the existence of convergent class of approximating functions. To be of maximal use, a class of functions A must satisfy the following conditions:

i. The members of A are sufficiently smooth functions.

ii. The members of A together with their derivatives and integrals are easy to implement on a computer.

iii. A is large enough to possibly contain approximating functions for a wide range of processes and phenomena.

For a long time it was believed that polynomials were the most suitable as a class of approximating functions. Later, though, it was found that there are classes of approximation, not necessarily polynomials, which are more efficient for certain problems. In recent years the use of a number of special classes of approximation; that is, the so-called class of spline functions proved to be very beneficial both in approximation theory and numerical analysis.

1.2 The Definition and Fundamental Properties of the Spline Functions

Since polynomials played a central role in approximation theory and numerical analysis for many years [1], [40]. We will start by defining polynomials.

Definition 1.1

We call the space $P_m = \left\{ p(x) : p(x) = \sum_{i=1}^{m} c_i x^{i-1}, \quad c_1,\ldots,c_m, x \text{ are real and } c_m \neq 0 \right\}$ the space of polynomials of order m.

Note that P_m has so many attractive features. Some of them are

i. P_m is finite dimensional linear space with a convenient basis

ii. Polynomials are smooth functions and they are easy to store, manipulate and evaluate on a digital computer

iii. The derivative and antiderivative of a polynomial is again a polynomial

iv. Various matrices (arising in interpolation and approximation by polynomials) are always nonsingular, and they have strong sign-regularity properties.

The main drawback of the space P_m for approximation purposes is that the class is relatively inflexible. Polynomials seems to do all right on sufficiently small intervals, but when we go to larger intervals, severe oscillations often appear in particular if m is more than 3 or 4. This suggest that in order to have a class of approximating functions with greater flexibility, we should work with polynomials of relatively low degree, and the interval of interest should be divided into smaller pieces [1], [7] and [40].

Definition 1.2

Let $a = x_0 < x_1 < \ldots < x_k < x_{k+1} = b$, and write $\Delta = \{ x_i \}_0^{k+1}$ the set Δ partitions the interval $[a,b]$ into k+1 subintervals, $I_i = [x_i, x_{i+1})$, $i = 0,1,\ldots,k-1$ and $\mathrm{I_k} = [x_k, x_{k+1}]$.

Given a positive integer m, let

$$PP_m(\Delta) = \left\{ \begin{array}{l} f : \text{there exist polynomiaks } p_0, p_1, p_2, \ldots, p_k \; in \; P_m \\ \text{with } f(x) = p_i(x) \; for \; x \in I_i \; i = 0, 1, \ldots, k \end{array} \right\}$$

We call $PP_m(\Delta)$ the space of piecewise polynomials of order m with knots $x_1, \ldots, x_k$.

While it is clear that we have gained flexibility by going over from polynomials to piecewise polynomials, we have lost smoothness, which is important, since piecewise polynomials functions are not necessarily smooth.

Definition 1. 3

Let Δ be a partition of the interval [a, b] as in Definition (1. 2), and let m be a positive integer. Let $\delta_m(\Delta) = PP_m(\Delta) \cap C^{m-2}[a,b]$ where $PP_m(\Delta)$ is the space of piecewise polynomials in Definition (1.2). We call $\delta_m(\Delta)$ the space of polynomial splines of order m with simple knots at the points $x_1, \ldots, x_k$.

Note that this space has the same attractive features of polynomials, in addition, low-order splines are very flexible and do not exhibit oscillations usually associated with polynomials. So, we can say that spaces of smooth piecewise polynomials (splines) should be useful for approximation purposes. Now, we will give some basic properties of such spaces.

Let $[a, b]$ be a finite closed interval, and let $\Delta = \{x_i\}_1^k \; with \; a = x_0 < x_1 < \ldots < x_k < x_{k+1} = b$ being a partition of $[a, b]$ into k subintervals $I_i = [x_i, x_{i+1})$, $i = 0, 1, 2, \ldots, k-1$, and

$I_k = [x_k, x_{k+1}]$. Let m be a positive integer, and let $M = (m_1, m_2, ..., m_k)$ be a vector of integers with $1 \leq m_i \leq m$, $i = 1,2,...,k$.

Definition 1.4

We call the space

$$\delta(p_m; M, \Delta) = \left\{ \begin{array}{l} S: \textit{there exist polynomials } s_0\ , s_1, ..., s_k \textit{ in } P_m \textit{ such that } S(x) = s_i(x) \\ \textit{for } x \in I_i\ , i = 0,1,...,k, \textit{ and } D^j_{s_{i-1}}(x_i) = D^j_{s_i}(x_i) \textit{ for } j = 0,1,...,m-1-m_i\ , \\ i = 1,...,k \end{array} \right\}$$

the space of polynomial splines of order m with knots $x_1, x_2, ..., x_k$ of multiplicities $m_1, m_2, ..., m_k$. We call M the multiplicity vector. It controls the nature of the space $\delta(p_m; m, \Delta)$ by controlling the smoothness of the splines at the knots.

Definition 1.5

Let $... \leq y_{-1} \leq y_0 \leq y_1 \leq y_2 \leq ...$ be a real sequence. Given the integers i and $m > 0$, for any $x \in R$ define the functions

$$Q_i^m(x) = \begin{cases} (-1)^m\ [y_i, y_{i+1}, ..., y_{i+m}]\quad (x-y)_+^{m-1} & \textit{if } y_i < y_{i+m} \\ 0 & \textit{otherwise.} \end{cases}$$

The functions Q_i^m are called B-spline functions of order m associated with nodes $y_i, y_{i+1}, ..., y_{i+m}$. These kinds of functions can also be defined with simple recursive definition as follows:

assume $m \geq 2$ and $y_i < y_{i+m}$. Then for any $x \in R$

$$Q_i^m(x) = 1/(y_{i+m} - y_i) \quad [(x - y_i)\, Q_i^{m-1}(x) + (y_{i+m} - x)\, Q_{i+1}^{m-1}]$$

where $Q_i^1(x) = \begin{cases} 1/(y_{i+1} - y_i) & \textit{if} \quad y_i \leq x \leq y_{i+m} \\ 0 & \text{otherwise.} \end{cases}$

B-spline functions have finite support. That is,

$Q_i^m(x) = 0$ for $x < y_i$ and $x > y_{i+m}$ also $Q_i^m(x) > 0$ for $y_i < x < y_{i+m}$.

For computational purposes, functions whose eigenvalues are neither too large nor too small are preferred. This is the reason why we will introduce the so-called normalization of B-spline functions.

Definition 1.6

Let $N_i^m(x) = (y_{i+m} - y_i)\, Q_i^m(x)$ with the B-spline functions Q_i^m as defined in Definition 1.5. The functions N_i^m are called the normalized B-spline functions associated with the nodes $y_i, \ldots, y_{i+m}$.

In many problems in applied mathematics we seek the approximation of functions that are known to be periodic. As it is desirable to work with periodic approximation functions in such cases. We will now give the definition of periodic spline functions.

Definition 1.7

Assume $a < b$, identifying b with a, we may regard the interval $[a,b)$ as a circle with circumference $L = b - a$. Now given $\Delta = \{ a < x_1 < x_2 < \ldots < x_k < b \}$, we may think of Δ as a partition of the circle into k subintervals, $I_i = [x_i, x_{i+1}), i = 1, 2, \ldots, k-1$, and $I_k = [x_k, x_{k+1})$. Let m be a positive integer, and let $M := (m_1, m_2, \ldots, m_k)$ be a vector of integers with $1 \le m_i \le m$, $i = 1, 2, \ldots, k$ we define

$$\delta^0(p_m; M, \Delta) = \left\{ \begin{array}{l} S : \textit{there exist polynomials } s_1, \ldots, s_k \textit{ in } p_m \textit{ such that } S(x) = s_i(x) \\ \textit{for } x \in I_i \text{ , } i = 1, \ldots, k, \textit{ and } D_{s_{i-1}}^{j-1}(x_i) = D_{s_i}^{j-1}(x_i) \textit{ for } j = 1, \ldots, m - m_i \text{ ,} \\ i = 1, \ldots, k \textit{ where we take } s_0 = s_k. \end{array} \right\}$$

We call δ^0 the space of periodic polynomial splines of order m with knots at $x_1, \ldots, x_k$.

There are several spaces of piecewise polynomial functions that have proved useful in applications. These include natural splines, g-spline, mono splines, and discrete spline. The definition and fundamental properties of these concepts can be found in [39].

Definition 1.8

The m^{th} order cardinal B-spline is defined by

$$N_m(x) \equiv (N_{m-1} * N_1)(x) = \int_0^1 N_{m-1}(x-t)\, dt, \quad m \geq 2, \tag{1.1}$$

where N_1 is the characteristic function of the interval $[0,1)$.

Theorem 1. 9

The m^{th} order cardinal B-spline N_m satisfies the following properties:

i. For every $f \in C$,

$$\int_{-\infty}^{\infty} f(x) N_m(x)\, dx = \int_0^1 \ldots \int_0^1 f(x_1 + \ldots + x_m)\, dx_1 \ldots dx_m . \tag{1.2}$$

ii. For every $g \in C^m$,

$$\int_{-\infty}^{\infty} g^{(m)}(x)\, N_m(x) dx = \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k}\, g(k). \tag{1.3}$$

iii. $N_m(x) = \dfrac{1}{(m-1)!} \sum\limits_{k=0}^{m} (-1)^k \binom{m}{k} (x-k)_+^{m-1} = \dfrac{1}{(m-1)!} \Delta^m x_+^{m-1}$.

iv. $N_m'(x) = (\Delta N_{m-1})(x) = N_{m-1}(x) - N_{m-1}(x-1)$.

v. The cardinal B-splines N_m and N_{m-1} are related by the identity:

$$N_m(x) = \frac{x}{m-1} N_{m-1}(x) + \frac{m-x}{m-1} N_{m-1}(x-1). \tag{1.4}$$

vi. $SuppN_m = [0,m]$. Where $Supp\, N_m = \{x : N_m(x) \neq 0\}$.

vii. $N_m(x) > 0$, for $0 < x < m$.

viii. N_m is symmetric with respect to the center of its support, namely

$$N_m(\frac{m}{2}+x) = N_m(\frac{m}{2}-x).$$

ix. $$\begin{cases} N_2(k) = \delta_{k,1}, \quad k \in Z \text{ and} \\ N_{n+1}(k) = \frac{k}{n} N_n(k) + \frac{n-k+1}{n} N_n(k-1), \quad k = 1,\ldots,n. \end{cases}$$

Proof. (i) Assertion (1.2) certainly holds for$m = 1$. Suppose it holds for $(m-1)$, then by the definition of N_m in (1.1) and this induction hypothesis, we have

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\, N_m(x) dx &= \int_{-\infty}^{\infty} f(x) \left\{ \int_0^1 N_{m-1}(x-t)\, dt \right\} dx \\ &= \int_0^1 \left\{ \int_{-\infty}^{\infty} f(x)\, N_{m-1}(x-t) dx \right\} dt \\ &= \int_0^1 \left\{ \int_{-\infty}^{\infty} f(y+t)\, N_{m-1}(y)\, dy \right\} dt \\ &= \int_0^1 \int_0^1 \ldots \int_0^1 f(x_1 + \ldots + x_{m-1} + t)\, dx_1 \ldots dx_{m-1} dt \\ &= \int_0^1 \int_0^1 \ldots \int_0^1 f(x_1 + \ldots + x_{m-1} + x_m)\, dx_1 \ldots dx_{m-1} dx_m.\end{aligned}$$

(ii) Assertion (1.3) follows from (1.2) since

$$\int_0^1 \ldots \int_0^1 g^{(m)}(x_1 + \ldots + x_m) dx_1 \ldots dx_m = \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} g(k)$$

by direct integration.

(iv) Using (1.1), we have

$$N_m'(x) = \int_0^1 N_{m-1}'(x-t)\, dt = -N_{m-1}(x-1) + N_{m-1}(x) = (\Delta N_{m-1})(x).$$

(v) To verify this identity, we represent x_+^{m-1} as the product of a monomial and a truncated power, namely:

$$x_+^{m-1} = x \,.\, x_+^{m-2}$$

and then apply the following "Leibniz Rule" for differences:

$$(\Delta^n fg)(x) = \sum_{k=0}^{n} \binom{n}{k} (\Delta^k f)(x)\,(\Delta^{n-k} g)(x-k) \qquad (1.5)$$

Now, if we set $f(x) = x$ and $g(x) = x_+^{m-2}$ in (1.5) and recall that $\Delta^k f(x) = 0$ for $k \geq 2$, we will have

$$\begin{aligned} N_m(x) &= \frac{1}{(m-1)!}\Delta^m x_+^{m-1} \\ &= \frac{1}{(m-1)!}\left\{x\,\Delta^m x_+^{m-2} + m\Delta^{m-1}(x-1)_+^{m-2}\right\} \\ &= \frac{1}{(m-1)!}\left\{x[\Delta^{m-1}x_+^{m-2} - \Delta^{m-1}(x-1)_+^{m-2}] + m\Delta^{m-1}(x-1)_+^{m-2}\right\} \\ &= \frac{x}{m-1}N_{m-1}(x) + \frac{m-x}{m-1}N_{m-1}(x-1). \end{aligned}$$

The rest of the assertions can be easily derived by induction, using the definition of N_m and (1.4).

1.3 The Numerical Solutions of Differential Equations by Spline Functions.

In this section we will give an example to see how spline functions could be used to approximate the solution of differential equations. A procedure for obtaining spline function approximations for solutions of initial value problems in ordinary differential equations is presented [34] and [36].

Example 1.10

Given the differential equation

$$y' = f(x,y), \qquad 0 \le x \le b, \tag{1.6}$$

Assume that $f(x,y) \in C^{m-2}$ in T where $T = \{(x, y) \mid 0 \le x \le b\}$ assume also that it satisfies the Lipschitsz condition

$$| f(x, y) - f(x, y_1) | \le L | y - y_1 | \quad \text{if} \quad 0 \le x \le b. \tag{1.7}$$

If $m \ge 3$ then (1.7) is equivalent to the boundedness of $\partial f / \partial y$ in T. These conditions on $f(x, y)$ guarantee the existence of a unique solution to (1.6) for any initial condition. Our construction of the approximate solution $S(x) = S_m(x)$ is as follows. Let $y(x)$ be the solution of (1.6) determined by the initial value $y(0) = y_0$. Let $n > m$ be an integer, $h = \frac{b}{n}$ and let $S(x), 0 \le x \le b,$ be a spline function of degree m, class C^{m-1} and having its knots at the points $x = h, 2h, ..., (n-1)h$. We define the first component of $S(x) = S_m(x)$ by

$$S(x) = y(0) + y'(0)\,x + \cdots + \frac{1}{(m-1)!} y^{(m-1')}(0)\,x^{m-1} + \frac{1}{m!} a_0\, x^m, \qquad 0 \le x \le h, \tag{1.8}$$

with the last coefficient a_0 as yet undetermined. We now determine a_0 by requiring that $S(x)$ should satisfy (1.6) for $x = h$. This gives the equation

$$S'(h) = f(h, S(h)) \tag{1.9}$$

to be solved for a_0. It is seen that (1.9) is an equation in ζ where $\xi = {a_0 \, h^{m-1}} \big/ {(m-1)!}$ which is conveniently solved by iteration. Having determined the polynomial (1.8), we repeat the same steps in the interval [h, 2h]; that is,

$$S(x) = \sum_{k=0}^{m-1} \frac{1}{k!} S^{(k)}(h)\,(x-h)^k + \frac{1}{m!} a_1\,(x-h)^m, \qquad h \le x \le 2h, \tag{1.10}$$

and determine a_1 so as to satisfy the equation

$$S'(2h) = f(2h, S(2h)).$$

Continuing in the same manner we obtain a spline function $S_m(x)$ satisfying the equation

$$S_m{}'(v\,h) = f(\,v\,h, S_m(v\,h)\,), \qquad v = 0, 1, \cdots, n-1. \tag{1.11}$$

Theorem 1.11

If $h < m/L$, then the spline function $S_m(x)$ exists and is uniquely defined by the above construction.

Proof.

Over the interval $[\,v\,h, (v+1)\,h]$ we define

$$\begin{aligned} S(x) &= \sum_{k=0}^{m-1} \frac{1}{k!} S^{(k)}(vh)\,(x-vh)^k + \frac{1}{m!} a_v\,(x-vh)^m \\ &= A_v(x) + \frac{1}{m!} a_v\,(x-vh)^m, \qquad v = 0, 1, \cdots, n-1. \end{aligned}$$

Thus $A_v(x)$ is uniquely determined by the spline continuity conditions, and a_v is to be found from relation (1.11) replacing v by $v+1$. Relation (1.11) will be satisfied if and only if

$$a_v = \frac{(m-1)!}{h^{m-1}} \left\{ f\left((v+1)\,h,\, A_v((v+1)\,h) + \frac{1}{m!} a_v\, h^m \right) - A_v{}'((v+1)\,h) \right\} = g_h(a_v). \tag{1.12}$$

One Lipschitz constant for $g_h(t)$ is $L\,h/m$ independent of v, where L is the Lipschitz constant for $f(x,y)$. Hence for $h < m/L$ we have that $g_h(t)$ is a strong contraction mapping, and (1.12) has a unique fixed point a_v which may be found by iteration.

Chapter 2

WAVELET TRANSFORMS

2.1 Fourier Analysis

The subject of Fourier analysis is one of the oldest subjects in mathematical analysis and is of great importance to mathematician and engineers alike [10], [20], [27], and [31]. From a practical point of view, when one thinks of Fourier analysis, one usually refers to (integral) Fourier transforms and Fourier series. A Fourier transform is the Fourier integral of some function f defined on the real line R. When f is thought of as an analog signal, then its domain of definition R is called the continuous time domain. In this case, the Fourier transform $\hat{f}$ of f describes the spectral behavior of the signal f. Since the spectral information is given in terms of frequency, the domain of definition of the Fourier transform $\hat{f}$, which is again R, is called the frequency domain. On the other hand, a Fourier series is a transformation of bi-infinite sequences to periodic functions. Hence, when a bi-infinite sequence is thought of as a digital signal, then its domain definition, which is the set Z of integers, is called the discrete time domain. In this case, its Fourier series again describes the spectral behavior of the digital signal, and the domain of definition of a Fourier series is again the real line R, which is the frequency domain. However, since Fourier series are 2π-periodic, the frequency domain R in this situation is usually identified with the unit circle.

2.2 Fourier and Inverse Fourier Transform

Throughout this chapter, all functions f defined on the real line R are assumed to be measurable. And for each $p, 1 \le p < \infty$, let $L^p(R)$ denote the class of measurable functions f on R such that the (Lebesgue) integral $\int_{-\infty}^{\infty} |f(x)|^p \, dx$ is finite. Also, let $L^\infty(R)$ be the collection of almost everywhere (a.e.) bounded functions; that is, functions bounded everywhere except on sets of (Lebesgue) measure zero. hence, endowed with the "norm"

$$\|f\|_p := \begin{cases} \left\{ \int_{-\infty}^{\infty} |f(x)|^p \, dx \right\}^{\frac{1}{p}} & \text{for } 1 \le p < \infty; \\ \operatorname*{ess\,sup}_{-\infty < x < \infty} |f(x)| & \text{for } p = \infty \end{cases}$$

each $L^p(R), 1 \le p < \infty$, is a Banach space.

Definition 2.1

Let $f, g \in L^2(R)$ then, the "inner product" is defined by

$$\langle f, g \rangle := \int_{-\infty}^{\infty} f(x) \, \overline{g(x)} \, dx. \qquad (2.1)$$

Endowed with this inner product, the Banach space $L^2(R)$ becomes a Hilbert space. Of course, it is clear that

$$\langle f, f\rangle = \| f \|_2^2, \quad f \in L^2(R). \tag{2.2}$$

In the following, we concentrate our attention on functions in $L^1(R)$.

Definition 2.2

The Fourier transform of a function $f \in L^1(R)$ is defined by

$$\hat{f}(\omega) = (Ff)(\omega) := \int_{-\infty}^{\infty} e^{-i\omega x} f(x)\, dx. \tag{2.3}$$

Some of the basic properties of $\hat{f}(\omega)$, for every $f \in L^1(R)$, are summarized in the following theorem.

Theorem 2.3

Let $f \in L^1(R)$. Then its Fourier transform $\hat{f}$ satisfies:

i. $\hat{f} \in L^\infty(R)$ with $\| \hat{f} \|_\infty \le \| f \|_1$;

ii. $\hat{f}$ is uniformly continuous on R;

iii. if the derivative f' of f also exists and is in $L^1(R)$, then

$$\hat{f}'(\omega) = i\omega \hat{f}(\omega); \qquad \text{and} \tag{2.4}$$

iv. $\hat{f}(\omega) \to 0$, as $\omega \to \pm\infty$.

Proof

Assertion (i) is obvious. To prove (ii), let δ be chosen arbitrary and consider

$$\sup_\omega \left| \hat{f}(\omega + \delta) - \hat{f}(\omega) \right| = \sup_\omega \left| \int_{-\infty}^{\infty} e^{-i\omega x} (e^{-i\delta x} - 1) f(x)\, dx \right|$$

$$\le \int_{-\infty}^{\infty} \left| e^{-i\delta x} - 1 \right| \left| f(x) \right| dx.$$

Now, since $\left|e^{-i\delta x}-1\right|\left|f(x)\right|\leq 2\left|f(x)\right|\in L^1(R)$ and $\left|e^{-i\delta x}-1\right|\to 0$ as $\delta\to 0$, the Lebesque Dominated Convergence Theorem implies that the quantity above tend to zero as $\delta\to 0$.

To establish (iii), we simply integrate (2.3) by parts, and use the fact that $f(x)\to 0$ as $x\to\pm\infty$. Indeed,

$$F\{f'(x)\}=\int_{-\infty}^{\infty}f'(x)e^{-i\omega x}\,dx$$

$$=\left[\ f(x)e^{-i\omega x}\Big|_{-\infty}^{\infty}-(-i\omega)\int_{-\infty}^{\infty}f(x)\,e^{-i\omega x}\,dx\ \right]$$

$$=i\omega\,\hat{f}(\omega).$$

Finally, the statement in (iv) is usually called the " Riemann-Lebesgue Lemma". To prove it, we first observe that if f' exists and is in $L^1(R)$, then by (iii) and (i), we have,

$$\left|\hat{f}(\omega)\right|=\frac{1}{|\omega|}\left|\hat{f}'(\omega\,)\right|\leq\frac{1}{|\omega|}\|f'\|_1\to 0,$$

as $\omega\to\pm\infty$. In general, for any given $\varepsilon>0$, we can find a function g such that $g, g'\in L^1$ and $\|f-g\|_1<\varepsilon$ then by (i), we have

$$|f'(\omega)|\leq\left|\hat{f}(\omega)-\hat{g}(\omega)\right|+|\hat{g}(\omega)|$$

$$\leq\|f-g\|_1+|\hat{g}(\omega)|<\varepsilon+|\hat{g}(\omega)|,$$

completing the proof of (iv).

If it happens that $\hat{f}$ is in $L^1(R)$, then we can usually "recover" f from $\hat{f}$, by using the " inverse Fourier transform" defined as follows.

Definition 2.4

Let $\hat{f} \in L^1(R)$ be the Fourier transform of some function $f \in L^1(R)$. Then the inverse Fourier transform of $\hat{f}$ is defined by

$$(F^{-1}\,\hat{f}\,)(x) := \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ix\omega}\,\hat{f}(\omega\,)\,d\omega\,. \tag{2.5}$$

So, the important question is: when can f be recovered from $\hat{f}$ by using the operator F^{-1}, or when is $(F^{-1}\,\hat{f})\,(x) = f(x)$? the answer is: at every point x where f is continuous. That is we have the following theorem.

Theorem 2.5

Let $f \in L^1(R)$ such that its Fourier transform $\hat{f}$ is also in $L^1(R)$. Then

$$f(x) = (F^{-1}\,\hat{f}\,)\,(x) \tag{2.6}$$

at every point where f is continuous.

Example 2.6

Let $a > 0$. Then

$$\int_{-\infty}^{\infty} e^{-i\omega x}\,e^{-ax^2}\,dx = \sqrt{\frac{\pi}{a}}\,e^{-\frac{\omega^2}{4a}}\,. \tag{2.7}$$

In particular, the Fourier transform of the Gaussian function e^{-x^2} is $\sqrt{\pi}\,e^{-\omega^2/4}$.

Definition 2.7

Let f and g be functions in $L^1(R)$. Then the (continuous-time) convolution of f and g is also in an $L^1(R)$ function h defined by

$$h(x) = (f * g) := \int_{-\infty}^{\infty} f(x-y)\,g(y)\,dy\,. \tag{2.8}$$

Theorem 2.8

Let f and g be in $L^1(R)$. Then $(f * g\hat{)}\ (\omega) = \hat{f}(\omega)\ \hat{g}(\omega)$

Proof

By definition and interchange of the order of integration we have

$$F(f*g) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x-y)\, g(y)\, e^{-i\omega x}\, dy\, dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x-y)\, g(y)\, e^{-i\omega x}\, dx\, dy.$$

Instead of x we now take $x - y = q$ as a new variable of integration. Then

$$F(f*g) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(q)\, g(y)\, e^{-i\omega(y+q)}\, dq\, dy$$

$$= \int_{-\infty}^{\infty} f(q)\, e^{-i\omega q}\, dq \int_{-\infty}^{\infty} g(y)\, e^{-i\omega y}\, dy$$

$$= F(f)\, F(g).$$

2.3 Fourier Series

We now turn to the study of $2\pi-$periodic functions. For each $p,\ 1 \le p \le \infty$, the following notation will be used:

$$\| f \|_{L^p(0,\,2\pi)} := \begin{cases} \left[\dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} |f(x)|^p\, dx \right]^{1/p} &, \text{ for } \ 1 \le p < \infty; \\ \operatorname*{ess\,sup}_{0 \le x \le 2\pi} |f(x)| &, \text{ for } \ p = \infty \end{cases} \tag{2.9}$$

for each $p, L^p(0, 2\pi)$ denotes the Banach space of functions f satisfying $f(x + 2\pi) = f(x)$ a.e. in R, and $\| f \|_{L^p(0,2\pi)} < \infty$.

Definition 2.9

Let $f, g \in L^2(0, 2\pi)$ then the inner product is defined by

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(x)\,\overline{g(x)}\,dx. \tag{2.10}$$

The companions of the spaces $L^p(0, 2\pi)$ are the (sequence) spaces $l^p = l^p(Z)$ of bi-infinite sequences $\{a_k\}, k \in Z,$ that satisfy: $\| \{a_k\} \|_{l^p} < \infty$, where

$$\| \{a_k\} \|_{l^p} := \begin{cases} \left\{ \sum_{k\in Z} |a_k|^p \right\}^{\frac{1}{p}}, & for \quad 1 \le p < \infty; \\ \sup_k |a_k|, & for \quad p = \infty. \end{cases} \tag{2.11}$$

Analogous to the Hilbert spaces $L^2(R)$ and $L^2(0, 2\pi)$, the space $l^2 = l^2(Z)$ is also a Hilbert space with inner product:

$$\langle \{a_k\}, \{b_k\} \rangle_{l^2} := \sum_{k\in Z} a_k\, \overline{b_k}\,. \tag{2.12}$$

Recall that the (integral) Fourier transform is used to describe the spectral behavior of an analog signal f with finite energy (i.e., $f \in L^2(R)$). Here, we introduce the " discrete Fourier transform" F^* of a "digital signal" $\{c_k\} \in l^p$ to describe its spectral behavior, as follows:

$$(F^*\{c_k\})\,(x) := \sum_{k\in Z} c_k\, e^{ikx}\,. \tag{2.13}$$

On the other hand, if f is any function in $L^p(0, 2\pi), 1 \le p \le \infty$, then we can define the "inverse discrete Fourier transform" F^{*-1} of f by:

$$(F^{*-1}\, f)\,(k) = c_k(f) := \frac{1}{2\pi}\int_0^{2\pi} f(x)\, e^{-ikx}\, dx. \tag{2.14}$$

That is, F^{*-1} takes $f \in L^p(0, 2\pi)$ to a bi-infinite sequence $\{c_k(f)\}, k \in Z$. This sequence, of course, defines a Fourier series

$$\sum_{k\in Z} c_k(f)\, e^{ikx} \tag{2.15}$$

and is called the sequence of"Fourier coefficients" of the Fourier series.

2.4 The Gabor Transform

A function f in $L^2(R)$ is used to represent an analog signal with finite energy, and its Fourier transform

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t}\, f(t)\, dt \tag{2.16}$$

reveals the spectral information of the signal. Unfortunately, formula (2.16) alone is not very useful for extracting information of the spectrum $\hat{f}$ from local observation of the signal f. What is needed is a "good" time-window.

The optimal window for time localization is achieved by using any Gaussian function

$$g_\alpha(t) := \frac{1}{2\sqrt{\pi\,\alpha}}\, e^{-\frac{t^2}{4\alpha}} \tag{2.17}$$

where $\alpha > 0$ is fixed, as window function.

Definition 2.10

A nontrivial function $\omega \in L^2(R)$ is called a window function if $x\,\omega(x)$ is also in $L^2(R)$.

The center t^* and radius Δ_ω of a window function ω are defined to be

$$t^* := \frac{1}{\|\omega\|_2^2} \int_{-\infty}^{\infty} x\,|\omega(x)|^2\, dx \tag{2.18}$$

and

$$\Delta_\omega := \frac{1}{\|\omega\|_2} \left\{ \int_{-\infty}^{\infty} (x - t^*)^2\, |\omega(x)|^2\, dx \right\}^{1/2} \tag{2.19}$$

respectively and the width of the window function ω is defined by $2\,\Delta_\omega$.

Definition 2.11

For any fixed value of $\alpha > 0$, the "Gabor transform" of an $f \in L^2(R)$ is defined by

$$(g_b^\alpha f)(\omega) = \int_{-\infty}^{\infty} (e^{-i\omega t} f(t))\, g_\alpha (t-b)\, dt. \tag{2.20}$$

That is, $(g_b^\alpha f)(\omega)$ localizes the Fourier transform of f around $t = b$. The "width" of the window is determined by the (fixed) positive constant α to be discussed below. Observe that from (2.7) in Example (2.6) with $\omega = 0$ and $a = (4\alpha)^{-1}$, we have

$$\int_{-\infty}^{\infty} g_\alpha (t-b)\, db = \int_{-\infty}^{\infty} g_\alpha (x)\, dx = 1, \tag{2.21}$$

so that

$$\int_{-\infty}^{\infty} (g_b^\alpha f)(\omega)\, db = \hat{f}(\omega), \quad \omega \in R.$$

That is, the set

$$\{ g_b^\alpha f : b \in R \}$$

of Gabor transform of f decomposes the Fourier transform $\hat{f}$ of f exactly to give its local spectral information. Note that since g_α is an even function, its center, defined by (2.18), is 0, and hence,

$$\Delta_{g_\alpha} := \frac{1}{\|g_\alpha\|_2} \left\{ \int_{-\infty}^{\infty} x^2\, g_\alpha^2(x)\, dx \right\}^{1/2}. \tag{2.22}$$

Theorem 2.12

For each $\alpha > 0$,

$$\Delta_{g_\alpha} = \sqrt{\alpha}. \tag{2.23}$$

That is, the width of the window function g_α is $2\sqrt{\alpha}$.

Proof

By setting $\omega = 0$ in (2.7), we have

$$\int_{-\infty}^{\infty} e^{-ax^2}\, dx = \sqrt{\pi}\, a^{-\frac{1}{2}}; \tag{2.24}$$

and differentiating both sides with respect to the parameter a yields

$$\int_{-\infty}^{\infty} x^2\, e^{-ax^2}\, dx = \frac{\sqrt{\pi}}{2} a^{-3/2}. \tag{2.25}$$

Hence, by setting $a = (2\alpha)^{-1}$ in (2.24) and (2.25), it follows that

$$\| g_\alpha \|_2 = (8\pi\alpha)^{-1/4}, \tag{2.26}$$

and consequently,

$$\Delta_{g_\alpha} = (8\pi\alpha)^{1/4} \left\{ \frac{1}{4\pi\alpha} \cdot \frac{\sqrt{\pi}}{2} (2\alpha)^{3/2} \right\}^{1/2} = \sqrt{\alpha}.$$

We may interpret the Gabor transform $g_b^\alpha f$ in (2.20) somewhat differently; namely, by setting

$$G_{b,\omega}^\alpha(t) := e^{i\omega t}\, g_\alpha(t-b) \tag{2.27}$$

we have

$$(g_b^\alpha f)(\omega) = \langle f, G_{b,\omega}^\alpha \rangle = \int_{-\infty}^{\infty} f(t)\, \overline{G_{b,\omega}^\alpha(t)}\, dt. \tag{2.28}$$

In other words, instead of considering $g_b^\alpha f$ as localization of the Fourier transform of f, we may interpret it as windowing the function (or signal) f by using the window function $G_{b,\omega}^\alpha$ in (2.27). We will follow this point of view in comparing it with the "integral wavelet transform" later.

One advantage of the formulation (2.28) is that the Parseval Identity can be used to relate the Gabor transform of f with the Gabor transform of $\hat{f}$. In fact, since

$$\hat{G}_{b,\omega}^\alpha(\eta) = e^{-ib(\eta-\omega)}\, e^{-\alpha(\eta-\omega)^2}, \tag{2.29}$$

which follows from (2.7) by letting $\alpha = 1/4a$, we have

$$(g_b^\alpha f)(\omega) = \langle f, G_{b,\omega}^\alpha \rangle = \frac{1}{2\pi} \langle \hat{f}, \hat{G}_{b,\omega}^\alpha \rangle \tag{2.30}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\eta)\, e^{ib(\eta-\omega)}\, e^{-\alpha(\eta-\omega)^2}\, d\eta$$

$$= \frac{e^{-ib\omega}}{2\sqrt{\pi\alpha}} \int_{-\infty}^{\infty} (e^{ib\eta}\, \hat{f}(\eta))\, g_{1/4\alpha}(\eta-\omega)\, d\eta$$

$$= \frac{e^{-ib\omega}}{2\sqrt{\pi\alpha}} (g_\omega^{1/4\alpha}\, \hat{f})\,(-b).$$

Let us interpret (2.30) from two points of view. First, we consider

$$\int_{-\infty}^{\infty} (e^{-i\omega t}\, f(t)\,)\, g_\alpha\, (t-b)\, dt \tag{2.31}$$

$$= \left(\sqrt{\frac{\pi}{\alpha}}\, e^{-ib\omega} \right) \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} (e^{ib\eta}\, \hat{f}(\eta))\, g_{1/4\alpha}(\eta-\omega)\, d\eta,$$

which says that, with the exception of the multiplicative term $\sqrt{\frac{\pi}{\alpha}}\, e^{-ib\omega}$, the "window Fourier transform" of f with window function g_α at $t = b$ agrees with the "window inverse Fourier transform" of $\hat{f}$ with window function $g_{1/4\alpha}$ at $\eta = \omega$. By theorem 2.12, the product of the widths of these two windows is

$$(2\Delta_{g_\alpha})\,(2\Delta_{g_{1/4\alpha}}) = 2. \tag{2.32}$$

On the other hand, by considering

$$H_{b,\omega}^\alpha(\eta) := \frac{1}{2\pi} \hat{G}_{b,\omega}^\alpha\, (\eta\,) = \left(\frac{e^{ib\omega}}{2\sqrt{\pi\,\alpha}} \right) e^{-ib\eta}\, g_{1/4\alpha}\, (\eta - \omega)\,, \tag{2.33}$$

we have

$$\left\langle f, G_{b,\omega}^{\alpha} \right\rangle = \left\langle \hat{f}, H_{b,\omega}^{\alpha} \right\rangle. \tag{2.34}$$

This identity says that the information obtained by investigating an analog signal $f(t)$ at $t = b$ by using the window function $G_{b,\omega}^{\alpha}$ as defined in (2.27) can also be obtained by observing the spectrum $\hat{f}(\eta)$ of the signal in a neighborhood of the frequency $\eta = \omega$ by using the window function $H_{b,\omega}^{\alpha}$ as defined in (2.33). Again the product of the width of the time-window $G_{b,\omega}^{\alpha}$ and of the frequency-window $H_{b,\omega}^{\alpha}$ is

$$\left(2\,\Delta_{G_{b,\omega}^{\alpha}}\right)\left(2\,\Delta_{H_{b,\omega}^{\alpha}}\right) = (2\,\Delta_{g_{\alpha}})\,(2\,\Delta_{g_{1/4\alpha}}) = 2\,. \tag{2.35}$$

The Cartesian product

$$\left[b - \sqrt{\alpha}, b + \sqrt{\alpha}\right] \times \left[\omega - \frac{1}{2\sqrt{\alpha}}, \omega + \frac{1}{2\sqrt{\alpha}}\right]$$

of these two windows is called a rectangular time-frequency window. It is usually plotted in time-frequency domain to show how a signal is localized. The width $2\sqrt{\alpha}$ of the time-window is called the "width of the time-frequency window", and the width $\frac{1}{\sqrt{\alpha}}$ of the frequency window is called the "height of the time-frequency window". Observe that the width of the time-frequency window is unchanged for observing the spectrum at all frequencies. This restricts the application of the Gabor transform to study signals with unusually high and low frequencies.

2.5 Short-Time Fourier Transform and the Uncertainty Principle

As discussed before, we have seen that the Gabor transform is a window Fourier transform with any Gaussian function g_α as the window function. For various reasons such as computational efficiency or convenience in implementation, other functions may also be used as window functions instead. Also as seen before, for a non-trivial function $\omega \in L^2(R)$ to qualify as a window function, it must satisfy the requirement that

$$t\,\omega(t) \in L^2(R). \tag{2.36}$$

From (2.36) and by an application of the Schwarz inequality to the product of $(1+|t|)^{-1}$ and $(1+|t|)\,\omega(t)$, it is clear that $\omega \in L^1(R)$ also. Hence, by Theorem 2.3, its Fourier transform $\hat{\omega}$ is continuous. However, although it follows from the Parseval identity that $\hat{\omega}$ is also in $L^2(R)$, it does not necessarily satisfy (2.36), and hence, may not be a (frequency) window function. Recall from the previous section that the importance of a Gaussian function g_α is that its Fourier transform is also a Gaussian function, so that g_α and $\hat{g}_\alpha$ can be used for time-frequency localization.

Example 2.12

Both the first order B-spline

$$N_1(t) := \begin{cases} 1 & \textit{for } 0 \le t < 1 \\ 0 & \textit{othrwise} \end{cases} \tag{2.37}$$

and the Haar function

$$\psi_1(t) = \psi_H(t) := \begin{cases} 1 & for\, 0 \le t < \frac{1}{2}; \\ -1 & for\, \frac{1}{2} \le t < 1; \\ 0 & otherwise, \end{cases} \tag{2.38}$$

are window functions; but their Fourier transforms $\hat{N}_1$ and $\hat{\psi}_1$ do not satisfy (2.36), and hence N_1 and ψ_1 can not be used for time-frequency localization.

In general, for any $\omega \in L^2(R)$ that satisfies (2.36), we define the center and radius of ω by

$$x^* := \frac{1}{\|\omega\|_2^2} \int_{-\infty}^{\infty} t\,|\omega(t)|^2\, dt \tag{2.39}$$

and

$$\Delta_\omega := \frac{1}{\|\omega\|_2} \left\{ \int_{-\infty}^{\infty} (t - x^*)^2\, |\omega(t)|^2\, dt \right\}^{1/2}. \tag{2.40}$$

We also use the value $2\Delta_\omega$ to measure the width of the window function ω. In signal analysis, if ω is considered as an analog signal itself, then Δ_ω is called the root mean square (RMS) duration of the analog signal, and $\Delta_{\hat{\omega}}$ is called its RMS bandwidth, provided that $\hat{\omega}$ also satisfies (2.36). The Gabor transform (2.20), can be generalized to any "window Fourier transform" of an $f \in L^2(R)$, by using a function ω that satisfies (2.36) as the window function, as follows:

$$(\tilde{g}_b f)(\omega) := \int \left(e^{-i\omega t} f(t) \right) \overline{\omega\,(t-b)}\, dt. \tag{2.41}$$

Hence by setting

$$W_{b,,\omega}(t) := e^{i\omega t}\, \omega\,(t-b) \tag{2.42}$$

we have

$$(\widetilde{g}_b f)(\omega) = \left\langle f, W_{b,\omega} \right\rangle = \int_{-\infty}^{\infty} f(t)\, \overline{W_{b,\omega}(t)}\, dt \tag{2.43}$$

so that $(\widetilde{g}_b\, f)\,(\omega)$ give local information of f in the time-window

$$[x^* + b - \Delta_\omega,\ x^* + b + \Delta_\omega]\,. \tag{2.44}$$

Now, suppose that the Fourier transform $\hat{\omega}$ of ω also satisfies (2.36). Then we can determine the center ω^* and radius $\Delta_{\hat{\omega}}$, of the window function $\hat{\omega}$, by using formulas analogous to (2.39) and (2.40). By setting

$$V_{b,\omega}(\eta) := \frac{1}{2\pi}\hat{W}_{b,\omega}(\eta) \tag{2.45}$$

$$= \left(\frac{e^{ib\omega}}{2\pi}\right) e^{-ib\eta}\ \hat{\omega}\,(\eta - \omega\,)\,.$$

Which is also a window function with center at $\omega^* + \omega$ and radius equal to $\Delta_{\hat{\omega}}$, we have, by the Parseval identity,

$$(\widetilde{g}_b f)(\,\omega\,) = \left\langle f, W_{b,\omega} \right\rangle = \left\langle \hat{f}, V_{b,\omega} \right\rangle. \tag{2.46}$$

Hence, $(\widetilde{g}_b\, f)(\omega)$ also gives local spectral information of f in the frequency-window

$$[\,\omega^* + \omega - \Delta_{\hat{\omega}},\ \omega^* + \omega + \Delta_{\hat{\omega}}]\,. \tag{2.47}$$

In summary, by choosing any $\omega \in L^2(R)$ such that both ω and $\hat{\omega}$ satisfy (2.36) to define the window Fourier transform in (2.41), we have a time-frequency window

$$[x^* + b - \Delta_\omega,\ x^* + b + \Delta_\omega] \times [\,\omega^* + \omega - \Delta_{\hat{\omega}},\ \omega^* + \omega + \Delta_{\hat{\omega}}]\ . \tag{2.48}$$

with width $2\Delta_\omega$(as determined by the width of the time-window) and constant window area

$$4\,\Delta_{\omega}\,\Delta_{\hat{\omega}}\,. \tag{2.49}$$

Again, the width of the time-frequency window remains unchanged for localizing signals with both high and low frequencies.

Definition 2.13

If $\omega \in L^2(R)$ is chosen in a way such that both ω and its Fourier transform $\hat{\omega}$ satisfy (2.36), then the window Fourier transform introduced in (2.41), by using ω as the window function, is called a "short-time Fourier transform" (STFT)

As observed earlier, since both ω and $\hat{\omega}$ satisfy (2.36), they must be continuous functions. In addition to the Gaussian functions, every B-spline of order higher than one can be used to define an STFT.

For accurate time-frequency localization, one chooses a window function ω such that The time-frequency window has sufficiently small area $4\,\Delta_{\omega}\,\Delta_{\hat{\omega}}$. We have already seen in (2.35) that if ω is any Gaussian function g_{α}, $\alpha > 0$, then the window area is 2. So, the first question to be answered is whether a smaller area can be achieved. The following theorem, known as the "Uncertainty Principle", says that it is not possible to find a window with size smaller than or equal to that of the Gaussian functions.

Theorem 2.14

Let $\omega \in L^2(R)$ be chosen such that both ω and its Fourier transform $\hat{\omega}$ satisfy (2.36). Then

$$\Delta_{\omega}\,\Delta_{\hat{\omega}} \geq \frac{1}{2}\,. \tag{2.50}$$

Furthermore, equality is attained if and only if

$$\omega\,(t) = c e^{i\,at}\, g_{\alpha}\,(t-b),$$

where $c \neq 0, \alpha > 0,$ and $a, b \in R$.

Hence, the Gabor transform introduced in the previous section is the STFT with the smallest time-frequency window. In some application, a larger window must be chosen in order to achieve other desirable properties. For example, a second or higher order B-spline facilitates computational and implementational effectiveness. The most important property not possessed by the Gabor transform is the additional condition:

$$\int_{-\infty}^{\infty} \psi(x)\, dx = 0,$$

where ψ is the window function. This property gives us an extra degree of freedom for introducing a dilation (or scale) parameter in order to make the time-frequency window flexible. With this dilation parameter, the time-localization integral transform to be discussed in the next section will be called an "integral wavelet transform" (IWT), and any window function for defining the IWT will be called "basic wavelet".

2.6 The Integral Wavelet Transform

We have seen that in analyzing a function (signal) with any STFT, the time-frequency window is rigid, in the sense that its width is unchanged in observing any frequency band (or octave)

$$[\, \omega^{*} + \omega \; - \Delta_{\hat{\omega}}, \omega^{*} + \omega \; + \Delta_{\hat{\omega}} \,]$$

with center frequency $\omega^{*} + \omega$. Since frequency is directly proportional to the number of cycles per unit time, it takes a narrow time-window to locate high-frequency phenomena more precisely and a wide time-window to analyze low-frequency behaviors more thoroughly. Hence, the STFT is not suitable for analyzing signals with both very high and very low frequencies. On the other hand, the integral wavelet transform (IWT), to be defined later, relative to some basic wavelet provides a flexible time-frequency window that automatically narrows when observing high-frequency phenomena and widens when studying low-frequency environments.

Definition 2.15

If $\psi \in L^2(R)$ satisfies the "admissibility" condition:

$$C_\psi := \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty, \tag{2.51}$$

then ψ is called a " basic wavelet". Relative to every basic wavelet ψ, the integral wavelet transform (IWT) on $L^2(R)$ is defined by

$$(W_\psi f)(b, a) := |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t)\, \overline{\psi\left(\frac{t-b}{a}\right)}\, dt, \qquad f \in L^2(R), \tag{2.52}$$

where a, b $\in$ R with $a \neq 0$.

If, in addition, both ψ and $\hat{\psi}$ satisfy (2.36), then the basic wavelet ψ provides a time-frequency window with finite area given by $4\,\Delta_\psi\,\Delta_{\hat{\psi}}$. In addition, under this additional assumption, it follows that $\hat{\psi}$ is a continuous function, so that the finiteness of C_ψ implies $\hat{\psi}\,(0) = 0$, or equivalently,

$$\int_{-\infty}^{\infty} \psi\,(t)\, dt = 0. \tag{2.53}$$

This is the reason that ψ is called a "wavelet". We will see that the admissibility condition is needed in obtaining the inverse of the IWT.

By setting

$$\psi_{b;a}(t) := | a |^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right), \tag{2.54}$$

the IWT defined in (2.52) can be written as

$$\left(W_{\psi} f\right)(b, a) = \left\langle f, \psi_{b;a} \right\rangle. \tag{2.55}$$

In the following discussion, we will assume that both ψ and $\hat{\psi}$ satisfy (2.36). Then if the center and radius of the window function ψ are given by t^{*} and Δ_{ψ}, respectively, the function $\psi_{b;a}$ is a window function with center at $b + a\,t^{*}$ and radius equal to $a\Delta_{\psi}$. Hence, the IWT, as formulated in (2.55), gives local information on an analog signal f with a time window

$$[\,b + at^{*} - a\Delta_{\psi}, b + at^{*} + a\Delta_{\psi}\,]. \tag{2.56}$$

This window narrows for small values of a and widens for allowing a being large.

Next, consider

$$\frac{1}{2\pi}\hat{\psi}_{b;a}(\omega) = \frac{|a|^{-\frac{1}{2}}}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega t}\, \psi\left(\frac{t-b}{a}\right) dt \tag{2.57}$$

$$= \frac{a\,|\,a\,|^{-\frac{1}{2}}}{2\pi} e^{-ib\omega}\, \hat{\psi}(a\,\omega)$$

and suppose that the center and radius of the window function $\hat{\psi}$ are given by ω^{*} and $\Delta_{\hat{\psi}}$, respectively. Then by setting

$$\eta(\omega) := \hat{\psi}(\omega + \omega^{*}), \tag{2.58}$$

we have a window function η with center at the origin and radius equal to $\Delta_{\hat{\psi}}$. Now from (2.55) and (2.57), and applying the Parseval identity, we have

$$(W_\psi f)(b,a) = \frac{a\left|a\right|^{-\frac{1}{2}}}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)\, e^{ib\omega}\, \eta\overline{\left(a\left(\omega - \frac{\omega^*}{a}\right)\right)}\, d\omega. \qquad (2.59)$$

Since it is clear that the window function $\eta\left(a\left(\omega - \frac{\omega^*}{\omega}\right)\right) = \eta\,(a\,\omega - \omega^*) = \hat{\psi}\,(a\,\omega)$ has radius given by $\frac{1}{a}\Delta_{\hat{\psi}}$, the expression in (2.59) says that, with the exception of a multiple of $a\left|a\right|^{-1/2}/2\pi$ and a linear phase-shift of $e^{ib\omega}$, the IWT $W_\psi\, f$ also gives local information of $\hat{f}$ with a frequency-window

$$\left[\frac{\omega^*}{a} - \frac{1}{a}\Delta_{\hat{\psi}}, \frac{\omega^*}{a} + \frac{1}{a}\Delta_{\hat{\psi}}\right]. \qquad (2.60)$$

In the following discussion, the center ω^* of $\hat{\psi}$ is assumed to be positive. In doing so, we may think of this window as frequency band (or octave) with center-frequency ω^*/a and bandwidth $2\,\Delta_{\hat{\psi}}/a$. The importance of this identification is that the ratio

$$\frac{\text{center frequency}}{\text{bandwidth}} = \frac{\omega^*/a}{2\,\Delta_{\hat{\psi}}/a} = \frac{\omega^*}{2\,\Delta_{\hat{\psi}}} \qquad (2.61)$$

is independent of the scaling a. Hence, if the frequency variable is identified as a constant multiple of a^{-1}, then an adaptive bandpass filter, with pass-band given by (2.60), has the property that the ratio of the center-frequency to the bandwidth is independent of the location of the center-frequency. This is called “constant-Q filtering”.

Now, if ω^*/a is considered to be frequency variable ω, then we may consider the $t-\omega$ plane as the time-frequency plane. Hence, with the time-window in (2.56) and the frequency-window in (2.60), we have a rectangular time-frequency window

$$[b+a\,t^{*}-a\Delta_{\psi},\, b+a\,t^{*}+a\Delta_{\psi}]\times\left[\frac{\omega^{*}}{a}-\frac{1}{a}\Delta_{\hat{\psi}},\, \frac{\omega^{*}}{a}+\frac{1}{a}\Delta_{\hat{\psi}}\right] \tag{2.62}$$

in the $t-\omega$ plane, with width $2a\,\Delta_{\psi}$ (determined by the width of the time-window). Hence, this window automatically narrows for detecting high-frequency phenomena (i.e., small $a>0$), and widens for investigating low-frequency behavior (i.e., large $a>0$).

We next derive a formula for reconstructing any finite-energy signal from its IWT values.

Theorem 2.16

Let ψ be a basic wavelet that defines an IWT W_{ψ}. Then

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[(W_{\psi}\,f)\,(b,a)\,\overline{(W_{\psi}\,f)\,(b,a)}\right]\frac{da}{a^{2}}\,db=C_{\psi}\,\langle f,g\rangle, \tag{2.63}$$

for all $f,g\in L^{2}(R)$. Furthermore, for any $f\in L^{2}(R)$ and $x\in R$ at which f is continuous,

$$f(x)=\frac{1}{C_{\psi}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[(W_{\psi}\,f)\,(b,a)\right]\psi_{b;a}(x)\,\frac{da}{a^{2}}\,db, \tag{2.64}$$

where $\psi_{b;a}$ is defined in (2.54).

Proof

By applying the Parseval Identity and (2.57), and using the notation

$$\begin{cases} F(x):=\hat{f}(x)\,\hat{\psi}\,(ax); \\ G(x):=\hat{g}(x)\,\hat{\psi}\,(ax), \end{cases} \tag{2.65}$$

we have

$$\int_{-\infty}^{\infty}\left[(W_{\psi}\,f)\,(b,a)\,\overline{(W_{\psi}\,f)\,(b,a)}\right]db$$

$$= \frac{1}{|a|}\int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f(t)\,\overline{\psi\left(\frac{t-b}{a}\right)}\,dt \int_{-\infty}^{\infty} \overline{g(s)}\,\psi\left(\frac{s-b}{a}\right) ds \right\} db$$

$$= \frac{a^2}{|a|}\int_{-\infty}^{\infty} \left\{ \frac{1}{2\pi}\,\overline{\int_{-\infty}^{\infty} \overline{F(x)}\,e^{-ibx}\,dx} \right\}\left\{ \frac{1}{2\pi}\int_{-\infty}^{\infty} \overline{G(y)}\,e^{-iby}\,dy \right\} db$$

$$= \frac{a^2}{2\pi|a|}\left\{ \frac{1}{2\pi}\int_{-\infty}^{\infty} \overset{\wedge}{G(b)}\;\overline{\overset{\wedge}{F(b)}}\,db \right\}$$

$$= \frac{a^2}{2\pi\,|\,a\,|}\int_{-\infty}^{\infty} \overline{G}(x)\,F(x)\,dx,$$

where the Parseval Identity is applied again to arrive at the last equality. Hence, by substituting (2.65) into the above expression, then integrating with respect to da/a^2 on $(-\infty, \infty)$, and recalling the definition of C_ψ from (2.51), we obtain

$$\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} \left[(W_\psi\, f)\,(b, a)\,\overline{(W_\psi\, f)\,(b, a)}\right] db\right\}\frac{da}{a^2} \tag{2.66}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \left\{ \hat{f}(x)\,\overline{\hat{g}(x)}\int_{-\infty}^{\infty} \frac{|\,\hat{\psi}(ax)|^2}{|\,a\,|}\,da \right\} dx$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \left\{ \hat{f}(x)\,\overline{\hat{g}(x)}\int_{-\infty}^{\infty} \frac{|\,\hat{\psi}(y)\,|^2}{|\,y\,|}\,dy \right\} dx$$

$$= C_\psi\,\frac{1}{2\pi}\left\langle \hat{f}, \hat{g}\right\rangle = C_\psi\,\langle f, g\rangle.$$

Furthermore, if f is continuous at x, then using the Gaussian function $g_\alpha(\cdot - x)$ for the function g and allowing α tending to 0 from above, we arrive at

$$\mathrm{f(x)} = \frac{1}{\mathrm{C}_\psi}\lim_{\alpha\to 0^+}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[(\mathrm{W}_\psi\,\mathrm{f})\,(\mathrm{b, a})\,\overline{\left\langle \mathrm{g}_\alpha(\cdot - \mathrm{x}), \psi_{\mathrm{b;a}}\right\rangle}\right]\frac{\mathrm{da}}{\mathrm{a}^2}\,\mathrm{db}$$

$$= \frac{1}{C_{\psi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[(W_{\psi} f)\,(b, a) \right] \psi_{b;a}(x) \, \frac{da}{a^2} \, db.$$

This completes the proof of the theorem.

Chapter 3

WAVELET ANALYSIS

3.1 Multiresolution Analysis and Wavelets

In this section, we will present some definitions and properties that are needed for the analysis of wavelets.

Definition 3.1

For $\varphi, \psi \in L^2(R)$ and $j, k \in Z$, define $\varphi_{j,k}, \psi_{j,k} \in L^2(R)$ by

$$\varphi_{j,k}(x) = 2^{j/2}\, \varphi\,(2^j x - k) \text{ and } \psi_{j,k}(x) = 2^{j/2}\, \psi\,(2^j x - k)\,. \tag{3.1}$$

Definition 3.2

A **Wavelet** is a function $\psi(t) \in L^2(R)$ such that the family of functions $\{\psi_{j,k}\}$ as defined in (3.1), is an orthonormal basis in the Hilbert space $L^2(R)$ [10], [27]. That is $\langle \psi_{j,k}, \psi_{l,m} \rangle = \delta_{j,l} . \delta_{k,m,}\ \ j, k, l, m \in Z$ and every $f \in L^2(R)$ can be written as

$$f(x) = \sum_{j,k=-\infty}^{\infty} c_{j,k}\psi_{j,k}(x). \tag{3.2}$$

The simplest example of an orthogonal wavelet is the Haar function ψ_H defined by

$$\psi_H(x) = \begin{cases} 1 & for \quad 0 \le x < \frac{1}{2} \\ -1 & for \quad \frac{1}{2} \le x < 1. \\ 0 & otherwise \end{cases}$$

By Definition (3.2) it is clear that any wavelet generates a direct sum decomposition of $L^2(R)$. For each $j \in Z$, let us consider the closed subspaces $V_j = \ldots \oplus W_{j-2} \oplus W_{j-1}$, $j \in Z$ of $L^2(R)$ where $W_{j,k} = clos_{L^2(R)}\langle \psi_{j,k} : k \in Z \rangle$, $j \in Z$.

These subspaces have the following properties:

i. $\ldots \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots,$

ii. $clos_{L^2}\left(\bigcup_{j \in Z} V_j \right) = L^2(R)$, where *clos* denotes the closure of the set.

iii. $\bigcap_{j \in z} V_j = \{0\},$

iv. $V_{j+1} = V_j \oplus W_j$, $j \in Z$, and

v. $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$, $j \in Z$.

In fact, if the reference subspace V_0, say, is generated by a single function $\phi \in L^2(R)$ in the sense that $V_0 = clos_{L^2(R)}\langle \phi_{0,k} : k \in Z \rangle$ where

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k).$$

Then all the subspaces V_j are also generated by the same function ϕ (just as the subspaces W_j are generated by ψ) where

$$V_j = clos_{L^2(R)} \langle \phi_{j,k} : k \in Z \rangle, \quad j \in Z. \tag{3.3}$$

Definition 3.3

A function $\phi \in L^2(R)$ is said to generate **a multiresolution analysis** (MRA) if it generates a nested sequence of closed subspaces V_j that satisfy (i), (ii), (iii), (v) in the sense of (3.3), such that $\{\phi_{0,k}\}$ forms a Riesz basis of V_0.

Definition 3.4

$\{\phi_{0,k}\}$ is said to form **a Riesz basis** of V_0. If there exist two constants A and B, with $0 < A \le B < \infty$, such that

$$A \| \{c_k\} \|_{l^2}^2 \le \left\| \sum_{k=-\infty}^{\infty} c_k \phi_{0,k} \right\|_2^2 \le B \| \{c_k\} \|_{l^2}^2$$

for all bi-infinite square summable sequences $\{c_k\}$; that is

$$\| \{c_k\} \|_{l^2}^2 = \sum_{k=-\infty}^{\infty} | c_k |^2 < \infty.$$

If ϕ generates an (MRA), then ϕ is called a "**scaling function**". Typical examples of scaling functions ϕ are the m^{th} order cardinal B-splines N_m.

3.2 Wavelet Decompositions and Reconstruction

Let $\{V_j\}$ be generated by some scaling function $\phi \in L^2(R)$ and $\{W_j\}$ is generated by some wavelet $\psi \in L^2(R)$. In this case, by property (ii), every function f in $L^2(R)$ can be approximated as closely as desired by an $f_N \in V_N$, for $N \in Z$. Since $V_j = V_{j-1} \oplus W_{j-1}$ for any $j \in Z$, f_N has a unique decomposition: $f_N = f_{N-1} + g_{N-1}$ where $f_{N-1} \in V_{N-1}$ and $g_{N-1} \in W_{N-1}$. By this process we have

$$f_N = g_{N-1} + g_{N-2} + g_{N-3} + \ldots + g_{N-M} + f_{N-M} \tag{3.4}$$

where $f_j \in V_j$ and $g_j \in W_j$ for any j, M here is chosen such that $\| f_{N-M} \|$ is smaller than some threshold. In what follows, we will discuss an algorithmic approach for expressing f_N as a direct sum of its components $g_{N-1}, \ldots, g_{N-M}$, and f_{N-M}, and recovering f_N from these components.

Since both the scaling function $\phi \in V_0$ and the wavelet $\psi \in W_0$ are in V_1, and since V_1 is generated by $\phi_{1,k}(x) = 2^{1/2}\phi(2x-k), k \in Z$ there exist two sequences $\{p_k\}$ and $\{q_k\} \in l^2$ such that

$$\phi(x) = \sum_k p_k \phi(2x-k) \tag{3.5}$$

$$\psi(x) = \sum q_k \phi(2x-k) \tag{3.6}$$

for all $x \in R$. The formulas (3.5) and (3.6) are called the **"two-scale relations"** of the scaling function and wavelet respectively. On the other hand, since both $\phi(2x)$ and

$\phi(2x-1)$ are in V_1 and $V_1 = V_0 \oplus W_0$, there are four l^2 sequences which we denote by $\{a_{-2k}\}$, $\{b_{-2k}\}$, $\{a_{1-2k}\}$, and $\{b_{1-2k}\}$, $k \in Z$ such that

$$\phi(2x) = \sum_k [a_{-2k}\phi(x-k) + b_{-2k}\psi(x-k)]; \tag{3.7}$$

$$\phi(2x-1) = \sum [a_{1-2k}\phi(x-k) + b_{1-2k}\psi(x-k)], \tag{3.8}$$

for all $x \in R$. The two formulas (3.7) and (3.8) can be combined into a single formula

$$\phi(2x-l) = \sum_k [a_{l-2k}\phi(x-k) + b_{l-2k}\psi(x-k)], \quad l \in Z. \tag{3.9}$$

Which is called the **"decomposition relation"** of ϕ and ψ. Now, we have two pairs of sequences $(\{p_k\},\{q_k\})$ and $(\{a_k\},\{b_k\})$, all of which are unique due to the direct sum relationship $V_1 = V_0 \oplus W_0$. These sequences are used to formulate the reconstruction and decomposition algorithms, which will follow. Hence, $\{p_k\}$ and $\{q_k\}$ are called reconstruction sequences, while $\{a_k\}$ and $\{b_k\}$ are called decomposition sequences.

To describe these algorithms, let us first recall that both $f_j \in V_j$ and $g_j \in W_j$ have unique series representations:

$$\begin{cases} f_j(x) = \sum_k c_k^j \phi(2^j x - k) \\ \text{with } c^j = \{c_k^j\} \in l^2 \end{cases} \tag{3.10}$$

and

$$\begin{cases} g_j(x) = \sum\limits_k d_k^j \psi(2^j x - k) \\ \text{with } d^j = \{d_k^j\} \in l^2 \end{cases} . \tag{3.11}$$

In the decomposition and reconstruction algorithms, the functions f_j and g_j will be represented by the sequences c^j and d^j as defined in (3.10) and (3.11).

Decomposition Algorithm

By applying (3.9)-(3.11), we have

$$\begin{cases} c_k^{j-1} = \sum\limits_l a_{l-2k} c_l^j \\ d_k^{j-1} = \sum b_{l-2k} c_l^j \end{cases} . \tag{3.12}$$

This algorithm can be described by the following schematic diagram:

$$\begin{tikzcd}
 & d^{N-1} & d^{N-2} & & d^{N-M} \\
c^N \arrow[ur] \arrow[r] & c^{N-1} \arrow[ur] \arrow[r] & c^{N-2} \arrow[ur] \arrow[r] & \dots \arrow[ur] \arrow[r] & c^{N-M} .
\end{tikzcd}$$

Reconstruction Algorithm

By applying (3.5), (3.6), (3.10), and (3.11), we have:

$$c_k^j = \sum [p_{k-2l} c_l^{j-1} + q_{k-2l} d_l^{j-1}]. \tag{3.13}$$

The following schematic diagram can also describe this algorithm:

$$\begin{tikzcd}
d^{N-M} \arrow[dr] & d^{N-M+1} \arrow[dr] & & d^{N-1} \arrow[dr] & \\
c^{N-M} \arrow[r] & c^{N-M+1} \arrow[r] & \cdots \arrow[r] & c^{N-1} \arrow[r] & c^N .
\end{tikzcd}$$

3.3 The Two Scale Relation for the Cardinal B-Spline

In this section we will study the relationship between any two consecutive subspaces of the nested sequence $\{V_j^m : j \in Z\}$ that are closed subspaces of $L^2(R)$,[37]. Now since for each $j \in Z$

$N_m(2^j.) \in V_j^m$ and $V_j^m \subset V_{j+1}^m$,

we will have

$$N_m(2^j x) = \sum_{k=-\infty}^{\infty} p_{m,k}\, N_m(2^{j+1}x - k), \tag{3.14}$$

where $\{p_{m,k} : k \in Z\}$ is some sequence in l^2. Taking the Fourier transform of both sides we get

$$\widehat{N}_m(w) = \frac{1}{2}\left(\sum p_{m,k}\, e^{-ikw/2}\right)\widehat{N}_m(\frac{w}{2}). \tag{3.15}$$

This formula can be applied to determine the sequence $\{p_{m,k}\}$. Indeed,

$$\widehat{N}_m(w) = \left(\frac{1-e^{-iw}}{iw}\right)^m \tag{3.16}$$

since N_m is the m-fold convolution of N_1 and $\widehat{N}_1(w) = \dfrac{1-e^{-iw}}{iw}$.

Substituting (3.16) into (3.15) we get,

$$\frac{1}{2}\sum_{k=-\infty}^{\infty} p_{m,k}\, e^{-ikw/2} = \left(\frac{1-e^{-iw}}{iw}\right)^m \left(\frac{iw/2}{1-e^{-iw/2}}\right)^m = \left(\frac{1+e^{-iw/2}}{2}\right)^m = 2^{-m}\sum_{k=0}^{m}\binom{m}{k}\, e^{-ikw/2}$$

and this yields

$$p_{m,k} = \begin{cases} 2^{-m+1} \binom{m}{k} & \text{for } 0 \le k \le m \\ 0 & \text{otherwise} \end{cases}.$$

Consequently, the precise formulation of (3.14) is given by

$$N_m(x) = \sum_{k=0}^{m} 2^{-m+1} \binom{m}{k} N_m(2x-k),$$

which is called the **"two-scale relation"** for the cardinal B-splines of order m.

3.4 Construction of Spline Interpolation Formulas

To construct a spline interpolation operator, it is very important to require the operator to reproduce polynomials at least up to some desirable degree. This requirement not only helps in achieving a tolerable order of approximation, but is also critical in preserving certain shapes of the given data. After all, to interpolate a set of constant data, one expects to use a (horizontal) straight line [1], [7], [38], [41].

We will first give a brief discussion of the cardinal spline interpolation problem, namely; for any given "admissible" data sequence $\{f_j\}$ we want to determine the solution $\{c_k\}$ in

$$\sum_{k=-\infty}^{\infty} c_k N_m(x + \frac{m}{2} - k)\Big|_{x=j} = f_j \, , \, j \in Z \, . \tag{3.17}$$

Here, $\{f_j\}$ is said to be admissible if it has at most polynomial growth.

Central to our discussion is the goal of constructing the so-called "Fundamental splines" that interpolate the data $\{\delta_{j,o}\}$. With a fundamental spline on hand, an interpolation operator may be readily obtained by using any given data sequence as the coefficient sequence of the spline series formed by integer translates of the fundamental spline.

Let us first investigate the cardinal spline interpolation stated in (3.17) with data sequence $\{\delta_{j,o}\}$. By solving the bi-infinite system

$$\sum_{k=-\infty}^{\infty} c_k^{(m)} N_m\left(\frac{m}{2}+j-k\right) = \delta_{j,o} \qquad j\in Z, \tag{3.18}$$

for $\{c_k^{(m)}\}$ we have an mth order "fundamental cardinal spline function"

$$L_m(x) = \sum_{k=-\infty}^{\infty} c_k^{(m)} N_m\left(x+\frac{m}{2}-k\right), \tag{3.19}$$

that has the interpolation property

$$L_m(x)=\delta_{j,o} \tag{3.20}$$

as given by (3.18). In contrast to the cardinal B-spline N_m which has compact support, we will see that the coefficient sequence $\{c_k^{(m)}\}$ is not finite for each $m\geq 3$, so that the fundamental cardinal spline L_m does not vanish identically outside any compact set. Hence, when it is applied to interpolate a given data sequence $\{f_j\}$, where $f_j=f(j)$ for some $f\in C$, say, one has to be careful about the convergence of the infinite spline series

$$(J_m f)(x):=\sum_{k=-\infty}^{\infty} f(k)L_m(x-k) . \tag{3.21}$$

Fortunately, as we will see in a moment, $\{c_k^m\}$ decays to zero exponentially fast as $k\to\pm\infty$. This implies that the fundamental cardinal spline function $L_m(x)$ also decays to zero at the same rate as $x\to\pm\infty$. Thus, if $\{f(k)\}$ is of at most polynomial growth, then the series in (3.21) certainly converges at every $x \in R$; and in view of the interpolation property (3.20), we have

$$(J_m f - f)(j) = 0, \quad j \in Z. \tag{3.22}$$

This means, the interpolation spline operator J_m gives a spline function $J_m f$ that interpolates the given data function f at every $x = j, \quad j \in Z$.

To study the fundamental cardinal spline functions $L_m(x)$, we must return to the system (3.18) of linear equations whose coefficients are given by the B-spline values $N_m\left(\frac{m}{2} + k\right)$. Now we consider the symbol

$$\tilde{N}_m(z) = \sum_k N_m\left(\frac{m}{2} + k\right) z^k,$$

and note that this symmetric Laurent polynomial can be easily transformed into an algebraic polynomial with integer coefficients by considering

$$E_{m-1}(z) := (m-1)! z^{\left[(m-1)/2\right]} \tilde{N}_m(z), \tag{3.23}$$

where [x] denotes the largest integer not exceeding x. This notion generalizes the definition of Euler-Frobenius polynomials from even-order cardinal B-splines to those of arbitrary orders. For more details see [10]. The most important property of the Euler-Frobenius polynomial E_{m-1} in (3.23) for our purpose here is that it does not vanish on the unit circle $|z| = 1$. Hence, it follows that $\tilde{N}_m(z) \neq 0$ for all $z = e^{-iw}, w \in R$. Now the system of linear equations (3.18) can be written as

$$\tilde{C}_m(z) = \frac{1}{\tilde{N}_m(z)} \tag{3.24}$$

where $\widetilde{C}_m(z)$ is the symbol of $\{c_k^{(m)}\}$. By using partial fractions, it is easy to see that the sequence $\{c_k^{(m)}\}$ exponentially decays as $k \to \pm\infty$, and the decay rate is given by the magnitude of the root of E_{m-1} in $|z|<1$ which is closest to the unit circle $|z|=1$. This formula can be used for computing $L_m(x)$.

Example 3.7

Determine the cubic fundamental cardinal spline $L_4(x)$.

By applying (ix) in Theorem (1.9), the non-zero values of $N_4(k)$, $k \in Z$, are found to be

$$\{N_4(1), N_4(2), N_4(3)\} = \left\{\frac{1}{4}, \frac{4}{6}, \frac{1}{6}\right\}$$

Hence, the corresponding Euler-Frobenius polynomial is given by

$$E_3(z) = 1 + 4z + z^2 = \left(z + 2 - \sqrt{3}\right)\left(z + 2 + \sqrt{3}\right).$$

Consequently, we have

$$C_4(z) = \frac{(4-1)!\, z^{\left[\frac{(4-1)}{2}\right]}}{E_3(z)} = \frac{6z}{(z+2-\sqrt{3})\,(z+2+\sqrt{3})} \tag{3.25}$$

$$= \frac{6}{(-2+\sqrt{3}) - (-2-\sqrt{3})}\left(\frac{-2+\sqrt{3}}{z+2-\sqrt{3}} - \frac{-2-\sqrt{3}}{z+2+\sqrt{3}}\right)$$

$$= \sqrt{3}\left(\sum_{n=0}^{\infty}\left(-2+\sqrt{3}\right)^{n+1} z^{-n-1} + \sum_{n=0}^{\infty}\left(-2-\sqrt{3}\right)^{-n} z^{n}\right)$$

$$= \sqrt{3}\sum_{n=-\infty}^{\infty}\left(-2+\sqrt{3}\right)^{|n|} z^{n},$$

so that the sequence $\{c_k^{(4)}\}$ is given by

$$c_k^{(4)}=(-1)^k\sqrt{3}\left(2-\sqrt{3}\right)^{|k|},\quad k\in Z. \tag{3.26}$$

This yields the cubic fundamental cardinal spline

$$L_4(x)=\sum_{k=-\infty}^{\infty}(-1)^k\sqrt{3}\left(2-\sqrt{3}\right)^{|k|}N_4(x+2-k) \tag{3.27}$$

Observe that the rate of decay of $L_4(x)$ is

$$O\left(\left(2-\sqrt{3}\right)^{|x|}\right),\quad as\quad x\to\pm\infty, \tag{3.28}$$

in view of the fact that $\mathrm{Supp}N_4(\cdot+2-k)=[k-2,k+2]$.

3.5 Interpolatory Spline-Wavelets

The only wavelet we are very familiar with so far, at least in explicit formulation, is the Haar wavelet $\psi_1=\psi_H$. On the one hand, its companion scaling function is the first order cardinal B-spline N1, namely:

$$\psi_H=N_1(2x)-N_1(2x-1) \tag{3.29}$$

while on the other hand, it is interesting to note that ψ_H is also related to the derivative of the second order cardinal B-spline N2, in the sense that

$$\psi_H(x)=N_2'(2x). \tag{3.30}$$

It is therefore natural to ask to what extent the observation in (3.30) would generalize. To answer this question, let us first remark that the second order cardinal B-spline N_2 can be viewed as a fundamental cardinal spline. In fact, the second order fundamental cardinal spline function L_2, defined as in (3.19) - (3.20), is given by

$$L_2(x) = N_2(x+1).$$

Hence, an equivalent statement of assertion (3.30) is

$$\psi_H(x) = L_2'(2x-1). \qquad (3.31)$$

If we follow this point of view, then we can get spline-wavelets of arbitrary orders. To be precise, let $\{V_j^m\}$ be the MRA of $L^2(R)$ generated by the m^{th} order cardinal B-spline, and let $\{W_j^m\}$, $j \in Z$, denote the sequence of orthogonal complementary (wavelet) spaces, in the sense that

$$V_{j+1}^m = V_j^m \oplus W_j^m, \quad j \in Z, \qquad (3.32)$$

where it should be recalled that the circle around the plus sign indicates orthogonal summation . In the following, for each positive integer m, L_m denotes the m^th^ order fundamental cardinal spline function introduced in (3.19) – (3.20).

Theorem 3.8

Let m be any positive integer, and define

$$\psi_{I,m} = L_{2m}^{(m)}\left(2x-1 \right), \qquad (3.33)$$

where L_{2m} is the $(2m)^{th}$ order fundamental cardinal spline. Then $\psi_{I,m}$ generates the (wavelet) spaces W_j^m, $j \in Z$, in the sense that

$$W_j^m = clos_{L^2(R)} \langle 2^{j/2} \psi_{I,m}\left(2^j x - k\right) : k \in Z \rangle, j \in Z. \qquad (3.34)$$

Proof

Let us first verify that $\psi_{I,m}$ is in W_0^m. For every $n \in Z$, by applying successive integration by parts and noting that the m^th^ derivative of the m^th^ order cardinal B-spline N_m is a finite linear combination of integer translates of the delta distribution, we have

$$\begin{aligned}\left\langle N_m(\cdot - n), \psi_{m,I} \right\rangle &= \int_{-\infty}^{\infty} N_m(x-n) L_{2m}^{(m)}(2x-1)\,dx \\ &= \frac{(-1)^m}{2^m} \int_{-\infty}^{\infty} L_{2m}(2x-1) N_m^{(m)}(x-n)\,dx \\ &= \sum_{k=0}^{m} \frac{1}{2^m} (-1)^{m-k} \binom{m}{k} \int_{-\infty}^{\infty} L_{2m}(2x-1)\,\delta(x-n-k)\,dx \\ &= \sum_{k=0}^{m} \frac{1}{2^m} (-1)^{m-k} \binom{m}{k} L_{2m}(2n+2k-1) = 0\end{aligned}$$

since $L_{2m}(\ell) = \delta_{\ell,o}$, $\ell \in Z$. Hence, $\psi_{I,m} \in W_0^m$.

Next, let us investigate the two-scale relation of $\psi_{I,m}$ with respect to $N_m(2x-k)$, $k \in Z$. That is, we are interested in studying the l^2 sequence $\{q_k\}$ for which

$$\psi_{I,m}(x) = L_{2m}^{(m)}(2x-1) = \sum_{k=-\infty}^{\infty} q_k\, N_m(2x-k). \tag{3.35}$$

Keeping the same notation as in (3.19), we write

$$L_{2m}(x) = \sum_{k=-\infty}^{\infty} c_k^{(2m)} N_{2m}(x+m-k). \tag{3.36}$$

On the other hand, by applying the cardinal B-spline identity (iv) in Theorem 1.8 repeatedly, it follows that

$$
\begin{aligned}
N_{2m}^{(m)}(x) &= \left(\Delta N_{2m-1}^{(m-1)}\right)(x) \\
&= \cdots = \left(\Delta^{m} N_{m}\right)(x) \\
&= \sum_{k=0}^{m} (-1)^{k} \binom{m}{k} N_{m}(x-k),
\end{aligned}
\tag{3.37}
$$

where Δ denotes the backward difference operator . Hence we obtain, from (3.33), (3.36) and (3.37),

$$
\begin{aligned}
\psi_{I,m}(x) = L_{2m}^{(m)}(2x-1) &= \sum_{k=-\infty}^{\infty} c_{k}^{(2m)} N_{2m}^{(m)}(2x-1+m-k) \\
&= \sum_{k=-\infty}^{\infty} c_{k}^{(2m)} \sum_{\ell=0}^{m} (-1)^{\ell} \binom{m}{l} N_{m}(2x-1+m-k-\ell) \\
&= \sum_{n=-\infty}^{\infty} q_{n} N_{m}(2x-n),
\end{aligned}
$$

with

$$
q_{n} := \sum_{\ell=0}^{m} (-1)^{\ell} \binom{m}{l} c_{m+n-1-\ell}^{(2m)} . \tag{3.38}
$$

The two-scale symbol Q corresponding to the two-scale sequence $\{q_k\}$ in (3.35), as given by (3.38), is now

$$
\begin{aligned}
Q(z) &= \frac{1}{2} \sum_{n=-\infty}^{\infty} \left(\sum_{\ell=0}^{m} (-1)^{\ell} \binom{m}{l} c_{m+n-1-\ell}^{(2m)} \right) z^{n} \\
&= \frac{z^{-m+1}}{2} (1-z)^{m} \sum_{n=-\infty}^{\infty} c_{n}^{(2m)} z^{n} .
\end{aligned}
\tag{3.39}
$$

Where it follows from (3.36) and the interpolatory property $L_{2m}(l) = \delta_{\ell,o.}$ that

$$
\sum_{n=-\infty}^{\infty} c_{n}^{(2m)} z^{n} = \frac{1}{F_{m}(z)}, \tag{3.40}
$$

with

$$
\begin{aligned}
F_m(z) := E_{N_m}(z) &= \sum_{k=-m+1}^{m-1} N_{2m}(m+k)\, z^k \\
&= \sum_{k=-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} N_m(k+x)\, N_m(x)\, dx \right\} z^k,
\end{aligned}
\tag{3.41}
$$

being the generalized Euler-Frobenius Laurent polynomial relative to the m$^{\text{th}}$ order cardinal B-spline N_m . In spline theory, where algebraic polynomials with integer coefficient are very desirable, the Euler-Frobenius polynomials of order $2m-1$ are defined by

$$E_{2m-1}(z) := (2m-1)!\, z^{m-1}\, F_m(z) \ .$$

Thus, substituting (3.14) into (3.39), we have found the formula for the two-scale symbol Q, namely;

$$Q(z) = \frac{z^{-m+1}}{2}(1-z)^m \frac{1}{F_m(z)}. \tag{3.42}$$

Note that F_m never vanishes on the unit circle, see [10]. Now, since the two-scale symbol of N_m is given by

$$P(z) = P_{N_m}(z) = \left(\frac{1+z}{2}\right)^m,$$

we can compute the determinant $\Delta_{P,Q}$ as follows:

$$
\begin{aligned}
\Delta_{P,Q}(z) &:= \det \begin{bmatrix} P(z) & Q(z) \\ P(-z) & Q(-z) \end{bmatrix} \\
&= \frac{(-z)^{-m+1}(1+z)^{2m}}{2^{m+1} F_m(-z)} - \frac{z^{-m+1}(1-z)^{2m}}{2^{m+1} F_m(z)} \\
&= 2^{m-1}\left[\frac{(P(z))^2}{\Pi_m(-z)} - \frac{(P(-z))^2}{\Pi_m(z)} \right]
\end{aligned}
\tag{3.43}
$$

$$= \quad 2^{m-1} \frac{z\Pi_m(z^2)}{\Pi_m(-z)\Pi_m(z)} = (-2)^{m-1} \frac{zF_m(z^2)}{F_m(-z)F_m(z)}.$$

Since F_m never vanishes on $|z|=1$, we have shown that

$$\Delta_{P,Q}(z) \neq 0, \quad |z| = 1.$$

Hence, $\{\psi(\cdot - k) : k \in Z\}$ is a Riesz basis of W_0 and this complete the proof.

Chapter 4

Differential Equations and Wavelets

In this chapter, we will present the numerical approximation of solutions of differential equations. The first section will be devoted to some basic definitions and results from linear algebra. The condition number of a matrix that measures the stability of the solution will be presented in the second section. Finally, we will introduce two methods for solving differential equations numerically. The first is called the finite difference method. It is based on approximating the derivatives that are involved in the differential equations using differences. The second, which will be presented in the fourth section, is a class of methods that is called the Galerkin-Wavelet methods. There, we will describe how wavelets could be used efficiently on the numerical solution of differential equations.

4.1 Basic Definitions and Results from Linear Algebra

Definition 4.1

Let V be a vector space over C. A (complex) inner product is a map $\langle \cdot, \cdot \rangle : V \times V \to C$ with the following properties:

i. $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$,

ii. $\langle \alpha\, u, v\rangle = \alpha \langle u, v\rangle$ for all $\alpha \in C$ and all $u, v \in V$,

iii. $\langle u, v\rangle = \overline{\langle v, u\rangle}$ for all $u, v \in V$,

iv. $\langle u, u\rangle \geq 0$ for all $u \in V$, and $\langle u, u\rangle = 0$ if and only if $u = 0$.

A vector space V with a complex inner product is called a (complex) inner product space.

Definition 4.2

Let V be a vector space over C with a complex inner product $\langle \cdot, \cdot \rangle$. For $v \in V$, define $\| v \| = \sqrt{\langle \cdot, \cdot \rangle}$. We call $\| v \|$ the norm or the length of v.

Definition 4.3

Let $A = [a_{ij}]$ be an $m \times n$ matrix over C.

The transpose A^t of A is the $n \times m$ matrix $B = [b_{ij}]$ defined by $b_{ij} = a_{ji}$, for all i, j

The conjugate transpose A^* of A is the $n \times m$ matrix $C = [c_{ij}]$ defined by $c_{ij} = \overline{a_{ji}}$, for all i, j.

Definition 4.4

Let A be an $n \times n$ matrix. A is unitary if A is invertible and $A^{-1} = A^*$.

Definition 4.5

An $n \times n$ matrix A is normal if $A^*A = A\,A^*$.

Definition 4.6

Let A be an $n \times n$ matrix. A is Hermition if $A^* = A$.

Definition 4.7

A matrix A is said to be circulant if $a_{m+k,\,n+k} = a_{m,\,n}$

Theorem 4.8

Let $T : l^2(Z_n) \rightarrow l^2(Z_n)$ be a linear transformation.

Then the following statements are equivalent:

a. T is translation invariant.

b. The matrix A representing T in the standard basis E is circulant.

c. The matrix A representing T in the Fourier basis F is diagonal.

Theorem 4.9

(Spectral theorem for matrices) Let A be an $n \times n$ matrix over C. Then the following statements are equivalent:

a. A is unitarily diagonalizable.

b. A is normal.

c. There is an orthonormal basis for C^n consisting of eigenvectors of A.

4.2 The Condition Number of a Matrix

Many applications of mathematics require the numerical approximation of solutions of differential equations. The methods for numerically solving a linear ordinary differential equation come down to solving a linear system of equations, or equivalently, a matrix equation $A\,x = y$. For such system to have a unique solution x for every y, the matrix A should be invertible. However, in applications there are further issues that are of crucial importance. One of these has to do with the condition number of the matrix.

Definition 4.10

Let A be an $n \times n$ matrix. Define $\|A\|$, called the norm of A by

$$\|A\| = \sup \frac{\|A\,z\|}{\|z\|},$$

where the supremum is taken over all nonzero vectors in C^n.

Definition 4.11

Let A be an invertible $n \times n$ matrix. Define $C_{\#}(A)$, the condition number of A, by $C_{\#}(A) = \|A\|\,\|A^{-1}\|$. If A is not invertible, set $C_{\#}(A) = \infty$.

Example 4.12

Consider the linear system $Ax = y$, where $x, y \in C^2$, and

$$A = \begin{bmatrix} 5.95 & -14.85 \\ 1.98 & -4.94 \end{bmatrix}.$$

The determinant of A is $.01$, which is not 0, so A is invertible. For

$$y = \begin{bmatrix} 3.05 \\ 1.02 \end{bmatrix},$$

the unique solution to $A\,x = y$ is

$$x = \begin{bmatrix} 8 \\ 3 \end{bmatrix}$$

now, if we suppose

$$y' = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

Then the solution to $A\,x' = y'$ is

$$x' = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

Note that y and y' are close but x and x' are far a part. A linear system for which this happens is called badly conditioned. In this situation, small errors in the data can lead to large errors in the solution. This is undesirable in applications, because in nearly all computations with real data there is an error either due to rounding off or due to imperfect measurement of the data. For a badly conditioned system, the apparent solution can be meaningless physically.

Lemma 4.13

Suppose that A is an $n \times n$ normal invertible matrix. Let

$$|\lambda|_{\max} = \max\left\{ |\lambda| : \lambda \ \textit{is an eigenvalue of } A \right\}$$

and

$$|\lambda|_{\min} = \min\left\{ |\lambda| : \lambda \ \textit{is an eigenvalue of } A \right\}$$

then

$$C_{\#}(A) = \frac{|\lambda|_{\max}}{|\lambda|_{\min}}.$$

The condition number of a matrix A measures the stability of the linear system $A\,x = y$ under perturbations of y. The stability of the linear system is most naturally described by comparing the relative size $\|\delta x\| / \|x\|$ of the change of the solution to the relative size $\|\delta y\| / \|y\|$ of the change in the given data. The condition number is the maximum value of this ratio.

Theorem 4.14

Suppose A is an $n \times n$ invertible matrix, $x, y, \delta x, \delta y \in C^n, x \neq 0, \quad A x = y,$ and $A \delta x = \delta y$. Then

$$\frac{\| \delta x \|}{\| x \|} \leq C_{\#}(A) \frac{\| \delta y \|}{\| y \|}. \tag{4.1}$$

Proof

To show this we need to show that

$$\| A x \| \leq \| A \| \| x \|. \tag{4.2}$$

By the definition of $\| A \|$ we have $\| A \| \geq \left\| A \frac{x}{\|x\|} \right\|$. Since $\left\| x / \|x\| \right\| = 1$ for $x \neq 0$

we get (4.2). Now, let $y = A x,$ using (4.2), we have

$$\| y \| \leq \| A \| \| x \| \tag{4.3}$$

Similarly, since $\delta x = A^{-1} \delta y$,

$$\| \delta x \| \leq \| A^{-1} \| \| \delta y \| \tag{4.4}$$

By multiplying inequalities (4.3) and (4.4) we get

$$\| y \| \| \delta x \| \leq C_{\#}(A) \| x \| \| \delta y \|,$$

which is equivalent to inequality (4.1).

The condition number of A measures how unstable the linear system $A x = y$ is under perturbation of the data y. In applications, therefore, a small condition number is desirable. If it happens that the condition number of A is high, it is recommended to replace the linear system $A\,x = y$ by an equivalent system whose matrix has a low condition number. Multiplying the system $A\,x = y$ by a preconditioning matrix B to obtain the equivalent system $B\,A\,x = B\,y$ such that $C_{\#}(BA)$ is smaller than $C_{\#}(A)$ could do this.

4.3 Finite Difference Methods for Differential Equations

For the sake of simplicity we will concentrate on the following problem.

$$\begin{cases} -u''(t) = f(t) & t \in (0,1) \\ u\,(0) = u(1) = 0, \end{cases} \tag{4.5}$$

where $f : [0,1] \to C$ is assumed to be continuos. Our target is to find a C^2 function u that satisfies (4.5). Note that at the end points 0 and 1, derivatives are interpreted in the one-sided sense.

Theoretically this equation is well known. And the unique solution u of the two point boundary value problem (4.5) is given by

$$u\,(x) = -\int_0^x\!\int_0^t f(s)\,ds\,dt + x\int_0^1\!\int_0^t f(s)\,ds\,dt\,. \tag{4.6}$$

However, if f is a function whose antiderivative can not be expressed in terms of elementary functions, it may not be possible to explicitly evaluate (4.6). One approach to approximate the solution u is to numerically estimate the integrals in this formula. Another way, which is more general, is the finite difference method.

Methods involving finite differences for solving boundary-value problems replace each of the derivatives in the differential equation by an appropriate difference-quotient evaluated on a finite set of points in the interval [0, 1].

By the definition of the derivative,

$$u'(t) \approx \frac{u(t+\Delta t)-u(t)}{\Delta t}$$

for small Δt, which is refered to as the forward difference formula. For reasons of symmetry, let $h>0$, consider both $\Delta t = h/2$ and $\Delta t = -h/2$, and average will produce the center difference formula

$$u'(t) \approx \frac{1}{2}\left[\frac{u(t+\frac{h}{2})-u(t)}{\frac{h}{2}} + \frac{u(t)-u(t-\frac{h}{2})}{\frac{h}{2}}\right]$$

$$= \frac{u(t+\frac{h}{2})-u(t-\frac{h}{2})}{h}.$$

Applying this to u' we obtain

$$u''(t) \approx \frac{u'(t+\frac{h}{2})-u'(t-\frac{h}{2})}{h} \approx \frac{u(t+h)-2u(t)+u(t-h)}{h^2}. \tag{4.7}$$

Now, consider the partition

$$t_j = \frac{j}{N}, \quad j = 0, 1, \ldots, N$$

and let $h = 1/N$.

Also set

$$x(j) = u\left(\frac{j}{N}\right) \text{ and } y(j) = \frac{1}{N^2} f\left(\frac{j}{N}\right), \text{ for } j = 0, 1, \ldots, N \tag{4.8}$$

To solve $-u''(t_j) = f(t_j)$, we approximate $u''(t_j)$ using (4.7). Now, consider the system of equations

$$-u\left(\frac{j+1}{N}\right)+2u\left(\frac{j}{N}\right)-u\left(\frac{j-1}{N}\right)=\frac{1}{N^2}f\left(\frac{j}{N}\right),$$

with boundary conditions $u(0)=u(1)=0$, to make sense we restrict ourselves to $1\le j\le N-1$. Thus we consider

$$-x(j+1)+2x(j)-x(j-1)=y(j) \text{ for } j=1,\dots,N-1, \tag{4.9}$$

with the boundary conditions

$$x(0)=0 \text{ and } x(N)=0. \tag{4.10}$$

Equation (4.9) is a linear system of $N-1$ equations in the $N-1$ unknowns $x(1),\dots,x(N-1)$ represented by the matrix equation

$$\begin{bmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & -1 & 2 & -1 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x(1) \\ x(2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x(N-1) \end{bmatrix} = \begin{bmatrix} y(1) \\ y(2) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y(N-1) \end{bmatrix} \tag{4.11}$$

which we denote by

$$A_N\,x = y.$$

We claim that $\det A_N = N$. We will use induction to show that

$$\det A_{N+1} = 2\det A_N - \det A_{N-1} \qquad \text{for} \quad N\ge 3 \tag{4.12}$$

Now for $N=3$

$$A_{N+1}=A_4=\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \Rightarrow \det A_4 = 2\begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} + \begin{vmatrix} -1 & -1 \\ 0 & 2 \end{vmatrix}$$

$$= 2\,(\,3\,) - 2 = 2 \det A_3 - \det A_2$$

so, (4.12) is true for $N = 3$. Assume that (4.12) is true for $N = k$, we need to show that it is true for $N = k+1$. Indeed,

$$\det A_{k+1} = \det \begin{bmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & 0 & -1 & 2 & -1 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & -1 & 2 \end{bmatrix}$$ expanding along the first row we get

$$\det A_{k+1} = 2 \begin{vmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & -1 & 2 & -1 \\ 0 & 0 & \cdot & \cdot & \cdot & -1 & 2 \end{vmatrix} + \begin{vmatrix} -1 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 2 & -1 & 0 & \cdot & \cdot & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdot & \cdot \\ \cdot & 0 & -1 & 2 & -1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & -1 & 2 & -1 \\ 0 & 0 & \cdot & \cdot & 0 & -1 & 2 \end{vmatrix}$$

Expanding the second term along the first column we obtain

$$\det A_{k+1} = 2 \begin{vmatrix} 2 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & -1 & 2 & -1 \\ 0 & 0 & \cdot & \cdot & \cdot & -1 & 2 \end{vmatrix} + (-1) \begin{vmatrix} 2 & -1 & 0 & \cdot & \cdot & 0 \\ -1 & 2 & -1 & 0 & \cdot & \cdot \\ 0 & -1 & 2 & -1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & -1 & 2 & -1 \\ 0 & \cdot & \cdot & 0 & -1 & 2 \end{vmatrix}$$

$= 2 \det A_k - \det A_{k-1}$. Hence, (4.12) holds for all $N \geq 3$. Now, using induction again it is clear that $\det A_N = N$. Hence, A_N is invertible, and the system $A_N x = y$ has a unique solution x for each vector y.

As we let $h \to 0$; that is, $N \to \infty$, we expect our solution to approximate the true values of u in (4.8) with greater accuracy. However, in general it is important numerically to have a well-conditioned linear system. So next we study the condition number of A_N.

Observe that A_N is real and symmetric, hence Hermitian. Therefore, it is normal. By Lemma 4.13, $C_{\#}(A) = |\lambda|_{\max} / |\lambda|_{\min}$. To compute the eigenvalues of A_N, we consider the matrix B_{N-1} that agrees with A_N except that the entries of B_{N-1} in the top right and lower left corners are -1 instead of 0. Then B_{N-1} is circulant. Hence, it is diagonalizable and we can find its eigenvalues using Theorem 4.8.

Another way to see the relation between A_N and the circulant variant is to observe that A_N is the $N-1 \times N-1$ submatrix obtained by deleting the first row and the first column of the $N \times N$ matrix B_N. We see that we can get information regarding the eigenvalues of A_N from those of B_N.

Suppose $x' = (x(0), x(1), \ldots, x(N-1))$ is an eigenvector of B_N such that $x(0) = 0$. Let λ be the associated eigenvalue. Let $x = (x(1), x(2), \ldots, x(N-1))$. Then $B_N\, x'(j) = A_N\, x(j)$ for $j = 1, 2, \ldots, N-1$ because the condition $x(0) = 0$ guarantees that the first column of B_N has no effect on the value of $B_N\, x'(j)$. Therefore,

$$A_N\, x(j) = B_N\, x'(j) = \lambda\, x'(j) = \lambda\, x(j),$$

for $j = 1, 2, \ldots, N-1$. So, x is an eigenvector of A_N with eigenvalue λ.

Because B_N is circulant, Theorem 4.8 says that its eigenvectors are the element of the Fourier basis $F_0, F_1, \ldots, F_{N-1}$. The eigenvalues of B_N are given by $\lambda_j = 4\sin^2(\pi j/N)$. It is not clear that there exists an eigenvector x' of B_N satisfies $x'(0) = 0$. However, if $1 \le j < N/2$, we have $\lambda_{N-j} = \lambda_j$, so the eigenspace corresponding to λ_j is two-dimensional, spanned by F_j and F_{N-j}. Therefore a linear combination of F_j and F_{N-j} belongs to this eigenspace. For $1 \le j < N/2$ and with $i = \sqrt{-1}$, define $K_j \in l^2(Z_N)$ by

$$K_j(n) = \frac{N}{2i}\left(F_j(n) - F_{N-j}(n)\right)$$

$$= \frac{1}{2i}\left(e^{2\pi i j n/N} - e^{-2\pi i j n/N}\right) = \sin\left(\frac{2\pi j n}{N}\right).$$

Then $K_j(0) = 0$ and $B_N K_j = \lambda_j K_j$. By the above discussion, this implies that the vector of length $N-1$ obtained by deleting the first component from K_j is an eigenvector of A_N with eigenvalue λ_j. Let $m = 2j$; define vectors G_m of length $N-1$ for $1 \le m \le N-1$ by

$$G_m(n) = \sin\left(\frac{\pi m n}{N}\right) \qquad \text{for } n = 1, 2, \ldots, N-1. \tag{4.13}$$

We claim that G_m for $1 \le m \le N-1$ are the eigenvectors of A_N with eigenvalues λ_m.

Indeed,

$$(A_N\, G_m) = -G_m(l-1) + 2\,G_m(l) - G_m(l+1)$$

$$= -\sin\left(\frac{\pi m(l-1)}{N}\right) + 2\sin\left(\frac{\pi m l}{N}\right) - \sin\left(\frac{\pi m(l+1)}{N}\right)$$

$$= -\left[\sin\frac{\pi m l}{N}\cos\frac{\pi m}{N} - \cos\frac{\pi m l}{N}\sin\frac{\pi m}{N}\right] + 2\sin\frac{\pi m l}{N}$$

$$-\left[\sin\frac{\pi m l}{N}\cos\frac{\pi m}{N} + \cos\frac{\pi m l}{N}\sin\frac{\pi m}{N}\right]$$

$$= \left[\ 2 - 2\cos\left(\frac{\pi m}{N}\right)\right]\sin\left(\frac{\pi m l}{N}\right) = 4\sin^2\left(\frac{\pi m}{2N}\right)G_m(l).$$

Thus, we see that the eigenvectors of A_N are G_m for $1 \le m \le N-1$ with corresponding eigenvalues $4\sin^2\left(\frac{\pi m}{2N}\right)$. These eigenvalues are distinct, so are the eigenvectors and we have a complete set of eigenvectors for the matrix A_N. Now, the condition number of A_N is given by

$$C_\#(A_N) = \frac{|\lambda|_{\max}}{|\lambda|_{\min}} = \frac{4\sin^2\left(\frac{\pi(N-1)}{2N}\right)}{4\sin^2\left(\frac{\pi}{2N}\right)}.$$

As $N \to \infty$, $\sin^2(\pi(N-1)/2N) \to 1$, whereas $\sin^2(\pi/2N)$ behaves like $\pi^2/4N^2$. Thus

$$C_\#(A_N) \approx \frac{4N^2}{\pi^2}.$$

So the condition number of A_N goes to $+\infty$ as fast as N^2. So although what should happen is that increasing N should increase the accuracy of the approximation to the solution u of equation (4.5). The linear $A_N\, x = y$ becomes increasingly unstable and the solution becomes more and more unreliable.

For the simple case (4.5) we were able to explicitly diagonalize the matrix A_N arising in the finite difference method. This is due to the fact that the operator L defined by $Lu=-u''$ is translation invariant. Consequently the matrix A_N was close to circulant, in the sense that A_N is closely related to the circulant matrix B_N. This and a bit of luck enabled us to find the eigenvectors of A_N. However, if L is a linear variable coefficient ordinary differential operator; that is, operator of the form

$$(Lu)(t)=L(u)(t)=\sum_{j=0}^{N} b_j \frac{d^j}{dt^j} u(t)$$

where the coefficients b_j are allowed to vary with t, L will not be translation invariant. Also the matrix A arising in the finite difference approximation to the solution of $Lu=f$ in $[0, 1]$ with the boundary conditions $u(0)=u(1)=0$, will not be closed to circulant. And even if A is diagonalizable (which is not clear), it is not clear how to explicitly diagonalize it. We also expect the condition number of A to be large because that is the case in the much simpler case (4.5).

An alternative approach will be using wavelets that includes the variable coefficient case. That is what we will consider in the next section. This approach leads to linear systems with bounded condition numbers.

4.4 Wavelet-Galerkin Methods

In this section, we will consider another approach to the numerical solution of ordinary differential equations, known as the Galerkin methods. We will see that using wavelets together with Galerkin method give the two main desired properties for the associated linear system namely, sparseness and low condition number, [20], [27] and [32].

We consider the class of ordinary differential equations of the form

$$Lu(t) = -\frac{d}{dt}\left(a(t)\frac{du}{dt} \right) + b(t)\,u(t) = f(t) \text{ for } 0 \le t \le 1 \tag{4.14}$$

with Dirichlet boundary conditions

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

where a, b and f are given real-valued functions. We assume that b and f are continuous and a has a continuous derivative on $[0, 1]$. We also assume that the operator L is uniformly elliptic; that is, there exist finite constants $C_1, C_2,$ and C_3 such that

$$0 < C_1 \le a(t) \le C_2 \text{ and } 0 \le b(t) \le C_3. \tag{4.15}$$

For the Galerkin method, we suppose that $\{v_j\}_j$ is a complete orthonormal system for $L^2([0, 1])$, and that every $v_j \in C^2[0, 1]$ and satisfies

$$v_j(0) = v_j(1) = 0. \tag{4.16}$$

We select some finite set Λ of indices j and consider the subspace

$$S = span\,\{v_j : j \in \Lambda\}. \tag{4.17}$$

We look for an approximation to the solution u of equation (4.14) of the form

$$u_s = \sum_{k \in \Lambda} x_k\, v_k \in S \tag{4.18}$$

where each x_k is a scalar. These coefficients should be determined such that u_s behaves like the true solution u on the subspace S; that is,

$$\langle Lu_s, v_j \rangle = \langle f, v_j \rangle \text{ for all } j \in \Lambda. \tag{4.19}$$

By linearity, it follows that

$$\langle Lu_s, g \rangle = \langle f, g \rangle \text{ for all } g \in S.$$

Substituting (4.18) in equation (4.19), we get

$$\left\langle L\left(\sum_{k\in\Lambda} x_k\, v_k\right), v_j \right\rangle = \langle f, v_j \rangle, \text{ for all } j \in \Lambda,$$

or

$$\sum_{k\in\Lambda} \langle L v_k, v_j \rangle x_k = \langle f, v_j \rangle, \text{ for all } j \in \Lambda. \tag{4.20}$$

Let x denote the vector $(x_k)_{k\in\Lambda}$, and let y be the vector $(y_k)_{k\in\Lambda}$, where $y_k = \langle f, v_k \rangle$. Let A be the matrix with rows and columns indexed by Λ, that is, $A = [a_{j,k}]_{j,k\in\Lambda}$, where

$$a_{j,k} = \langle L v_k, v_j \rangle. \tag{4.21}$$

Thus, equation (4.20) is the linear system of equations

$$\sum_{k\in\Lambda} a_{j,k}\; x_k = y_j \quad \text{, for all } j \in \Lambda,$$

or

$$A\,x = y. \tag{4.22}$$

In the Galerkin method, for each subset Λ, we obtain an approximation $u_s \in S$ to u, by solving the linear system (4.22) for x and using these components to find u_s by equation

(4.18). If we increase our set Λ in some systematic way, we expect that our approximation u_s will converge to the actual solution u.

As before we should be concerned about the nature of the linear system (4.22) that results from choosing a wavelet basis for the Galerkin method. We would like the matrix A to have two main properties. First, we would like the matrix A to have small condition number. Second, we would like A to be sparse.

There is a way of modifying the wavelet system for $L^2(R)$ so as to obtain a complete orthonormal system

$$\{\psi_{j,k}\}_{(j,k)\in\Gamma} \tag{4.23}$$

for $L^2([0,1])$. The set Γ is a certain subset of $Z\times Z$. Now $\psi_{j,k}$ is C^2 for each $(j,k)\in\Lambda$, and satisfies the boundary conditions

$$\psi_{j,k}(0)=\psi_{j,k}(1)=0.$$

The wavelet system $\{\psi_{j,k}\}_{(j,k)\in\Gamma}$ also satisfies the following estimate: there exist constants $C_4, C_5>0$ such that for all functions g of the form

$$g=\sum c_{j,k}\,\psi_{j,k} \tag{4.24}$$

for which the sum is finite, we have

$$C_4\sum_{j,k}2^{2j}\left|c_{j,k}\right|^2\le\int_0^1\left|g'(t)\right|^2dt\le C_5\sum_{j,k}2^{2j}\left|c_{j,k}\right|^2. \tag{4.25}$$

The notation used for applying the Galerkin method with these wavelets is somewhat confusing due to the fact that the wavelets are indexed by two integers. Thus using wavelets, we write equation (4.18) as

$$u_s = \sum_{(j,k)\in\Lambda} x_{j,k}\, \psi_{j,k}$$

and as a result equation (4.20) becomes

$$\sum_{(j,k)\in\Lambda} \left\langle L\psi_{j,k}, \psi_{l,m}\right\rangle x_{j,k} = \left\langle f, \psi_{l,m}\right\rangle \text{ for all } (l,m)\in\Lambda, \tag{4.26}$$

where Λ is some finite set of indices. We can write (4.26) as a matrix equation of the form $Ax = y$, where the vectors $x = (x_{j,k})_{(j,k)\in\Lambda}$ and $y = (y_{j,k})_{(j,k)\in\Lambda}$ are indexed by the pairs $(j, k)\in\Lambda$. while the matrix $A = [a_{l,m;j,k}]_{(l,m),(j,k)\in\Lambda}$ defined by

$$a_{l,m;j,k} = \left\langle L\psi_{j,k}, \psi_{l,m}\right\rangle \tag{4.27}$$

has its rows indexed by the pairs $(l, m)\in\Lambda$ and its columns indexed by the pairs $(j, k)\in\Lambda$.

As suggested, we would like A to be sparse and have a low condition number. Actually A itself does not have a low condition number, but we can replace the system $Ax = y$ by an equivalent system $Mz = v$, for which the new matrix M has the desired properties. To show this, we define the matrix $D = [d_{l,m;j,k}]_{(l,m),(j,k)\in\Lambda}$ by

$$d_{l,m;j,k} = \begin{cases} 2^j \text{ if } (l,m) = (j,k) \\ 0 \quad \text{if } (l,m) \neq (j,k). \end{cases} \tag{4.28}$$

Define the matrix $M = [m_{l,m;j,k}]_{(l,m),(j,k)\in\Lambda}$ by

$$M = D^{-1}\, A\, D^{-1}. \tag{4.29}$$

This means that, we have componentwise

$$m_{l,m;j,k} = 2^{-j-l}\, a_{l,m;j,k} = 2^{-j-l}\left\langle L\psi_{j,k}, \psi_{l,m}\right\rangle. \tag{4.30}$$

The system $Ax = y$ is equivalent to

$$D^{-1}A\,D^{-1}D\,x = D^{-1}y,$$

if we set $z = D\,x$ and $v = D^{-1}y,$ we have

$$M\,z = v. \tag{4.31}$$

The norm equivalence (4.25) has the consequence that the system (4.31) is well conditioned. The following theorem is needed to prove the main result. It explains the need for the uniform ellipticity assumption (4.15).

Theorem 4.15

Let L be a uniformly elliptic Sturm-Liouville operator. Suppose $g \in L^2([0,1])$ is C^2in [0, 1] and satisfies $g(0) = g(1) = 0$. Then

$$C_1 \int_0^1 |\,g'(t)|^2\,dt \le \langle L\,g,\,g\rangle \le (C_2 + C_3) \int_0^1 |\,g'(t)|^2 dt, \tag{4.32}$$

where C_1, C_2, and C_3 are the constants in relation (4.15).

Proof

Observe that

$$\langle -(a\,g')',\,g\rangle = \int_0^1 -(a\,g')'(t)\,\overline{g}(t)\,dt$$

$$= \int_0^1 a(t)g'(t)\,\overline{g'(t)}\,dt = \langle ag',\,g'\rangle.$$

Therefore,

$\langle Lg,\,g\rangle = \langle -(a\,g')' + bg,\,g\rangle = \langle a\,g',\,g'\rangle + \langle b\,g,\,g\rangle$. So by relation (4.15),

$$C_1 \int_0^1 |g'(t)|^2\,dt \le \int_0^1 a(t)\,|g'(t)|^2\,dt = \int_0^1 a(t)\,g'(t)\overline{g'(t)}\,dt = \langle ag',\,g'\rangle. \tag{4.33}$$

Also by relation (4.15) we have,

$$0 \le \int_0^1 b(t)|\,g(t)\,|^2\,dt = \langle bg,\,g\rangle.$$

Adding the above inequalities gives

$$C_1 \int_0^1 |g'(t)|^2\, dt \le \langle Lg, g\rangle,$$

which is the left half of the relation (4.32). For the other half, note that by (4.15) we have,

$$\langle a g', g'\rangle = \int_0^1 a(t)\,|g'(t)|^2\, dt \le C_2 \int_0^1 |g'(t)|^2\, dt. \tag{4.34}$$

Also since $g(0) = 0$,

$$g(t) = \int_0^t g'(s)\, ds,$$

by the fundamental theorem of calculus. Now, if we apply the Cauchy-Schwarz inequality to the functions $g'\chi_{[0,t]}$ and $\chi_{[0,t]}$ where $\chi_{[0,t]}(x) = \begin{cases} 0 & x \notin [0,t] \\ 1 & x \in [0,t] \end{cases}$, we get

$$|g(t)|^2 \le \left(\int_0^t |g'(s)|^2\, ds\right)\left(\int_0^t 1\, ds\right) \le \int_0^1 |g'(s)|^2\, ds$$

for every $t \in [0,1]$. Therefore

$$\int_0^1 |g(t)|^2\, dt \le \int_0^1 |g'(s)|^2\, ds \int_0^1 dt = \int_0^1 |g'(s)|^2\, ds. \tag{4.35}$$

Hence, by (4.15),

$$\langle bg, g\rangle = \int_0^1 b(t)\, g(t)\overline{g(t)}\, dt \le C_3 \int_0^1 |g(t)|^2\, dt \le C_3 \int_0^1 |g'(t)|^2\, dt.$$

This result and (4.34) give the right hand side of (4.32).

Theorem 4.16

Let L be a uniformly elliptic Sturm-Liouvilly operator. Let $\{\psi_{j,k}\}_{(j,k)\in\Gamma}$ be a complete orthonormal system for $L^2([0,1])$ such that each $\psi_{j,k}$ is C^2, with $\psi_{j,k}(0) = \psi_{j,k}(1) = 0$, and such that the norm equivalence (4.25) holds. Let Λbe a finite subset of Γ. Let M be the matrix defined in equation (4.29). Then the condition number of M satisfies

$$C_{\#}(M) \leq \frac{(C_2 + C_3)\, C_5}{C_1\, C_4} \tag{4.36}$$

for any finite set Λ, where C_1, C_2, and C_3 are as given in (4.15), and C_4 and C_5 are the constants in relation (4.25).

Proof

Let $z = (z_{j,k})_{(j,k)\in\Lambda}$ be any vector with $\| z \| = 1$. Let $w = D^{-1} z$ where the matrix D is as given in (4.28); that is, $w = (w_{j,k})_{(j,k)\in\Lambda}$, where

$$w_{j,k} = 2^{-j} z_{j,k}.$$

Define

$$g = \sum_{(j,k)} w_{j,k}\, \psi_{j,k}.$$

Then by equation (4.30),

$$\begin{aligned}
\langle M z, z\rangle &= \sum_{(l,m)\in\Lambda} (M z)_{l,m}\, \overline{z_{l,m}} \\
&= \sum_{(l,m)\in\Lambda} \sum_{(j,k)\in\Lambda} \langle L\psi_{j,k}, \psi_{l,m}\rangle\, 2^{-j} z_{j,k}\, 2^{-l}\, \overline{z_{l,m}} \\
&= \left\langle L\left(\sum_{(j,k)\in\Lambda} w_{j,k}\, \psi_{j,k} \right), \sum_{(l,m)} w_{l,m}\, \psi_{l,m} \right\rangle = \langle Lg, g\rangle,
\end{aligned}$$

since $2^{-j} z_{j,k} = w_{j,k}$ and $2^{-l} z_{l,m} = w_{l,m}$. Applying Theorem 4.15 and relation (4.25) gives

$$\langle M z, z\rangle = \langle Lg, g\rangle \leq (C_2 + C_3) \int_0^1 |g'(t)|^2 dt \leq (C_2 + C_3) C_5 \sum_{(j,k)\in\Lambda} 2^{2j} \left| w_{j,k} \right|^2,$$

and

$$\langle M z, z\rangle = \langle L g, g\rangle \geq C_1 \int_0^1 |g'(t)|^2 \, dt \geq C_1 \, C_4 \sum_{(j,k)\in\Lambda} 2^{2j} \left| w_{j,k} \right|^2 .$$

However,

$$\sum_{(j,k)\in\Lambda} 2^{2j} \left| w_{j,k} \right|^2 = \sum_{(j,k)\in\Lambda} \left| z_{j,k} \right|^2 = \| z \|^2 = 1 .$$

So for any z with $\| z \| = 1$,

$$C_1 \, C_4 \leq \langle M z, z\rangle \leq (C_2 + C_3) \, C_5 .$$

If λ is an eigenvalue of M, we can normalize the corresponding eigenvector z so that $\| z \| = 1$, to obtain

$$\langle M z, z\rangle = \langle \lambda z, z\rangle = \lambda \langle z, z\rangle = \lambda \| z \|^2 = \lambda .$$

Hence, every eigenvalue of M satisfies

$$C_1 \, C_4 \leq \lambda \leq (C_2 + C_3) \, C_5 .$$

Note that M is Hermitian, so it is normal and $C_{\#}(M)$ is given by Lemma (4.13) to be the ratio of the largest eigenvalue to the smallest. And the result follows since all of the eigenvalues are positive.

So the matrix in the preconditioned system $M z = v$ has a bounded condition number independent of the set Λ. As a result, if we increase Λ to approximate our solution with more accuracy, the condition number stays bounded. This is an advantage of the Galerkin method over the finite difference method where the condition number grows as N^2. To see the advantage of the wavelets over the Fourier system we should consider the other feature of the matrix M that is desirable: we would like M to be sparse. Note that $\psi_{j,k}$ is 0 outside an interval of length $c2^{-j}$ around the point $2^{-j}k$, for some constant c. Because L involves only differentiation and multiplication by another function, it does not change this localization property. As a result, $L\psi_{j,k}$ is 0 outside this interval. Similarly, $\psi_{l,m}$ is 0 outside an interval of length $c2^{-l}$ around the point $2^{-l}m$. As we let j and l get large, fewer and fewer of these intervals intersect, so more and more of the matrix entries $m_{l,m;j,k}$ are zeros. This means M is sparse. The basic reason for this sparseness is the compact support of the wavelet.

Although the matrices we obtained using finite differences were sparse, they have large condition numbers. Using the Galerkin method with the Fourier system, one obtains a bounded condition number, but the matrix will be no longer sparse. Using the Galerkin method with a wavelet system, will guarantee both advantages.

The fact that a wavelet system nearly diagonalizes a very broad class of operators is one of the key properties of wavelets. We have seen that this is very important in applications to numerical differential equations. Another main property of wavelets is their combination of spatial and frequency localization. This property is usually used in signal analysis. A third key property of wavelets is that norm equivalencies for wavelets such as relation (4.25) hold for a much larger class of function spaces than the Fourier system.

4.5 Further Research

Consider the following second order boundary value problem [34], [36]

$$\begin{aligned} &y''(t) = f(t,\, y(t),\, y'(t)) \qquad 0 \le t \le 1 \\ &y(0) = y(1) = 0 \end{aligned} \tag{4.37}$$

where $f : [0,1] \times R^2 \to R$, $(t, u, v) \to f(t, u, v)$ is continuous in t, u, and v. Assume that f satisfies the following Lipschitz condition

$$\begin{aligned} &\left| f^{(q)}(t, u_1, v_1) - f^{(q)}(t, u_2, v_1) \right| \le L\left[|u_1 - u_2| + |\, v_1 - v_2 \,| \right] \\ &(t, u_1, v_1), (t, u_2, v_2) \in [0,1] \times R^2. \end{aligned} \tag{4.38}$$

Let Δ be a partition of $[0, 1]$ given by

$$\Delta : 0 = t_0 < t_1 < \cdots < t_k < t_{k+1} < \cdots < t_n = 1,$$

where $t_{k+1} - t_k = h < 1$ and $k = 0, 1, \ldots, n-1$.

The functions $f^{(q)}(t, y, y')$, $q = 1, 2, \ldots, r$ are generated as follows:

$f^{(0)} = f(t, y, y')$ and from $f^{(q-1)}$ we obtain recursively

$$f^{(q)} = \frac{\partial f^{(q-1)}}{\partial t} + \frac{\partial f^{(q-1)}}{\partial y} y' + \frac{\partial f^{(q-1)}}{\partial y'} y''. \tag{4.39}$$

Now define the spline function approximation $y(t)$ by $S_\Delta(t)$:

$$\begin{aligned} S_\Delta(t) = S_k(t) = S_{k-1}(t) + S'_{k-1}(t_k)\,(t - t_k) + S''_{k-1}(t_k)\,\frac{(t-t_k)^2}{2!} \\ + \sum_{j=0}^{r} f^{(j)}\left[t_k, S_{k-1}(t_k), S'_{k-1}(t_k)\right] \frac{(t-t_k)^{j+3}}{(j+3)!} \end{aligned} \tag{4.40}$$

where $S_{-1}^{(i)}(t_0) = y_0^{(i)} = y(0) = 0$. By construction it is clear that

$$S_\Delta(t) \in C^2\big([0,1] \times R^2 \,\big).$$

For all $t \in [t_k, t_{k+1}]$, $k = 0\,(1)\, n-1$, let the exact solution of (4.37) be written in the form

$$y(t) = \sum_{j=0}^{r+2} \frac{y_k^{(j)}}{j!}(t-t_k)^j + y^{(r+3)}(\xi_k)\frac{(t-t_k)^{r+3}}{(r+3)!}$$

where $\xi_k \in (t_k, t_{k+1})$ and $k = 0, 1, \ldots, n-1$.

Before proceeding to the discussion of the convergence of the spline approximation, let us introduce the following notations:

$$\begin{aligned}
&e(t) = |\, y(t) - S_\Delta(t)|, \text{ the exact error} \\
&e_k = |\, y_k - S_\Delta(t_k)|, \text{ the error in the computed solution,} \\
&f_k^{(j)} = f^{(j)}\left[t_k, S_{k-1}(t_k), S'_{k-1}(t_k)\right] \\
&f^{*(j)} = f^{(j)}(t_k, y_k, y'_k) \\
&\text{where } j = 0, 1, \ldots, r \text{ and } k = 0, 1, \ldots, n-1.
\end{aligned} \tag{4.42}$$

First we estimate $|\, y(t) - S_k(t)\,|$. Using (4.40), (4.41), the Lipschitz condition (4.38) and the notation (4.42) we obtain

$$\begin{aligned}
e(t) \le\ & |\, y_k - S_{k-1}(t_k)\,| + |\, y'_k - S'_{k-1}(t_k)\,|\,|\, t - t_k| + |\, y''_k - S''_{k-1}(t_k)\,| \frac{(t-t_k)^2}{2!} \\
& + \sum_{j=0}^{r-1} |\, y_k^{(j+3)} - f_k^{(j)}\,| \frac{|\, t - t_k|^{j+3}}{(j+3)!} + |\, y^{(r+3)}(\xi_k) - f_k^{(r)}\,| \frac{|t - t_k|^{r+3}}{(r+3)!} \\
& \le e_k + h\, e'_k + \frac{h^2}{2!} e''_k + \sum_{j=0}^{r-1} |\, y_k^{(j+3)} - f_k^{(j)}| \frac{h^{j+3}}{(j+3)!} + |\, y^{(r+3)}(\xi_k) - f_k^{(r)}\,| \frac{h^{r+3}}{(r+3)!}.
\end{aligned} \tag{4.43}$$

If we denote

$$F = |\, y_k^{(j+3)} - f_k^{(j)}\,| \text{ and } \hat{F} = |y^{(r+3)}(\xi_k) - f_k^{(r)}| \tag{4.44}$$

then,

$$F \le L\,(e_k + e'_k) \tag{4.45}$$

and

$$\widehat{F} \le w\,(y^{(r+3)}, h\,) + L\,(e_k + e'_k) \tag{4.46}$$

where $w\left(y^{(r+3)}, h\right)$ is the modulus of continuity of the function $y^{(r+3)}$. Thus using (4.45) and (4.46) we may obtain $\sum_{j=0}^{r-1} \frac{h^{j+3}}{(j+3)!} < e^h - 2 < e$. With this result we can easily get

$$e(t) \le (1 + c_0\ h)\, e_k + (1 + c_0)\, h\, e'_k + \frac{h^2}{2!} e''_k + \frac{h^{r+3}}{(r+3)!} w\,(\,y^{(r+3)}, h\,) \tag{4.47}$$

where $c_0 = L\left(e + \frac{1}{(r+3)!}\right)$ is a constant independent of h. In a similar manner, we can easily get

$$e'(t) \le c_1\ h\, e_k + (1 + c_1\ h)\, e'_k + h\, e''_k + \frac{h^{r+2}}{(r+2)!} w\,(y^{(r+3)}, h) \tag{4.48}$$

where $c_1 = L\left(e + \frac{1}{(r+2)!}\right)$ which is independent of h.

Now the estimate of $|\ y''(t) - S''_k(t)|$, we use the Lipschitz condition (4.38) and notations (4.42) and utilizing the inequality

$$\sum_{j=0}^{r-1} \frac{h^j}{(j+1)!} < e\,,$$

we can get

$$e''(t) \le c_2\ h\, e_k + c_2\ h\, e'_k + e''_k + \frac{h^{r+1}}{(r+1)!} w\,(y^{(r+3)}, h) \tag{4.49}$$

where $c_2 = L\left(e + \frac{1}{(r+1)!}\right)$ is constant and independent of h.

If we let $|E(x)| = \max\{e(x), e'(x), e''(x)\}$. It can be determined that

$$E(x) \le O(h^{\alpha+r}).$$

To find the estimate for $\left|y^{(q)}(t) - S_k^{(q)}\right|$, $q = 3, 4, \ldots,\ r+2$, using (4.38)-(4.49) we get

$$\left| y^{(q)}(t) - S_k^{(q)}(t) \right| \le \sum \left| \ y_k^{(j+3)} - f_k^{(j)} \right| \frac{\left| t - t_k \ \right|^{j+3-q}}{(j+3-q)!} + \left| y^{(r+3)}(\xi_k) - f_k^{(r)} \right| \frac{\left| t - t_k \right|^{r+3-q}}{(r+3-q)!}$$

$$\le O\left(h^{\alpha+r+3-q}\right) \tag{4.50}$$

thus we proved the following theorem.

Theorem 4.17

Let $S_\Delta(t)$ be the approximate solutions of the problem (4.37), given by equations (4.40)-(4.41) and $f \in C^r\left([0,1] \times R^2\right)$, then for all $t \in [t_k, t_{k+1}] \subset [0,1]$, $\quad h = 0, 1, \ldots, n-1$ we have $\left| y^{(i)}(t) - S_k^{(i)}(t) \right| \le K\, h^{r+3-j}\, w(h)$ where $j = 3, 4, \ldots,\ r+3$, and K is a constant independent of h.

With the help of spline approximation, the problem has been converted into the form given in (4.5)

$$\begin{aligned} &u''(t) = f(t),\ t \in [0,1] \\ &u(0) = u(1) = 0. \end{aligned}$$

This means wavelet representation for the solution of BVP procedure can be easily applied. Using the above procedure, delay neutrol and functional differential equations may also be considered.

Bibliography

1. J. H. Ahlberg, E. N. Nilson and J. L. Walsh, *the theory of Splines and their Applications*, Academic Press, New York, 1967.

2. E. Bacry, S. Mallat, and G. Papanicolaou, A wavelet based space-time adaptive numerical method for partial differential equations, *Math. Modelling Number. Anal.* **26** (1992), 793-834.

3. G. Battle. A block spin construction of ondelettes. Part: Lemarie function, *Comm. Math. Phys.* **110** (1987), 601-615.

4. S. Bertoluzza. A posteriori error estimates for wavelet Galerkin methods, *Appl. Math. Lett.* **8** (1995).

5. G. Beylkin, On wavelet-based algorithms for solving differential equations, Preprint, (1992).

6. G. Beylkin, R. Coifman, and V. Rokhlin, Fast wavelet transform and numerical algorithms I, *Comm. Pure and Appl. Math.* **44** (1991), 141-183.

7. C. de Boor, *A Practical Guide to Splines*, Spinger Verlag, New York 1978.

8. C. de Boor, R. DeVore, and A. Ron, On the construction of multivariate prewavelet, *Constr. Approx.* **9 (2)** (1993), 123-126.

9. J.M. Camicer, W. Dahmen, and J.M. Pena, Local decomposition of refinable spaces and wavelets, *Appl. Comput. Harmonic Anal.* **3** (1996), 127-153.

10. C. K. Chui, *An Introduction to Wavelets*, Academic Press, Boston, (1992).

11. C. K. Chui and C. Li, A general framework of multivariate wavelets with duals, Appl. *Comput. Harmonic Anal.* **1** (1994), 368-390.

12. C. K. Chui and E. Quak, Wavelets on a bounded interval, in: "Numerical Methods of Approximation Theory, Vol. 9" (D. Braess and L.L Schumaker, Eds.), Birkhauser, Basel, (1992).

13. C. K. Chui, J. Stockler, and J.D. Ward, Compactly supported box spline wavelets, *Approx. Theory Appl.* **8** (1992), 77-100.

14. C.K.Chui and J.Z. Wang, On compactly supported spline-wavelets and a duality principle, *Trans. Amer. Math. Soc.* **330** (1992), 903-915.

15. C. K.Chui and J.Z. Wang, A general framework for compactly supported spline and wavelets, J. Approx. Theory **71** (1992), 263-304.

16. C. K.Chui and J.Z. Wang, A cardinal spline approach to wavelet, *American mathematical society*, **113** (1991) 785-793

17. A. Cohen, I. Daubechies, and J. Feauveau, Biorthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **45** (1992), 485-560.

18. A. Cohen and I. Daubechies, Non-separable bidimensional wavelet bases, *Rev. Mat. Iberoamericana* **9** (1993), 51-137.

19. A. Cohen, I. Daubechies, and P. Vial, Wavelets and fast wavelet transforms on the interval, Appl. Comput. Harmonic Anal. **1** (1993), 54-81.

20. S. Dahlke, Wavelets: Construction Principles and Applications to the Numerical treatment of operator Equations, Shaker Verlag, Achen, 1996.

21. S. Dahlke, W. Dahmen, and V. Latour, Smooth refinable functions and wavelets obtained by convolution products, *Appl. Comput. Harmonic Anal.* **2**(1995), 68-84.

22. S. Dahlke and A. Kunoth, Biorthogonal wavelets and multigrid, in: Proceedings of the 9th GAMM-Seminar "Adaptive Methods: Algorithms, Theory and Applications", (W. Hackbusch and G.Wittum, Eds.), NNFM Series Vol. 46,Vieweg, (1994), pp. 99-119.

23. S. Dahlke, V. Latour, and M. Neeb, Generalized cardinal B-splines: stability, linear independenc, and appropriate scaling matrices, *Constr. Approx.* **13** (1997), 29-56.

24. S.Dahlike and I. Weinreich, Wavelet-Galerkin methods: An adapted biorthogonal wavelet basis, *Constr. Approx.* **9** (1993), 237-262.

25. W. Dahmen, S. Prossdorf, and R. Schneider, Wavelet approximation methods for pseudodifferential equations II: Matrix compression and fast solutions, *Adv. Comp. Math.* **1** (1993), 259-335.

26. I. Daubechies, Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **41** (1987), 909-996.

27. I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Math. **61**, SIAM, Philadelphia, (1992).

28. G. David, ***Wavelets and Singular Integrals on Curves and Surfaces***, Lecture Notes in Mathematics 1465, Springer Verlag, Berlin Heidelberg (1991).

29. R. DeVore, B. Jawerth, and V. Popov, Compression of wavelet decomposition, *Amer. J. Math.* **114** (1992), 737-785.

30. B. Engquist, S. Osher, and S. Zhong, Fast wavelet based algorithms for linear evolution equations, *SIAM J. Sci. Comput.* **15** (1994), 755-775.

31. M. W. Frazier, An Introduction to Wavelet Through Linear Algebra, Springer Verlag, New York, 1999.

32. K. Grochenig and W.R. Madych, Multiresolution analysis, Haar basses and self-similar tilings of IR^n, *IEEE Trans. Inform.* Th. **38 (2)** (1992), 556-568.

33. S. Jaffard, Wavelet methods for fast resolution of elliptic problems, *SIAM J. Numer. Anal.* **29** (1992), 965-986.

34. F. R. Loscalzo and T. D. Talbot, Spline function approximations for solution of ordinary differential equations, *SIAM J. Num.* Anal. **4** (1967), 433-445.

35. S. Mallat, Multiresolution approximation and wavelet orthonormal bases of $L^2(R)$, *Trans. Amer. Math. Soc.* **315** (1989), 69-88.

36. G. Micula and H. Akca, Approximate solutions of the second order differential equations with deviating argument by Spline functions, *Mathematica Revue D Analyse Numerique*, **1** (1988), 37-46.

37. G. Micula and S. Micula, Handbook of Splines, kluwer Academic Publishers, Dordrecht, 1999.

38. G. Micula and H. Akca, Approximate solutions of the second order differential equations with deviating argument by Spline functions, *Studia Mathematica*, **2** (1988), 45-57

39. I. J. Schoenberg, Contribution to the problem of approximation of equidisdant data by analitic functions, *Quart. Appl. Math.*, **4** (1946), 55-88 and 112-441.

40. L. Schumaher, Spline Functions Basic Theory, Willy, New York 1981.

41. L. Villemoes, Wavelet analysis of refinement equation, SIAM J. M*ath, Anal.* **25** (1994), 1433-1460.