Survey on wavelet transform and application in ODE and wavelet networks

3 authors:

Haydar Akca
Abu Dhabi University

Mohammed H. Al-Lail
King Fahd University of Petroleum and Minerals

Valéry Covachev
Bulgarian Academy of Sciences

Advances in Dynamical Systems and Applications.
ISSN 0973-5321 Volume 1 Number 2 (2006), pp. 129–162

http://www.ripublication.com/adsa.htm

Survey on Wavelet Transform and Application in ODE and Wavelet Networks

Haydar Akca
United Arab Emirates University, Faculty of Sciences,
Mathematical Sciences Department, P.O. Box 17551, Al Ain, UAE
E-mail: hakca@uaeu.ac.ae

Mohammed H. Al-Lail
King Fahd University of Petroleum and Minerals,
Department of Mathematical Science,
Dhahran 31261, Saudi Arabia

Valéry Covachev
Department of Mathematics & Statistics,
College of Science, Sultan Qaboos University,
Sultanate of Oman
E-mail: vcovachev@hotmail.com

Abstract

We study historical development of wavelets and introduce basic definitions and formulations. Wavelets are mathematical tools that cut up data or functions or operators into different frequency components, and then study each component with a resolution matching to its scale. We discuss different approaches of using wavelets in the solution of boundary value problems for ordinary differential equations. We also introduce convenient wavelet representations for the derivatives for certain functions. A wavelet network is a network combining the idea of the feed-forward neural networks and the wavelet decomposition. Recent developments and wavelet network algorithm are discussed.

AMS subject classification: 42C40, 34B05, 65T60.
Keywords: Wavelet, networks, multiresolution analysis, orthonormal basis, Fourier analysis, boundary value problems.

Received November 11, 2006; Accepted December 13, 2006

1. Introduction

Wavelet theory involves representing general functions in terms of simpler building blocks at different scales and positions. The fundamental idea behind wavelets is to analyze according to scale. Wavelets are mathematical tools that cut up data or functions or operators into different frequency components, and then study each component with a resolution matching to its scale. In the history of mathematics, wavelet analysis shows many different origins. Much of the work was performed in the 1930s [18]. Before 1930, the main branch of mathematics leading to wavelets began with Joseph Fourier with his theory of frequency analysis. In 1909, Haar discovered the simplest solution and at the same time opened a route leading to wavelets.

The rest of this paper is organized as follows. In Section 1, basic definitions of multiresolution analysis and construction of wavelets is introduced. The relations between wavelets and differential equations (ODE) and Wavelet–Galerkin methods for differential equations is introduced in Section 2. Differential and integral equations are discussed in Section 3. Section 4 is devoted to difference equations. In Section 5, for certain functions derivative applications are introduced. The theory of wavelet networks and the idea of combining wavelets and neural networks are discussed in Section 6.

1.1. Multiresolution Analysis and Construction of Wavelets

The objective of this section is to construct a wavelet system, which is a complete orthonormal set in $L^2(\mathbb{R})$. The idea of multiresolution analysis is to represent a function (or signal) f as a limit of successive approximations, each of which is a finer version of the function f. The basic principle of the multiresolution analysis (MRA) deals with the decomposition of the whole function space into individual subspaces $V_n \subset V_{n+1}$.

Definition 1.1. (Multiresolution analysis). A multiresolution analysis (MRA) of $L^2(\mathbb{R})$ is defined as a sequence of closed subspaces V_j of $L^2(\mathbb{R})$, $j \in \mathbb{Z}$, that satisfy the following properties:

1. Monotonicity $\quad V_j \subset V_{j+1}, \quad$ for all $j \in \mathbb{Z}$.
2. Dilation property $\quad f(x) \in V_j \Longleftrightarrow f(2x) \in V_{j+1} \quad$ for all $j \in \mathbb{Z}$.
3. Intersection property $\quad \bigcap_{j\in\mathbb{Z}} V_j = \{0\}$.
4. Density property $\quad \bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$.
5. Existence of a scaling function. There exists a function $\phi \in V_0$ such that $\{\phi(x - n) : n \in \mathbb{Z}\}$ is an orthonormal basis for V_0,

$$V_0 = \left\{ \sum_{k\in\mathbb{Z}} \alpha_k \phi(x-k) : \{\alpha_k\}_{k\in\mathbb{Z}} \in l^2(\mathbb{Z}) \right\}.$$

Density property means that for any $f \in L^2(\mathbb{R})$, there exists a sequence $\{f_n\}_{n=1}^{\infty}$ such that each $f_n \in \bigcup_{j\in\mathbb{Z}} V_j$ and $\{f_n\}_{n=1}^{\infty}$ converges to f in $L^2(\mathbb{R})$, that is, $\|f_n - f\| \to 0$ as $n \to \infty$.

The function ϕ is called the *scaling function* or *father wavelet* of the given MRA. Sometimes condition 5 is relaxed by assuming that $\{\phi(x-n) : n \in \mathbb{Z}\}$ is a Riesz basis for V_0. In this case, we have a multiresolution analysis with a Riesz basis. Dilation condition 2 implies that $f\ (x) \in V_j \Longleftrightarrow f(2^m x) \in V_{j+m}$ for all $j, m \in \mathbb{Z}$. In particular, $f(x) \in V_0 \Longleftrightarrow f(2^j x) \in V_j$. Let

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

The orthonormality of the set $\{\phi(x-n) : n \in \mathbb{Z}\}$ implies that for each $j \in \mathbb{Z}$, $\{\phi_{j,k}(x), k \in \mathbb{Z}\}$ is an orthonormal set, because changing variables shows that for $j, k, m \in \mathbb{Z}$,

$$\langle \phi_{j,k}, \phi_{j,m} \rangle = \langle \phi_{0,k}, \phi_{0,m} \rangle.$$

Then $\{\phi_{j,k}(x), k \in \mathbb{Z}\}$ is an orthonormal basis for V_j. It follows that for each $j \in \mathbb{Z}$,

$$V_j = \left\{ \sum_{k\in\mathbb{Z}} \alpha_k \phi_{j,k}(x) : \{\alpha_k\}_{k\in\mathbb{Z}} \in l^2(\mathbb{Z}) \right\}.$$

Define the orthogonal projection operator P_j from $L^2(\mathbb{R})$ onto V_j by

$$P_j(f)(x) = \sum_{k\in\mathbb{Z}} \langle f, \phi_{j,k} \rangle \phi_{j,k}(x).$$

Then we have

$$\lim_{j\to\infty} P_j(f) = f \ \text{ and } \lim_{j\to-\infty} P_j(f) = 0.$$

The projection $P_j(f)$ can be considered as an approximation of f at the scale 2^{-j}. Therefore, the successive approximations of a given function f are defined as the orthogonal projections $P_j(f)$ onto the space V_j. We can choose $j \in \mathbb{Z}$ such that $P_j(f)$ is a good approximation of f [8, 12, 26].

The real importance of a multiresolution analysis lies in the simple fact that it enables us to construct an orthonormal basis for $L^2(\mathbb{R})$ [8, 12, 14, 17].

In order to prove this statement, we first assume that $\{V_m\}$ is a multiresolution analysis. Since $V_0 \subset V_1$, we define W_0 as the orthogonal complement of V_0 in V_1; that is, $V_1 = V_0 \bigoplus W_0$. Since $V_m \subset V_{m+1}$, we define W_m as the orthogonal complement of V_m in V_{m+1} for every $m \in \mathbb{Z}$ so that we have

$$V_{m+1} = V_m \bigoplus W_m \quad \text{for each } m \in \mathbb{Z}.$$

Since V$_m \to \{0\}$ as $m \to -\infty$, we see that

$$V_{m+1} = V_m \bigoplus W_m = \bigoplus_{l=-\infty}^{m} W_l \quad \text{for all } m \in \mathbb{Z}.$$

Since $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$, we may take the limit as $m \to \infty$ to obtain

$$L^2(\mathbb{R}) = \bigoplus_{l=-\infty}^{\infty} W_l.$$

To find an orthonormal wavelet, therefore, all we need to do is to find a function $\psi \in W_0$ such that $\{\psi(x-k) : k \in \mathbb{Z}\}$ is an orthonormal basis for W_0. In fact, if this is the case, then $\left\{\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) : k \in \mathbb{Z}\right\}$ is an orthonormal basis for W_j for all $j \in \mathbb{Z}$ due to the condition in the definition of multiresolution analysis and definition of W_j. Hence

$$\left\{\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) : k,\ j \in \mathbb{Z}\right\}$$

is an orthonormal basis for $L^2(\mathbb{R})$, which shows that ψ is an orthonormal wavelet on $\mathbb{R}$.

Daubechies has constructed, for an arbitrary integer N, an orthonormal basis for $L^2(\mathbb{R})$ of the form

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k),\ j,k \in \mathbb{Z},$$

that satisfies the following properties:

1. The support of ψ is contained in $[-N+1, N]$. To emphasize this point, ψ is often denoted by ψ_N.

2. ψ_N has γN continuous derivatives, where $\gamma = \left(1 - \frac{1}{2}\log_2 3\right) = 0.20752$, for large N [17]. Hence, a C^N compactly supported wavelet has a support whose measure is, roughly, $5N$.

3. ψ_N has N vanishing moments

$$\int_{-\infty}^{\infty} x^k \psi(x)\, dx = 0 \quad \text{for} \quad k = 0, 1, \ldots, N.$$

Or, equivalently,

$$\left[\frac{d^k\hat{\psi}(\xi)}{d\xi^k}\right]_{\xi=0} = 0 \quad \text{for} \quad k = 0, 1, \ldots, N.$$

The multiresolution analysis (MRA) is well adapted to image analysis. The spaces V_j that appeared in the definition of an MRA can be interpreted as spaces where an approximation to the image at the j^{th} level is obtained. In addition, the detail in the approximation occurring in V_j, that is not in V_{j-1}, is stored in the spaces W_{j-1} which satisfy $V_j = V_{j-1} \bigoplus W_{j-1}$. This leads to efficient decomposition and reconstruction algorithms [5,8,14,26]. Choose an MRA with scaling ϕ and wavelet ψ.

The orthogonality property puts a strong limitation on the construction of wavelets. It is known that the Haar wavelet is the only real valued wavelet that is compactly supported, symmetric and orthogonal [11].

Definition 1.2. (Biorthogonal wavelets). Two functions $\psi, \tilde{\psi} \in L^2(\mathbb{R})$ are called *biorthogonal wavelets* if each one of the set $\{\psi_{j,k} : j,k \in \mathbb{Z}\}$ and $\left\{\tilde{\psi}_{j,k} : j,k \in \mathbb{Z}\right\}$ is a Riesz basis of $L^2(\mathbb{R})$ and they are biorthogonal,

$$\left\langle \psi_{j,k}, \tilde{\psi}_{l,m} \right\rangle = \delta_{j,l}\delta_{k,m} \quad \text{for all } j,l,k,m \in \mathbb{Z}.$$

Now let us replace condition 5 in Definition 1.1 by the following

$5'$. *Existence of a scaling function.* There exists a function $\phi \in V_0$, such that the set of functions$\left\{\phi_{j,l}(x) = 2^{j/2}\phi(2^j x - l) : l \in \mathbb{Z}\right\}$ is a Riesz basis of V_j.

As a result, there is a sequence $\{h_k : k \in \mathbb{Z}\}$ such that the scaling function satisfies a refinement equation

$$\phi(x) = 2\sum_{n\in\mathbb{Z}} h_n \phi(2x - n).$$

Define W_j as a complementary space of V_j in V_{j+1}, such that $V_{j+1} = V_j \bigoplus W_j$, and consequently,

$$L^2(\mathbb{R}) = \bigoplus_{l=-\infty}^{\infty} W_l.$$

A function ψ is a wavelet if the set of function $\{\psi(x-l) : l \in \mathbb{Z}\}$ is a Riesz basis of W_0. Then the set of wavelet functions $\left\{\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) : j,k \in \mathbb{Z}\right\}$ is a Riesz basis of $L^2(\mathbb{R})$. Since the wavelet is an element of V_1, it satisfies the relation

$$\psi(x) = 2\sum_{n\in\mathbb{Z}} g_n \phi(2x - n).$$

There are dual functions $\tilde{\phi}_{j,l}$ and $\tilde{\psi}_{j,l}$ so that the projection operators P_j and Q_j onto V_j and W_j, respectively, are given by

$$P_j(f)(x) = \sum_{k\in\mathbb{Z}} \langle f, \tilde{\phi}_{j,k}\rangle \phi_{j,k}(x),$$

and

$$Q_j(f)(x) = \sum_{k\in\mathbb{Z}} \langle f, \tilde{\psi}_{j,k}\rangle \psi_{j,k}(x).$$

Then we have

$$f = \sum_{j,k\in\mathbb{Z}} \langle f, \tilde{\psi}_{j,k}\rangle \psi_{j,k}.$$

Here the definitions of $\tilde{\phi}_{j,k}$ and $\tilde{\psi}_{j,k}$ are similar to those for $\phi_{j,k}$ and $\psi_{j,k}$. Then, the basis functions and dual functions are biorthogonal [19],

$$\langle \phi_{j,l}, \tilde{\phi}_{j,k}\rangle = \delta_{l,k} \quad \text{and} \quad \langle \psi_{j,l}, \tilde{\psi}_{m,k}\rangle = \delta_{j,m}\delta_{l,k}.$$

Note that if the basis functions are orthogonal, they coincide with the dual function and the projections are orthogonal.

2. Wavelets and Differential Equations

Many applications of mathematics require the numerical approximation of solutions of differential equations. In this section we will present different approaches of using wavelets in the solution of boundary value problems for ordinary differential equations. We consider the class of ordinary differential equation of the form

$$Lu(x) = f(x) \quad \text{for } x \in [0,1],$$

where $L = \sum_{j=0}^{m} a_j(x) D^j$, and with appropriate boundary conditions on $u(x)$ for $x = 0, 1$.

There are two major solution techniques. First, if the coefficients $a_j(x)$ of the operator are constants, then the Fourier transform is well suited for solving these equations because the complex exponentials are eigenfunctions of a constant coefficient operator and they form an orthogonal system. As a result, the operator becomes diagonal in the Fourier basis and can be inverted trivially. If the coefficients are not constant, finite element or finite difference methods can be used [15].

2.1. Wavelet–Galerkin Methods for Differential Equations

The classical Galerkin methods have the disadvantage that the stiffness matrix becomes ill conditioned as the problem size grows. To overcome this disadvantage, we use wavelets as basis functions in a Galerkin method. Then, the result is a linear system that is sparse because of the compact support of the wavelets, and that, after preconditioning, has a condition number independent of problem size because of the multiresolution structure [1, 6, 13, 15, 19, 29].

The methods for numerically solving a linear ordinary differential equation come down to solving a linear system of equations, or equivalently, a matrix equation $Ax = y$. The system has a unique solution x for every y if and only if A is an invertible matrix. However, in applications there are further issues that are of crucial importance. One of these has to do with the condition number of a matrix A which measures the stability of the linear system $Ax = y$. Let us see an example [15].

Definition 2.1. Let A be an $n \times n$ matrix. Define $\|A\|$ called the operator norm, or just the norm, of A by

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

where the supremum is taken over all nonzero vector in $\mathbb{C}^n$.

Equivalently,

$$\|A\| = \sup \left\{ \|Az\| : \|z\| = 1, z \in \mathbb{C}^n \right\}.$$

Definition 2.2. (Condition number of a matrix). Let A be an $n \times n$ matrix. Define $C_{\#}(A)$, the condition number of the matrix A, by

$$C_{\#}(A) = \|A\| \left\|A^{-1}\right\|.$$

If A is not invertible, set $C_{\#}(A) = \infty$.

Note that the condition number $C_{\#}(A)$ is scale invariant [7], that is, for $c \neq 0$,

$$C_{\#}(cA) = C_{\#}(A).$$

Lemma 2.3. Suppose that A is an $n \times n$ normal invertible matrix. Let

$$|\lambda|_{\max} = \max \{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$$

and

$$|\lambda|_{\min} = \min \{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

Then

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

The condition number of A measures how unstable the linear system $Ax = y$ is under perturbation of the data y. In applications, a small condition number (i.e., near 1) is desirable [15]. If the condition number of A is high, we would like to replace the linear system $Ax = y$ by an equivalent system $Mz = v$ whose matrix M has a low condition number.

We consider the class of ordinary differential equations (known as Sturm–Liouville equations) of the form

$$\begin{aligned} Lu(t) &= -a(t)u''(t) - a'(t)u'(t) + b(t)u(t) \\ &= -\frac{d}{dt}\left(a(t)\frac{du}{dt}\right) + b(t)u(t) = f(t) \end{aligned}$$

for $0 \leq t \leq 1$ with Dirichlet boundary conditions

$$u(0) = u(1) = 0.$$

Here a, b, and f are given real-valued functions and we wish to solve for u. We assume f and b are continuous and a has a continuous derivative on $[0, 1]$ (this always means a one-sided derivative at the endpoints). Note that L may be a variable coefficient differential operator because $a(t)$ and $b(t)$ are not necessarily constant. We assume that the operator is uniformly elliptic which means that there exist finite constants C_1, C_2, and C_3 such that

$$0 < C_1 \leq a(t) \leq C_2 \text{ and } 0 \leq b(t) \leq C_3 \tag{2.1}$$

for all $t \in [0, 1]$. By a result in the theory of ordinary differential equations, there is a unique function u satisfying the differential equation and the boundary conditions $u(0) = u(1) = 0$.

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 is C^2 on $[0, 1]$ and it satisfies

$$v_j(0) = v_j(1) = 0.$$

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

$$S = \operatorname{span}\{v_j;\ j \in \Lambda\}.$$

We look for an approximate solution u of the form

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

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,

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

By linearity, it follows that

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

Note that the approximate solution u_s obviously satisfies the boundary conditions $u_s(0) = u_s(1) = 0$.

Using these results we get

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

or

$$\sum_{k\in\Lambda} \langle Lv_k, v_j\rangle\, x_k = \langle f, v_j\rangle \quad \text{for all } j \in \Lambda.$$

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

$$a_{j,k} = \langle Lv_k, v_j\rangle.$$

Thus, we get a linear system of equations

$$\sum_{k\in\Lambda} a_{j,k} x_k = y_j \quad \text{for all } j \in \Lambda, \quad \text{or} \quad Ax = y.$$

In the Galerkin method, for each subset Λ we obtain an approximation $u_s \in S$, by solving the linear system for x and using these components to determine u_s. We expect that as we increase our set Λ in some systematic way, our approximation u_s will converge to the exact solution u.

The nature of the linear system results from choosing a wavelet basis for the Galerkin method. There are two properties that the matrix A in the linear system should have. First, A should have a small condition number to obtain stability of the solution under small perturbations in the data. Second, A should be sparse for quick calculations [15, 16].

There is a way of modifying the wavelet system for $L^2(\mathbb{R})$ so as to obtain a complete orthonormal system $\{\psi_{j,k}\}_{(j,k)\in\Gamma}$ for $L^2[0,1]$. See for more details [2, 13, 16] and references given therein. The set Γ is a certain subset of $\mathbb{Z}\times\mathbb{Z}$. For each $(j,k)\in\Lambda$, $\psi_{j,k} \in C^2$ 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

$$g=\sum_{j,k} c_{j,k}\psi_{j,k},$$

where the sum is finite, we have

$$C_4\sum_{j,k} 2^{2j}\left|c_{j,k}\right|^2 \le \int_0^1 \left|g'(t)\right|^2 dt \le C_5\sum_{j,k} 2^{2j}\left|c_{j,k}\right|^2. \tag{2.2}$$

An estimate of this form is called a norm equivalence. It states that up to the two constants, the quantities $\sum\limits_{j,k} 2^{2j}\left|c_{j,k}\right|^2$ and $\int\limits_0^1 \left|g'(t)\right|^2 dt$ are equivalent.

For wavelets we write an equation as

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

and

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

for some finite set of indices Λ. We can write the system 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 $(j,k)\in\Lambda$, and the matrix

$$A=\left[a_{l,m;j,k}\right]_{(l,m),(j,k)\in\Lambda}$$

defined by

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

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. A itself does not have a low condition number, however, we can replace the system $Ax=y$ by an equivalent system $Mz=v$, for which the new matrix M has low condition number. To get this, first define the diagonal matrix

$$D=\left[d_{l,m;j,k}\right]_{(l,m),(j,k)\in\Lambda}$$

by

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

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

$$M = D^{-1}AD^{-1}.$$

By writing this out, we get

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

Then, the system $Ax = y$ is equivalent to $D^{-1}AD^{-1}Dx = D^{-1}y$. If we let $z = Dx$ and $v = D^{-1}y$, we get $Mz = v$.

Lemma 2.4. Let L be a uniformly elliptic Sturm–Liouville operator. Suppose g is C^2 on $[0,1]$ and satisfies $g(0) = g(1) = 0$. Then

$$C_1 \int_0^1 |g'(t)|^2\,dt \le \langle Lg, g\rangle \le (C_2 + C_3)\int_0^1 |g'(t)|^2\,dt,$$

where $C_1,\ C_2$, and C_3 are the constants in relation (2.1).

Theorem 2.5. ([15, 17]) Let L be a uniformly elliptic Sturm–Liouville 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, satisfies $\psi_{j,k}(0) = \psi_{j,k}(1) = 0$, and such that the norm equivalence holds. Let Λ be a finite subset of Γ. Let M be the matrix defined in the above equation. Then the condition number of M satisfies

$$C_{\#}(M) \le \frac{(C_2 + C_3)\,C_5}{C_1C_4}$$

for any finite set Λ, where C_1, C_2, and C_3 are the constants in relation (2.1), and C_4, C_5 are the constants in relation (2.2).

Note that the matrices obtained by using finite differences are sparse, but they have large condition numbers [15]. Using the Galerkin method with the Fourier system, we can obtain a bounded condition number but the matrix is not sparse. Using the Galerkin method with a wavelet system, we obtain both advantages [4, 6, 13, 15].

The derivative operator is not diagonal in a wavelet basis [3, 13, 27]. However, we can make the differential operator diagonal by using two pairs of biorthogonal or dual bases of compactly supported wavelets [13]. In this case, we have two related multiresolution spaces $\{V_j\}$ and $\{\tilde{V}_j\}$ such that

$$V_{j+1} \subset V_j, \quad \text{and} \quad \tilde{V}_{j+1} \subset \tilde{V}_j, \quad \text{for all } j \in \mathbb{Z},$$

corresponding to two scaling functions $\phi, \tilde{\phi}$ and two wavelets $\psi, \tilde{\psi}$. They are defined by two trigonometric polynomials m_0 and $\tilde{m}_0$, satisfying

$$m_0(w)\overline{\tilde{m}_0(w)} + m_0(w+\pi)\overline{\tilde{m}_0(w+\pi)} = 1.$$

Then we have

$$\widehat{\phi}(w) = \frac{1}{\sqrt{2\pi}} \prod_{j=1}^{\infty} m_0(2^{-j}w),$$

$$\widehat{\tilde{\phi}}(w) = \frac{1}{\sqrt{2\pi}} \prod_{j=1}^{\infty} \tilde{m}_0(2^{-j}w),$$

also, we have

$$\hat{\psi}(w) = e^{-\frac{iw}{2}} \overline{\tilde{m}_0\left(\frac{w}{2}+\pi\right)} \hat{\phi}\left(\frac{w}{2}\right)$$

and

$$\widehat{\tilde{\psi}}(w) = e^{-\frac{iw}{2}} \overline{m_0\left(\frac{w}{2}+\pi\right)} \widehat{\tilde{\phi}}\left(\frac{w}{2}\right)$$

with $\langle \psi_{j,k}, \tilde{\psi}_{m,n} \rangle = \delta_{j,m}\delta_{k,n}$, where

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x-k), \qquad \tilde{\psi}_{j,k}(x) = 2^{-j/2}\tilde{\psi}(2^{-j}x-k).$$

It follows that if we construct two pairs of biorthogonal wavelet bases, one using $\psi, \tilde{\psi}$, and the other using $\psi^*, \tilde{\psi}^*$, then we have

$$\psi' = 4\psi^*, \quad \text{or} \quad (\psi_{j,k})' = 2^{-j}4\psi^*_{j,k},$$

and hence

$$\left\langle \frac{d}{dx}\psi_{j,k}, \tilde{\psi}^*_{m,n} \right\rangle = 2^{-j}4\delta_{j,m}\delta_{k,n}.$$

This means that we have diagonalized the derivative operator. Note that this is not a "true" diagonalization because we use two different bases. However, this means that we can find the wavelet coefficients of f', i.e.,

$$\left\langle \frac{d}{dx} f, \psi_{j,k} \right\rangle = 2^{-j}4\left\langle f, \psi^*_{j,k} \right\rangle.$$

For more details see [13] and the references therein.

Another approach of diagonalizing the differential operator, using wavelets, is by constructing biorthogonal wavelets with respect to the inner product defined by the operator [18, 19].

We consider the class of ordinary differential equation of the form

$$Lu(x) = f(x) \quad \text{for } x \in [0,1],$$

where $L = \sum_{j=0}^{m} a_j(x) D^j$, and with appropriate boundary conditions on $u(x)$ for $x = 0, 1$. Define the operator inner product associated with an operator L by

$$\langle\langle u, v \rangle\rangle = \langle Lu, v \rangle .$$

An approximate solution u can be found with a Petrov–Galerkin method, i.e., consider two spaces S and S^* and look for a solution $u \in S$ such that

$$\langle\langle u, v \rangle\rangle = \langle f, v \rangle$$

for all v in S^*. If S and S^* are finite dimensional spaces with the same dimension, this leads to a linear system of equations. The matrix of this system, also referred to as the stiffness matrix, has as elements the operator inner products of the basis functions of S and S^*.

We assume that L is self-adjoint and positive definite and, in particular, we can write

$$L = V^* V,$$

where V^* is the adjoint of V. We call V the square root operator of L. Suppose that $\{\Psi_{j,l}\}$ and $\{\Psi^*_{j,l}\}$ are bases for S and S^* respectively. The entries of the stiffness matrix are then given by

$$\left\langle \left\langle \Psi_{j,l}, \Psi^*_{m,n} \right\rangle \right\rangle = \left\langle L\Psi_{j,l}, \Psi^*_{m,n} \right\rangle = \left\langle V\Psi_{j,l}, V\Psi^*_{m,n} \right\rangle .$$

Now, the idea is to let

$$\Psi_{j,l} = V^{-1}\psi_{j,l} \quad \text{and} \quad \Psi^*_{j,l} = V^{-1}\tilde{\psi}_{j,l},$$

where ψ and $\tilde{\psi}$ are the wavelets of a classical multiresolution analysis. We will call the Ψ and Ψ^* functions the operator wavelets. Then the operator wavelets are biorthogonal with respect to the operator inner product. We want the operator wavelets to be compactly supported and to be able to construct compactly supported operator scaling functions $\Phi_{j,l}$. The analysis is relatively straightforward for simple constant coefficient operators such as the Laplace and polyharmonic operator [19].

Example 2.6. (Laplace operator). Consider the one dimensional Laplace operator

$$L = -D^2.$$

Then the square root operator V is $V = D$. The associated operator inner product is

$$\langle\langle u, v\rangle\rangle = \langle Lu, v\rangle = \langle Vu, Vv\rangle = \langle u', v'\rangle .$$

Since the action of V^{-1} is taking the antiderivative, we define the operator wavelets as

$$\Psi(x) = \int_{-\infty}^{x} \psi(t)\, dt, \quad \text{and} \quad \Psi^*(x) = \int_{-\infty}^{x} \tilde{\psi}(t)\, dt.$$

Note that the operator wavelets $\Psi(x)$ and $\Psi^*(x)$ are compactly supported because the integral of the original wavelets has to vanish. Also translation and dilation invariance is preserved, so we define

$$\Psi_{j,l}(x) = \Psi(2^j x - l) \quad \text{and} \quad \Psi^*_{j,l}(x) = \Psi^*(2^j x - l).$$

Now,

$$\begin{aligned}\left\langle\left\langle \Psi^*_{j,l}(x), \Psi_{m,n}(x)\right\rangle\right\rangle &= \left\langle V\Psi^*_{j,l}(x), V\Psi_{m,n}(x)\right\rangle \\ &= 2^j \delta_{j,m}\delta_{l,n} \quad \text{for } j, l, m, n \in \mathbb{Z}.\end{aligned}$$

This means that the stiffness matrix is diagonal with powers of 2 on its diagonal. We now need to find an operator scaling function Φ. The antiderivative of the original scaling function is not compactly supported and hence not suited. To find an operator scaling function Φ convolute of the original scaling function with the indicator function $\chi_{[0,1]}$, $\Phi = \phi * \chi_{[0,1]}$, and define $\Phi_{j,l}(x) = \Phi(2^j x - l)$. Similarly for the dual functions $\Phi^* = \tilde{\phi} * \chi_{[0,1]}$. Now, define

$$V_j = \text{clos span } \{\Phi_{j,k} : k \in \mathbb{Z}\} \quad \text{and} \quad W_j = \text{clos span } \{\Psi_{j,k} : k \in \mathbb{Z}\} .$$

We want to show that $V_j \subset V_{j+1}$ and W_j complements V_j in V_{j+1}. By taking the Fourier transform we get

$$\hat{\Phi}(w) = \frac{1 - e^{-iw}}{iw}\hat{\phi}(w) \quad \text{and} \quad \hat{\Psi}(w) = \frac{1}{iw}\hat{\psi}(w).$$

A simple calculation shows that the operator scaling function satisfies the following equation

$$\hat{\Phi}(w) = H\left(\frac{w}{2}\right)\hat{\Phi}\left(\frac{w}{2}\right), \quad \text{with} \quad H(w) = \frac{1 + e^{-iw}}{2}h(w).$$

Consequently, $V_j \subset V_{j+1}$. Also

$$\hat{\Psi}(w) = G\left(\frac{w}{2}\right)\hat{\Phi}\left(\frac{w}{2}\right), \quad \text{with} \quad G(w) = \frac{1}{2(1 - e^{-iw})}g(w),$$

where $h(w)$ and $g(w)$ are defined by the previous equations respectively.

This implies that $W_j \subset V_{j+1}$. To prove that W_j complements V_j in V_{j+1}, we have to prove that

$$\Delta(w) = \det \begin{bmatrix} H(w) & H(w+\pi) \\ G(w) & G(w+\pi) \end{bmatrix}$$

does not vanish. In fact,

$$\begin{aligned} \Delta(w) &= H(w)G(w+\pi) - H(w+\pi)G(w) \\ &= \frac{1+e^{-iw}}{2} h(w) \frac{1}{2(1-e^{-i(w+\pi)})} g(w+\pi) \\ &\quad - \frac{1+e^{-i(w+\pi)}}{2} h(w+\pi) \frac{1}{2(1-e^{-iw})} g(w) \\ &= \frac{1}{4} h(w) g(w+\pi) - \frac{1}{4} h(w+\pi) g(w) = \frac{1}{4}\delta(w), \end{aligned}$$

where $\delta(w) = h(w)g(w+\pi) - h(w+\pi)g(w)$, and this cannot vanish since ϕ and ψ generate a multiresolution analysis. Then W_j complements V_j in V_{j+1}. The construction of the dual functions Φ^* and Ψ^* from $\tilde{\phi}$ and $\tilde{\psi}$ is completely similar. The coefficients of the trigonometric functions H, H^*, G and G^* now define a fast wavelet transform.

Now, we will describe the algorithm in the case of periodic boundary conditions. This implies that the basis functions on the interval $[0,1]$ are just the periodization of the basis functions on the real line.

Let $S = V_n$ and consider the basis $\{\Phi_{n,l} : 0 \le l \le 2^n\}$. Define vectors b and x such that

$$b_l = \left\langle f, \Phi^*_{n,l} \right\rangle, \quad \text{and} \quad u = \sum_{l=0}^{2^n-1} x_l \Phi_{n,l}.$$

The Galerkin method with this basis then yields a system

$$Ax = b \quad \text{with} \quad A_{k,l} = \langle \langle \Phi_{n,l}, \Phi_{n,k} \rangle \rangle .$$

The matrix A is not diagonal and the condition number grows as $O(2^{2n})$ [19]. Now, consider the decomposition

$$V_n = V_0 \oplus W_0 \oplus \cdots \oplus W_{n-1},$$

and the corresponding wavelet basis. The space V_0 has dimension one and contains constant functions. We now switch to a one-index notation such that the sets

$$\left\{1, \Psi_{j,l} : 0 \le j < n, 0 \le l < 2^j\right\} \text{ and } \left\{\Psi_k : 0 \le k < 2^n\right\}$$

coincide. Now, define the vectors $\tilde{b}$ and $\tilde{x}$ such that

$$\tilde{b} = \langle f, \Psi_l^* \rangle \quad \text{and} \quad u = \sum_{l=0}^{2^n-1} \tilde{x}_l \Psi_l.$$

There exist matrices T and T^* [19] such that $\tilde{b} = T^* b$ and $x = T\tilde{x}$.

The matrix T^* corresponds to the fast wavelet transform decomposition with filters H^* and G^*, and T corresponds to reconstruction with filters H and G. In the wavelet basis the system becomes

$$\tilde{A}\tilde{x} = \tilde{b} \quad \text{with} \quad \tilde{A} = T^* A T \quad \text{and} \quad \tilde{A}_{k,l} = \langle\langle \Psi_{n,l}, \Psi_{n,k} \rangle\rangle \,.$$

Since $\tilde{A}$ is diagonal, it can be trivially inverted and the solution is then given by

$$x = T\tilde{A}^{-1}T^* b.$$

Example 2.7. (Helmholtz operator). The one-dimensional Helmholtz operator is defined by

$$L = -D^2 + k^2.$$

Without loss of generality assume that $k = 1$ which can always be obtained by a transformation. The square root operator is

$$V = D + 1 = e^{-x} D e^{x} \quad \text{and} \quad V^{-1} = e^{-x} D^{-1} e^{x}.$$

Note that $V^{-1}\psi$ will not necessarily give a compactly supported function because $e^x \psi_{j,l}$ in general does not have a vanishing integral. Therefore we let

$$\Psi_{j,l} = V^{-1} e^{-x} \psi_{j,l} = e^{-x} D^{-1} \psi_{j,l}.$$

If $\psi_{j,l}$ has a vanishing integral, then $\Psi_{j,l}$ is compactly supported.

In order to diagonalize the stiffness matrix, the original wavelets now need to be orthogonal with respect to a weighted inner product with weight function e^{-2x}:

$$\begin{aligned} \left\langle \left\langle \Psi_{j,l}, \Psi^*_{m,n} \right\rangle \right\rangle &= \left\langle V\Psi_{j,l}, V\Psi^*_{m,n} \right\rangle \\ &= \left\langle e^{-x}\psi_{j,l}, e^{-x}\psi_{m,n} \right\rangle \\ &= \int_{-\infty}^{\infty} e^{-2x} \psi_{j,l}(x) \tilde{\psi}_{m,n}(x)\, dx. \end{aligned}$$

To find the wavelet let supp $\psi_{j,l} = [2^{-j}l, 2^{-j}(l+1)]$.

Then the orthogonality of the wavelets on each level immediately follows from their disjoint support. To get orthogonality between two different levels, we need that V_j is orthogonal to W_m for $m \geq j$ or

$$\int_{-\infty}^{\infty} e^{-2x} \phi_{j,l}(x) \tilde{\psi}_{m,n}(x)\, dx = 0 \quad \text{for } \ m \geq j.$$

Now, let the scaling function coincide with e^{2x} on the support of the finer scale wavelets,

$$\phi_{j,l}(x) = e^{2x} \chi_{[j,l]},$$

where $\chi_{[j,l]}$ is the indicator function on the interval $[2^{-j}l, 2^{-j}(l+1)]$, normalized so that the integral of the scaling functions is constant. As in the Haar case we choose the wavelets as

$$\psi_{j,l} = \phi_{j+1,2l} - \phi_{j+1,2l+1}$$

so that they have vanishing integral. The orthogonality between levels now follows from the fact that the scaling functions coincide with e^{2x} on the support of the finer scale wavelets, and from the vanishing integral of the wavelets

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-2x}\phi_{j,l}(x)\tilde{\psi}_{m,n}(x)\,dx &= \int_{-\infty}^{\infty} \chi_{[j,l]}\tilde{\psi}_{m,n}(x)\,dx\\ &= \int_{-\infty}^{\infty} \tilde{\psi}_{m,n}(x)\,dx = 0.\end{aligned}$$

One can see that the operator wavelets are now piecewise combinations of e^{x} and e^{-x}. The operator scaling functions are chosen as

$$\Phi_{j,l} = e^{-x}D^{-1}(\phi_{j,l} - \phi_{j,l+1}) \quad \text{so that} \quad \Psi_{j,l} = \Phi_{j+1,2l}.$$

The operator scaling functions on one level are translates of each other but the ones on different levels are no longer dilates of each other. They are supported on the same sets. The operator scaling functions satisfy a relation

$$\Phi_{j,l} = \sum_{k=0}^{2} H_k^j \Phi_{j+1,2l+k},$$

where

$$H_0^j = H_2^j = \frac{\sinh(2^{-j-1})}{\sinh(2^{-j})} \quad \text{and} \quad H_1^j = 1.$$

The Helmholtz operator in this bases of hyperbolic wavelets is diagonal. So we can conclude that a wavelet transform can diagonalize constant coefficient operators similar to the Fourier transform. The resulting algorithm is faster ($O(N)$ instead of $O(N\log N)$) [19].

Now, how should we use wavelets for variable coefficient operators? Consider the following operator

$$L = -Dp^2(x)D,$$

where p is sufficiently smooth and positive. The square root operator now is

$$V = pD \quad \text{and} \quad V^{-1} = D^{-1}\frac{1}{p}.$$

The analysis is similar to the case of the Helmholtz operator. Applying V^{-1} directly to a wavelet does not yield a compactly supported function. Therefore we take

$\Psi_{j,l} = V^{-1} p\psi_{j,l} = D^{-1}\psi_{j,l}$. Then,

$$\left\langle\left\langle \Psi_{j,l}, \Psi^*_{m,n}\right\rangle\right\rangle = \left\langle V\Psi_{j,l}, V\Psi^*_{m,n}\right\rangle = \langle p\psi_{j,l}, p\psi_{m,n}\rangle$$
$$= \int_{-\infty}^{\infty} p^2 \psi_{j,l}(x)\tilde{\psi}_{m,n}(x)\, dx,$$

which implies that the wavelets need to be biorthogonal with respect to a weighted inner product with p^2 as weight function. We use the same trick as for the Helmholtz operator. Let the scaling function $\phi_{j,l}$ coincide with $\frac{1}{p^2}$ on the interval $[2^{-j}l, 2^{-j}(l+1)]$,

$$\phi_{j,l} = \frac{1}{p^2}\chi_{[j,l]},$$

and normalize them so that they have a constant integral. We then take the wavelets

$$\psi_{j,l} = \phi_{j+1,2l} - \phi_{j+1,2l+1},$$

so they have vanishing integral and the operator wavelets are compactly supported. The operator wavelets $\Psi_{j,l}$ are now piecewise functions that locally look like

$$AP + B,$$

where P is the antiderivative of $\frac{1}{p^2}$. The operator wavelets are neither dilates nor translates of one function, since their behavior locally depends on p [19]. The coefficients in the fast wavelet transform are now different everywhere and they depend in a very simple way on the Haar wavelet transform of $\frac{1}{p^2}$. Then, the entries of the diagonal stiffness matrix can be calculated from the wavelets transform of $\frac{1}{p^2}$. We refer for more details to [19] and the references cited therein.

3. Differential and Integral Equations

Differential equations can be transformed into integral equations by using the continuous wavelet transform. An abstract proof of the following lemma can be found in [24] but here we present our proof.

Lemma 3.1. Let $\psi \in L^2(\mathbb{R})$, with

$$0 < C_\psi = \int_{-\infty}^{\infty} \frac{\left|\hat{\psi}(\xi)\right|^2}{|\xi|}\, d\xi < \infty.$$

Then for any $f \in L^2(\mathbb{R})$ we have

$$f^{(k)}(x) = \frac{1}{C_\psi} \int\limits_{-\infty}^{\infty} \frac{da}{a^2} \int\limits_{-\infty}^{\infty} \langle f, \psi_{a,b} \rangle \, a^{-k} \psi_{a,b}^{(k)}(x) \, db,$$

where

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right), \quad a, b \in \mathbb{R},\ a \neq 0.$$

Now, consider the following class of differential equations

$$\sum_{k=0}^{n} a_k(x) y^{(k)} = b(x),$$

where $\{a_k(x);\ k = 0, 1, \ldots, n\} \subset L^\infty(\mathbb{R})$, $\{y^{(k)};\ k = 0, 1, \ldots, n\} \subset L^2(\mathbb{R})$, and $b(x) \in L^2(\mathbb{R})$. Let $\{\psi^{(k)};\ k = 0, 1, \ldots, n\} \subset L^2(\mathbb{R})$ with $\mathrm{supp}(\psi) \subset [-L, L]$.

Using the result of Lemma 3.1, we have

$$y^{(k)}(x) = \frac{1}{C_\psi} \int\limits_{-\infty}^{\infty} \frac{da}{a^2} \int\limits_{-\infty}^{\infty} \langle y, \psi_{a,b} \rangle \, a^{-k} \psi_{a,b}^{(k)}(x) \, db$$

(for $k = 0, 1, \ldots, n$). Then the equation becomes

$$\frac{1}{C_\psi} \int\limits_{-\infty}^{\infty} \frac{da}{a^2} \int\limits_{-\infty}^{\infty} \langle f, \psi_{a,b} \rangle \sum_{k=0}^{n} a^{-k} a_k(x) \psi_{a,b}^{(k)}(x) \, db = b(x).$$

Then the differential equation is equivalent to an integral equation.

Example 3.2. Consider the differential equation

$$\sum_{k=0}^{n} a_k(x) y^{(k)} = b(x),$$

$\{b(x),\ a_k(x);\ k = 0, 1, \ldots, n\} \subset C[-\pi, \pi]$, $\{y^{(k)};\ k = 0, 1, \ldots, n\} \subset L^2(\mathbb{R})$. If $x \notin [-\pi, \pi]$, let $b(x) = a_k(x) = y^{(k)} = 0$ for $k = 0, 1, \ldots, n$. Then $\{b(x), a_k(x);\ k = 0, 1, \ldots, n\} \subset L^\infty(\mathbb{R})$ and $\{b(x),\ a_k(x),\ y^{(k)};\ k = 0, 1, \cdots, n\} \subset L^2(\mathbb{R})$. Define ψ by

$$\psi(x) = \begin{cases} \cos x, & x \in [-\pi, \pi], \\ 0, & x \notin [-\pi, \pi]. \end{cases}$$

Thus ψ is a wavelet because

$$\hat{\psi}(w) = \frac{1}{\sqrt{2\pi}} \left[\frac{\sin(w+1)\pi}{w+1} + \frac{\sin(w-1)\pi}{w-1}\right],$$

and

$$0 < C_\psi < \infty.$$

Then the continuous wavelet transform of y with respect to the wavelet ψ is

$$\begin{aligned}(T_\psi y)(a,b) &= \int_{-\infty}^{\infty} y(z)\overline{\psi_{a,b}(z)}\,dz \\ &= \frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty} y(z)\psi\left(\frac{z-b}{a}\right)\,dz \\ &= \frac{1}{\sqrt{|a|}}\int_{-|a|\pi+b}^{|a|\pi+b} y(z)\cos\left(\frac{z-b}{a}\right)\,dz.\end{aligned}$$

Now, by using Lemma 3.1 we have

$$\begin{aligned}b(x) &= \frac{1}{C_\psi}\int\limits_{-\infty}^{\infty}\frac{da}{a^2}\left(\int\limits_{-|a|\pi+x}^{|a|\pi+x}\left[\frac{1}{\sqrt{|a|}}\int_{-|a|\pi+b}^{|a|\pi+b} y(z)\cos\left(\frac{z-b}{a}\right)dz\right]\right) \\ &= \frac{1}{C_\psi}\int\limits_{-\infty}^{\infty}\frac{da}{a^2}\left(\sum_{k=0}^{n} a_k(x)\frac{1}{\sqrt{|a|}}a^{-k}\cos\left(\frac{x-b}{a}+k\frac{\pi}{2}\right)\,db\right).\end{aligned}$$

Then in order to solve the differential equation we only need to solve the integral equation [24].

4. Using Difference Equations

Suppose ϕ is a scaling function for a multiresolution analysis $\{V_j\}_{j\in\mathbb{Z}}$,

$$V_j = \left\{\sum_{k\in\mathbb{Z}} \alpha_k\phi_{j,k}(x) : \{\alpha_k\}_{k\in\mathbb{Z}} \in l^2(\mathbb{Z})\right\}$$

where

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

The orthogonal projection operator P_j from $L^2(\mathbb{R})$ onto V_j is defined by

$$P_j(f)(x) = \sum_{k\in\mathbb{Z}}\langle f,\phi_{j,k}\rangle\phi_{j,k}(x),$$

also we have

$$\lim_{j\to\infty} P_j(f) = f.$$

The projection $P_j(f)$ can be considered as an approximation of f at the scale 2^{-j}. Therefore, the successive approximations of a given function f are defined as the orthogonal projections $P_j(f)$ onto the space V_j. We can choose $j \in \mathbb{Z}$ such that $P_j(f)$ is a good approximation of f. For very large j we can approximate $f(x)$ by $P_j(f)$, that is,

$$f(x) \approx P_j(f)(x) = \sum_{k\in\mathbb{Z}} \alpha_{j,k}\phi_{j,k}(x),$$

where

$$\alpha_{j,k} = \langle f, \phi_{j,k}\rangle, \quad \text{and} \quad \phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k).$$

From the definition of the derivative we have

$$f'(x) = \lim_{j\to\infty} \frac{f\left(x + \frac{1}{2^j}\right) - f(x)}{\frac{1}{2^j}}.$$

Again for large j we can approximate $f'(x)$ by

$$f'(x) \approx 2^j \left[f\left(x + \frac{1}{2^j}\right) - f(x)\right].$$

Substituting the above values we get

$$\begin{aligned} f'(x) &\approx 2^j \left\{ f\left(x + \frac{1}{2^j}\right) - f(x)\right\} \\ &= 2^j \left\{ \sum_{k\in\mathbb{Z}} \alpha_{j,k} 2^{j/2}\phi\left(2^j\left(x + \frac{1}{2^j}\right) - k\right) - \sum_{k\in\mathbb{Z}} \alpha_{j,k}2^{j/2}\phi(2^j x - k)\right\} \\ &= 2^j \left\{ \sum_{k\in\mathbb{Z}} \alpha_{j,k}2^{j/2}\phi(2^j x + 1 - k) - \sum_{k\in\mathbb{Z}} \alpha_{j,k}2^{j/2}\phi(2^j x - k)\right\} \\ &= 2^j \left\{ \sum_{k\in\mathbb{Z}} \left(\alpha_{j,k+1} - \alpha_{j,k}\right)\phi_{j,k}(x)\right\}. \end{aligned}$$

Let $V_j = \{f \in L^2(\mathbb{R}) : f = \text{constant on } I_{j,k}, \forall k \in \mathbb{Z}\}$ be the space of all functions in $L^2(\mathbb{R})$ which are constants on intervals of the form $I_{j,k} = [2^{-j}k, 2^{-j}(k+1)]$, $k \in \mathbb{Z}$. Then $\{V_j, j \in \mathbb{Z}\}$ is an MRA. The scaling function is given by $\phi = \chi_{[0,1]}$. Now, consider a simple differential equation

$$f'(x) + bf(x) = 0, \quad f(0) = f_0,$$

where b is a constant real number. The exact solution of the differential equation is

$$f(x) = f_0 e^{-bx}.$$

Now, substituting these results yields

$$2^j \left[\sum_{k\in\mathbb{Z}} (\alpha_{j,k+1} - \alpha_{j,k})\, \phi_{j,k} \right] + b \sum_{k\in\mathbb{Z}} \alpha_{j,k}\phi_{j,k}(x)$$
$$= \sum_{k\in\mathbb{Z}} \left(2^j \alpha_{j,k+1} + (b - 2^j)\alpha_{j,k}\right) \phi_{j,k} = 0,$$

taking the inner product with $\phi_{j,n}$ we get

$$2^j \alpha_{j,n+1} + (b - 2^j)\alpha_{j,n} = \alpha_{j,n+1} = \left(1 - \frac{b}{2^j}\right) \alpha_{j,n}.$$

Solving the difference equation we get

$$\alpha_{j,n} = \left(1 - \frac{b}{2^j}\right)^n \alpha_{j,0,},$$

where

$$\alpha_{j,0} = \langle f, \phi_{j,0}\rangle = \int_{-\infty}^{\infty} f(x)\phi_{j,0}(x)\, dx = \int_0^{2^{-j}} f(x) 2^{j/2} \phi(2^j x)\, dx$$
$$= 2^{j/2} f(0) \int_0^{2^{-j}} \phi(2^j x)\, dx = 2^{j/2} f(0) \int_0^{2^{-j}} 1\, dx = 2^{-j/2} f(0) = 2^{-j/2} f_0.$$

Since $f(x)$ is continuous and the integration is taken over a small interval $[0, 2^{-j}]$, we can approximate $f(x)$ by $f(0)$ for very large j. Similarly for $\alpha_{j,k}$ we have

$$\alpha_{j,k} = \langle f, \phi_{j,k}\rangle = \int_{-\infty}^{\infty} f(x)\phi_{j,k}(x)\, dx$$
$$= \int_{2^{-j}k}^{2^{-j}(k+1)} f(x) 2^{j/2} \phi(2^j x - k)\, dx = 2^{-j/2} f(2^{-j}k).$$

Then, we get

$$f(2^{-j}k) = f_0 \left(1 - \frac{b}{2^j}\right)^k.$$

Let $k \to 2^j x$, then

$$f(x) = f_0 \left(1 - \frac{b}{2^j}\right)^{2^j x}$$

for very large j. Taking the limit as $j \to \infty$ we get

$$f(x) = \lim_{j\to\infty} f_0 \left(1 - \frac{b}{2^j}\right)^{2^j x} = f_0 e^{-bx}$$

which coincides with the exact solution.

5. Application to Derivatives

In this subsection we will prove that for certain functions the derivative can be written as

$$f'(x) = \sum_{n\in\mathbb{Z}} t_n f(x-n),$$

where $t_n \in \mathbb{R}$ for all $n \in \mathbb{Z}$.

Let P_k be the space of polynomials which have degree less than or equal to k. Then t_n can be found by solving a system of linear equations. For example, for $f \in P_2$ one can prove that

$$f'(x) = \frac{1}{2}f(x+1) - \frac{1}{2}f(x-1).$$

For $f \in P_4$ we have

$$\begin{aligned} f'(x) = {} & \frac{-1}{12}f(x+2) + \frac{2}{3}f(x+1) \\ & - \frac{2}{3}f(x-1) + \frac{1}{12}f(x-2). \end{aligned}$$

Lemma 5.1. Let $f \in L^2(\mathbb{R})$, and $\hat{f}$ does not vanish in $[-\pi,\pi]$ almost everywhere, then f' can be written in the following form:

$$f'(x) = \sum_{n\in\mathbb{Z}} t_n f(x-n),$$

where

$$t_n = \begin{cases} \dfrac{(-1)^n}{n}, & n \neq 0,\ n \in \mathbb{Z}, \\ 0, & n = 0. \end{cases}$$

Proof. Taking the Fourier transform we get

$$(iw)\hat{f}(w) = \sum_{n\in\mathbb{Z}} t_n e^{-iwn}\hat{f}(w).$$

Since $\hat{f}(w) \neq 0$ a.e. for $w \in (-\pi,\pi)$, then by cancelling $\hat{f}(w)$ from both sides we get

$$(iw) = \sum_{n\in\mathbb{Z}} t_n e^{-iwn}.$$

Taking inner product with $e^{-iwm},\ m \neq 0$, we get

$$\int_{-\pi}^{\pi}(iw)e^{iwm}\,dw = \int_{-\pi}^{\pi}\sum_{n\in\mathbb{Z}} t_n e^{-iwn}e^{iwm}\,dw.$$

Then

$$\frac{2\pi\,(\cos\pi m)\,m - 2\sin\pi m}{m^2} = \sum_{n\in\mathbb{Z}} t_n \int_{-\pi}^{\pi} e^{-iwn}e^{iwm}\,dw.$$

Since $\{e^{-iwn}\}$ are orthogonal in $(-\pi, \pi)$ and m is an integer, we have

$$\frac{2\pi(-1)^m}{m} = 2\pi t_m \quad \text{and} \quad t_m = \frac{(-1)^m}{m}.$$

If $m = 0$, then $t_0 = 0$. ∎

Example 5.2. Let $f(x) = \dfrac{\sin x}{x}$. Using the definition introduced,

$$\begin{aligned}\sum_{n\in\mathbb{Z}} t_n f(x-n) &= \sum_{n\neq 0} \frac{(-1)^n}{n} \frac{\sin(x-n)}{x-n} \\ &= \frac{\cos x}{x} - \frac{\sin x}{x^2} = f'(x).\end{aligned}$$

Similar results can be found with higher derivatives. For example, for the second derivative we have

$$f''(x) = \sum_{n\in\mathbb{Z}} r_n f(x-n),$$

where

$$r_n = \begin{cases} \dfrac{2(-1)^{n+1}}{n^2}, & n \neq 0, n \in \mathbb{Z}, \\ -\dfrac{\pi^2}{3}, & n = 0. \end{cases}$$

For the third derivative,

$$f'''(x) = \sum_{n\in\mathbb{Z}} r_n f(x-n),$$

where

$$r_n = \begin{cases} (-1)^n \left(\dfrac{6}{n3} - \dfrac{\pi^2}{n}\right), & n \neq 0,\ n \in \mathbb{Z}, \\ 0, & n = 0. \end{cases}$$

For the fourth derivative we get

$$f^{(4)}(x) = \sum_{n\in\mathbb{Z}} r_n f(x-n),$$

where

$$r_n = \begin{cases} 4(-1)^n \left(\dfrac{\pi^2}{n^2} - \dfrac{6}{n^4}\right), & n \neq 0, n \in \mathbb{Z}, \\ \dfrac{\pi^4}{5}, & n = 0. \end{cases}$$

Higher derivatives can also be obtained in a similar procedure.

There are four main properties of wavelets; namely:

1. They are local in both space and frequency.
2. They satisfy biorthogonality conditions.
3. They provide a multiresolution structure.
4. They provide fast transform algorithms.

Because of these properties wavelets have proven to be useful in the solution of ordinary differential equations. As proposed by several researchers, wavelets can be used as basis functions in Galerkin method.

This has proven to work and results in a linear system that is sparse because of the compact support of wavelets, and that, after preconditioning, has a condition number independent of problem size because of the multiresolution structure. By using two pairs of biorthogonal compactly supported wavelets, derivative operator can be diagonalized [13]. Like the Fourier transform, wavelets can diagonalize constant coefficient operators.

The resulting algorithm is slightly faster [19]. Even non constant coefficient operators can be diagonalized with the right choice of basis which yields a much faster algorithm than classical iterative methods.

6. Wavelet Networks

The idea of combining wavelets and neural networks has resulted in the formulation of wavelet networks — a feed-forward neural network with one hidden layer of nodes, whose basis functions are drawn from a family of orthonormal wavelets.

The similarities between the discrete inverse wavelet transform (WT) and a one-hidden-layer neural network, universal approximation properties of neural network, and a rich theoretical basis of wavelet and neural networks have resulted in heightened activity in wavelet research and applications.

The field of wavelet networks is new, although some sporadic and isolated attempts have taken place in the recent years to build a theoretical basis and application to various fields. There is tremendous potential for its application to new and existing areas. The use of wavelet network can be traced to application of Gabor wavelets for image classification. Recent popularity of applications of wavelet networks in speech segmentation, speaker recognition, face tracking, real environment characterization for haptic displays, forecasting, and prediction of chaotic signals. Zhang [30] used the wavelet networks to control a robot arm and used a mother wavelet of the form

$$\Psi(x) = \big(x^T x - \dim(x)\big)e^{-1/2x^T x}.$$

In Szu's work with regard to the classification of problems and speaker recognition, a mother wavelet is of the form

$$\cos(1.75t)e^{-1/2t^2}.$$

In the literature; combination of sigmoid function has been generalized to polynominal approximation of sigmoid functions. Wavelet networks are inspired by both the feed-forward neural networks and the theory of underlying wavelet decompositions. The basis idea of wavelet decomposition is to expand a generic signal $f(x) \in L^2(\mathbb{R}^n)$ into a series of functions by dilating and translating a single mother wavelet. These functions are then provided to a perception type neural network.

Most work done in wavelet network uses simple wavelets and takes advantage of only a reduced part of wavelet theory. Zhang and Benveniste [30] have found a link between the wavelet decomposition theory and neural networks and present a basic back-propagation wavelet network learning algorithm. Their wavelet network preserves the universal approximation properties of traditional feed-forward neural networks and presents an explicit link between the network coefficients and some appropriate transform. It has been shown that a three-layer neural network can approximate any arbitrary continuous function on a compact set within a predetermined precision 0. In particular, any arbitrary function, $f : \mathbb{R} \to \mathbb{R}$, with p continuous derivatives on $(0, 1)$ and $|f^p(x)| \leq \Gamma$, such that the function is equal to zero in some neighborhood of the end-points, can be approximated by

$$f(x) = \sum_{k=1}^{N} \xi_k T(\omega_k x + b_k),$$

where N is the number of neurons in the hidden layer, is the weight between the input and hidden layer, and is the weight between the hidden and output layer. The function is the nonlinearity, for example, sigmoid or Gaussian. The most popular wavelet networks are based on perception structure, so we can safely say that wavelet networks are a subset of perception architecture. Wavelet decomposition for approximation is very similar to a layer neural network. Wavelet networks learning is generally performed by means of standard learning algorithms; therefore, their study is similar to the study of any other multilayer neural networks layer. Most wavelet network properties, when used with standard learning algorithms, are due to the wavelet localization property.

In addition to forming an orthogonal basis, wavelets have the capacity to explicitly represents the behavior of a function at different resolution of input variables. Therefore a wavelet network can first be trained to learn the mapping at the coarsest resolution level and trained to include elements of higher resolutions. The process may be repeated for finer granularity. This hierarchical, multiresolution training can result in more meaningful interpretation of the resulting mapping and adaptation of network that are more efficient compared to conventional methods. In addition, the wavelet theory provides useful guidelines for the construction and initialization of networks and, consequently, the training times are significantly reduced.

There are two main approaches to form wavelet networks. In the first approach, the wavelet component is decoupled from the learning component of the perception architecture. In essence, a signal is decomposed on some wavelet and wavelet coefficients are fed to the neural network. In the second approach, the wavelet theory and neural networks are combined into a single method. In wavelet networks, both the position and dilation of the wavelet as well as the weights are optimized. Originally, wavelets referred to neural networks using dyadic wavelets; in wavenets, the position and dilation of wavelet are fixed and the weights are optimized through a learning process.

The basic neuron of a wavelet network is a multidimensional wavelet in which the dilation and the translation coefficients are considered as neuron parameters. The output of a wavelet is therefore a linear combination of several multidimensional wavelets. The expression

$$H(x) = \prod_i h_{d_i t_i}(x_i)$$

is a multidimensional wavelet and

$$h_{d,t}(x) = h(d * (x - t))$$

represents a derived wavelet, where h is the main wavelet. Note that the wavelet neuron is equivalent to a multidimensional wavelet. $x_1, x_2, \dots, x_k$ are assets of input values, H represent a mother wavelet, and each of the $h_{d,t}$ values is a derived wavelet.

Here each H_i is a wavelet neuron, the output is given by

$$\theta + \sum_{j=1}^{N} \omega_j H_j(x_1, x_2, \dots, x_k).$$

The inputs are labeled $x_1, x_2, \dots, x_k$ and the weights are labeled $\omega_1, \omega_2, \dots, \omega_n$. The hidden layer consist of $H_1, H_2, \dots, H_n$ wavelet neurons. The output of three layer perception is given by

$$y(x) = \sum_{i=1}^{N} \omega_i f(a_i x + b_i).$$

Here f is as activation functions, and a_i, b_i, ω_i are the network weight parameters that are optimized during learning. A wavelet network has the same structure except that the function f is replaced by a wavelet represented by H.

A wavenet in its simplest form corresponds to a feed-forward neural network using wavelets an activation function

$$y(x) = \sum_{i,j} d_{i,j} f_{i,j}(x) + \bar{y},$$

where $\bar{y}$ is the average of y, $d_{i,j}$ are the coefficients of the neural network, and f is a wavelet. For orthogonal wavelets, a simple gradient descent rule will lead to a global minimum under the following conditions, if the weights $d_{i,j}$ are optimized. Select an input data point (o_k, p_k) such that $y(p_k) = o_k$ to the network, then the error $E(k)$ is given by

$$E(k) = \big(\hat{y}(x) - \sum_{d_{i,j}} f_{i,j}(x)\big)^2 = \sum_{i,j} ((\hat{d}_{i,j} - d_{i,j}) f_{i,j}(x))^2.$$

Due to orthogonality, the diagonal term vanishes. Differentiating $E(k)$ with respect to $\hat{d}_{i,j}$, we get

$$\left.\frac{\partial E(k)}{\partial d_{i,j}}\right|_{x_k} = 2\hat{d}_{i,j}(\hat{y}(x_k) - y(x_k)).$$

6.1. Theory of Wavelet Networks

A neural network is a parallel dynamic system whose state response to input carries out processing. Neural networks can be viewed as approximation tools for fitting models (linear or nonlinear) based on input and output data. They are tool for general approximation and have been used in black box identification of nonlinear systems.

One of the important concepts in neural network is the interaction between small scale or large scale phenomena. For example, in neural networks for recognizing a pattern in digital image, the global (large scale) properties are patterns and the local (small scale) properties are the values of individual pixel in the image. The situation is similar for some theorems in calculus which show relationship between infinitesimal properties, local properties, and global properties. For example, the well-known theorem in calculus: "If the derivative of a function f is positive at every point on an interval (a, b), then f is monotone increasing on the interval" or "If a differentiable functions f on the interval (a, b) has a local maximum or local minimum at c, then its derivative $f'(c)$ is zero." The former statement relates the infinitesimal property of f at every point of the interval (derivative being positive) to the global property of pattern of f (being monotone increasing). The second statement relates the local property of f (having a local maximum at c, that is, $f(c) \geq f(x)$ for all x in neighborhood of c) to the infinitesimal property of f at one point ($f'(c) = 0$).

Wavelet decomposition of functions is similar to Fourier decomposition of functions but wavelet theory provides a new theory for hierarchical decomposition of functions and multiscale approximation of functions. The basis functions $e^{2\pi ikx}$ of Fourier decomposition have many desirable properties, such as being orthogonal,

$$\langle e^{2\pi ikx}, e^{2\pi imx} \rangle = \int e^{2\pi ikx} e^{-2\pi imx} dx = 0, \quad \text{for } k \neq m,$$

and being eigenvectors of the differential operator d/dx, that is,

$$\frac{d}{dx}\left(e^{2\pi ikx}\right) = 2\pi ik e^{2\pi ikx}.$$

However, they are not local in space (or time if x is the time variable). The basis functions used in wavelet decompositions do not behave so nicely as $e^{2\pi ikx}$ under the differential operator d/dx (some wavelet functions are not differentiable and the Haar wavelet is not even continuous), but they are localized in both space (or time) and frequency. The wavelet decomposition of a function is similar to $\left(1 + \dfrac{1}{2}\right)$-layer neural network, which is a network of the form

$$g(\mathbf{x}) = \sum_{i=1}^{N} \omega_i \sigma(\mathbf{a}_i \cdot \mathbf{x} + b_i),$$

where ω_i and b_i are real numbers, $\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{in})$ and $\mathbf{x} = (x_1, x_2, \dots, x_n)$ are vectors in $\mathbb{R}^n$, $\mathbf{a}_i \cdot \mathbf{x}$ is the usual Euclidean inner product $\sum_{k=1}^{n} a_{ik}x_k$ and σ is a certain continuous monotone increasing function (usually a sigmoid function), with

$$\lim_{t\to-\infty} \sigma(t) = 0, \quad \lim_{t\to\infty} \sigma(t) = 1.$$

A sigmoid function can be viewed as a smoothed version of the sign function which returns -1 for negative input and 1 for nonnegative input.

Some examples for sigmoid functions are

$$\begin{aligned}
f_i(x) &= \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \\
f_i(x) &= \frac{1 - e^{-x}}{1 + e^{-x}} = \tanh(x/2), \\
f_i(x) &= \frac{2}{\pi} \arctan\left(\frac{\pi}{2}x\right), \\
f_i(x) &= \frac{x^2}{1 + x^2} \mathrm{sgn}(x),
\end{aligned}$$

where $\mathrm{sgn}(\cdot)$ is a signum function and all the above nonlinear functions are bounded, monotonic and nondecreasing functions.

Any bounded continuous sigmoid function is discriminatory, which means,

$$\int_{[0,1]^n} \sigma(\mathbf{ax} + b)d\mu(\mathbf{x}) = 0, \quad \forall \mathbf{a} \in \mathbb{R}^n \text{ and } \forall b \in \mathbb{R},$$

implies $\mu = 0$. If σ is continuous and discriminatory, then the finite sums of the form

$$\sum_{i=1}^{N} \omega_i \sigma(\mathbf{a}_i \cdot \mathbf{x} + b_i)$$

can approximate any continuous function defined on the n-dimensional cube $[0, 1]^n$. That means given $\varepsilon > 0$ and a continuous function f defined on $[0, 1]^n$, there is a finite sum

$$g(\mathbf{x}) = \sum_{i=1}^{N} \omega_i \sigma(\mathbf{a}_i \cdot \mathbf{x} + b_i)$$

such that

$$|f(\mathbf{x}) - g(\mathbf{x})| < \varepsilon \quad \forall \mathbf{x} \in [0, 1]^n.$$

In such a case, the set of all finite sums of the form $\sum_{i=1}^{N} \omega_i \sigma(\mathbf{a}_i \cdot \mathbf{x} + b_i)$ is said to be dense in $C([0,1]^n)$, the space of all continuous functions on $[0,1]^n$.

6.2. Wavelet Network Structure

Wavelet network is a network combining the idea of the feed-forward neural networks and the wavelet decompositions. Zhang and Benveniste [30] provide an alternative to the feed-forward neural networks for approximating functions. Wavelet networks use simple wavelets, and wavelet network learning is performed by the standard back propagation type algorithm as the traditional neural network. The localization property of wavelet decomposition is reflected in the important properties of wavelet networks.

Wavelet networks can approximate any continuous function on $[0,1]^n$ and have certain advantage such as the use of wavelet coefficient as the initial value for back propagation training and possible reduction of the network size while achieving the same level of approximation. In a feed-forward network, neurons take their inputs from the previous layer only and send the output to the next layer only. Since the signal goes in one direction only, the network can compute a result vary quickly.

Basic neurons of a wavelet network are multidimensional wavelets and the neurons parameters are the dilation and translation coefficient. The output of a wavelet network is the linear combination of the values of several multidimensional wavelets.

Suppose there is a function ψ defined on $\mathbb{R}^n$ such that there is a countable set Ψ of the form

$$\Psi = \{\psi(D_i(\mathbf{x} - \mathbf{t}_i))\},$$

D_i is an $n \times n$ diagonal matrix with the diagonal vector $d_i \in \mathbb{R}^n$, and $\mathbf{x}$, $\mathbf{t}_i \in \mathbb{R}^n$, is a frame which means there exist constants A and B such that

$$A\|f\|^2 \leq \sum_{\alpha \in \Psi} |\langle \alpha, f \rangle|^2 \leq B\|f\|^2$$

for any $f \in L^2(\mathbb{R}^n)$. It follows from the frame property that the set S of all linear combinations of the elements in Ψ is dense in $L^2(\mathbb{R}^n)$. Obviously, the set of all linear combinations of the form

$$\sum_{i=1}^{N} \omega_i \psi(D_i(\mathbf{x} - \mathbf{t}_i)),$$

where D_i and t_i are not restricted to those in Ψ is a subset of S and is also dense in $L^2(\mathbb{R}^n)$. For example, we can use ψ given by

$$\psi(\mathbf{x}) = \psi(x_1, \ldots, x_n) = \psi_s(x_1)\psi_s(x_2)\cdots\psi_s(x_n),$$

where $\psi_s : \mathbb{R} \to \mathbb{R}$ is given by

$$\psi_s(x) = -xe^{-x^2/2},$$

the first derivative of the Gaussian function $e^{-x^2/2}$. Note the derivative of ψ_s,

$$\frac{d}{dt}\psi_s(x) = -e^{-x^2/2}(1-x^2).$$

The wavelet network structure will be of the form

$$h(\omega, x) = \sum_{i=1}^{N} a_i \psi(D_i(\mathbf{x} - \mathbf{t}_i)) + \bar{g},$$

where $a_i \in \mathbb{R}$, ψ is a given wavelet function, D_i an $n \times n$ diagonal matrix, $\mathbf{x}$ and $\mathbf{t}_i \in \mathbb{R}^n$, and $\bar{g}$ is the average value of $g(x)$, ω represents all the parameters

$$a_1, a_2, \dots, a_n, D_1, D_2, \dots, D_n, \mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_n, \text{ and } \bar{g}.$$

The matrix D_i and $\mathbf{t}_i$ are set by the wavelet decomposition and the weights ω_i are initially set to be zero. It should be noted that the wavelet decomposition uses the given D_i and t_i and finds the weight coefficients ω_i, while the wavelet network tries to adjust $D_i, \mathbf{t}_i$, and the weight coefficients ω_i altogether to fit the data. Wavelet networks can be used for concept learning for a concept $S \subset [0,1]^n$ by using $f = \chi_S$, the characteristic function of S, that is,

$$f(x) = \begin{cases} 1 & \text{for } \mathbf{x} \in S, \\ 0 & \text{for } \mathbf{x} \notin S. \end{cases}$$

Given $\varepsilon > 0$, there exist

$$g(x) = h(\omega, \mathbf{x}) = \sum_{i=1}^{N} a_i \psi(D_i(\mathbf{x} - \mathbf{t}_i)) + \bar{g} \tag{6.1}$$

and $D \subset [0,1]^n$ with measure $\geq 1 - \varepsilon$ such that $|g(x) - f(x)| < \varepsilon, \quad \forall \mathbf{x} \in D$. The learning algorithm of a wavelet network modifies the dilation and the translation coefficient of every wavelet neuron and the coefficient (weight) of a linear combination of the neurons so that the network closely fits the data. We assume the data is contained with noise, so the learning algorithm should not seek to interpolate the data points. The network g_θ, where θ represents all the parameters $D_i, \mathbf{t}_i$ and ω_i, will be adjusted by the learning algorithm to minimize a suitable objective function, so that it becomes an optimization problem. A simple objective function we consider is

$$C(\theta) = E(|g_\theta(\mathbf{x}) - \mathbf{y}|^2),$$

where $\mathbf{x}_k$ and $\mathbf{y}_k$ are data pairs, that is, $f(\mathbf{x}_k) = \mathbf{y}_k + \eta_k$, where η_k is a random noise. Though a standard gradient descent algorithm can be used, a heavy computation requirement makes it not practical in some situations. In practice some other more efficient algorithms, such as stochastic gradient method, are used. The function computed by the basic wavelet network model is differentiable with respect to all parameters (dilation and translation parameter and the weights).

6.3. Wavelet Network Algorithm

The following algorithm follows the paper of Zhang and Benveniste [30]: Description of the algorithm to approximate a real valued function $f(x)$ defined on a closed interval $[a, b]$ by a neural network of the form

$$\sum_{i=1}^{N} \omega_i \psi\left(\frac{x - t_i}{s_i}\right) + \bar{g}.$$

The algorithm is based on samples of input and output pairs $x_k,\ y_k = f(x_k) + v_k$, where v_k is the random measurement noise and f is the function the network is to approximate. Let θ be the set of all parameters $\bar{g}, t_i, s_i, D_i$ be the network defined by

$$g_\theta(x) = \sum_{i=1}^{N} \omega_i \psi(D_i(x - t_i)) + \bar{g}.$$

The function

$$C(\theta) = E(|g_\theta(x) - y|^2)$$

is the objective function to be minimized. The minimization can be done by various optimization methods. One method is a stochastic gradient algorithm which recursively minimizes $C(\theta)$ by modifying in the opposite direction of the gradient of

$$c(\theta, x_k, y_k) = \frac{1}{2}(g_\theta(x_k) - y_k)^2$$

after each sample (x_k, y_k). The factor $1/2$ is put to simplify the formulas in taking the gradient. The gradient of c is the vector

$$\nabla c = \left(\frac{\partial c}{\partial \theta_1}, \frac{\partial c}{\partial \theta_2}, \dots\right),$$

more explicitly,

$$\begin{aligned}
\frac{\partial c}{\partial \bar{g}} &= g_\theta(x_k) - y_k, \\
\frac{\partial c}{\partial \omega_i} &= (g_\theta(x_k) - y_k)\psi(D_i(x - t_i)), \\
&\vdots
\end{aligned}$$

The objective function may have a number of local extrema and we have the usual difficulty of avoiding being trapped at a local minimum. This problem is addressed in general optimization literature.

The initialization of the network parameter $(\omega_i, t_i, s_i, \bar{g})$ comes from the wavelet decomposition of the function using input and output measurements $(x, f(x))$. The integral in the formulas for wavelet decomposition are roughly estimated from the discrete data. For example, $\bar{g}$ is set to the average value of $f(x)$ computed from the measurement. Select a point $p \in [a, b]$ as follows. Let

$$\varrho(x) = \left|\frac{df}{dx}\right|, \quad \rho(x) = \frac{\varrho(x)}{\int_a^b \varrho(x)\,dx},$$

which is estimated from the measurement $(x, f(x))$. The point p is taken by

$$p = \int_a^b x\rho(x)\,dx.$$

The ρ can be considered as probability density function, p the mean value which is well defined except the trivial case when $\frac{df}{dx} \equiv 0$, that is, when f is a constant function. In practice, taken the values $t_1 = p, \quad s_1 = \frac{b-a}{2}$, the point t_1 divides the interval $[a, b]$ into two parts. We initialize the rest of t_i and s_i recursively in each subinterval until all the parameters are initialized. This requires N to be a power of 2. If N is not a power of 2 in practice, we apply the recursion as far as possible, then initialize the rest of t_i at random for the finest scale.

References

[1] B. Al-Humaidi. Spline-wavelets in numerical solutions to differential equations, Master's thesis, KFUPM, 2001.

[2] Silvia Bertoluzza, Giovanni Naldi, and Jean-Christophe Ravel. Wavelet methods for the numerical solution of boundary value problems on the interval, In *Wavelets: theory, algorithms, and applications (Taormina, 1993)*, volume 5 of *Wavelet Anal. Appl.*, pages 425–448. Academic Press, San Diego, CA, 1994.

[3] G. Beylkin. On the representation of operators in bases of compactly supported wavelets, *SIAM J. Numer. Anal.*, 29(6):1716–1740, 1992.

[4] G. Beylkin. On wavelet-based algorithms for solving differential equations, In *Wavelets: mathematics and applications*, *Stud. Adv. Math.*, pages 449–466. CRC, Boca Raton, FL, 1994.

[5] G. Beylkin, M.E. Brewster, and A.C. Gilbert. A multiresolution method for numerical reduction and homogenization of nonlinear ODEs, *Appl. Comput. Harmon. Anal.*, 5(4):450–486, 1998.

[6] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms, I. *Comm. Pure Appl. Math.*, 44(2):141–183, 1991.

[7] Richard L. Burden, J. Douglas Faires, and Albert C. Reynolds. *Numerical analysis*, Prindle, Weber & Schmidt, Boston, Mass., fifth edition, 1993.

[8] Charles K. Chui. *An introduction to wavelets*, volume 1 of *Wavelet Analysis and its Applications*, Academic Press Inc., Boston, MA, 1992.

[9] A. Cohen, Ingrid Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.*, 45(5):485–560, 1992.

[10] Stephan Dahlke. *Wavelets: construction principles and applications to the numerical treatment of operator equations*, Shaker Verlag, Aachen, 1997.

[11] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.*, 41(7):909–996, 1988.

[12] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[13] Ingrid Daubechies. Two recent results on wavelets: wavelet bases for the interval, and biorthogonal wavelets diagonalizing the derivative operator, In *Recent advances in wavelet analysis*, volume 3 of *Wavelet Anal. Appl.*, pages 237–257. Academic Press, Boston, MA, 1994.

[14] Lokenath Debnath. *Wavelet transforms and their applications*, Birkhäuser Boston Inc., Boston, MA, 2002.

[15] Michael W. Frazier. *An introduction to wavelets through linear algebra*, Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1999.

[16] Carla Guerrini and Manuela Piraccini. Parallel wavelet-Galerkin methods using adapted wavelet packet bases, In *Approximation theory VIII, Vol. 2 (College Station, TX, 1995)*, volume 6 of *Ser. Approx. Decompos.*, pages 133–142. World Sci. Publ., River Edge, NJ, 1995.

[17] Eugenio Hernández and Guido Weiss. *A first course on wavelets*, Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1996. With a foreword by Yves Meyer.

[18] Stéphane Jaffard, Yves Meyer, and Robert D. Ryan. *Wavelets*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, revised edition, 2001, Tools for science & technology.

[19] Björn Jawerth and Wim Sweldens. Wavelet multiresolutions analyses adapted for the fast solution of boundary value ordinary differential equations, In N. Melson, T. Manteuffel, and S. McCormick, editors, *Sixth Copper mountain conference on multigrid methods*, volume 3224 of *NASA conference publication*, pages 257–273, 1993.

[20] Xue-zhang Liang and Ming-cai Liu. Wavelet-Galerkin methods for second kind integral equations. In *Wavelet analysis and applications (Guangzhou, 1999)*, volume 25 of *AMS/IP Stud. Adv. Math.*, pages 207–216. *Amer. Math. Soc.*, Providence, RI, 2002.

[21] Stephane G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbf{R})$, *Trans. Amer. Math. Soc.*, 315(1):69–87, 1989.

[22] Yves Meyer. *Wavelets and operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press, Cambridge, 1992, Translated from the 1990 French original by D. H. Salinger.

[23] Yves Meyer. *Wavelets*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1993, Algorithms & applications, Translated from the French and with a foreword by Robert D. Ryan.

[24] Han-zhang Qu, Chen Xu, and Ruizhen Zhao. An application of continuous wavelet transform in differential equations, In *Wavelet analysis and its applications*, volume 2251 of *Lecture Notes in Comput. Sci.*, pages 107–116, Springer, Berlin, 2001.

[25] N. Roland. Fourier and wavelet representations of functions, *Electronic Journal of Undergraduate Mathematics, Furman University*, 6:1–2, 2000.

[26] Abul Hasan Siddiqi. *Applied functional analysis*, volume 258 of *Monographs and Textbooks in Pure and Applied Mathematics*, Marcel Dekker Inc., New York, 2004, Numerical methods, wavelet methods, and image processing.

[27] Wim Sweldens and Robert Piessens. Calculation of the wavelet decomposition using quadrature formulae. In *Wavelets: an elementary treatment of theory and applications*, volume 1 of *Ser. Approx. Decompos.*, pages 139–160. World Sci. Publ., River Edge, NJ, 1993.

[28] Wim Sweldens and Robert Piessens. Quadrature formulae and asymptotic error expansions for wavelet approximations of smooth functions, *SIAM J. Numer. Anal.*, 31(4):1240–1264, 1994.

[29] Philippe Tchamitchian. Wavelets and differential operators. In *Different perspectives on wavelets (San Antonio, TX, 1993)*, volume 47 of *Proc. Sympos. Appl. Math.*, pages 77–88, *Amer. Math. Soc.*, Providence, RI, 1993.

[30] Qinghua Zhang and Albert Benveniste. Wavelet networks, *IEEE Trans. Neural Networks*, 3:889–898, 1992.