

RICE UNIVERSITY

**The Mathematical Theory and Applications of
Biorthogonal Coifman Wavelet Systems**

by

Jun Tian

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Raymond O. Wells, Jr., Chairman
Professor of Mathematics and Education

John C. Polking
Professor of Mathematics

C. Sidney Burrus
Professor of Electrical and Computer
Engineering

Houston, Texas

May, 1996

The Mathematical Theory and Applications of Biorthogonal Coifman Wavelet Systems

Jun Tian

Abstract

In this thesis, we present a theoretical study of biorthogonal Coifman wavelet systems, a family of biorthogonal wavelet systems with vanishing moments equally distributed between scaling functions and wavelet functions. One key property of these wavelet systems is that they provide nice wavelet sampling approximation with exponential decay. Moreover they are compactly supported, symmetric, have growing smoothness with large degrees, and converge to the sinc wavelet system. Using a time domain design method, the exact formulas of the coefficients of biorthogonal Coifman wavelet systems of all degrees are obtained. An attractive feature behind it is that all the coefficients are dyadic rational, which means that we can implement a very fast multiplication-free discrete wavelet transform, which consists of only addition and shift operations, on digital computers. The transform coding performance of biorthogonal Coifman wavelet systems is quite comparable to other widely used wavelet systems. The orthogonal counterparts, orthogonal Coifman wavelet systems, are also discussed in this thesis.

In addition we develop a new wavelet-based embedded image coding algorithm, the Wavelet-Difference-Reduction algorithm. Unlike zerotree type schemes which use spatial orientation tree structures to implicitly locate the significant wavelet transform coefficients, this new algorithm is a direct approach to find the positions of significant coefficients. It combines the discrete wavelet transform, differential coding, binary reduction, ordered bit plane transmission, and adaptive arithmetic coding. The encoding can be stopped at any point, which allows a target rate or distortion metric to be met exactly; the decoder can also terminate the decoding at any point, and produce a corresponding reconstruction image. Our algorithm provides a fully embedded code to successively approximate the original image source; thus it's well suited for progressive image transmission. It is very simple in its form (which will

make the encoding and decoding very fast), and has a clear geometric structure, which enables us to process the image data in the compressed wavelet domain. The image coding results of it are quite competitive with almost all previous reported image compression algorithms on standard test images.

Acknowledgments

I would like to take this opportunity to express my high appreciation to Ronny Wells, my academic advisor. His broad knowledge, profound insight, long-lasting enthusiasm has always guided me into the right direction to do the right job. Also under his instruction, my English has been improved a lot. I am really enjoying working with him.

John C. Polking was my first advisor at Rice University. He helped me in every way to get started for graduate study. My second advisor was Bob Hardt. Whenever I had a question, he was always there and ready to help me out. I learned geometric analysis from him, harmonic analysis from Stephen Semmes, and partial differential equations from Leon Simon. All these have provided a solid foundation for what I am working on now. I would like to thank all the professors at mathematics department of Rice University. Their lectures have benefited graduate students a lot and they have created such a nice academic environment to us to learn mathematics. A special thanks to Sharon F. McDonough and Maxine Turner. They just make our life at Rice so smooth and comfortable.

Most of the work in the thesis was done in the Computational Mathematics Laboratory, Rice University. It was there that Ronny led me into the fantastic world of wavelet analysis. Xiaodong Zhou was my “big brother” in the lab. He always shared his creative ideas with me and we had a very good time together. I learned digital signal processing from Sid Burrus. He has given me lots of feedback to keep moving forward. From Houston to Bremen, Markus Lang not only gave me numerous assistance in research, it was also a great fun to chat with him. Haitao Guo provided his nice compression program to me. Without his help, the discovery of the Wavelet-Difference-Reduction algorithm would not have been possible. Richard A. Tapia explained the Newton-Kantorovich Theorem to me. Dong Wei helped me with a complete transform coding evaluation of biorthogonal Coifman wavelet systems. Also I would like to thank Roland Glowinski, David W. Scott, Richard G. Baraniuk, Jan E. Odegard, and other members at the lab.

I spent about half a year at the Center for Medical Visualization and Diagnostic Systems, Germany. The visit was both fruitful and colorful. I would like to thank Heinz-Otto Peitgen, Carl Evertsz, Hartmut Juergens, Aska Loffroy, Thomas Netsch, and all others there for their hospitality.

Meanwhile I want to thank Howard L. Resnikoff and Peter N. Heller at Aware, Inc. They have broadened my eyes and the discussions with them have been very helpful. Amir Said at State University of Campinas explained their codetree algorithms to me in details. I have gained much benefit from his joint paper with William A. Pearlman at Rensselaer Polytechnic Institute and the paper by Jerome M. Shapiro at Aware, Inc. I would like to take this opportunity to thank all these three authors. Also thanks to Alistair Moffat of University of Melbourne, Mladen V. Wickerhauser of Washington University, Xuguang Yang of University of Illinois at Urbana-Champaign, Lucas Monzon of Yale University for their valuable help.

And thanks to all the members in my thesis committee. Their comments have enhanced the thesis considerably.

A final thanks goes to all the people caring about me and helping me out in this or that way.

Contents

Abstract	ii
Acknowledgments	iv
List of Illustrations	viii
List of Tables	ix
1 Introduction	1
2 General Theory of Wavelet Analysis	4
2.1 The Advent of the Wavelet Transform	4
2.2 Orthogonal Wavelet Systems	5
2.2.1 The Multiresolution Analysis	7
2.2.2 Daubechies' Work	9
2.2.3 Orthonormality	11
2.2.4 Smoothness	13
2.3 The Mallat Algorithm	13
2.4 Biorthogonal Wavelet Systems	14
3 Biorthogonal Coifman Wavelet Systems	17
3.1 A Wavelet Approximation Theorem	17
3.2 Definition of Biorthogonal Coifman Wavelet Systems	27
3.3 Construction of BCWs	34
3.4 Properties of BCWs	38
3.4.1 Approximation	38
3.4.2 Compact Support	40
3.4.3 Symmetry	41
3.4.4 Smoothness	42
3.4.5 Unconditional Bases	43
3.4.6 Multiplication-Free Discrete Wavelet Transform	45
3.4.7 Convergence to Sinc Wavelet System	49

3.5	Conclusions	53
4	Orthogonal Coifman Wavelet Systems	54
4.1	Defintion of Orthogonal Coifman Wavelet Systems	54
4.2	Vanishing Moments and Wavelet Approximation	55
4.3	Existence and Construction	57
4.4	Conclusions	61
5	Image Coding	62
5.1	Some Background	63
5.2	Shapiro's Embedded Zerotree Wavelet Algorithm	64
5.3	Said and Pearlman's Codetree Algorithm	65
5.4	Wavelet-Difference-Reduction Algorithm	67
5.4.1	Differential Coding	67
5.4.2	Binary Reduction	67
5.4.3	Coding the Significant Maps	68
5.4.4	Outline of the Algorithm	69
5.5	Comparisons of Three Algorithms	72
5.6	Applications of WDR Algorithm	73
5.6.1	Still Images	74
5.6.2	Medical Images	76
5.6.3	Synthetic Aperture Radar Images	76
5.7	Image Processing in the Compressed Wavelet Domain	77
5.7.1	Denoising	80
5.7.2	Speckle Reduction	81
5.7.3	Zooming	81
5.8	Image Coding Evaluation of BCWs	82
5.8.1	Image-Independent Measures	82
5.8.2	Image-Dependent Measures	84
5.9	Conclusions	85
6	Summary	86
	Bibliography	87

Illustrations

2.1	The Haar Wavelet Function	6
2.2	The Haar Scaling Function	7
3.1	The Biorthogonal Coifman Wavelet System of Degree 2	39
3.2	The Biorthogonal Coifman Wavelet System of Degree 3	39
3.3	The Biorthogonal Coifman Wavelet System of Degree 4	41
4.1	Orthogonal Coifman Wavelet Systems of Degree 1 and 2	60
4.2	Orthogonal Coifman Wavelet Systems of Degree 3 and 4	60
5.1	Parent-Child Dependencies in the Wavelet Decomposition Domain with 3 Scales.	65
5.2	Scanning Order of the Wavelet Transform Coefficients with 3 Scales.	66
5.3	Example of 3-scale Wavelet Transform of an 8×8 Image.	71
5.4	Illustration of ICompress	75
5.5	Coding Performance of the New Algorithm on “Lena”, Compared with SPC and EZW	77
5.6	Original 8 bpp, 512×512 , Grayscale “Lena”	78
5.7	8:1 Compression, PSNR = 39.84 dB	78
5.8	16:1 Compression, PSNR = 36.59 dB	79
5.9	32:1 Compression, PSNR = 33.44 dB	79

Tables

3.1	The Coefficients of Biorthogonal Coifman Wavelet Systems	40
3.2	Sobolev Smoothness of Biorthogonal Coifman Wavelet Systems	44
5.1	Comparison Using Image-Independent Measures	83
5.2	Comparison Using the Weighted SBC Gain	85

Chapter 1

Introduction

The theory of wavelet analysis has grown explosively in the last decade. The terminology “wavelet” was first introduced, in the context of a mathematical transform, in 1984 by A. Grossmann and J. Morlet [19]. In 1988, I. Daubechies, in her celebrated paper [8], introduced a class of compactly supported orthogonal wavelet systems in general, as well as a family with growing smoothness for large support, the Daubechies wavelet systems. In 1989, S. Mallat [32] presented the theory of multiresolution analysis and the Mallat algorithm. The spline family was introduced and studied by G. Battle [2], P. G. Lemarié [30], and C. K. Chui [4]. The necessary and sufficient conditions for an orthogonal wavelet system were given by A. Cohen [6] and W. Lawton [29]. Except for the Haar wavelet system, orthogonal wavelet systems can’t be symmetric, though symmetry is highly desired, for example, in the applications in signal processing, where symmetry corresponds to linear phase. To obtain symmetry and keep the property of perfect reconstruction, A. Cohen, I. Daubechies, and J.-C. Feauveau [7] replaced the orthogonality condition with biorthogonality and thus established the theory of biorthogonal wavelet systems. At the same time, lots of pioneer work has been done by many scientists from mathematics, physics and engineering. For more details of wavelet theory, we refer to [5], [9], [35] and [39].

Along with the rapid development of its theoretical aspects, wavelet analysis immediately found its application in mathematical modeling, neural networks, numerical analysis, and signal processing. Meanwhile, it is still keeping spreading its influence in other untouched areas. Since Fourier analysis has played a big role in science, it will not be a surprise that wavelet analysis will be on the stage of scientific research and applications for a long time, and will give us a better understanding of the world we are living.

In this thesis, we will focus on *biorthogonal Coifman wavelet systems*, a family of biorthogonal wavelet systems with very nice properties both in the theoretical sense and application sense. The original idea goes back to R. Coifman of Yale University. In the spring of 1989, he suggested that it might be worthwhile to construct orthogo-

nal wavelet systems with vanishing moments not only for the wavelet functions (which is the hypothesis posed on the Daubechies wavelet systems), but also for the scaling functions. This turned out to be a big success. One key property of these orthogonal wavelet systems with vanishing moments equally distributed between scaling functions and wavelet functions (which are called *orthogonal Coifman wavelet systems*) is that they have very nice approximation properties with exponential decay. This result was proved by us in 1993 and it is a natural extension of the result of R. O. Wells, Jr. and X. Zhou [56]. In 1994, we introduced *biorthogonal Coifman wavelet systems*, the biorthogonal counterparts of orthogonal Coifman wavelet systems. These biorthogonal Coifman wavelet systems also have fast approximation properties with exponential decay. Moreover they are symmetric, compactly supported, have growing smoothness, and converge to the *sinc wavelet system*. Another attractive feature of biorthogonal Coifman wavelet systems is that all the scaling vectors are dyadic rational, which means we can implement a multiplication-free discrete wavelet transform. In 1995, D. Wei et al. [55] did the coding performance evaluation of biorthogonal Coifman wavelet systems. It turns out that biorthogonal Coifman wavelet systems are very useful for image transform coding and seem to be quite comparable to the wavelet systems used in the state-of-the-art compression systems.

Perhaps the biggest success of wavelet applications has been claimed in signal processing, in particular, image coding. With good localization properties in both the spatial domain and the frequency domain, the wavelet transform can handle non-stationary signals pretty well. Current research on wavelet based image coding [1], [12], [31], [45], and [52], etc, has shown the high promise of this relatively new yet almost mature technology.

In the second part of the thesis, we propose a new image coding method. Unlike zerotree type schemes, such as J. Shapiro's embedded zerotree wavelet algorithm [45], and A. Said and W. A. Pearlman's codetree algorithm [43], all use spatial orientation tree structures to implicitly locate the significant wavelet transform coefficients, this new algorithm is a direct approach to find the positions of significant coefficients. It combines the discrete wavelet transform, differential coding, binary reduction, ordered bit plane transmission, and adaptive arithmetic coding. The encoding can be stopped at any point, which allows a target rate or distortion metric to be met exactly. The bits in the bit stream are generated in the order of importance, yielding a fully embedded code to successively approximate the original image source; thus it's well suited for progressive image transmission. The decoder can also terminate the decoding at any

point, and produce a lower (bit) rate reconstruction image. Our algorithm is very simple in its form (which will make the encoding and decoding very fast), requires no training of any kind or prior knowledge of image sources, and has a clear geometric structure which enables us to process the image data in the compressed wavelet domain. The image coding results are quite competitive with almost all previous reported image compression algorithms (including [45] and [43]) on standard test images.

This thesis is organized as follows. In Chapter 2 we give an overview of the general theory of wavelet analysis. This provides the underneath background where biorthogonal Coifman wavelet systems reside in and also it is an attempt to make the thesis self-contained. Chapter 3 studies biorthogonal Coifman wavelet systems. The definition, construction, and properties will be discussed in this chapter. The existence and construction of orthogonal Coifman wavelet systems will be studied in Chapter 4. The image coding algorithm, the *Wavelet-Difference-Reduction* algorithm, is presented in Chapter 5. We will compare its coding performance with other algorithms and evaluate it in various applications. We conclude the thesis in Chapter 6.

Some of the work in the thesis was announced and developed in [48], [38], [55], [50], and [49].

Chapter 2

General Theory of Wavelet Analysis

The first orthogonal wavelet system, the Haar wavelet system, was constructed by A. Haar [21] in 1910. The Haar wavelet system consists of piecewise constant functions and provides an orthonormal bases of $L^2(\mathbf{R})$. Seventy years later, A. Grossmann and J. Morlet introduced the notion “wavelet transform” [19] in 1984. A big breakthrough of wavelet analysis was brought by I. Daubechies in 1988. In her classical paper [8], she introduced a class of compactly supported orthogonal wavelet systems in general, as wells as a family with growing smoothness for large support, the Daubechies wavelet systems. Her work immediately stimulated a rapid development in the theory and applications of wavelet analysis. In 1989, S. Mallat presented the theory of multiresolution analysis [32]. With the multiresolution analysis, we can now construct the wavelet system with desired property. The discrete wavelet transform can be computed by the Mallat algorithm [32]. Usually the wavelet system constitutes a frame of $L^2(\mathbf{R})$. To obtain an orthogonal system, it has to satisfy the orthogonality conditions given by A. Cohen [6] and W. Lawton [29]. Later A. Cohen, I. Daubechies, and J.-C. Feauveau [7] established the theory of biorthogonal wavelet systems. The main advantage of biorthogonal wavelet systems over orthogonal ones is that biorthogonal systems can be symmetric, while orthogonal systems can’t except the Haar system. On the other hand, orthogonal systems preserve the L^2 norm, biorthogonal systems don’t. So there are tradeoffs between using biorthogonal systems and using orthogonal systems, depending on the specific applications.

In the chapter, we will go through some basic theory of wavelet analysis. For more details we refer to the original papers and [5], [9], [35], and [39].

2.1 The Advent of the Wavelet Transform

The wavelet transform was first introduced by A. Grossmann and J. Morlet [19] in 1984. It is a tool that cuts up data, functions or operators into different frequency components, and then studies each component with a resolution matched to its scale.

Similar to the Fourier transform, there are continuous wavelet transform (CWT) and discrete wavelet transform (DWT).

Definition 2.1 The *continuous wavelet transform* (CWT) of a function $f(x) \in L^2(\mathbf{R})$ with respect to $\psi(x) \in L^2(\mathbf{R})$ is

$$CWT(f)(a, b) = |a|^{1/2} \int_{-\infty}^{\infty} f(x)\psi(a(x - b)) dx ,$$

where $a, b \in \mathbf{R}, a \neq 0$.

Definition 2.2 The *discrete wavelet transform* (DWT) of a function $f(x) \in L^2(\mathbf{R})$ with respect to $\psi(x) \in L^2(\mathbf{R})$ is

$$DWT(f)(j, k) = |a_0|^{j/2} \int_{-\infty}^{\infty} f(x)\psi(a_0^j x - kb_0) dx ,$$

where $a_0, b_0 \in \mathbf{R}, a_0 \neq 0, j, k \in \mathbf{Z}$.

For the discrete wavelet transform, a_0 and b_0 are usually set to be 2 and 1. In such a case, the DWT becomes

$$DWT(f)(j, k) = 2^{j/2} \int_{-\infty}^{\infty} f(x)\psi(2^j x - k) dx .$$

2.2 Orthogonal Wavelet Systems

The Haar wavelet function is a piecewise constant function (see Figure 2.1)

$$\psi^{Haar}(x) = \begin{cases} 1, & 0 \leq x < \frac{1}{2}, \\ -1, & \frac{1}{2} \leq x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

The family of functions $\psi^{Haar}_{j,k}(x)$,

$$\psi^{Haar}_{j,k}(x) = 2^{j/2} \psi^{Haar}(2^j x - k) = \begin{cases} 1, & \frac{k}{2^j} \leq x < \frac{k+1/2}{2^j}, \\ -1, & \frac{k+1/2}{2^j} \leq x < \frac{k+1}{2^j}, \\ 0, & \text{otherwise,} \end{cases}$$

generated from $\psi^{Haar}(x)$ by the operation of dilations and translations, constitute an orthonormal basis of $L^2(\mathbf{R})$, where the expansion coefficients are exactly the discrete

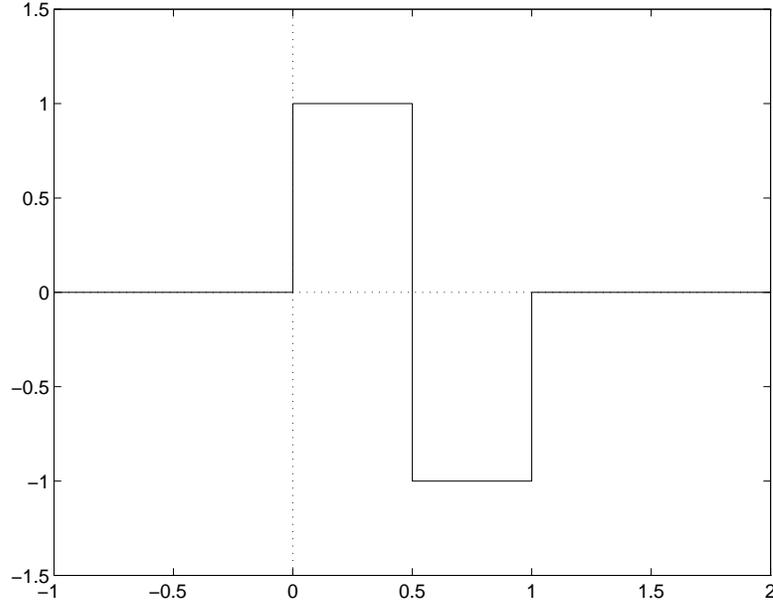


Figure 2.1 The Haar Wavelet Function

wavelet transform with respect to $\psi^{Haar}(x)$. A critical insight is to look at another piecewise constant function, the so-called Haar scaling function (see Figure 2.2)

$$\phi^{Haar}(x) = \begin{cases} 1, & 0 \leq x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

The linear subspaces V_j spanned by $\{\phi^{Haar}_{j,k}, k \in \mathbf{Z}\}$, where

$$\phi^{Haar}_{j,k}(x) = 2^{j/2} \phi^{Haar}(2^j x - k) = \begin{cases} 1, & \frac{k}{2^j} \leq x < \frac{k+1}{2^j}, \\ 0, & \text{otherwise,} \end{cases}$$

are a sequence of nested subspaces of $L^2(\mathbf{R})$,

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \cdots,$$

and their union is a dense subset of $L^2(\mathbf{R})$,

$$\overline{\bigcup_{j \in \mathbf{Z}} V_j} = L^2(\mathbf{R}).$$

If W_j are the linear subspace spanned by $\{\psi^{Haar}_{j,k}, k \in \mathbf{Z}\}$, the fact is that W_j is the orthogonal complement of V_j in V_{j+1} . Thus it follows that

$$L^2(\mathbf{R}) = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus W_{j_0+2} \oplus \cdots,$$

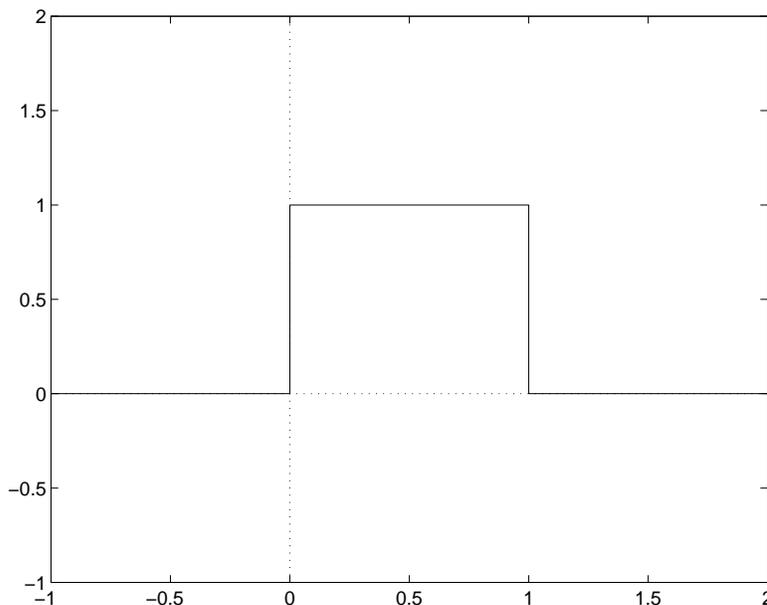


Figure 2.2 The Haar Scaling Function

where j_0 is some integer. So the collection of functions $\{\phi^{Haar}_{j_0,k}, k \in \mathbf{Z}\}$ and $\{\psi^{Haar}_{j,k}, k \in \mathbf{Z}, j \in \mathbf{Z}, j \geq j_0\}$, is also an orthonormal basis of $L^2(\mathbf{R})$, and the expansion coefficients are exactly the discrete wavelet transform with respect to ϕ^{Haar} and ψ^{Haar} .

The Haar wavelet system explained above illustrates the core idea of the multiresolution analysis, which is the starting point of the wavelet analysis.

2.2.1 The Multiresolution Analysis

In this subsection we will state the main results of the multiresolution analysis. For all the proofs and more details, we refer to [34], [32], and [24].

A multiresolution analysis consists of a sequence of nested closed subspaces $V_j \subset L^2(\mathbf{R})$, $j \in \mathbf{Z}$,

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \cdots,$$

such that

$$\overline{\bigcup_{j \in \mathbf{Z}} V_j} = L^2(\mathbf{R}),$$

and

$$\bigcap_{j \in \mathbf{Z}} V_j = \{0\}.$$

There must exist a L^2 function $\phi(x) \in V_0$ so that

$$\{\phi_{0,k}(x), k \in \mathbf{Z}\} \text{ is an orthonormal basis in } V_0,$$

where $\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k)$. The last requirement for a multiresolution analysis is that

$$f(\cdot) \in V_0 \iff f(2^j \cdot) \in V_j.$$

In a multiresolution analysis, since

$$\phi(x) \in V_0 \subset V_1,$$

and $\{\phi_{1,n}(x), n \in \mathbf{Z}\}$ is an orthonormal basis in V_1 , there exists $\{a_k, k \in \mathbf{Z}\}$ such that

$$\phi(x) = \sum_{k \in \mathbf{Z}} a_k \phi(2x - k).$$

Define a function $\psi(x)$

$$\psi(x) = \sum_{k \in \mathbf{Z}} (-1)^k a_{-k+1} \phi(2x - k),$$

and assume W_j is the orthogonal complement of V_j in V_{j+1} , then the fundamental result is that

$$\{\psi_{j,k}(x), k \in \mathbf{Z}\} \text{ is an orthonormal basis in } W_j,$$

where $\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$. Thus the family $\{\psi_{j,k}(x), j, k \in \mathbf{Z}\}$ constitute an orthonormal basis of $L^2(\mathbf{R})$, and the expansion coefficients are exactly the discrete wavelet transform with respect to $\psi(x)$.

We call $\phi(x)$, $\psi(x)$, and $\{a_k\}$ the *scaling function*, the *wavelet function*, and the *scaling vector*, respectively. Sometimes we will also use the following two vectors $\{h_k, k \in \mathbf{Z}\}$ and $\{g_k, k \in \mathbf{Z}\}$,

$$h_k = 2^{-1/2} a_k, \quad g_k = (-1)^k 2^{-1/2} a_{-k+1}.$$

With these two vectors, we have

$$\phi(x) = \sum_{k \in \mathbf{Z}} h_k \phi_{1,k}(x),$$

and

$$\psi(x) = \sum_{k \in \mathbf{Z}} g_k \phi_{1,k}(x).$$

We call $\{h_k, k \in \mathbf{Z}\}$, $\{g_k, k \in \mathbf{Z}\}$ the *scaling vector* and the *wavelet vector*, respectively. Note that the difference between $\{a_k, k \in \mathbf{Z}\}$ and $\{h_k, k \in \mathbf{Z}\}$ is that they have different normalization.

2.2.2 Daubechies' Work

I. Daubechies [8] constructed a class of compactly supported orthogonal wavelet systems in general, as well as a family with growing smoothness for large support, the Daubechies wavelet system. In this subsection, we will review in brief her method to construct orthonormal bases of compactly supported wavelets. For more details, we refer to [8] and [9].

Let's start with the scaling vector $\{a_k, k \in \mathbf{Z}\}$. As we know, the scaling function $\phi(x)$ satisfies the 2-scale difference equation

$$\phi(x) = \sum_{k \in \mathbf{Z}} a_k \phi(2x - k).$$

Taking the Fourier transform, we have

$$\hat{\phi}(\xi) = m_0(\xi/2) \hat{\phi}(\xi/2), \quad (2.1)$$

where

$$m_0(\xi) = \frac{1}{2} \sum_{k \in \mathbf{Z}} a_k e^{-ik\xi}.$$

Studying on the equation (2.1), Daubechies found that in order to obtain an orthonormal wavelet system, the $\{a_k\}$ must satisfy the linear condition

$$\sum_{k \in \mathbf{Z}} a_k = 2. \quad (2.2)$$

and the quadratic condition

$$\sum_{k \in \mathbf{Z}} a_k a_{k+2l} = 2\delta_{0,l} = \begin{cases} 2, & l = 0, \\ 0, & \text{otherwise.} \end{cases} \quad (2.3)$$

Yet these two equations (2.2) and (2.3) are not sufficient to imply an orthonormal wavelet system.

At the same time, we are mainly interested in the case when the scaling vector $\{a_k\}$ has finite length, i.e., there exists a positive number K , such that $a_k = 0$, when $|k| > K$, because it guarantees the existence of the scaling function $\phi(x)$ due to the lemma by Deslauriers and Dubuc [11].

Lemma 2.1 If $m(\xi) = \sum_{k=N_1}^{N_2} a_k e^{-ik\xi}$, with $\sum_{k=N_1}^{N_2} a_k = 1$, then the infinite product $\prod_{j=1}^{\infty} m(2^{-j}\xi)$ is an entire function of exponential type. In particular, it is the Fourier transform of a distribution with support in $[N_1, N_2]$.

With the above conditions on $\{a_k\}$, W. Lawton [28] prove that $\psi(x)$ will generate a tight frame. We combine all these result as the following single theorem.

Theorem 2.1 Let $\{a_k, k \in \mathbf{Z}\}$ be a sequence having finite length, $a_k = 0$ when $|k| > K$ for some number K . Assume $\{a_k\}$ satisfy

$$\sum_{k \in \mathbf{Z}} a_k = 2,$$

$$\sum_{k \in \mathbf{Z}} a_k a_{k+2l} = 2\delta_{0,l}.$$

Define

$$m_0(\xi) = \frac{1}{2} \sum_{k \in \mathbf{Z}} a_k e^{-ik\xi},$$

then the infinite product $\prod_{j=1}^{\infty} m_0(2^{-j}\xi)$ is the Fourier transform of an L^2 function with compact support, i.e., set

$$\hat{\phi}(\xi) = (2\pi)^{-1/2} \prod_{j=1}^{\infty} m_0(2^{-j}\xi),$$

then $\phi(x) \in L^2(\mathbf{R})$ and $\phi(x)$ has compact support. In addition, define

$$\psi(x) = \sum_{k \in \mathbf{Z}} (-1)^k a_{-k+1} \phi(2x - k),$$

then $\psi(x)$ is also compactly supported and its dilations and translations constitute a tight frame for $L^2(\mathbf{R})$, i.e., for all $f(x) \in L^2(\mathbf{R})$,

$$\sum_{j,k \in \mathbf{R}} |\langle f, \psi_{j,k} \rangle|^2 = \|f\|^2,$$

where $\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$. Thus the L^2 norm of $f(x)$ is preserved in the discrete wavelet transform with respect to $\psi(x)$.

2.2.3 Orthonormality

As we have seen, the linear condition (2.2) and the quadratic condition (2.3) are not sufficient to produce an orthogonal system. It can be shown that the wavelet function $\psi(x)$ in Theorem 2.1 provides an orthogonal system, i.e., $\psi_{j,k}$ constitute an orthonormal basis of $L^2(\mathbf{R})$, if and only if

$$\int_{-\infty}^{\infty} \phi(x)\phi(x-k) dx = \delta_{0,k}, \quad \forall k \in \mathbf{Z}.$$

The first necessary and sufficient condition for the above equation was identified by A. Cohen [6]. He introduced the concept “congruent set” and used this concept to give an equivalent condition of orthogonality.

Definition 2.3 A compact set K is called *congruent* to $[-\pi, \pi]$ modulo 2π if

1. The Lebesgue measure $|K| = 2\pi$;
2. For all $\xi \in [-\pi, \pi]$, there exists $l \in \mathbf{Z}$ so that $\xi + 2l\pi \in K$.

We are now ready to state Cohen’s theorem.

Theorem 2.2 Assume all the conditions of Theorem 2.1. Then the following three conditions are equivalent:

- 1.

$$\int_{-\infty}^{\infty} \phi(x)\phi(x-k) dx = \delta_{0,k}, \quad \forall k \in \mathbf{Z}.$$

2. There exists a compact set K congruent to $[-\pi, \pi]$ modulo 2π and containing a neighborhood of 0 so that

$$\inf_{k>0} \inf_{\xi \in K} |m_0(2^{-k}\xi)| > 0.$$

3. There is no non-trivial cycle $\{\xi_1, \dots, \xi_n\}$ for the operation $\xi \mapsto 2\xi \pmod{2\pi}$ such that $|m_0(\xi_j)| = 1$ for all $j = 1, \dots, n$.

The following corollary is quite useful in practice.

Corollary 2.1 Assume all the conditions of Theorem 2.1. If m_0 has no zeros in $[-\pi/3, \pi/3]$, then

$$\int_{-\infty}^{\infty} \phi(x)\phi(x-k) dx = \delta_{0,k}, \quad \forall k \in \mathbf{Z}.$$

Another equivalent condition of orthogonality was given by W. Lawton [29]. It is a very simple criterion based on the multiresolution matrix.

Theorem 2.3 Assume all the conditions of Theorem 2.1. Define a multiresolution matrix T by

$$t_{l,n} = \frac{1}{2} \sum_k a_k a_{n-2l+k}.$$

If T has 1 as a nondegenerate eigenvalue, then

$$\int_{-\infty}^{\infty} \phi(x)\phi(x-k) dx = \delta_{0,k}, \quad \forall k \in \mathbf{Z}.$$

Note that since only finite a_k are nonzero, the multiresolution matrix T has only finite nonzero entries. Thus we can check only a small submatrix, which contains all the nonzero entries of T .

Combining Theorem 2.1 and Theorem 2.3, we can construct an orthogonal wavelet system from the scaling vector $\{a_k, k \in \mathbf{Z}\}$.

Theorem 2.4 Let $\{a_k, k \in \mathbf{Z}\}$ be a sequence having finite length, $a_k = 0$ when $|k| > K$ for some number K . Assume $\{a_k\}$ satisfy

$$\begin{aligned} \sum_{k \in \mathbf{Z}} a_k &= 2, \\ \sum_{k \in \mathbf{Z}} a_k a_{k+2l} &= 2\delta_{0,l}. \end{aligned}$$

Define

$$\begin{aligned} m_0(\xi) &= \frac{1}{2} \sum_{k \in \mathbf{Z}} a_k e^{-ik\xi}, \\ \hat{\phi}(\xi) &= (2\pi)^{-1/2} \prod_{j=1}^{\infty} m_0(2^{-j}\xi), \\ \psi(x) &= \sum_{k \in \mathbf{Z}} (-1)^k a_{-k+1} \phi(2x-k). \end{aligned}$$

If the multiresolution matrix T

$$t_{l,n} = \frac{1}{2} \sum_k a_k a_{n-2l+k}.$$

has 1 as a nondegenerate eigenvalue, then the $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ define a multiresolution analysis, and the $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ are the associated orthonormal wavelet basis, where the expansion coefficients are exactly the discrete wavelet transform with respect to $\psi(x)$.

2.2.4 Smoothness

T. Eirola [14] developed a method to calculate the Sobolev smoothness of wavelet systems. The equation he worked on is actually the two-scale difference equation

$$\phi(x) = \sum_k a_k \phi(2x - k),$$

no matter how the $\phi(x)$ will be the scaling function of a wavelet system or not. Thus his result is quite general, no restricted to the wavelet analysis.

The main idea is to estimate the multiresolution operator. For more details, we refer to [14] and [22].

2.3 The Mallat Algorithm

The Mallat algorithm [32] tells us the relation between the discrete wavelet transform of different scales. Assume that

$$\begin{aligned} c_{j+1,k} &= \int_{-\infty}^{\infty} f(x) 2^{(j+1)/2} \phi(2^{j+1}x - k) dx, \\ c_{j,k} &= \int_{-\infty}^{\infty} f(x) 2^{j/2} \phi(2^j x - k) dx, \\ d_{j,k} &= \int_{-\infty}^{\infty} f(x) 2^{j/2} \psi(2^j x - k) dx. \end{aligned}$$

Then we have

$$\begin{aligned} c_{j,k} &= \int_{-\infty}^{\infty} f(x) 2^{j/2} \phi(2^j x - k) dx \\ &= \int_{-\infty}^{\infty} f(x) 2^{j/2} \left(\sum_{n \in \mathbf{Z}} a_n \phi(2^{j+1}x - 2k - n) \right) dx \\ &= \sum_{n \in \mathbf{Z}} a_n \int_{-\infty}^{\infty} f(x) 2^{j/2} \phi(2^{j+1}x - 2k - n) dx \\ &= \sum_{n \in \mathbf{Z}} a_n 2^{-1/2} c_{j+1, 2k+n} \\ &= \sum_{n \in \mathbf{Z}} h_n c_{j+1, 2k+n} \\ &= \sum_{n \in \mathbf{Z}} h_{n-2k} c_{j+1, n} \end{aligned}$$

Similarly

$$\begin{aligned}
d_{j,k} &= \int_{-\infty}^{\infty} f(x)2^{j/2}\psi(2^jx - k) dx \\
&= \int_{-\infty}^{\infty} f(x)2^{j/2} \left(\sum_{n \in \mathbf{Z}} (-1)^n a_{-n+1} \phi(2^{j+1}x - 2k - n) \right) dx \\
&= \sum_{n \in \mathbf{Z}} (-1)^n a_{-n+1} \int_{-\infty}^{\infty} f(x)2^{j/2} \phi(2^{j+1}x - 2k - n) dx \\
&= \sum_{n \in \mathbf{Z}} (-1)^n a_{-n+1} 2^{-1/2} c_{j+1,2k+n} \\
&= \sum_{n \in \mathbf{Z}} g_n c_{j+1,2k+n} \\
&= \sum_{n \in \mathbf{Z}} g_{n-2k} c_{j+1,n}
\end{aligned}$$

Thus we can compute the discrete wavelet transform coefficients at level j from the discrete wavelet transform coefficients at level $j + 1$ through the following Mallat algorithm,

$$c_{j,k} = \sum_{n \in \mathbf{Z}} h_{n-2k} c_{j+1,n} ,$$

and

$$d_{j,k} = \sum_{n \in \mathbf{Z}} g_{n-2k} c_{j+1,n} .$$

It can be shown that in an orthogonal wavelet system, the following reconstruction formula holds,

$$c_{j+1,k} = \sum_{n \in \mathbf{Z}} (h_{2k-n} c_{j,n} + g_{2k-n} d_{j,n}) .$$

2.4 Biorthogonal Wavelet Systems

In a biorthogonal wavelet system, the decomposition function and the reconstruction function can be different. Thus we will have an expansion which looks like

$$f(x) = \lim_{J \rightarrow \infty} \sum_{j=-J}^J \sum_k < f, \psi_{j,k} > \tilde{\psi}_{j,k} .$$

In 1992, A. Cohen, I. Daubechies, and J.-C. Feauveau [7] established the theory of biorthogonal wavelet systems. Basically we have two scaling vectors, the analysis

scaling vector $\{a_k, k \in \mathbf{Z}\}$ and the synthesis scaling vector $\{\tilde{a}_k, k \in \mathbf{Z}\}$. They satisfy the linear conditions

$$\sum_{k \in \mathbf{Z}} a_k = 2, \quad (2.4)$$

$$\sum_{k \in \mathbf{Z}} \tilde{a}_k = 2, \quad (2.5)$$

and the bilinear condition

$$\sum_{k \in \mathbf{Z}} a_k \tilde{a}_{k+2l} = 2\delta_{0,l}, \quad \forall l \in \mathbf{Z}. \quad (2.6)$$

With $\{a_k\}$, $\{\tilde{a}_k\}$ satisfying (2.4), (2.5), and (2.6), assume both $\{a_k\}$ and $\{\tilde{a}_k\}$ have finite length, we can define two compactly supported functions $\phi(x)$ and $\tilde{\phi}(x)$ by their Fourier transform,

$$\hat{\phi}(\xi) = (2\pi)^{-1/2} \prod_{j=1}^{\infty} m_0(2^{-j}\xi),$$

$$\hat{\tilde{\phi}}(\xi) = (2\pi)^{-1/2} \prod_{j=1}^{\infty} \tilde{m}_0(2^{-j}\xi),$$

where

$$m_0(\xi) = \frac{1}{2} \sum_{k \in \mathbf{Z}} a_k e^{-ik\xi},$$

$$\tilde{m}_0(\xi) = \frac{1}{2} \sum_{k \in \mathbf{Z}} \tilde{a}_k e^{-ik\xi}.$$

We also define two compactly supported functions $\psi(x)$ and $\tilde{\psi}(x)$ by

$$\psi(x) = \sum_{k \in \mathbf{Z}} (-1)^k \tilde{a}_{-k+1} \phi(2x - k),$$

$$\tilde{\psi}(x) = \sum_{k \in \mathbf{Z}} (-1)^k a_{-k+1} \tilde{\phi}(2x - k).$$

For $\{a_k\}$ and $\{\tilde{a}_k\}$ satisfying (2.4), (2.5), (2.6) and having finite length, the $\psi_{j,k}$ and $\tilde{\psi}_{j,k}$ constitute a weak dual frame of $L^2(\mathbf{R})$, i.e., for any $f(x), g(x) \in L^2(\mathbf{R})$,

$$\lim_{J \rightarrow \infty} \sum_{j=-J}^J \sum_k \langle f, \psi_{j,k} \rangle \langle \tilde{\psi}_{j,k}, g \rangle = \langle f, g \rangle,$$

or, in a weak sense,

$$f(x) = \lim_{J \rightarrow \infty} \sum_{j=-J}^J \sum_k \langle f, \psi_{j,k} \rangle \tilde{\psi}_{j,k}, \quad (2.7)$$

where $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, $\tilde{\psi}_{j,k}(x) = 2^{j/2}\tilde{\psi}(2^j x - k)$. And we call $\phi(x)$, $\psi(x)$, $\tilde{\phi}(x)$, $\tilde{\psi}(x)$ the analysis scaling function, the analysis wavelet function, the synthesis scaling function and the synthesis wavelet function, respectively.

If

$$|\hat{\phi}(\xi)| \leq C(1 + |\xi|)^{-1/2-\epsilon}, \quad |\hat{\tilde{\phi}}(\xi)| \leq C(1 + |\xi|)^{-1/2-\epsilon},$$

for some constant C , then the limit in (2.7) converges strongly in $L^2(\mathbf{R})$.

In a biorthogonal system, we can't get the orthogonality, but we can impose some conditions to establish a dual Riesz bases. The sufficient and necessary condition of a dual Riesz bases for the scaling vectors $\{a_k\}$ and $\{\tilde{a}_k\}$ satisfying (2.4), (2.5), (2.6) and having finite length is similar to the orthogonal cases, which states

$$\int_{-\infty}^{\infty} \phi(x)\tilde{\phi}(x - k) dx = \delta_{0,k}, \quad \forall k \in \mathbf{Z}.$$

Chapter 3

Biorthogonal Coifman Wavelet Systems

Based on the general theory of wavelet analysis, we will present the mathematical theory of biorthogonal Coifman wavelet systems in this chapter. It all begins with a wavelet approximation theorem, which is valid for both biorthogonal and orthogonal Coifman wavelet systems. The definition of biorthogonal Coifman wavelet systems follows right after it and the problem of existence is considered next. Here we introduce a time domain design method which is very straight forward. Thus the existence of biorthogonal Coifman wavelet systems is proved by direct construction for all degrees. Beside the approximation property, biorthogonal Coifman wavelet systems are symmetric, have compact support and growing smoothness with large degrees, and converge to the sinc wavelet system. An attractive feature of biorthogonal Coifman wavelet systems is that all the scaling vectors are dyadic rational, which means we can have a very fast multiplication-free discrete wavelet transform implemented on digital computers.

3.1 A Wavelet Approximation Theorem

The Mallat Algorithm tells us how to compute the discrete wavelet transform coefficients from one level to the next finer level. Namely, if $\{c_{j+1,k}, k \in \mathbf{Z}\}$ are the coefficients at the $(j + 1)$ -th level, $\{c_{j,k}, k \in \mathbf{Z}\}$ and $\{d_{j,k}, k \in \mathbf{Z}\}$ are the coefficients at the j -th level,

$$c_{j+1,k} = \int_{-\infty}^{\infty} f(x) 2^{(j+1)/2} \phi(2^{j+1}x - k) dx ,$$

$$c_{j,k} = \int_{-\infty}^{\infty} f(x) 2^{j/2} \phi(2^j x - k) dx ,$$

$$d_{j,k} = \int_{-\infty}^{\infty} f(x) 2^{j/2} \psi(2^j x - k) dx ,$$

then

$$c_{j,k} = \sum_n h_{n-2k} c_{j+1,n} ,$$

$$d_{j,k} = \sum_n g_{n-2k} c_{j+1,n},$$

where $\{h_n\}$ and $\{g_n\}$ are the scaling vector and the wavelet vector, respectively. And we can further decompose $\{c_{j,k}, k \in \mathbf{Z}\}$ into $\{c_{j-1,k}, k \in \mathbf{Z}\}$ and $\{d_{j-1,k}, k \in \mathbf{Z}\}$, decompose $\{c_{j-1,k}, k \in \mathbf{Z}\}$ into $\{c_{j-2,k}, k \in \mathbf{Z}\}$ and $\{d_{j-2,k}, k \in \mathbf{Z}\}$, and so on. This leaves one problem, that is, how to get the coefficients at the starting level, which are those $\{c_{j+1,k}, k \in \mathbf{Z}\}$ in the above decomposition. Without knowing the coefficients at the starting level, there is no meaning to talk about the Mallat Algorithm. Thanks to the following theorem, proved by R. O. Wells, Jr. and X. Zhou [56] in 1991, we know how to solve this problem easily. This theorem is stated in \mathbf{R}^2 for simplicity, but it is true in \mathbf{R}^n as well. The proof can be found in [56].

Theorem 3.1 Assume $\phi(x)$ to be the scaling function of an orthonormal wavelet system with a finite length scaling vector $\{a_k, k \in \mathbf{Z}\}$, $a_k = 0$ when $|k| > K$ for some positive integer K ,

$$\phi(x) = \sum_{k \in \mathbf{Z}} a_k \phi(2x - k). \quad (3.1)$$

We define a constant c by

$$c := \frac{1}{2} \sum_{k \in \mathbf{Z}} k a_k. \quad (3.2)$$

Assume the function $f(x, y) \in C^2(\bar{\Omega})$, where Ω is a bounded open set in \mathbf{R}^2 . Let, for $j \in \mathbf{N}$,

$$S^j(f)(x, y) := \frac{1}{2^j} \sum_{(k,l) \in \Lambda} f\left(\frac{k+c}{2^j}, \frac{l+c}{2^j}\right) \phi_{j,k}(x) \phi_{j,l}(y), \quad (x, y) \in \Omega,$$

where the index set $\Lambda = \{(k, l) \in \mathbf{Z}^2 : (\text{supp}(\phi_{j,k}(x)) \times \text{supp}(\phi_{j,l}(y))) \cap \Omega \neq \emptyset\}$, and $\phi_{j,i}(x) = 2^{j/2} \phi(2^j x - i)$. Then

$$\|f(x, y) - S^j(f)(x, y)\|_{L^2(\Omega)} \leq C(1/2^j)^2,$$

and if $\phi(x) \in C^1(\mathbf{R})$,

$$\|f(x, y) - S^j(f)(x, y)\|_{H^1(\Omega)} \leq C/2^j,$$

where C is a constant depending only on K , the diameter of Ω , and the maximum modulus of the first and second order derivatives of $f(x, y)$ on $\bar{\Omega}$.

Theorem 3.1 provides a second order wavelet approximation result for L^2 functions. Sample values of a sufficient smooth (more precisely, C^2) function can be used as the wavelet transform coefficients and the corresponding wavelet approximation function $S^j(f)$ converges in Sobolev norms of first order to the original function. So in practice, we can take the sample values as the discrete wavelet transform coefficients at the starting level and apply the Mallat Algorithm on them. Also from Theorem 3.1, we know that the finer the starting level is, the closer the approximation is. Thus we will always take sample values at the finest level whenever possible.

In Theorem 3.1, sample values are taken on the points $\left(\frac{k+c}{2^j}\right)$, a translate of constant c from the dyadic rational $\frac{k}{2^j}$. Let's have a close look at the constant c . It is defined in (3.2) and actually it is closely related to the zero moment and the first moment of the scaling function $\phi(x)$.

Lemma 3.1 Assume c to be a constant as defined in (3.2), where $\{a_k\}$ is a finite length sequence satisfying the two-scale difference equation (3.1) and

$$\sum_{k \in \mathbf{Z}} a_k = 2.$$

Then

$$c \int_{-\infty}^{\infty} \phi(x) dx = \int_{-\infty}^{\infty} x \phi(x) dx. \quad (3.3)$$

Proof Using (3.1), we have

$$\begin{aligned} \int_{-\infty}^{\infty} x \phi(x) dx &= \int_{-\infty}^{\infty} x \left(\sum_k a_k \phi(2x - k) \right) dx \\ &= \sum_k a_k \int_{-\infty}^{\infty} x \phi(2x - k) dx \\ &= \frac{1}{4} \sum_k a_k \int_{-\infty}^{\infty} (x + k) \phi(x) dx \\ &= \frac{1}{4} \sum_k a_k \int_{-\infty}^{\infty} x \phi(x) dx + \frac{1}{4} \sum_k k a_k \int_{-\infty}^{\infty} \phi(x) dx \\ &= \frac{1}{2} \int_{-\infty}^{\infty} x \phi(x) dx + \frac{1}{2} c \int_{-\infty}^{\infty} \phi(x) dx \end{aligned}$$

Thus,

$$c \int_{-\infty}^{\infty} \phi(x) dx = \int_{-\infty}^{\infty} x \phi(x) dx.$$

□

The integrals on the left hand side and the right hand side of (3.3) are the zero moment and the first moment of $\phi(x)$, respectively. It follows from Lemma 3.1 that if the first moment of $\phi(x)$ equals zero, and the zero moment is not equal to zero, then c must be zero. In a wavelet system, the zero moment of the scaling function is always one. In this case, c is exactly the first moment of $\phi(x)$. When the first moment is zero, we may sample on dyadic rationals to get a good approximation. It turns out the moment values of $\phi(x)$ plays an important role in the wavelet approximation, not just the first moment. The wavelet approximation theorem tells us how we can impose more vanishing moments on the scaling function $\phi(x)$ to produce better approximation result.

Theorem 3.2 (*Wavelet Approximation Theorem*) Suppose $\phi(x)$ is an $L^2(\mathbf{R})$ solution of the two-scale difference equation (3.1), where $\{a_k\}$ is a finite length sequence satisfying the vanishing moment conditions up to degree N , i.e.,

$$\sum_{k \in \mathbf{Z}} (2k)^p a_{2k} = \sum_{k \in \mathbf{Z}} (2k+1)^p a_{2k+1} = 0, \quad \text{for } p = 1, \dots, N, \quad (3.4)$$

$$\sum_{k \in \mathbf{Z}} a_{2k} = \sum_{k \in \mathbf{Z}} a_{2k+1} = 1. \quad (3.5)$$

For a function $f(x) \in C_0^{N+1}(\mathbf{R})$, define

$$S^j(f)(x) := 2^{-j/2} \sum_{k \in \mathbf{Z}} f\left(\frac{k}{2^j}\right) \phi_{j,k}(x), \quad (3.6)$$

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

$$\|f(x) - S^j(f)(x)\|_{L^2} \leq C 2^{-j(N+1)}, \quad (3.7)$$

where C depends only on f and the sequence $\{a_k\}$.

The conditions (3.4) and (3.5) are equivalent to the vanishing moments of the scaling function $\phi(x)$ and the wavelet function $\psi(x)$, as we will see in Section 3.2. From (3.7) it follows that with more vanishing moments on $\phi(x)$ and $\psi(x)$, the convergence rate will be improved with an exponential decay.

The proof of Theorem 3.2 is based on the following lemma.

Lemma 3.2 Assume $\{a_k\}$ and $\phi(x)$ satisfy the same conditions as in Theorem 3.2, then

$$\sum_{k \in \mathbf{Z}} (x-k)^p \phi(x-k) = \delta_{0,p} = \begin{cases} 1 & \text{if } p = 0 \\ 0 & \text{else} \end{cases}, \quad \text{for } p = 0, \dots, N. \quad (3.8)$$

Lemma 3.2 will be proved in Section 3.2. The equality (3.8) holds in L^2 , thus, it is true for almost all $x \in \mathbf{R}$, but not for every $x \in \mathbf{R}$. If $\phi(x)$ is continuous, then (3.8) is valid for every $x \in \mathbf{R}$.

Based on Lemma 3.2, we are now ready to prove the wavelet approximation theorem.

Proof of Theorem 3.2 Using the Taylor expansion of f at the point x ,

$$f\left(\frac{k}{2^j}\right) = \sum_{p=0}^{N-1} \left(\frac{1}{p!} f^{(p)}(x) \left(\frac{k}{2^j} - x\right)^p \right) + \frac{1}{N!} f^{(N)}(\theta_k) \left(\frac{k}{2^j} - x\right)^N,$$

for some θ_k on the line segment connecting x and $\frac{k}{2^j}$.

From Lemma 3.2, for $1 \leq p \leq N$,

$$\sum_{k \in \mathbf{Z}} \left(\frac{k}{2^j} - x\right)^p \phi_{j,k}(x) = 2^{j/2-jp} \sum_{k \in \mathbf{Z}} (k - 2^j x)^p \phi(2^j x - k) = 0,$$

and

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

Since both $f(x)$ and $\phi(x)$ have compact support, we can find a positive number A , such that $\text{supp}(f) \subset [-A, A]$, and $\text{supp}(\phi) \subset [-A, A]$. Then

$$\sum_{|k| \leq (2^j+1)A} \left(\frac{k}{2^j} - x\right)^p f^{(p)}(x) \phi_{j,k}(x) = 0, \quad \text{for } 1 \leq p \leq N,$$

$$\sum_{|k| \leq (2^j+1)A} f(x) \phi_{j,k}(x) = 2^{j/2} f(x),$$

$$S^j(f)(x) = 2^{-j/2} \sum_{k \in \mathbf{Z}} f\left(\frac{k}{2^j}\right) \phi_{j,k}(x) = 2^{-j/2} \sum_{|k| \leq (2^j+1)A} f\left(\frac{k}{2^j}\right) \phi_{j,k}(x).$$

Back to the Taylor expansion,

$$S^j(f)(x) = 2^{-j/2} \sum_{|k| \leq (2^j+1)A} \left(\sum_{p=0}^{N-1} \left(\frac{1}{p!} f^{(p)}(x) \left(\frac{k}{2^j} - x\right)^p \right) \right)$$

$$\begin{aligned}
& + \frac{1}{N!} f^{(N)}(\theta_k) \left(\frac{k}{2^j} - x \right)^N \phi_{j,k}(x) \\
= & 2^{-j/2} \left(\sum_{p=0}^{N-1} \sum_{|k| \leq (2^j+1)A} \frac{1}{p!} f^{(p)}(x) \left(\frac{k}{2^j} - x \right)^p \phi_{j,k}(x) \right) \\
& + 2^{-j/2} \sum_{|k| \leq (2^j+1)A} \left(\frac{1}{N!} f^{(N)}(\theta_k) \left(\frac{k}{2^j} - x \right)^N \phi_{j,k}(x) \right) \\
= & f(x) + 2^{-j/2} \sum_{|k| \leq (2^j+1)A} \left(\frac{1}{N!} f^{(N)}(\theta_k) \left(\frac{k}{2^j} - x \right)^N \phi_{j,k}(x) \right) \\
= & f(x) + 2^{-j/2} \sum_{|k| \leq (2^j+1)A} \left(\frac{1}{N!} (f^{(N)}(\theta_k) - f^{(N)}(x)) \left(\frac{k}{2^j} - x \right)^N \phi_{j,k}(x) \right).
\end{aligned}$$

Thus,

$$S^j(f)(x) - f(x) = 2^{-j/2} \sum_{|k| \leq (2^j+1)A} \left(\frac{1}{N!} (f^{(N)}(\theta_k) - f^{(N)}(x)) \left(\frac{k}{2^j} - x \right)^N \phi_{j,k}(x) \right).$$

But

$$\left| f^{(N)}(\theta_k) - f^{(N)}(x) \right| \leq C |\theta_k - x| \leq C \left| \frac{k}{2^j} - x \right|,$$

for some constant C depending only on f . So

$$\begin{aligned}
& \|f(x) - S^j(f)(x)\|_{L^2} \\
& \leq \left\| 2^{-j/2} \sum_{|k| \leq (2^j+1)A} \left| \frac{C}{N!} \left(\frac{k}{2^j} - x \right)^{N+1} \phi_{j,k}(x) \right| \right\|_{L^2} \\
& = \frac{2^{-j(N+3/2)} C}{N!} \left\| \sum_{|k| \leq (2^j+1)A} |(k-y)^{N+1} \phi(y-k)| \right\|_{L^2},
\end{aligned}$$

where we make the substitution $y = 2^j x$. Define

$$g_k(y) := |(k-y)^{(N+1)} \phi(y-k)|,$$

then $g_k(y)$ has compact support $[-A+k, A+k]$, and $\|g_k(y)\|_{L^2}$ is equal to the L^2 norm of $x^{N+1} \phi(x)$, which is another constant C depending only on $\phi(x)$. Using this we obtain

$$\|f(x) - S^j(f)(x)\|_{L^2}$$

$$\begin{aligned}
&\leq \frac{2^{-j(N+3/2)}C}{N!} \left(\int_{-\infty}^{\infty} \left(\sum_{|k_1|, |k_2| \leq (2^j+1)A} g_{k_1}(y) g_{k_2}(y) \right) dy \right)^{1/2} \\
&= \frac{2^{-j(N+3/2)}C}{N!} \left(\int_{-\infty}^{\infty} \left(\sum_{|k_1|, |k_2| \leq (2^j+1)A, |k_1-k_2| \leq 2A} g_{k_1}(y) g_{k_2}(y) \right) dy \right)^{1/2} \\
&= \frac{2^{-j(N+3/2)}C}{N!} \left(2(2^j+1)A \cdot 4A \cdot \left(\max_{k_1, k_2 \in \mathbf{Z}} \left\{ \left| \int_{-\infty}^{\infty} g_{k_1}(y) g_{k_2}(y) dy \right| \right\} \right) \right)^{1/2} \\
&= \frac{2^{-j(N+3/2)}C}{N!} \left(8(2^j+1)A^2 \cdot C^2 \right)^{1/2} \\
&\leq C2^{-j(N+1)},
\end{aligned}$$

where C depends only on $f(x)$ and $\phi(x)$, i.e., C depends only on f and $\{a_k\}$. \square

We call $S^j(f)(x)$ defined in (3.6) the *wavelet sampling approximation* of the function $f(x)$ at the level j . It is similar to but distinct from the *wavelet orthogonal projection*

$$P^j(f) := \sum_{k \in \mathbf{Z}} \left(\int_{-\infty}^{\infty} f(x) \phi_{j,k}(x) dx \right) \cdot \phi_{j,k}(x),$$

which has been studied by various authors (see [3], [16], [17], [46], and others). The wavelet sampling approximation is what is used in most applications of wavelets, as it is the easiest approximation to compute. (Simply let the sample values of the given function be the corresponding expansion coefficients.) The value of the above result is that for biorthogonal and orthogonal Coifman wavelet systems (which will be defined in Section 3.2 and 4), the degree of approximation is much better than that obtained using Daubechies wavelet systems and the orthogonal projection.

As it can be seen in the proof, the smooth condition $f(x) \in C_0^{N,1}(\mathbf{R})$ is sufficient for the above argument. We formulate the similar result for Hölder space in the following corollary.

Corollary 3.1 Assume $\{a_k\}$ and $\phi(x)$ satisfy the same conditions as in Theorem 3.2. If $f(x) \in C_0^{N,\alpha}(\mathbf{R})$, $0 < \alpha \leq 1$, then

$$\|f(x) - S^j(f)(x)\|_{L^2} \leq C2^{-j(N+\alpha)},$$

where $S^j(f)(x)$ is the wavelet sampling approximation at the level j , and C is a constant depending only on f and $\{a_k\}$.

For smooth functions, the L^2 norm estimate is not enough, sometimes. One of the best candidates is the H^n norm. The H^n norm also measure the difference of the (weak) derivatives. Since the wavelet sampling approximation $S^j(f)$ is a linear combination of $\phi_{j,k}$, we require some regularity condition on the scaling function $\phi(x)$.

Theorem 3.3 Assume the same conditions as in Theorem 3.2. If in addition $\phi(x) \in C^n(\mathbf{R})$, where $n \in \mathbf{Z}, 0 \leq n \leq N$, then

$$\|f(x) - S^j(f)(x)\|_{H^n} \leq C2^{-j(N+1-n)}, \quad (3.9)$$

where C depends only on f and the sequence $\{a_k\}$.

To get the H^n estimate of the difference between $f(x)$ and $S^j(f)(x)$, we need a variation of Lemma 3.2.

Lemma 3.3 Assume the same conditions as in Lemma 3.2, if in addition $\phi(x) \in C^n(\mathbf{R})$, where n is a nonnegative integer, then

$$\sum_{k \in \mathbf{Z}} (x-k)^p \phi^{(m)}(x-k) = (-1)^p (p!) \delta_{0, m-p}, \quad (3.10)$$

where $p = 0, \dots, N$, and $0 \leq m \leq n$.

Proof We prove this by induction on m . The $m = 0$ case is given in Lemma 3.2. Now assume (3.10) holds for $0 \leq m \leq l$, where $l \leq n - 1$. We will prove that when $m = l + 1$, (3.10) still holds. Define

$$s(x, p, m) = \sum_{k \in \mathbf{Z}} (x-k)^p \phi^{(m)}(x-k). \quad (3.11)$$

Again, since $\phi(x)$ has compact support, $s(x, p, m)$ is well-defined. And we can interchange the sum and the differentiation because for any point x , only finite terms in the right hand side sum of (3.11) count. So

$$\begin{aligned} s(x, p, l+1) &= \sum_{k \in \mathbf{Z}} (x-k)^p \phi^{(l+1)}(x-k) \\ &= \sum_{k \in \mathbf{Z}} \left(\left((x-k)^p \phi^{(l)}(x-k) \right)' - p \cdot (x-k)^{(p-1)} \phi^{(l)}(x-k) \right) \\ &= \sum_{k \in \mathbf{Z}} \left((x-k)^p \phi^{(l)}(x-k) \right)' - \sum_{k \in \mathbf{Z}} p \cdot (x-k)^{(p-1)} \phi^{(l)}(x-k) \end{aligned}$$

$$\begin{aligned}
&= \left(\sum_{k \in \mathbf{Z}} (x-k)^p \phi^{(l)}(x-k) \right)' - p \sum_{k \in \mathbf{Z}} (x-k)^{(p-1)} \phi^{(l)}(x-k) \\
&= (s(x, p, l))' - p \cdot s(x, p-1, l) \\
&= ((-1)^p (p!) \delta_{0, l-p})' - p \cdot (-1)^{(p-1)} ((p-1)!) \delta_{0, l-p+1} \\
&= 0 + (-1)^p (p!) \delta_{0, l-p+1} \\
&= (-1)^p (p!) \delta_{0, l+1-p}.
\end{aligned}$$

The lemma is proved. \square

The proof of Theorem 3.3 is almost the same as Theorem 3.2. We will look at the Taylor expansion again.

Proof of Theorem 3.3 The case $n = 1$ will be proved here. The proof for the general case will be then apparent.

We now show that

$$\|f'(x) - (S^j(f)(x))'\|_{L^2} \leq C 2^{-jN},$$

for some constant C , independent of j . Using the same notation as in the proof of Theorem 3.2, we have

$$\begin{aligned}
&(S^j(f)(x))' \\
&= \left(2^{-j/2} \sum_{k \in \mathbf{Z}} f\left(\frac{k}{2^j}\right) \phi_{j,k}(x) \right)' \\
&= \left(2^{-j/2} \sum_{|k| \leq (2^j+1)A} f\left(\frac{k}{2^j}\right) \phi_{j,k}(x) \right)' \\
&= 2^{-j/2} \sum_{|k| \leq (2^j+1)A} f\left(\frac{k}{2^j}\right) (\phi_{j,k}(x))' \\
&= 2^j \sum_{|k| \leq (2^j+1)A} f\left(\frac{k}{2^j}\right) \phi'(2^j x - k) \\
&= 2^j \sum_{|k| \leq (2^j+1)A} \left(\sum_{p=0}^{N-1} \left(\frac{1}{p!} f^{(p)}(x) \left(\frac{k}{2^j} - x\right)^p \right) \right. \\
&\quad \left. + \frac{1}{N!} f^{(N)}(\theta_k) \left(\frac{k}{2^j} - x\right)^N \right) \phi'(2^j x - k) \\
&= 2^j \left(\sum_{p=0}^{N-1} \sum_{|k| \leq (2^j+1)A} \left(\frac{1}{p!} f^{(p)}(x) \left(\frac{k}{2^j} - x\right)^p \phi'(2^j x - k) \right) \right)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{|k| \leq (2^j+1)A} \left(\frac{1}{N!} f^{(N)}(\theta_k) \left(\frac{k}{2^j} - x \right)^N \phi'(2^j x - k) \right) \\
= & 2^j \sum_{|k| \leq (2^j+1)A} \left(f(x) \phi'(2^j x - k) \right) \\
& + 2^j \sum_{|k| \leq (2^j+1)A} \left(f'(x) \left(\frac{k}{2^j} - x \right) \phi'(2^j x - k) \right) \\
& + 2^j \left(\sum_{p=2}^{N-1} \sum_{|k| \leq (2^j+1)A} \frac{1}{p!} f^{(p)}(x) \left(\frac{k}{2^j} - x \right)^p \phi'(2^j x - k) \right) \\
& + 2^j \sum_{|k| \leq (2^j+1)A} \left(\frac{1}{N!} f^{(N)}(\theta_k) \left(\frac{k}{2^j} - x \right)^N \phi'(2^j x - k) \right) \\
= & f'(x) + 2^j \sum_{|k| \leq (2^j+1)A} \left(\frac{1}{N!} f^{(N)}(\theta_k) \left(\frac{k}{2^j} - x \right)^N \phi'(2^j x - k) \right) \\
= & f'(x) + 2^j \sum_{|k| \leq (2^j+1)A} \left(\frac{1}{N!} (f^{(N)}(\theta_k) - f^{(N)}(x)) \left(\frac{k}{2^j} - x \right)^N \phi'(2^j x - k) \right).
\end{aligned}$$

Thus,

$$\begin{aligned}
& \left\| f'(x) - (S^j(f)(x))' \right\|_{L^2} \\
= & 2^j \left\| \sum_{|k| \leq (2^j+1)A} \left(\frac{1}{N!} (f^{(N)}(\theta_k) - f^{(N)}(x)) \left(\frac{k}{2^j} - x \right)^N \phi'(2^j x - k) \right) \right\|_{L^2}.
\end{aligned}$$

Now applying the same estimate as in Theorem 3.2, we get

$$\|f'(x) - (S^j(f)(x))'\|_{L^2} \leq C 2^{-jN},$$

where C depends only on f and $\{a_k\}$. Then (3.9) follows readily from above and Theorem 3.2. \square

If Ω is a bounded open set in \mathbf{R} , a function $f \in C^N(\bar{\Omega})$ can be extended to $\tilde{f} \in C_0^N(\mathbf{R})$. So the all previous results on $C_0^N(\mathbf{R})$ function is certainly true for $C^N(\bar{\Omega})$ function while the $L^2(\mathbf{R})$ norm is replaced with the $L^2(\Omega)$ norm.

When dealing with higher dimension \mathbf{R}^m other than \mathbf{R} , we may take the tensor product of the one dimensional wavelet systems to construct higher dimensional wavelet systems. And it is straightforward to generalize the one dimensional results to higher dimensional cases.

Theorem 3.4 Suppose $\{a_k^1, k \in \mathbf{Z}\}, \{a_k^2, k \in \mathbf{Z}\}, \dots, \{a_k^m, k \in \mathbf{Z}\}$ are a family of sequences with finite length (i.e., there exists a positive integer K , $a_k^l = 0$ for $1 \leq l \leq m$, and $|k| > K$). Assume that $\{a_k^l\}$ satisfies the vanishing moments conditions up to degree N , i.e., for $1 \leq l \leq m$

$$\sum_{k \in \mathbf{Z}} (2k)^p a_{2k}^l = \sum_{k \in \mathbf{Z}} (2k+1)^p a_{2k+1}^l = 0, \quad \text{for } p = 1, \dots, N,$$

$$\sum_{k \in \mathbf{Z}} a_{2k}^l = \sum_{k \in \mathbf{Z}} a_{2k+1}^l = 1.$$

And for $1 \leq l \leq m$, $\phi^l(x)$ is a $L^2(\mathbf{R})$ solution of the two-scale difference equation

$$\phi^l(x) = \sum_{k \in \mathbf{Z}} a_k^l \phi^l(2x - k).$$

Then for any function $f(x_1, x_2, \dots, x_m) \in C_0^{N,1}(\mathbf{R}^m)$,

$$\|f(x_1, x_2, \dots, x_m) - S^j(f)(x_1, x_2, \dots, x_m)\|_{L^2} \leq C2^{-j(N+1)},$$

where C depends only on f and the family of sequences $\{a_k^l\}$, and

$$\begin{aligned} S^j(f)(x_1, x_2, \dots, x_m) &:= 2^{-jm/2} \sum_{k_1, k_2, \dots, k_m \in \mathbf{Z}} f\left(\frac{k_1}{2^j}, \frac{k_2}{2^j}, \dots, \frac{k_m}{2^j}\right) \\ &\quad \cdot \phi_{j, k_1}^1(x_1) \cdot \phi_{j, k_2}^2(x_2) \cdots \phi_{j, k_m}^n(x_m). \end{aligned}$$

If in addition, $\phi^l(x) \in C^n(\mathbf{R})$ for all $1 \leq l \leq m$, where n is a nonnegative integer not greater than N , then

$$\|f(x_1, x_2, \dots, x_m) - S^j(f)(x_1, x_2, \dots, x_m)\|_{H^n} \leq C2^{-j(N+1-n)}.$$

The proof is an easy modification of the one dimensional case and will be omitted here. In practice we may choose $\{a_k^l\}$ to be the same sequence, then $\phi^1(x) = \phi^2(x) = \dots = \phi^m(x)$.

3.2 Definition of Biorthogonal Coifman Wavelet Systems

The wavelet approximation theorem requires the sequence $\{a_k\}$ satisfying the linear conditions (3.4) and (3.5). When working in the wavelet system, we see that these two conditions are exactly the vanishing moment conditions on the scaling function $\phi(x)$ and $\psi(x)$.

Lemma 3.4 Suppose $\phi(x), \psi(x) \in L^2(\mathbf{R})$ satisfying

$$\phi(x) = \sum_{k \in \mathbf{Z}} a_k \phi(2x - k), \quad (3.12)$$

$$\psi(x) = \sum_{k \in \mathbf{Z}} (-1)^k a_{-k+1} \phi(2x - k), \quad (3.13)$$

where $\{a_k\}$ is a finite length sequence, $\phi(x) \in L^1(\mathbf{R})$ and it is normalized

$$\int_{-\infty}^{\infty} \phi(x) dx = 1. \quad (3.14)$$

Then the following two conditions are equivalent,

1. The vanishing moments of $\phi(x)$ and $\psi(x)$ are both of degree N , i.e.,

$$\text{Mom}_0(\psi) := \int_{-\infty}^{\infty} \psi(x) dx = 0,$$

$$\text{Mom}_p(\phi) := \int_{-\infty}^{\infty} x^p \phi(x) dx = 0, \quad \text{for } p = 1, \dots, N,$$

$$\text{Mom}_p(\psi) := \int_{-\infty}^{\infty} x^p \psi(x) dx = 0, \quad \text{for } p = 1, \dots, N.$$

2. The sequence $\{a_k\}$ satisfying

$$\sum_{k \in \mathbf{Z}} (2k)^p a_{2k} = \sum_{k \in \mathbf{Z}} (2k+1)^p a_{2k+1} = 0, \quad \text{for } p = 1, \dots, N,$$

$$\sum_{k \in \mathbf{Z}} a_{2k} = \sum_{k \in \mathbf{Z}} a_{2k+1} = 1.$$

Proof 1 \implies 2: We have

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \phi(x) dx \\ &= \int_{-\infty}^{\infty} \left(\sum_k a_k \phi(2x - k) \right) dx \\ &= \sum_k a_k \int_{-\infty}^{\infty} \phi(2x - k) dx \\ &= \frac{1}{2} \sum_k a_k \end{aligned}$$

Similarly,

$$\begin{aligned}
0 &= \int_{-\infty}^{\infty} \psi(x) dx \\
&= \int_{-\infty}^{\infty} \left(\sum_k (-1)^k a_{-k+1} \phi(2x - k) \right) dx \\
&= \sum_k (-1)^k a_{-k+1} \int_{-\infty}^{\infty} \phi(2x - k) dx \\
&= \sum_k (-1)^k a_{-k+1} \cdot \frac{1}{2} \\
&= \frac{1}{2} \left(\sum_k a_{2k+1} - \sum_k a_{2k} \right)
\end{aligned}$$

Then it follows

$$\sum_k a_{2k} = \sum_k a_{2k+1} = 1.$$

Now assume that the equality

$$\sum_k (2k)^p a_{2k} = \sum_k (2k+1)^p a_{2k+1} = 0 \quad (3.15)$$

holds for $1 \leq p \leq l$, where $0 \leq l \leq N-1$. (When $l=0$, there will be no assumption on (3.15).) We want to prove that (3.15) also holds for $p=l+1$. Then by induction, (3.15) will be valid for $p=1, \dots, N$.

By the vanishing moments on $\phi(x)$ and $\psi(x)$,

$$\begin{aligned}
0 &= \int_{-\infty}^{\infty} x^{l+1} \phi(x) dx \\
&= \int_{-\infty}^{\infty} x^{l+1} \left(\sum_k a_k \phi(2x - k) \right) dx \\
&= \sum_k a_k \int_{-\infty}^{\infty} x^{l+1} \phi(2x - k) dx \\
&= \frac{1}{2^{l+2}} \sum_k a_k \int_{-\infty}^{\infty} (x+k)^{l+1} \phi(x) dx \\
&= \frac{1}{2^{l+2}} \sum_k a_k \int_{-\infty}^{\infty} \left(\sum_{m=0}^{l+1} \binom{l+1}{m} x^m k^{l+1-m} \right) \phi(x) dx \\
&= \frac{1}{2^{l+2}} \sum_k a_k \sum_{m=0}^{l+1} \binom{l+1}{m} k^{l+1-m} \int_{-\infty}^{\infty} x^m \phi(x) dx \\
&= \frac{1}{2^{l+2}} \sum_k a_k k^{l+1}
\end{aligned}$$

and

$$\begin{aligned}
0 &= \int_{-\infty}^{\infty} x^{l+1} \psi(x) dx \\
&= \int_{-\infty}^{\infty} x^{l+1} \left(\sum_k (-1)^k a_{-k+1} \phi(2x - k) \right) dx \\
&= \sum_k (-1)^k a_{-k+1} \int_{-\infty}^{\infty} x^{l+1} \phi(2x - k) dx \\
&= \frac{1}{2^{l+2}} \sum_k (-1)^k a_{-k+1} \int_{-\infty}^{\infty} (x+k)^{l+1} \phi(x) dx \\
&= \frac{1}{2^{l+2}} \sum_k (-1)^k a_{-k+1} \int_{-\infty}^{\infty} \left(\sum_{m=0}^{l+1} \binom{l+1}{m} x^m k^{l+1-m} \right) \phi(x) dx \\
&= \frac{1}{2^{l+2}} \sum_k (-1)^k a_{-k+1} \sum_{m=0}^{l+1} \binom{l+1}{m} k^{l+1-m} \int_{-\infty}^{\infty} x^m \phi(x) dx \\
&= \frac{1}{2^{l+2}} \sum_k (-1)^k a_{-k+1} k^{l+1} \\
&= \frac{1}{2^{l+2}} \sum_k (-1)^{1-k} a_k (1-k)^{l+1} \\
&= \frac{1}{2^{l+2}} \sum_k (-1)^{1-k} a_k (-1)^{l+1} k^{l+1} \\
&= \frac{1}{2^{l+2}} \sum_k (-1)^{l-k} a_k k^{l+1} \\
&= \frac{(-1)^l}{2^{l+2}} \left(\sum_k (2k)^{l+1} a_{2k} - \sum_k (2k+1)^{l+1} a_{2k+1} \right)
\end{aligned}$$

Thus it follows immediately that

$$\sum_k (2k)^{l+1} a_{2k} = \sum_k (2k+1)^{l+1} a_{2k+1} = 0.$$

2 \implies 1: This part is apparent from the above argument. \square

Based on Lemma 3.4, here is the proof of Lemma 3.2.

Proof of Lemma 3.2 Define

$$s(x) = \sum_{k \in \mathbf{Z}} (x-k)^p \phi(x-k).$$

Since $\{a_k\}$ has finite length, $\phi(x)$ has compact support. So $s(x)$ is well-defined and periodic with 1 as a period. We have

$$2^p s(x) = 2^p \sum_{k \in \mathbf{Z}} (x-k)^p \phi(x-k)$$

$$\begin{aligned}
&= 2^p \sum_{k \in \mathbf{Z}} \sum_{m \in \mathbf{Z}} (a_m(x-k)^p \phi(2x-2k-m)) \\
&= 2^p \sum_{k \in \mathbf{Z}} \sum_{i \in \mathbf{Z}} (a_{i-2k}(x-k)^p \phi(2x-i)) \\
&= \sum_{i \in \mathbf{Z}} \sum_{k \in \mathbf{Z}} (a_{i-2k}(2x-i+i-2k)^p \phi(2x-i)) \\
&= \sum_{i \in \mathbf{Z}} \sum_{k \in \mathbf{Z}} \sum_{l=0}^p \left(a_{i-2k} \binom{p}{l} (2x-i)^l (i-2k)^{p-l} \phi(2x-i) \right) \\
&= \sum_{i \in \mathbf{Z}} \sum_{l=0}^p \left(\binom{p}{l} \left(\sum_{k \in \mathbf{Z}} (i-2k)^{p-l} a_{i-2k} \right) (2x-i)^l \phi(2x-i) \right)
\end{aligned}$$

From (3.4) and (3.5),

$$\sum_{k \in \mathbf{Z}} (i-2k)^{p-l} a_{i-2k} = \delta_{0,p-l}.$$

Thus

$$2^p s(x) = \sum_{i \in \mathbf{Z}} (2x-i)^p \phi(2x-i) = s(2x).$$

If we know $\phi(x)$ is continuous, then since

$$\int_0^1 s(x) dx = \int_0^1 \left(\sum_{k \in \mathbf{Z}} (x-k)^p \phi(x-k) \right) dx = \int_{-\infty}^{\infty} x^p \phi(x) dx = \delta_{0,p},$$

the lemma follows easily. Without assuming the continuity, we can prove the lemma in the following way. For any $i, j \in \mathbf{Z}, j \geq 0$,

$$\begin{aligned}
\int_{\frac{i}{2^j}}^{\frac{i+1}{2^j}} s(x) dx &= \int_{\frac{i}{2^j}}^{\frac{i+1}{2^j}} 2^{-p} s(2x) dx \\
&= 2^{-p-1} \int_{\frac{i}{2^{(j-1)}}}^{\frac{i+1}{2^{(j-1)}}} s(x) dx \\
&= \dots \\
&= 2^{(-p-1)j} \int_i^{i+1} s(x) dx \\
&= 2^{(-p-1)j} \int_i^{i+1} \left(\sum_{k \in \mathbf{Z}} (x-k)^p \phi(x-k) \right) dx \\
&= 2^{(-p-1)j} \int_{-\infty}^{\infty} x^p \phi(x) dx.
\end{aligned}$$

The conditions (3.4) and (3.5) imply that

$$\int_{-\infty}^{\infty} x^p \phi(x) dx = \delta_{0,p}.$$

Then,

$$\int_{\frac{i}{2^j}}^{\frac{i+1}{2^j}} s(x) dx = 2^{(-p-1)j} \delta_{0,p}. \quad (3.16)$$

As we know, the restriction of the Haar wavelet system on $[0,1]$ is an orthonormal basis of $L^2([0,1])$, and the left hand side of (3.16) is exactly the coefficients for the wavelet orthogonal projection at the level j . So it follows immediately that

$$s(x) = \sum_{k \in \mathbf{Z}} (x - k)^p \phi(x - k) = \delta_{0,p}.$$

The lemma follows. \square

It is now clear why the two conditions (3.4) and (3.5) are imposed on $\{a_k\}$ in the wavelet approximation theorem. All we want is that the vanishing moments of $\phi(x)$ and $\psi(x)$ are both of some degree N . Thus the vanishing moment conditions not only imply the smoothness of the scaling function (and hence the wavelet function) but also provide a very neat approximation. Based on this observation, we introduce the biorthogonal Coifman wavelet system.

Definition 3.1 A biorthogonal wavelet system with compact support is called a *biorthogonal Coifman wavelet system* (in short, *BCW*) of degree N if the following two conditions are satisfied,

- the vanishing moments of the scaling function $\tilde{\phi}(x)$ and the wavelet function $\tilde{\psi}(x)$ are both of degree N , i.e.,

$$\text{Mom}_p(\tilde{\phi}) = \int_{-\infty}^{\infty} x^p \tilde{\phi}(x) dx = \delta_{0,p}, \quad \text{for } p = 0, \dots, N, \quad (3.17)$$

$$\text{Mom}_p(\tilde{\psi}) = \int_{-\infty}^{\infty} x^p \tilde{\psi}(x) dx = 0, \quad \text{for } p = 0, \dots, N, \quad (3.18)$$

- the vanishing moment of the wavelet function $\psi(x)$ is of degree N ,

$$\text{Mom}_p(\psi) = \int_{\mathbf{R}} x^p \psi(x) dx = 0, \quad \text{for } p = 0, \dots, N. \quad (3.19)$$

Note that in the definition of the biorthogonal Coifman wavelet system, although there is no vanishing moment requirement on the analysis scaling function $\phi(x)$, it follows that $\phi(x)$ also has vanishing moments up to degree N , because of the perfect reconstruction condition.

Lemma 3.5 For a biorthogonal Coifman wavelet system of degree N , the vanishing moments' degree of the analysis scaling function $\phi(x)$ is also N ,

$$\text{Mom}_p(\phi) = \int_{-\infty}^{\infty} x^p \phi(x) dx = 0, \quad \text{for } p = 1, \dots, N.$$

Proof From Lemma 3.4, it is sufficient to prove that

$$\sum_{k \in \mathbf{Z}} a_k k^p = 0, \quad \text{for } p = 1, \dots, N.$$

From the vanishing moment condition of the synthesis part, we have

$$\sum_{k \in \mathbf{Z}} \tilde{a}_{2k} (2k)^m = \sum_{k \in \mathbf{Z}} \tilde{a}_{2k+1} (2k+1)^m = \delta_{0,m}, \quad \text{for } 0 \leq m \leq p.$$

Thus

$$\begin{aligned} \sum_{k,l \in \mathbf{Z}} k^{p-m} (k+2l)^m a_k \tilde{a}_{k+2l} &= \sum_{k \in \mathbf{Z}} k^{p-m} a_k \sum_{l \in \mathbf{Z}} (k+2l)^m \tilde{a}_{k+2l} \\ &= \sum_{k \in \mathbf{Z}} k^{p-m} a_k \cdot \delta_{0,m} \end{aligned}$$

So

$$\begin{aligned} \sum_{k \in \mathbf{Z}} a_k k^p &= \sum_{k,l \in \mathbf{Z}} k^p a_k \tilde{a}_{k+2l} \\ &= \sum_{m=0}^p \left((-1)^m \binom{p}{m} \sum_{k,l \in \mathbf{Z}} k^{p-m} (k+2l)^m a_k \tilde{a}_{k+2l} \right) \\ &= \sum_{k,l \in \mathbf{Z}} a_k \tilde{a}_{k+2l} \sum_{m=0}^p \left((-1)^m \binom{p}{m} k^{p-m} (k+2l)^m \right) \\ &= \sum_{k,l \in \mathbf{Z}} a_k \tilde{a}_{k+2l} (k - (k+2l))^p \\ &= \sum_{k,l \in \mathbf{Z}} (-2l)^p a_k \tilde{a}_{k+2l} \end{aligned}$$

The perfect reconstruction condition states that

$$\sum_{k \in \mathbf{Z}} a_k \tilde{a}_{k+2l} = 2\delta_{0,l}, \quad \forall l \in \mathbf{Z}.$$

It follows that

$$\sum_{k \in \mathbf{Z}} a_k k^p = \sum_{k,l \in \mathbf{Z}} (-2l)^p a_k \tilde{a}_{k+2l}$$

$$\begin{aligned}
&= \sum_{l \in \mathbf{Z}} (-2l)^p \sum_{k \in \mathbf{Z}} a_k \tilde{a}_{k+2l} \\
&= \sum_{l \in \mathbf{Z}} (-2l)^p \cdot 2\delta_{0,l} \\
&= \sum_{l \in \mathbf{Z}} 0 \\
&= 0
\end{aligned}$$

□

Thus in a biorthogonal Coifman wavelet system, both the analysis pair $\{\phi, \psi\}$ and the synthesis pair $\{\tilde{\phi}, \tilde{\psi}\}$ have vanishing moments up to some degree N . In the decomposition process, based on the wavelet approximation theorem, we can sample on dyadic rationals and take these values as the discrete wavelet transform coefficients at the starting level. We can apply the Mallat Algorithm on these sample values and analyse the data (compression, denoising, etc). In the reconstruction process, again, based on the wavelet approximation theorem, after applying the Mallat Algorithm on the data, we can take the inverse discrete wavelet transform coefficients as the sample values on dyadic rationals and reconstruct the original data. That's why we impose vanishing moments on both the analysis pair and the synthesis pair. Now the natural question is whether biorthogonal Coifman wavelet systems exist. If yes, how can one design them? We will study on the existence problem in the next section.

3.3 Construction of BCWs

In this section, we will construct biorthogonal Coifman wavelet systems and obtain the exact formula for each degree. As is obvious, if we can construct them, then the existence problem is solved simultaneously.

Now let's see how many conditions are there imposed on biorthogonal Coifman wavelet systems? From the theory of biorthogonal wavelet systems, we know that the analysis scaling vector $\{a_k\}$ and synthesis scaling vector $\{\tilde{a}_k\}$ must satisfy the linear condition

$$\sum_{k \in \mathbf{Z}} a_k = \sum_{k \in \mathbf{Z}} \tilde{a}_k = 2, \quad (3.20)$$

and the perfect construction condition

$$\sum_{k \in \mathbf{Z}} a_k \tilde{a}_{k+2l} = 2\delta_{0,l}, \quad \forall l \in \mathbf{Z}. \quad (3.21)$$

The vanishing moment conditions in the definition of the biorthogonal Coifman wavelet system are equivalent to

$$\sum_{k \in \mathbf{Z}} (2k)^p a_{2k} = \sum_{k \in \mathbf{Z}} (2k+1)^p a_{2k+1} = \delta_{0,p}, \quad \text{for } p = 0, \dots, N, \quad (3.22)$$

$$\sum_{k \in \mathbf{Z}} (2k)^p \tilde{a}_{2k} = \sum_{k \in \mathbf{Z}} (2k+1)^p \tilde{a}_{2k+1} = \delta_{0,p}, \quad \text{for } p = 0, \dots, N. \quad (3.23)$$

In practice, we prefer that the length of the synthesis scaling vector be shorter than the length of the analysis scaling vector. The reason is that for longer analysis scaling vectors, we will get more redundant information after the discrete wavelet transform, which will be ideal for the further processing in the wavelet domain; and for shorter synthesis scaling vectors, the inverse discrete wavelet transform will provide a compact representation of the original data, which is very useful, for example, for data storage and transmission. Based on this thought, we will first look at the synthesis scaling vector $\{\tilde{a}_k\}$ and try to find the minimum length solution of it at each degree.

The linear condition (3.20) for $\{\tilde{a}_k\}$ is already included in (3.23) and we don't need to worry about the bilinear condition (3.21) right now. Thus there are totally $2N + 2$ linear conditions on the scaling vector $\{\tilde{a}_k\}$. And the minimum length solution will have, of course, $2N + 2$ elements in the finite length sequence $\{\tilde{a}_k\}$. For symmetric reason, we will distribute these $2N + 2$ elements in $\{\tilde{a}_k\}$ as symmetric as possible. More precisely, we are looking for a solution of the form $\{\tilde{a}_{-N}, \tilde{a}_{-N+1}, \dots, \tilde{a}_{N+1}\}$. Let's have another look at the linear equation system (3.23). By its form, these $2N + 2$ linear equations can be naturally divided into two parts, those on the even terms \tilde{a}_{2k} and those on the odd terms \tilde{a}_{2k+1} . For those $N + 1$ linear equations on the even terms, it is easy to see that the solution is exactly that all are zero except that $\tilde{a}_0 = 1$. For those $N + 1$ linear equations on the odd terms, the coefficient matrix of these $N + 1$ variables \tilde{a}_{2k+1} is a Vandermonde's matrix (so is for the even terms). Thus the solution for these $N + 1$ odd terms always exist and it is unique. Using the determinant formula of a Vandermonde's matrix, we get the exact formulas for these odd terms,

- if N is even, $N = 2n$,

$$\tilde{a}_{2k+1} = \frac{\prod_{j \neq k, j=-n}^n (2j+1)}{2^N \prod_{j \neq k, j=-n}^n (j-k)} = \frac{(-1)^k}{2k+1} \binom{2n-1}{n-1} \binom{2n}{n+k} \frac{2n+1}{2^{4n-1}}$$

- if N is odd, $N = 2n - 1$,

$$\tilde{a}_{2k+1} = \frac{\prod_{j \neq k, j=-n}^{n-1} (2j+1)}{2^N \prod_{j \neq k, j=-n}^{n-1} (j-k)} = \frac{(-1)^k}{2k+1} \binom{2n-2}{n-1} \binom{2n-1}{n+k} \frac{2n-1}{2^{4n-3}}$$

here we define $\binom{m}{l} := 0$ if $l > m$ or $l < 0$, which is standard in combinatorial theory.

So we have got the formula for the minimum length synthesis scaling vector $\{\tilde{a}_k\}$ for every degree N . As it can be seen in the above discussion, the minimum length synthesis scaling vector is unique. The next theorem tells us how to construct the analysis scaling vector $\{a_k\}$ from the synthesis scaling vector $\{\tilde{a}_k\}$.

Theorem 3.5 Assume \tilde{a}_k to be defined as above, depending on whether N is even or odd. Set

$$a_{2k+1} = \tilde{a}_{2k+1}, \quad (3.24)$$

and

$$a_{2k} = 2\delta_{0,k} - \sum_{l \in \mathbf{Z}} \tilde{a}_{2l+1} \tilde{a}_{2l+1-2k}. \quad (3.25)$$

Then the resulting $\{a_k\}$ and $\{\tilde{a}_k\}$ will constitute a biorthogonal Coifman wavelet system of degree N .

Proof We need to check (3.21) and (3.22).

First,

$$\begin{aligned} \sum_{k \in \mathbf{Z}} a_k \tilde{a}_{k+2l} &= \sum_{k \in \mathbf{Z}} a_{2k} \tilde{a}_{2k+2l} + \sum_{k \in \mathbf{Z}} a_{2k+1} \tilde{a}_{2k+1+2l} \\ &= a_{-2l} + \sum_{k \in \mathbf{Z}} \tilde{a}_{2k+1} \tilde{a}_{2k+1+2l} \\ &= 2\delta_{0,-l} - \sum_{k \in \mathbf{Z}} \tilde{a}_{2k+1} \tilde{a}_{2k+1+2l} + \sum_{k \in \mathbf{Z}} \tilde{a}_{2k+1} \tilde{a}_{2k+1+2l} \\ &= 2\delta_{0,-l} \\ &= 2\delta_{0,l} \end{aligned}$$

All remaining is to show that

$$\sum_{k \in \mathbf{Z}} (2k)^p a_{2k} = \delta_{0,p}, \quad \text{for } p = 0, \dots, N.$$

We have

$$\begin{aligned}
& \sum_{k \in \mathbf{Z}} (2k)^p a_{2k} \\
&= \sum_{k \in \mathbf{Z}} \left((2k)^p \left(2\delta_{0,k} - \sum_{n \in \mathbf{Z}} \tilde{a}_{2n+1} \tilde{a}_{2n+1-2k} \right) \right) \\
&= 2\delta_{0,p} - \sum_{k,n \in \mathbf{Z}} (2k)^p \tilde{a}_{2n+1} \tilde{a}_{2n+1-2k} \\
&= 2\delta_{0,p} - \sum_{m,n \in \mathbf{Z}} (2n-2m)^p \tilde{a}_{2n+1} \tilde{a}_{2m+1} \\
&= 2\delta_{0,p} - \sum_{m,n \in \mathbf{Z}} ((2n+1) - (2m+1))^p \tilde{a}_{2n+1} \tilde{a}_{2m+1} \\
&= 2\delta_{0,p} - \sum_{m,n \in \mathbf{Z}} \sum_{l=0}^p \binom{p}{l} (-1)^l (2n+1)^{p-l} (2m+1)^l \tilde{a}_{2n+1} \tilde{a}_{2m+1} \\
&= 2\delta_{0,p} - \sum_{l=0}^p (-1)^l \left(\binom{p}{l} \sum_{m \in \mathbf{Z}} (2m+1)^l \tilde{a}_{2m+1} \sum_{n \in \mathbf{Z}} (2n+1)^{p-l} \tilde{a}_{2n+1} \right) \\
&= 2\delta_{0,p} - \sum_{l=0}^p (-1)^l \left(\binom{p}{l} \sum_{m \in \mathbf{Z}} (2m+1)^l \tilde{a}_{2m+1} \delta_{l,p} \right) \\
&= 2\delta_{0,p} - (-1)^p \sum_{m \in \mathbf{Z}} (2m+1)^p \tilde{a}_{2m+1} \\
&= 2\delta_{0,p} - (-1)^p \delta_{0,p} \\
&= 2\delta_{0,p} - \delta_{0,p} \\
&= \delta_{0,p}
\end{aligned}$$

□

Note that the analysis scaling vector given by Theorem 3.5 is the minimum length solution for the analysis scaling vector. Because the bilinear condition (3.21) is exactly

$$a_{2k} = 2\delta_{0,k} - \sum_{l \in \mathbf{Z}} a_{2l+1} \tilde{a}_{2l+1-2k},$$

the minimum length analysis scaling vector must be the analysis scaling vector having the minimum number of nonzero odd terms. From (3.22) and (3.23), such an analysis scaling vector will have the same odd terms as the minimum length synthesis scaling vector. Thus we obtain the biorthogonal Coifman wavelet system with minimum length.

In the remainder of this thesis, a biorthogonal Coifman wavelet system of degree N will always be referred to as biorthogonal Coifman wavelet system of degree N with minimum length, which is given by in Theorem 3.5, unless it is otherwise stated. For convenience, we will call it *BCW- N* .

The scaling vectors of the minimum length biorthogonal Coifman wavelet systems with degrees $N = 0, 1, 2, 3$, and 4 are listed in Table 3.1. The biorthogonal Coifman wavelet system with degree 0 (BCW-0) is exactly the Haar wavelet system, which is orthogonal. The biorthogonal Coifman wavelet system with degree 1 (BCW-1) is a spline system. We include the illustrations for degree 2, 3 and 4 (BCW-2, BCW-3 and BCW-4) in Figure 3.1, 3.2, and 3.3.

3.4 Properties of BCWs

We now know how to construct biorthogonal Coifman wavelet systems. and the scaling vectors of degrees up to 4 are listed in Table 3.1. In this section we will look at some properties of these wavelet systems.

3.4.1 Approximation

The wavelet approximation theorem is the starting point of biorthogonal Coifman wavelet systems (and also orthogonal Coifman wavelet systems, which will be studied in Chapter 4). From Definition 3.1, the following theorem is just stating the wavelet approximation theorem in the language of biorthogonal wavelet systems.

Theorem 3.6 For a biorthogonal Coifman wavelet system of degree N with the analysis scaling function $\phi(x)$ and synthesis scaling function $\tilde{\phi}(x)$, if $f(x) \in C^{N,1}(\mathbf{R})$, define, for $j \in \mathbf{Z}$,

$$f^j(x) := 2^{-j/2} \sum_{k \in \mathbf{Z}} f\left(\frac{k}{2^j}\right) \tilde{\phi}_{j,k}(x),$$

where $\tilde{\phi}_{j,k}(x) = 2^{j/2} \tilde{\phi}(2^j x - k)$. Then

$$\|f(x) - f^j(x)\|_{L^2} \leq C 2^{-j(N+1)},$$

where C depends only on f and $\tilde{\phi}$.

If in addition $\tilde{\phi} \in C^n(\mathbf{R})$, where $n \in \mathbf{Z}, 0 \leq n \leq N$, then

$$\|f(x) - f^j(x)\|_{H^n} \leq C 2^{-j(N+1-n)},$$

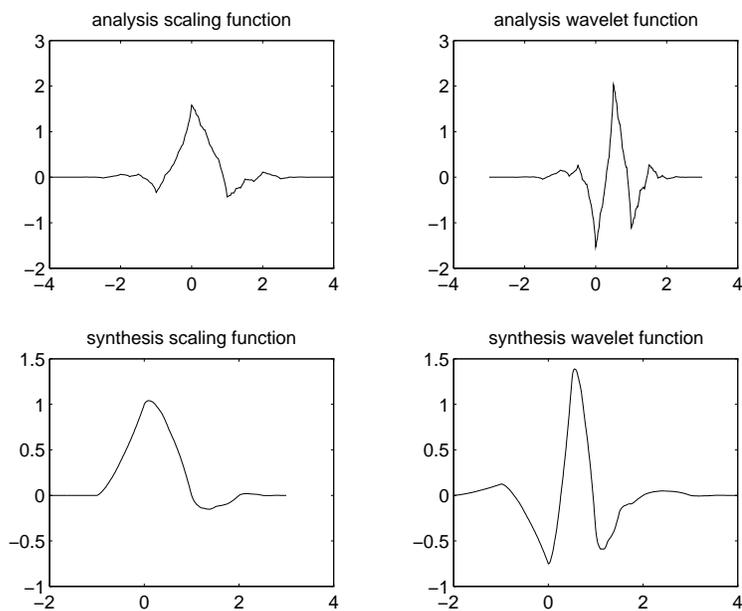


Figure 3.1 The Biorthogonal Coifman Wavelet System of Degree 2

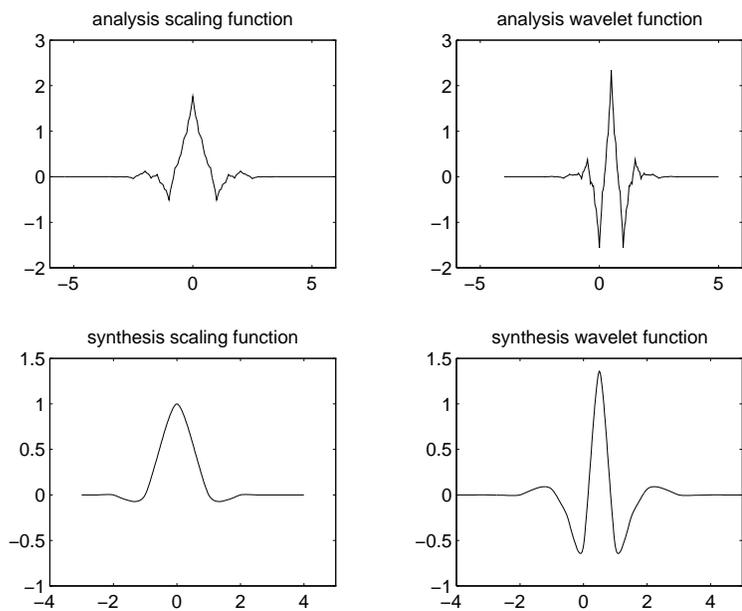


Figure 3.2 The Biorthogonal Coifman Wavelet System of Degree 3

Table 3.1 The Coefficients of Biorthogonal Coifman Wavelet Systems

N	a_k	\tilde{a}_k	N	a_k	\tilde{a}_k
N = 0	$a_0 = 1$	$\tilde{a}_0 = 1$	N = 3	$a_1 = 9/16$	$\tilde{a}_1 = 9/16$
	$a_1 = 1$	$\tilde{a}_1 = 1$		$a_2 = -63/256$	$\tilde{a}_2 = 0$
N = 1	$a_{-2} = -1/4$			$a_3 = -1/16$	$\tilde{a}_3 = -1/16$
	$a_{-1} = 1/2$	$\tilde{a}_{-1} = 1/2$		$a_4 = 9/128$	$\tilde{a}_4 = 0$
	$a_0 = 3/2$	$\tilde{a}_0 = 1$		$a_5 = 0$	
	$a_1 = 1/2$	$\tilde{a}_1 = 1/2$		$a_6 = -1/256$	
	$a_2 = -1/4$	$\tilde{a}_2 = 0$		$a_7 = 0$	
	$a_3 = 0$		N = 4	$a_{-8} = 15/16384$	
N = 2	$a_{-4} = 3/64$			$a_{-7} = 0$	
	$a_{-3} = 0$			$a_{-6} = -35/2048$	
	$a_{-2} = -3/16$	$\tilde{a}_{-2} = 0$		$a_{-5} = 0$	
	$a_{-1} = 3/8$	$\tilde{a}_{-1} = 3/8$		$a_{-4} = 345/4096$	$\tilde{a}_{-4} = 0$
	$a_0 = 41/32$	$\tilde{a}_0 = 1$		$a_{-3} = -5/128$	$\tilde{a}_{-3} = -5/128$
	$a_1 = 3/4$	$\tilde{a}_1 = 3/4$		$a_{-2} = -405/2048$	$\tilde{a}_{-2} = 0$
	$a_2 = -3/16$	$\tilde{a}_2 = 0$		$a_{-1} = 15/32$	$\tilde{a}_{-1} = 15/32$
	$a_3 = -1/8$	$\tilde{a}_3 = -1/8$		$a_0 = 10317/8192$	$\tilde{a}_0 = 1$
	$a_4 = 3/64$			$a_1 = 45/64$	$\tilde{a}_1 = 45/64$
	$a_5 = 0$			$a_2 = -405/2048$	$\tilde{a}_2 = 0$
N = 3	$a_{-6} = -1/256$			$a_3 = -5/32$	$\tilde{a}_3 = -5/32$
	$a_{-5} = 0$			$a_4 = 345/4096$	$\tilde{a}_4 = 0$
	$a_{-4} = 9/128$			$a_5 = 3/128$	$\tilde{a}_5 = 3/128$
	$a_{-3} = -1/16$	$\tilde{a}_{-3} = -1/16$		$a_6 = -35/2048$	
	$a_{-2} = -63/256$	$\tilde{a}_{-2} = 0$		$a_7 = 0$	
	$a_{-1} = 9/16$	$\tilde{a}_{-1} = 9/16$	$a_8 = 15/16384$		
	$a_0 = 87/64$	$\tilde{a}_0 = 1$	$a_9 = 0$		

where C depends only on f and $\tilde{\phi}$.

The same results hold when replacing $\tilde{\phi}$ with ϕ .

3.4.2 Compact Support

From Definition 3.1, biorthogonal Coifman wavelet systems are always compactly supported. All the analysis scaling vectors and synthesis scaling vectors have finite length. This finiteness property is extremely useful when implementing a discrete wavelet transform on digital computers.

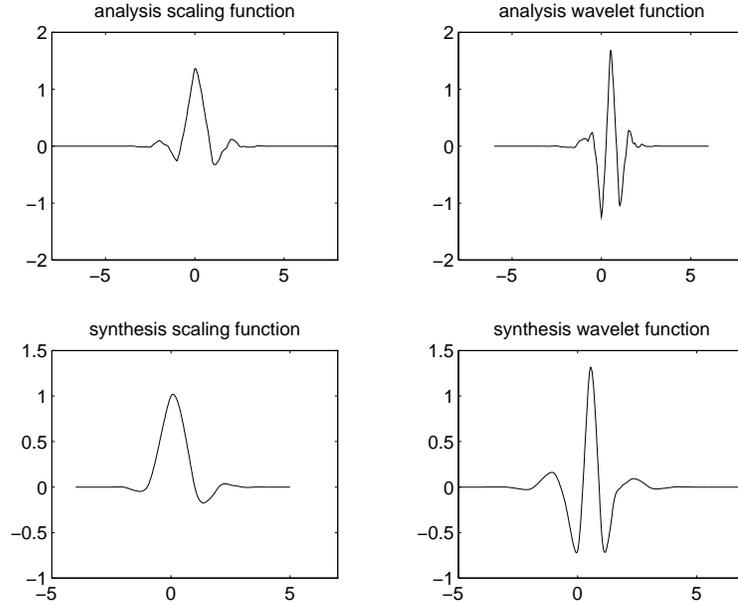


Figure 3.3 The Biorthogonal Coifman Wavelet System of Degree 4

3.4.3 Symmetry

As we know, in a wavelet system, the symmetry of the scaling vector corresponds to the symmetry of the scaling function. Thus we call a wavelet system is *symmetric* if the scaling vector is symmetric. One big advantage of biorthogonal wavelet systems over orthogonal wavelet systems is that biorthogonal wavelet systems can be symmetric, while orthogonal ones can't, except for the Haar wavelet system. Symmetry is always pursued whenever possible in applications. For example, in image coding, if we have a symmetric biorthogonal wavelet system, then the image data can be reflectively extended to reduce the edge effect.

The even terms of biorthogonal Coifman wavelet systems are always symmetric, which is apparent from Theorem 3.5. Since the analysis scaling vector and synthesis scaling vector have the same odd terms, the symmetry of a biorthogonal Coifman wavelet system will depend solely on the symmetry of the odd terms of scaling vectors.

Lemma 3.6 If N is odd, then in a biorthogonal Coifman wavelet system of degree N ,

$$a_{2k+1} = a_{-2k-1}, \quad \tilde{a}_{2k+1} = \tilde{a}_{-2k-1}.$$

Proof Assume $N = 2n - 1$, for some $n \in \mathbf{N}$, then

$$\begin{aligned}
\tilde{a}_{2k+1} &= \frac{(-1)^k}{2k+1} \binom{2n-2}{n-1} \binom{2n-1}{n+k} \frac{2n-1}{2^{4n-3}} \\
&= \frac{(-1)^k}{(2k+1)(n-k-1)!(n+k)!} \binom{2n-2}{n-1} \frac{(2n-1)(2n-1)!}{2^{4n-3}} \\
&= \frac{(-1)^{k+1}}{(-2k-1)(n-k-1)!(n+k)!} \binom{2n-2}{n-1} \frac{(2n-1)(2n-1)!}{2^{4n-3}} \\
&= \frac{(-1)^{-k-1}}{(2(-k-1)+1)(n+(-k-1))!(n-(-k-1)-1)!} \\
&\quad \cdot \binom{2n-2}{n-1} \frac{(2n-1)(2n-1)!}{2^{4n-3}} \\
&= \frac{(-1)^{-k-1}}{2(-k-1)+1} \binom{2n-2}{n-1} \binom{2n-1}{n+(-k-1)} \frac{2n-1}{2^{4n-3}} \\
&= \tilde{a}_{2(-k-1)+1} \\
&= \tilde{a}_{-2k-1}
\end{aligned}$$

And

$$a_{2k+1} = \tilde{a}_{2k+1} = \tilde{a}_{-2k-1} = a_{-2k-1}.$$

□

Thus all odd degrees' biorthogonal Coifman wavelet systems are symmetric.

Theorem 3.7 If N is odd, then the biorthogonal Coifman wavelet system of degree N is symmetric, i.e.,

$$a_k = a_{-k}, \quad \tilde{a}_k = \tilde{a}_{-k}.$$

This theorem can also be tested for BCW-0, BCW-1, BCW-2, BCW-3, and BCW-4 in Table 3.1.

3.4.4 Smoothness

Smoothness has been paid lots of attention in wavelet analysis since smooth wavelet systems will be more appropriate for smooth functions. That is one reason why the Haar wavelet system is not widely used in practice. Theorem 3.3 states that if

the scaling function belongs to some smooth function space, then we can have an H^n estimate on the approximation error. So we would like to have some asymptotic estimate of the smoothness of biorthogonal Coifman wavelet systems. By the Sobolev Embedding Theorem, we know that it is sufficient to work on the Sobolev smoothness. Using Eirola's method [14], it is straightforward to prove the following theorem.

Theorem 3.8 For a biorthogonal Coifman wavelet system with degree N , its Sobolev smoothness is

$$s_N = \left(1 - \frac{\log 3}{2 \log 2}\right) N + O(\log N) \approx 0.2075N .$$

We borrowed a Matlab program provided by P. Heller to calculate the Sobolev smoothness of biorthogonal Coifman wavelet systems. The results of the first ten are listed in Table 3.2. The BCW-0, which is the Haar wavelet system, and the BCW-1, which is a piecewise linear spline system, are omitted from the table.

3.4.5 Unconditional Bases

As we know, in a biorthogonal wavelet system with the analysis wavelet function $\psi(x)$ and synthesis wavelet function $\tilde{\psi}(x)$, the wavelet expansion

$$f = \lim_{J \rightarrow \infty} \sum_{j=-J}^J \sum_{k \in \mathbf{Z}} \langle f, \psi_{j,k} \rangle \tilde{\psi}_{j,k}$$

holds only in the weak L^2 sense. Without any other assumption, the $\psi_{j,k}$ or $\tilde{\psi}_{j,k}$ may even fail to constitute frames. Though we can't derive orthonormal bases from symmetric biorthogonal wavelet systems, we do desire to have some reasonable bases, such as unconditional bases where we can interchange the order of the summation in the wavelet expansion. In a Hilbert space, an unconditional basis is also called a Riesz basis. The $\psi_{j,k}, \tilde{\psi}_{j,k}$ constitute two dual Riesz bases if and only if

$$\int_{-\infty}^{\infty} \phi(x) \tilde{\phi}(x - k) dx = \delta_{0,k} \quad \forall k \in \mathbf{Z}. \quad (3.26)$$

In [7], A. Cohen, I. Daubechies, and J.-C. Feauveau proved several equivalent conditions of (3.26). But none of them are easy to verify. Here we give a simple criterion which is a variation of Lawton's condition in the orthogonal wavelet system.

Table 3.2 Sobolev Smoothness of Biorthogonal Coifman Wavelet Systems

N	biorthogonal Coifman wavelet system	
	analysis scaling function	synthesis scaling function
2	1.200	1.839
3	1.179	2.441
4	1.773	2.714
5	1.772	3.175
6	2.292	3.409
7	2.305	3.793
8	2.793	4.004
9	2.815	4.344

Lemma 3.7 In a compactly supported biorthogonal wavelet system with the analysis pair $\{\phi(x), \psi(x)\}$ and synthesis pair $\{\tilde{\phi}(x), \tilde{\psi}(x)\}$, define a multiresolution matrix $T = (t_{k,l})$,

$$t_{k,l} = \frac{1}{2} \sum_{m \in \mathbf{Z}} a_m \tilde{a}_{m+l-2k},$$

where $\{a_k\}$ and $\{\tilde{a}_k\}$ are the scaling vectors with finite length. If

$$\int_{-\infty}^{\infty} \phi(x) dx = \int_{-\infty}^{\infty} \tilde{\phi}(x) dx = 1,$$

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

and T has 1 as a nondegenerate eigenvalue, then

$$\int_{-\infty}^{\infty} \phi(x) \tilde{\phi}(x - k) dx = \delta_{0,k} \quad \forall k \in \mathbf{Z}.$$

Proof Define

$$c_k = \int_{-\infty}^{\infty} \phi(x) \tilde{\phi}(x - k) dx.$$

We have

$$c_k = \int_{-\infty}^{\infty} \left(\sum_{m \in \mathbf{Z}} a_m \phi(2x - m) \right) \left(\sum_{n \in \mathbf{Z}} \tilde{a}_n \tilde{\phi}(2x - 2k - n) \right) dx$$

$$\begin{aligned}
&= \sum_{m,n \in \mathbf{Z}} a_m \tilde{a}_n \int_{-\infty}^{\infty} \phi(2x - m) \tilde{\phi}(2x - 2k - n) dx \\
&= \frac{1}{2} \sum_{m,n \in \mathbf{Z}} a_m \tilde{a}_n \int_{-\infty}^{\infty} \phi(x) \tilde{\phi}(x + m - 2k - n) dx \\
&= \frac{1}{2} \sum_{m,n \in \mathbf{Z}} a_m \tilde{a}_n c_{2k-m+n} \\
&= \frac{1}{2} \sum_{m,l \in \mathbf{Z}} a_m \tilde{a}_{m+l-2k} c_l \\
&= \sum_{l \in \mathbf{Z}} t_{k,l} c_l,
\end{aligned}$$

i.e.,

$$Tc = c.$$

Thus c is an eigenvector of the multiresolution matrix T with eigenvalue 1. Since 1 is a nondegenerate eigenvalue, and $(\cdots, 0, \cdots, 0, 1, 0, \cdots, 0, \cdots)'$ is also an eigenvector with eigenvalue 1, it follows that $c = \gamma(\cdots, 0, \cdots, 0, 1, 0, \cdots, 0, \cdots)'$, or $c_k = \gamma \delta_{0,k}$ for some constant γ . Using the argument similar to the one in Section 2.2.3, we will get $\gamma = 1$. So

$$\int_{-\infty}^{\infty} \phi(x) \tilde{\phi}(x - k) dx = \delta_{0,k}, \quad \forall k \in \mathbf{Z}.$$

□

Based on the lemma, it is easy to check that for BCW-0, BCW-1, BCW-2, BCW-3, and BCW-4, $\psi(x)$ and $\tilde{\psi}(x)$ all constitute dual Riesz bases. A more general theoretical discussion can be found in [38].

3.4.6 Multiplication-Free Discrete Wavelet Transform

Though we didn't expect it, it turned out that the scaling vectors of biorthogonal Coifman wavelet systems are all dyadic rationals, i.e., all the elements in the scaling vectors are of the form $(2p + 1)/2^q$, for some $p, q \in \mathbf{Z}$. This is a really attractive feature since we can therefore implement a very fast multiplication-free discrete wavelet transform on digital computers. It is one of the main advantage of biorthogonal Coifman wavelet systems over other widely used biorthogonal wavelet systems, such as the Cohen-Daubechies-Feauveau 9-7 biorthogonal wavelet system (CDF-97) [7].

Theorem 3.9 In a biorthogonal Coifman wavelet system of degree N , the scaling vectors are dyadic rationals, i.e., $\forall k \in \mathbf{Z}$, there must exist $p_1, p_2, q_1, q_2 \in \mathbf{Z}$, such that

$$a_k = \frac{2p_1 + 1}{2^{q_1}}, \quad \tilde{a}_k = \frac{2p_2 + 1}{2^{q_2}},$$

whenever a_k or \tilde{a}_k are nonzero.

Proof It is clear that the addition, subtraction, or multiplication between two dyadic rationals is still a dyadic rational. Or we can say these three operations are close in dyadic rationals. So all we need to prove is that \tilde{a}_{2k+1} is a dyadic rational, for all nonzero \tilde{a}_{2k+1} . There are two possible cases.

1. if N is even, $N = 2n$,

$$\tilde{a}_{2k+1} = \frac{(-1)^k}{2k+1} \binom{2n-1}{n-1} \binom{2n}{n+k} \frac{2n+1}{2^{4n-1}}.$$

Both $\binom{2n-1}{n-1}$ and $\binom{2n}{n+k}$ are integers. Thus it is sufficient to prove that $(2k+1)$ can divide $\binom{2n-1}{n-1} \binom{2n}{n+k} (2n+1)$. All possible choices for \tilde{a}_{2k+1} being nonzero are $-n \leq k \leq n$. When $k = -1, 0$, or n , the proof is trivial. So first let's look at $0 < k < n$.

Recall that for two positive integers a, b ,

$$a^{d(a,b)} \mid b!,$$

the function $d(a, b)$ is defined by

$$d(a, b) := \sum_{k=1}^{\infty} \lfloor \frac{b}{a^k} \rfloor = \lfloor \frac{b}{a} \rfloor + \lfloor \frac{b}{a^2} \rfloor + \cdots + \lfloor \frac{b}{a^k} \rfloor + \cdots$$

where $\lfloor \cdot \rfloor$ is the integer part of a real number. Since

$$\binom{2n-1}{n-1} \binom{2n}{n+k} (2n+1) = \frac{(2n-1)!(2n+1)!}{(n-1)!n!(n+k)!(n-k)!},$$

it suffices to show that

$$\lfloor \frac{2n-1}{2k+1} \rfloor + \lfloor \frac{2n+1}{2k+1} \rfloor \geq \lfloor \frac{n-1}{2k+1} \rfloor + \lfloor \frac{n}{2k+1} \rfloor + \lfloor \frac{n+k}{2k+1} \rfloor + \lfloor \frac{n-k}{2k+1} \rfloor + 1. \quad (3.27)$$

Assume $n-k = t(2k+1) + r$, where $t, r \in \mathbf{Z}, t \geq 0, 0 \leq r \leq 2k$.

- if $r = 0$, then

$$\begin{aligned} n + k &= t(2k + 1) + 2k, \\ 2n + 1 &= (2t + 1)(2k + 1). \end{aligned}$$

So

$$\lfloor \frac{2n + 1}{2k + 1} \rfloor = 2t + 1 = \lfloor \frac{n - k}{2k + 1} \rfloor + \lfloor \frac{n + k}{2k + 1} \rfloor + 1.$$

We know that

$$\lfloor \frac{2n - 1}{2k + 1} \rfloor \geq \lfloor \frac{n - 1}{2k + 1} \rfloor + \lfloor \frac{n}{2k + 1} \rfloor,$$

Thus (3.27) follows immediately.

- if $1 \leq r \leq k$, then

$$\begin{aligned} n &= t(2k + 1) + (k + r), \\ n - 1 &= t(2k + 1) + (k + r - 1), \\ 2n - 1 &= (2t + 1)(2k + 1) + (2r - 2). \end{aligned}$$

Thus

$$\lfloor \frac{2n - 1}{2k + 1} \rfloor = 2t + 1 = \lfloor \frac{n - 1}{2k + 1} \rfloor + \lfloor \frac{n}{2k + 1} \rfloor + 1.$$

It implies (3.27).

- if $k + 1 \leq r \leq 2k$, then

$$\begin{aligned} n + k &= (t + 1)(2k + 1) + (r - 1), \\ 2n + 1 &= (2t + 2)(2k + 1) + (2r - 2k - 1). \end{aligned}$$

So

$$\lfloor \frac{2n + 1}{2k + 1} \rfloor = 2t + 2 = \lfloor \frac{n - k}{2k + 1} \rfloor + \lfloor \frac{n + k}{2k + 1} \rfloor + 1.$$

Again, we get (3.27) from the above equality.

Now suppose $-n \leq k \leq -2$. Set $l = -k, 2 \leq l \leq n$. We want to show that

$$\lfloor \frac{2n - 1}{2l - 1} \rfloor + \lfloor \frac{2n + 1}{2l - 1} \rfloor \geq \lfloor \frac{n - 1}{2l - 1} \rfloor + \lfloor \frac{n}{2l - 1} \rfloor + \lfloor \frac{n + l}{2l - 1} \rfloor + \lfloor \frac{n - l}{2l - 1} \rfloor + 1. \quad (3.28)$$

Assume $n - l = t(2l - 1) + r$, where $t, r \in \mathbf{Z}, t \geq 0, 0 \leq r \leq 2l - 2$.

- if $0 \leq r \leq l - 2$, then

$$\begin{aligned} n &= t(2l - 1) + (l + r), \\ n - 1 &= t(2l - 1) + (l + r - 1), \\ 2n - 1 &= (2t + 1)(2l - 1) + 2r. \end{aligned}$$

So

$$\lfloor \frac{2n - 1}{2l - 1} \rfloor = 2t + 1 = \lfloor \frac{n - 1}{2l - 1} \rfloor + \lfloor \frac{n}{2l - 1} \rfloor + 1.$$

And (3.28) follows.

- if $l - 1 \leq r \leq 2l - 3$, then

$$\begin{aligned} n + l &= (t + 1)(2l - 1) + (r + 1), \\ 2n + 1 &= (2t + 2)(2l - 1) + (2r - 2l + 3). \end{aligned}$$

So

$$\lfloor \frac{2n + 1}{2l - 1} \rfloor = 2t + 2 = \lfloor \frac{n + l}{2l - 1} \rfloor + \lfloor \frac{n - l}{2l - 1} \rfloor + 1.$$

- if $r = 2l - 2$, then

$$\begin{aligned} n + l &= (t + 2)(2l - 1), \\ 2n + 1 &= (2t + 3)(2l - 1). \end{aligned}$$

So

$$\lfloor \frac{2n + 1}{2l - 1} \rfloor = 2t + 3 = \lfloor \frac{n + l}{2l - 1} \rfloor + \lfloor \frac{n - l}{2l - 1} \rfloor + 1.$$

2. if N is odd, $N = 2n - 1$,

$$\tilde{a}_{2k+1} = \frac{(-1)^k}{2k + 1} \binom{2n - 2}{n - 1} \binom{2n - 1}{n + k} \frac{2n - 1}{2^{4n-3}}.$$

Since $\tilde{a}_{2k+1} = \tilde{a}_{-2k-1}$, it will be sufficient to only prove for the case $1 \leq k \leq n - 2$.

Assume $n - k - 1 = t(2k + 1) + r$, where $t, r \in \mathbf{Z}, t \geq 0, 0 \leq r \leq 2k$.

- if $r = 0$, then

$$\begin{aligned} n &= t(2k + 1) + (k + 1), \\ 2n - 1 &= (2t + 1)(2k + 1), \end{aligned}$$

i.e.,

$$(2k + 1) | (2n - 1).$$

- if $1 \leq r \leq 2k$, since

$$\binom{2n-2}{n-1} \binom{2n-1}{n+k} = \frac{(2n-2)!(2n-1)!}{(n-1)!(n-1)!(n+k)!(n-k-1)!},$$

we will prove the inequality

$$\lfloor \frac{2n-2}{2k+1} \rfloor + \lfloor \frac{2n-1}{2k+1} \rfloor \geq \lfloor \frac{n-1}{2k+1} \rfloor + \lfloor \frac{n-1}{2k+1} \rfloor + \lfloor \frac{n-k-1}{2k+1} \rfloor + \lfloor \frac{n+k}{2k+1} \rfloor + 1.$$

- i. $1 \leq r \leq k$, then

$$n-1 = t(2k+1) + (k+r),$$

$$2n-2 = (2t+1)(2k+1) + (2r-1).$$

So

$$\lfloor \frac{2n-2}{2k+1} \rfloor = 2t+1 = \lfloor \frac{n-1}{2k+1} \rfloor + \lfloor \frac{n-1}{2k+1} \rfloor + 1.$$

- ii. $k+1 \leq r \leq 2k$, then

$$n+k = (t+1)(2k+1) + r,$$

$$2n-1 = (2t+2)(2k+1) + (2r-2k-1).$$

So

$$\lfloor \frac{2n-1}{2k+1} \rfloor = 2t+2 = \lfloor \frac{n-k-1}{2k+1} \rfloor + \lfloor \frac{n+k}{2k+1} \rfloor + 1.$$

□

3.4.7 Convergence to Sinc Wavelet System

The *sinc wavelet system* is a basic wavelet system whose scaling vector $\{a_k^{sinc}, k \in \mathbf{Z}\}$ is defined by

$$a_{2k}^{sinc} = \delta_{0,k}, \quad a_{2k+1}^{sinc} = \frac{(-1)^k 2}{(2k+1)\pi}.$$

It had been a problem for some time to find a sequence of scaling functions with compact supports which approximate the function $\text{sinc}(\pi x) = \frac{\sin \pi x}{\pi x}$, the scaling function of the sinc wavelet system. This problem is important because of the special relation of the sinc function to signal processing applications. The family of biorthogonal Coifman wavelet systems just provides a very suitable candidate which also has growing smoothness.

Theorem 3.10 Suppose $\{\tilde{a}_k^N, k \in \mathbf{Z}\}$ to be the synthesis scaling vector of the biorthogonal Coifman wavelet system of degree N . Then

$$\lim_{N \rightarrow \infty} \left\| (\tilde{a}^N) - (a^{sinc}) \right\|_{l^2} = \lim_{N \rightarrow \infty} \left(\sum_{k \in \mathbf{Z}} (\tilde{a}_k^N - a_k^{sinc})^2 \right)^{1/2} = 0.$$

Before proving Theorem 3.10, we first prove that \tilde{a}^N converges to a^{sinc} termwise.

Lemma 3.8 Assume the same condition as in Theorem 3.10, then

$$\lim_{N \rightarrow \infty} \tilde{a}_k^N = a_k^{sinc}.$$

Proof From the definition, we only need to check

$$\lim_{N \rightarrow \infty} \tilde{a}_{2k+1}^N = a_{2k+1}^{sinc}. \quad (3.29)$$

First let's look at the case when N is even, $N = 2n$. Assume $|k| < n$ (otherwise we can choose a larger N), we have

$$\begin{aligned} \tilde{a}_{2k+1}^N &= \frac{(-1)^k}{2k+1} \binom{2n-1}{n-1} \binom{2n}{n+k} \frac{2n+1}{2^{4n-1}} \\ &= \frac{(-1)^k}{2k+1} \cdot \frac{(2n-1)!(2n+1)!}{2^{4n-1}(n-1)!n!(n-k)!(n+k)!} \end{aligned}$$

We want to show that

$$\lim_{n \rightarrow \infty} \frac{(2n-1)!(2n+1)!}{2^{4n-1}(n-1)!n!(n-k)!(n+k)!} = \frac{2}{\pi}. \quad (3.30)$$

Recall that Stirling's formula states

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi} n^{n+1/2} e^{-n}} = 1$$

So

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{(2n-1)!(2n+1)!}{2^{4n-1}(n-1)!n!(n-k)!(n+k)!} \\ &= \lim_{n \rightarrow \infty} \frac{(2n-1)^{2n-1/2} (2n+1)^{2n+3/2} e^{-1}}{2^{4n} \pi (n-1)^{n-1/2} n^{n+1/2} (n-k)^{n-k+1/2} (n+k)^{n+k+1/2}} \end{aligned}$$

$$\begin{aligned}
&= \lim_{n \rightarrow \infty} \frac{e^{-1}}{2^{4n}\pi} \left(\frac{2n-1}{n-1}\right)^{n-1/2} \left(\frac{2n+1}{n}\right)^{n+1/2} \left(\frac{2n-1}{n-k}\right)^n \\
&\quad \cdot \left(\frac{2n+1}{n+k}\right)^{n+1} \left(\frac{n-k}{n+k}\right)^{k-1/2} \\
&= \lim_{n \rightarrow \infty} \frac{e^{-1}}{2^{4n}\pi} \cdot 2^{n-1/2} e^{1/2} \cdot 2^{n+1/2} e^{1/2} \cdot 2^n e^{1/2} \cdot 2^{n+1} e^{-1/2} \cdot 1 \\
&= \frac{2}{\pi}
\end{aligned}$$

We have used

$$\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x = e, \quad \lim_{x \rightarrow \infty} \left(1 - \frac{1}{x}\right)^x = e^{-1}$$

in the third step. Then (3.30) is proved.

When N is odd, $N = 2n - 1$, the ratio

$$\frac{\tilde{a}_{2k+1}^N}{\tilde{a}_{2k+1}^{N-1}} = \frac{2n-1}{2n+2k} \longrightarrow 1 \quad \text{as } N \rightarrow \infty$$

Then (3.29) also holds when N is odd. □

The next weapon we need is Lebesgue's Dominated Convergence Theorem of $l^2(\mathbf{R})$.

Theorem 3.11 (*Lebesgue's Dominated Convergence Theorem*) Assume $\{c_k^1\}, \{c_k^2\}, \dots$ are a family of $l^2(\mathbf{R})$ sequences with the l^2 norm $\|(c_k)\|_{l^2} = (\sum_k c_k^2)^{1/2}$. Suppose $(d_k) \in l^2(\mathbf{R})$, and $\forall k \in \mathbf{Z}, d_k \geq 0$. If

$$\lim_{n \rightarrow \infty} c_k^n \text{ exists for all } k \in \mathbf{Z},$$

and

$$|c_k^n| \leq d_k \text{ for all } k \in \mathbf{Z}.$$

Then $(\lim_{n \rightarrow \infty} c_k^n)_{k \in \mathbf{Z}} \in l^2(\mathbf{R})$ and

$$\left\| \left(\lim_{n \rightarrow \infty} c_k^n \right)_{k \in \mathbf{Z}} \right\|_{l^2} = \lim_{n \rightarrow \infty} \|(c_k^n)\|_{l^2}.$$

Proof of Theorem 3.10 Since

$$\lim_{N \rightarrow \infty} (\tilde{a}_k^N - a_k^{\text{sinc}}) = 0 \text{ for all } k,$$

to apply Theorem 3.11, we just need to find a $l^2(\mathbf{R})$ dominating sequence $\{d_k, k \in \mathbf{Z}\}$ such that

$$|\tilde{a}_k^N - a_k^{\text{sinc}}| \leq d_k \text{ for all } k.$$

If N is even, $N = 2n$, we can assume $|k| \leq n$ since otherwise $\tilde{a}_{2k+1}^N = 0$ is always bounded. Note that

$$(n-k)!(n+k)! \geq (n!)^2$$

then

$$\begin{aligned} |\tilde{a}_{2k+1}^N| &= \left| \frac{(2n-1)!(2n+1)!}{(2k+1)2^{4n-1}(n-1)!n!(n-k)!(n+k)!} \right| \\ &\leq \left| \frac{(2n-1)!(2n+1)!}{(2k+1)2^{4n-1}(n-1)!(n!)^3} \right| \end{aligned}$$

Set $k = 0$ in (3.30),

$$\lim_{n \rightarrow \infty} \frac{(2n-1)!(2n+1)!}{2^{4n-1}(n-1)!(n!)^3} = \frac{2}{\pi}$$

So for N large enough,

$$\begin{aligned} \frac{(2n-1)!(2n+1)!}{2^{4n-1}(n-1)!(n!)^3} &\leq 1 \\ |\tilde{a}_{2k+1}^N| &\leq \left| \frac{1}{2k+1} \right| \end{aligned}$$

If N is odd, $N = 2n - 1$, since

$$(n+k)!(n-1-k)! \geq (n-1)!n!,$$

we have

$$\begin{aligned} |\tilde{a}_{2k+1}^N| &= \left| \frac{1}{2k+1} \binom{2n-2}{n-1} \binom{2n-1}{n+k} \frac{2n-1}{2^{4n-3}} \right| \\ &= \left| \frac{((2n-1)!)^2}{2^{4n-3}(2k+1)((n-1)!)^2(n+k)!(n-1-k)!} \right| \\ &\leq \left| \frac{((2n-1)!)^2}{2^{4n-3}(2k+1)((n-1)!)^2(n-1)!n!} \right| \\ &= \left| \frac{1}{2k+1} \tilde{a}_1^N \right| \end{aligned}$$

Because

$$\lim_{N \rightarrow \infty} \tilde{a}_1^N = a_1^{\text{sinc}} = \frac{2}{\pi},$$

it follows that

$$\tilde{a}_1^N \leq 1 \text{ for } N \text{ large enough.}$$

Then

$$|\tilde{a}_{2k+1}^N| \leq \left| \frac{1}{2k+1} \right|.$$

Set

$$d_{2k+1} = \left| \frac{1}{2k+1} \right| + |a_{2k+1}^{\text{sinc}}| \leq \left| \frac{2}{2k+1} \right|, \quad d_{2k} = 0$$

we have $(d_k)_{k \in \mathbf{Z}} \in l^2(\mathbf{R})$ and

$$|\tilde{a}_k^N - a_k^{\text{sinc}}| \leq d_k,$$

By Theorem 3.11 and Lemma 3.8,

$$\lim_{N \rightarrow \infty} \left\| \tilde{\alpha}^N - \alpha^{\text{sinc}} \right\|_{l^2} = 0.$$

The theorem follows. \square

The Mallat Algorithm only involves scaling vectors. Using the Cauchy Inequality, one can show that the discrete wavelet transform of biorthogonal Coifman wavelet systems converge to the discrete wavelet transform of the sinc wavelet system.

Corollary 3.2 If $\{s_k, k \in \mathbf{Z}\}$ is a $l^2(\mathbf{R})$ sequence, $(\sum_k s_k^2)^{1/2} < \infty$, then

$$\lim_{N \rightarrow \infty} \text{DWT}(\text{BCW-}N, (s_k)) = \text{DWT}(\text{sinc}, (s_k)).$$

3.5 Conclusions

In this chapter, we have studied various properties of biorthogonal Coifman wavelet systems. In practice these properties really show why biorthogonal Coifman wavelet systems are considered one of the best wavelet systems available today. The image coding evaluation of biorthogonal Coifman wavelet systems is included in Section 5.8.

In [55], a generalization of biorthogonal Coifman wavelet systems is discussed. In it, the condition that the analysis wavelet function and synthesis wavelet function must have the same degree of vanishing moments is relaxed. Interested readers may refer to that paper.

Chapter 4

Orthogonal Coifman Wavelet Systems

As we know, orthogonal wavelet systems not only provide orthonormal bases of $L^2(\mathbf{R})$, but also provide unconditional bases of $L^p(\mathbf{R})$, for $1 < p < \infty$, while biorthogonal ones can't. And orthogonal wavelet transform will preserve the L^2 norm. Thus we can have exact error estimates in the wavelet decomposition domain. So only orthogonal wavelet systems can be included in the wavelet packets, while biorthogonal ones can't. These are several advantages of orthogonal wavelet systems over biorthogonal ones. One main defect of orthogonal wavelet systems is that they can't be symmetric, except for the Haar wavelet systems.

Orthogonal Coifman wavelet systems were first studied by I. Daubechies in [10] (she called these wavelet systems *Coiflets*). In [10] a method to construct these wavelet systems of even degrees (which will be odd degrees in our definition, see below) was proposed and the general existence problem is still open. Orthogonal Coifman wavelet systems seems more "symmetric", more smooth than the Daubechies wavelet systems. In this chapter we will study these orthogonal Coifman wavelet systems.

4.1 Defintion of Orthogonal Coifman Wavelet Systems

Similar to biorthogonal Coifman wavelet systems, orthogonal Coifman wavelet systems are compactly supported orthogonal wavelet systems with vanishing moments equally distributed for the scaling function and wavelet function.

Definition 4.1 An orthogonal wavelet system with compact support is called an *orthogonal Coifman wavelet system* (in short, *OCW*) of degree N if the vanishing moments of the scaling function $\phi(x)$ and the wavelet function $\psi(x)$ are both of degree N , i.e.,

$$\text{Mom}_p(\phi) = \int_{-\infty}^{\infty} x^p \phi(x) dx = \delta_{0,p}, \quad \text{for } p = 0, \dots, N,$$

$$\text{Mom}_p(\psi) = \int_{-\infty}^{\infty} x^p \psi(x) dx = 0, \quad \text{for } p = 0, \dots, N.$$

4.2 Vanishing Moments and Wavelet Approximation

Lemma 3.4 states that the vanishing moment conditions on the scaling function and wavelet function are equivalent to those on the scaling vectors $\{a_k\}$. So Theorem 3.2 holds for orthogonal Coifman wavelet systems. Since these are orthogonal systems, we can prove it in a more direct way.

Theorem 4.1 For an orthogonal Coifman wavelet system of degree N with the scaling function $\phi(x)$, if $f(x) \in C_0^{N,1}(\mathbf{R})$, define, for $j \in \mathbf{Z}$,

$$f^j(x) := 2^{-j/2} \sum_{k \in \mathbf{Z}} f\left(\frac{k}{2^j}\right) \phi_{j,k}(x),$$

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

$$\|f(x) - f^j(x)\|_{L^2} \leq C 2^{-j(N+1)},$$

where C depends only on $f(x)$ and the scaling vector $\{a_k\}$.

Proof In an orthogonal wavelet system with the scaling function $\phi(x)$ and wavelet function $\psi(x)$, we have

$$\|f - f^j\|_{L^2} = \|f - P^j(f)\|_{L^2} + \|P^j(f) - f^j\|_{L^2},$$

where the wavelet orthogonal projection

$$P^j(f) = \sum_{k \in \mathbf{Z}} \left(\int_{-\infty}^{\infty} f(x) \phi_{j,k}(x) dx \right) \cdot \phi_{j,k}(x).$$

We will prove

$$\|f - P^j(f)\|_{L^2} \leq C 2^{-j(N+1)}, \quad (4.1)$$

and

$$\|P^j(f) - f^j\|_{L^2} \leq C 2^{-j(N+1)}. \quad (4.2)$$

By the orthonormality of the wavelet system,

$$\|f - P^j(f)\|_{L^2} = \left\| \sum_{l \geq j} \sum_{k \in \mathbf{Z}} \langle f, \psi_{l,k} \rangle \psi_{l,k} \right\|_{L^2} = \left(\sum_{l \geq j} \sum_{k \in \mathbf{Z}} (\langle f, \psi_{l,k} \rangle)^2 \right)^{1/2},$$

where $\psi_{l,k} = 2^{l/2}\psi(2^l x - k)$. Using the vanishing moments of the wavelet function $\psi(x)$,

$$\begin{aligned}
\langle f, \psi_{l,k} \rangle &= \int_{-\infty}^{\infty} f(x) \cdot 2^{l/2} \psi(2^l x - k) dx \\
&= 2^{-l/2} \int_{-\infty}^{\infty} f\left(\frac{x+k}{2^l}\right) \psi(x) dx \\
&= 2^{-l/2} \int_{-\infty}^{\infty} \left(f\left(\frac{k}{2^l}\right) + f^{(1)}\left(\frac{k}{2^l}\right) \frac{x}{2^l} + \frac{f^{(2)}\left(\frac{k}{2^l}\right)}{2!} \left(\frac{x}{2^l}\right)^2 + \dots \right. \\
&\quad \left. + \frac{f^{(N-1)}\left(\frac{k}{2^l}\right)}{(N-1)!} \left(\frac{x}{2^l}\right)^{N-1} + \frac{f^{(N)}\left(\frac{\theta k + (1-\theta)x}{2^l}\right)}{N!} \left(\frac{x}{2^l}\right)^N \right) \psi(x) dx \\
&= 2^{-l/2} \int_{-\infty}^{\infty} \frac{f^{(N)}\left(\frac{\theta k + (1-\theta)x}{2^l}\right)}{N!} \left(\frac{x}{2^l}\right)^N \psi(x) dx \\
&= \frac{2^{-l(N+\frac{1}{2})}}{N!} \int_{-\infty}^{\infty} f^{(N)}\left(\frac{\theta k + (1-\theta)x}{2^l}\right) x^N \psi(x) dx \\
&= \frac{2^{-l(N+\frac{1}{2})}}{N!} \int_{-\infty}^{\infty} \left(f^{(N)}\left(\frac{\theta k + (1-\theta)x}{2^l}\right) - f^{(N)}\left(\frac{k}{2^l}\right) \right) x^N \psi(x) dx,
\end{aligned}$$

where $0 \leq \theta \leq 1$. Now using the same estimate as in the proof of Theorem 3.2, one can show that

$$\|f - P^j(f)\|_{L^2} \leq C 2^{-j(N+1)},$$

where C depends only on $f(x)$ and $\psi(x)$.

For (4.2), we have

$$\begin{aligned}
\|P^j(f) - f^j\|_{L^2} &= \left\| \left(\sum_{k \in \mathbf{Z}} \langle f, \phi_{j,k} \rangle \phi_{j,k}(x) \right) - \left(2^{-j/2} \sum_{k \in \mathbf{Z}} f\left(\frac{k}{2^j}\right) \phi_{j,k}(x) \right) \right\|_{L^2} \\
&= \left\| \sum_{k \in \mathbf{Z}} \left(\langle f, \phi_{j,k} \rangle - 2^{-j/2} f\left(\frac{k}{2^j}\right) \right) \phi_{j,k}(x) \right\|_{L^2} \\
&= \left(\sum_{k \in \mathbf{Z}} \left(\langle f, \phi_{j,k} \rangle - 2^{-j/2} f\left(\frac{k}{2^j}\right) \right)^2 \right)^{1/2}
\end{aligned}$$

Using the vanishing moments of $\phi(x)$,

$$\begin{aligned}
\langle f, \phi_{j,k} \rangle &= \int_{-\infty}^{\infty} f(x) \cdot 2^{j/2} \phi(2^j x - k) dx \\
&= 2^{-j/2} \int_{-\infty}^{\infty} f\left(\frac{x+k}{2^j}\right) \phi(x) dx
\end{aligned}$$

$$\begin{aligned}
&= 2^{-j/2} \int_{-\infty}^{\infty} \left(f\left(\frac{k}{2^j}\right) + f^{(1)}\left(\frac{k}{2^j}\right) \frac{x}{2^j} + \frac{f^{(2)}\left(\frac{k}{2^j}\right)}{2!} \left(\frac{x}{2^j}\right)^2 + \dots \right. \\
&\quad \left. + \frac{f^{(N-1)}\left(\frac{k}{2^j}\right)}{(N-1)!} \left(\frac{x}{2^j}\right)^{N-1} + \frac{f^{(N)}\left(\frac{\theta k + (1-\theta)x}{2^j}\right)}{N!} \left(\frac{x}{2^j}\right)^N \right) \phi(x) dx \\
&= 2^{-j/2} f\left(\frac{k}{2^j}\right) + 2^{-j/2} \int_{-\infty}^{\infty} \frac{f^{(N)}\left(\frac{\theta k + (1-\theta)x}{2^j}\right)}{N!} \left(\frac{x}{2^j}\right)^N \phi(x) dx,
\end{aligned}$$

where $0 \leq \theta \leq 1$. So

$$\begin{aligned}
\langle f, \phi_{j,k} \rangle - 2^{-j/2} f\left(\frac{k}{2^j}\right) &= 2^{-j/2} \int_{-\infty}^{\infty} \frac{f^{(N)}\left(\frac{\theta k + (1-\theta)x}{2^j}\right)}{N!} \left(\frac{x}{2^j}\right)^N \phi(x) dx, \\
\|P^j(f) - f^j\|_{L^2} &\leq C 2^{-j(N+1)}.
\end{aligned}$$

Combining (4.1) and (4.2), the theorem follows. \square

The above proof gives us insight of the different roles of the vanishing moments of the scaling function $\phi(x)$ and wavelet function $\psi(x)$. The vanishing moments of $\psi(x)$ will reduce the error in the wavelet orthogonal projection, or the distance from $f(x)$ to the projection space, which is spanned by $\{\phi_{j,k}(x), k \in \mathbf{Z}\}$. The vanishing moments of $\phi(x)$, on the other hand, will reduce the distance between the wavelet orthogonal projection and the wavelet sampling approximation in the projection space.

4.3 Existence and Construction

A big problem concerning orthogonal Coifman wavelet systems is the existence problem. Till now we still don't know whether orthogonal Coifman wavelet systems exist for an arbitrary degree. I. Daubechies discussed the construction of odd degrees with a preimposed forms. But even for the odd degrees, the existence problem is not solved. Here we propose a numerical method starting with biorthogonal Coifman wavelet systems.

Biorthogonal Coifman wavelet systems and orthogonal Coifman wavelet systems are connected via the same vanishing moments imposed on scaling functions and wavelet functions. The only difference between these two is that the quadratic condition in orthogonal Coifman wavelet systems

$$\sum_{k \in \mathbf{Z}} a_k a_{k+2l} = 2\delta_{0,l}, \quad \forall l \in \mathbf{Z}$$

is replaced by the bilinear condition

$$\sum_{k \in \mathbf{Z}} a_k \tilde{a}_{k+2l} = 2\delta_{0,l}, \quad \forall l \in \mathbf{Z}.$$

How to construct orthogonal Coifman wavelet systems from biorthogonal Coifman wavelet systems is the problem we are tackling on. One of the most powerful method to approximation a solution of some system with a known starting point is the well-known Newton's method. One drawback of Newton's method is that even though numerically Newton's iterates converge to some point, we still can't assert the limit will be exactly a solution. By utilizing a fundamental result due to L. Kantorovich, the existence for some orthogonal Coifman wavelet system can be proved theoretically.

Theorem 4.2 (*Newton-Kantorovich Theorem*) Assume D is a bounded, open subset of \mathbf{R}^n , $f : D \rightarrow \mathbf{R}^n$ is C^1 on a convex set $D_0 \subset D$ such that

$$\|f'(x) - f'(y)\| \leq \gamma \|x - y\|, \quad \forall x, y \in D_0.$$

Suppose that there exists an $x_0 \in D_0$ such that $\|(f'(x_0))^{-1}\| \leq \beta$ and $\alpha = \beta\gamma\eta \leq 1/2$, where $\eta \geq \|(f'(x_0))^{-1} f(x_0)\|$. Set

$$t^* = (\beta\gamma)^{-1}[1 - (1 - 2\alpha)^{1/2}], \quad t^{**} = (\beta\gamma)^{-1}[1 + (1 - 2\alpha)^{1/2}],$$

and assume that $\overline{B}(x_0, t^*) \subset D_0$. Then the Newton iterates

$$x_{k+1} = x_k - (f'(x_k))^{-1} f(x_k), \quad k = 0, 1, \dots,$$

are well-defined, remain in $\overline{B}(x_0, t^*)$ and converge to a solution x^* of $f(x) = 0$ which is unique in $B(x_0, t^{**}) \cap D_0$. Here $B(x_0, r)$ denotes the open ball of radius r about the point x_0 , and $\overline{B}(x_0, r)$ is the close ball.

The basic idea of the proof is to construct a majorizing sequence for x_k . For a complete proof and some applications of the Newton-Kantorovich theorem, see [26] and [27].

The Newton-Kantorovich Theorem is just one example that one can get theoretical results from numerical computation. In our case, this theorem will enable us to prove the existence of orthogonal Coifman wavelet systems of degrees up to 9.

It is clear from the definition that the degree 0 orthogonal Coifman wavelet system is exactly the Haar wavelet system with $\{a_0 = 1, a_1 = 1\}$. Next we will go for orthogonal Coifman wavelet system with higher degrees.

a. $N = 1$

b. $N = 2$

The significance of the Newton-Kantorovich Theorem is that even if we don't know the existence of the solution of $f(x) = 0$, we can still apply the Newton's iteration method. If the numerical results are good enough (i.e., the conditions of the Newton-Kantorovich Theorem are satisfied), then it follows that $f(x) = 0$ has a solution and the numerical results will give a fast approximation.

Thus we propose our Newton's method algorithm to get orthogonal Coifman wavelet systems, starting from biorthogonal Coifman wavelet systems. Note that since the vanishing moment conditions are just linear equations on the scaling vector $\{a_k\}$, every Newton iterate x_k will always satisfy the vanishing moment conditions.

Newton's method algorithm:

1. Take the synthesis scaling vector of the biorthogonal Coifman wavelet system of degree N as the initial point x_0 .
2. Compute Newton iterates starting from x_0 .
3. If for some x_{k_0} , the Newton-Kantorovich condition

$$\gamma \cdot \|(f'(x_{k_0}))^{-1}\| \cdot \|(f'(x_{k_0}))^{-1} f(x_{k_0})\| \leq 1/2$$

is satisfied, where γ is the Lipschitz constant of $f'(x)$, then define $y_0 = x_{k_0}$.

4. An error estimate is needed if we are trying to prove the above inequality from numerical results.
5. From the Newton-Kantorovich Theorem, the Newton iterates starting from y_0 converge to a solution of $f(x) = 0$.
6. Check that 1 is not a degenerate eigenvalue of the corresponding multiresolution operator T_α . Then the iterate y_k we choose will be a good approximation to the scaling vector of the orthogonal Coifman wavelet system of degree N .

Based on the above algorithm, here we give the scaling vectors of orthogonal Coifman wavelet systems of degree through 0 to 9.

Some illustrations of orthogonal Coifman wavelet systems are included in Figure 4.1 and 4.2

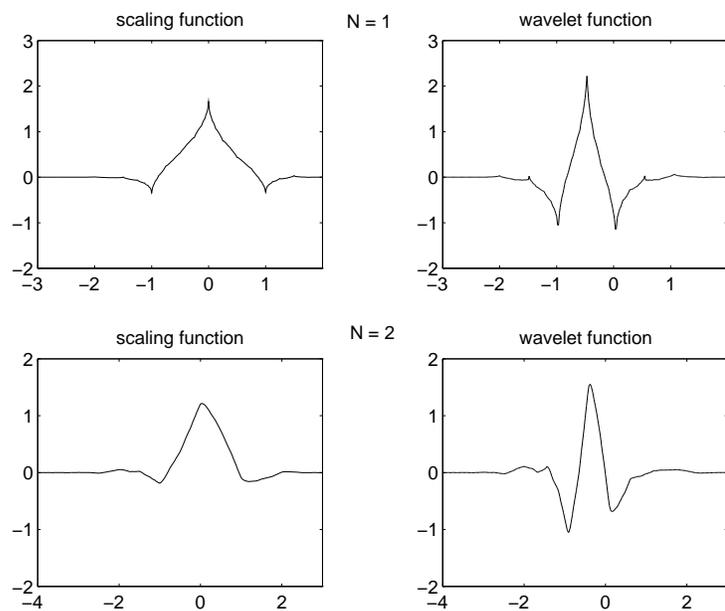


Figure 4.1 Orthogonal Coifman Wavelet Systems of Degree 1 and 2

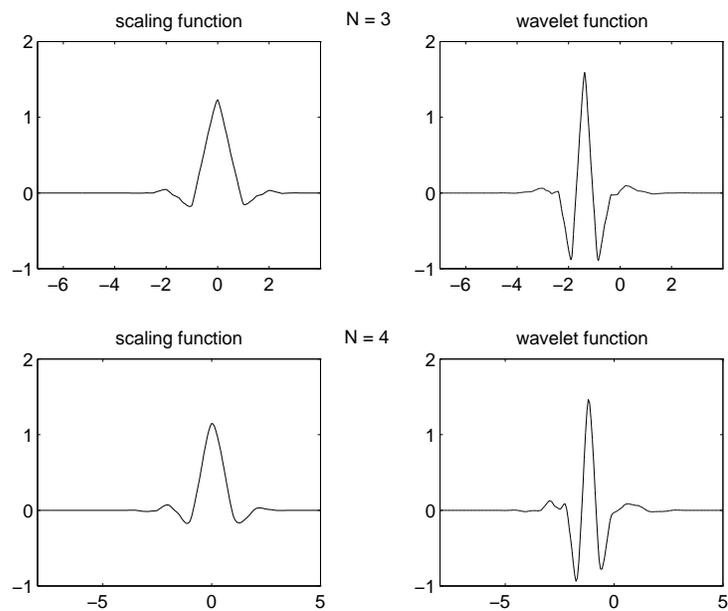


Figure 4.2 Orthogonal Coifman Wavelet Systems of Degree 3 and 4

4.4 Conclusions

In the chapter we have seen that the vanishing moments of the scaling function and the wavelet function play different roles in the wavelet sampling approximation. Depending on the specific application, we can distribute more vanishing moments on the scaling function, or more vanishing moments on the wavelet function, or let them have the same vanishing moments. The third case is much more valuable, due to the wavelet approximation theorem. Also a numerical method is proposed to construct orthogonal Coifman wavelet systems. Usually it is impossible to verify a Cauchy sequence by numerical computation. But with the Newton-Kantorovich Theorem, we can prove the existence of orthogonal Coifman wavelet systems by numerical computation and get a good approximate solution by the Newton's method.

Chapter 5

Image Coding

The advent of multimedia computing has led to an increased demand for digital images. To make widespread use of digital imagery practical, some form of data compression must be used, since the storage and manipulation of these images in their raw form is very expensive. Thus image data compression has been widely used in still images, medical imaging, seismic waves, synthetic aperture radar images, FBI fingerprints, etc. It has been a hot topic both in research and in application for a long time. In the last ten years, the wavelet analysis has become a cutting-edge technology in this area. Since the wavelet transform has very good localization properties in both the spatial domain and the frequency domain, it can handle non-stationary signals very efficiently and provide efficient compression algorithms.

In this chapter, we will present our work in image compression. After briefly viewing some background material, we will first study two of the best image compression algorithms. One is J. Shapiro's embedded zerotree wavelet algorithm. The most significant contribution of Shapiro's work is that it is a whole new idea and sparked lots of other interesting compression algorithms. Another one we will look at is A. Said and W. A. Pearlman's codetree algorithm. This algorithm is a combination of Shapiro's algorithm and their set partitioning sorting algorithm. Motivated by these three people's nice work, we propose a new embedded wavelet-based image compression algorithm, the *Wavelet-Difference-Reduction* algorithm. It combines the discrete wavelet transform, differential coding, binary reduction, ordered bit plane transmission and adaptive arithmetic coding. After the detailed description of this new algorithm, we will compare it with the other two studied early this chapter and see various applications of the new algorithm. And we propose a novel method to process image data in the compressed wavelet domain.

5.1 Some Background

A typical transform image coder consists of three subsystems: a transform subsystem, a quantization subsystem, and an entropy coding subsystem. In the transform subsystem, the input image data is stored in another form by some invertible transformation. The purpose of the transform subsystem is to remove the redundancies in the original image data to a large extent. In the quantization subsystem, the quantizer (scalar quantizer or vector quantizer) allocates different number of bits for the representation of different transform coefficients. The quantization is the only step that some information of the original image data will be lost. The entropy coding subsystem is a lossless data compression process on the stream of quantized transform coefficients. Typical choices of the lossless data compression algorithms are Huffman coding, or arithmetic coding [57].

Right after I. Daubechies's work on compactly supported orthogonal wavelet systems [8], wavelet analysis has been immediately applied to image compression and showed its big potential, since wavelet transform can remove the spatial and spectral redundancies of the image data pretty well. Thus it becomes the ideal choice for the transform subsystem. The main difference between different wavelet-based image coder is in the quantization subsystem. We can choose either scalar quantizer or vector quantizer. And in each quantizer, lots of different methods have been proposed to improve the compression performance.

Embedded image coding is a way to successively approximate the original image. At the beginning of the bit steam, the embedded code contains all lower rate (here the rate is the bit rate, or the number of bits per pixel) codes. As the encoding continues, more and more fine detail will be sent to the embedded code, in the order of importance. The word "embedded" means that all the information in the lower rate codes is contained in the higher rate codes. As J. Shapiro pointed out in his paper [45], embedded coding is similar in spirit to binary finite precision representation of real numbers. And using embedded coding, the encoder can stop at any point when some target rate or distortion metric is met. Also the decoder can stop at any point and give a reconstruction image, as if the encoder had terminated its encoding job at the same lower rate. Two examples of embedding image codings are J. Shapiro's embedded zerotree wavelet algorithm [45] and A. Said and W. A. Pearlman's codetree algorithm [43], which we will study next.

5.2 Shapiro's Embedded Zerotree Wavelet Algorithm

In his celebrated paper [45], J. Shapiro presented the embedded zerotree wavelet algorithm (EZW). First he defined a spatial orientation tree structure. In a wavelet decomposition domain* with N scales, except for the high-pass coefficients at the finest scale, HL_1, LH_1, HH_1 , and the low-pass coefficients at the coarsest scale, LL_N , every coefficient at a given scale is related to a set of four coefficients at the next finer scale of similar orientation. The coefficient at the coarse scale is called the *parent* and the four coefficients corresponding to the same spatial location at the next finer scale of same orientation are called *children*. For every coefficient in LL_N , we define three coefficients, from HL_N, LH_N, HH_N , respectively, with the same spatial location, as its children. A parent-child relationship is illustrated in Figure 5.1. A scanning of the coefficients is performed in such a way that no child node is scanned before its parent. A typical scanning pattern is indicated in Figure 5.2. The scan begins at LL_N , then HL_N, LH_N, HH_N , at which point it moves on to scale $N-1$, etc. A wavelet coefficient x is defined as *insignificant* with respect to a threshold T if $|x| < T$, otherwise x is said to be *significant*. And x is an element of a zerotree if x and all its descendants are insignificant. An element x is called a *zerotree root* if it is an element of zerotree but its parent is not an element of zerotree. Moreover, x is said to be an isolated zero if x is insignificant but has some significant descendant.

With these settings, we are ready to encode the image using EZW. First select a threshold T such that $|x_j| < 2T$ for all the wavelet transform coefficients x_j , and for some $j_0, T \leq |x_{j_0}|$. Initially put all the wavelet transform coefficients on the dominant list, with the order shown in Figure 5.2. We begin with the dominant pass. For each x_j on the dominant list, if x_j is significant with respect to T , then output its sign and move x_j to the subordinate list. If x_j is insignificant, then output a symbol (say "I") if it is an isolated zero, or output another symbol (say "Z") if it is a zerotree root. In other cases, nothing will be output since x_j is predictably insignificant. The dominant pass is followed by a subordinate pass. First, the threshold T is divided by 2. And the refinement value (the meaning of refinement values will be given in Section 5.4) of x_j in the subordinate list with respect to T will be output. The subordinate pass is followed by the dominant pass of the next round. And the cycle keeps going until some target rate is met.

*We assume a two-dimensional setting with wavelets scaled by power of 2, although this is not essential for these algorithms.

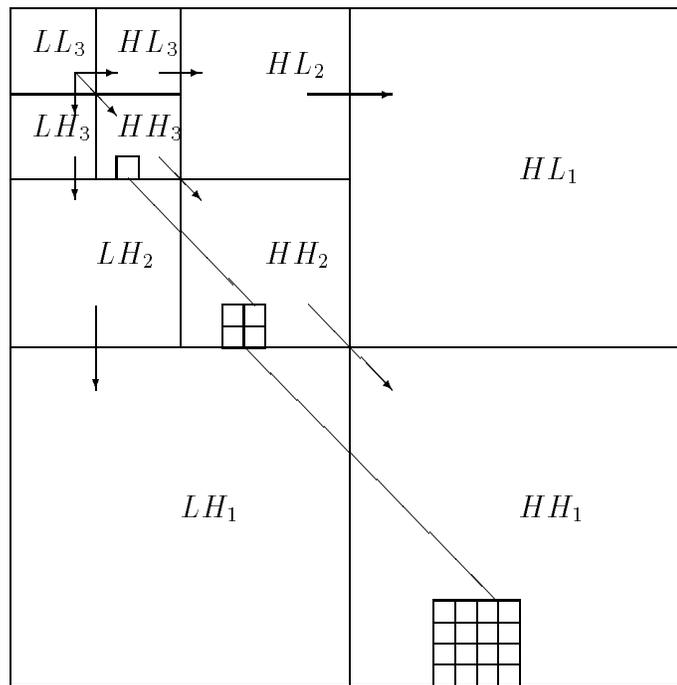


Figure 5.1 Parent-Child Dependencies in the Wavelet Decomposition Domain with 3 Scales.

The EZW takes advantage of the predictably insignificant coefficients and the adaptive arithmetic coding can encode the symbol stream very efficiently.

5.3 Said and Pearlman's Codetree Algorithm

A. Said and W. A. Pearlman introduced a set partitioning sorting algorithm and developed their state-of-the-art Said-Pearlman-Codetree algorithm (SPC) [43] after Shapiro's work. Using the same spatial orientation tree structure, they defined three ordered lists, **LIS**, list of insignificant sets, **LIP**, list of insignificant pixels, and **LSP**, list of significant pixels. **LIS** is the list of descendants sets, with two types **A** and **B**, and each element of **LIS** is represented by the ancestor. Type **A** is a set of all descendants of some coefficient; and type **B** is a set of all descendants excluding its children, so it looks like the union of four (or three) elements in **A**, yet it isn't generated in this manner.

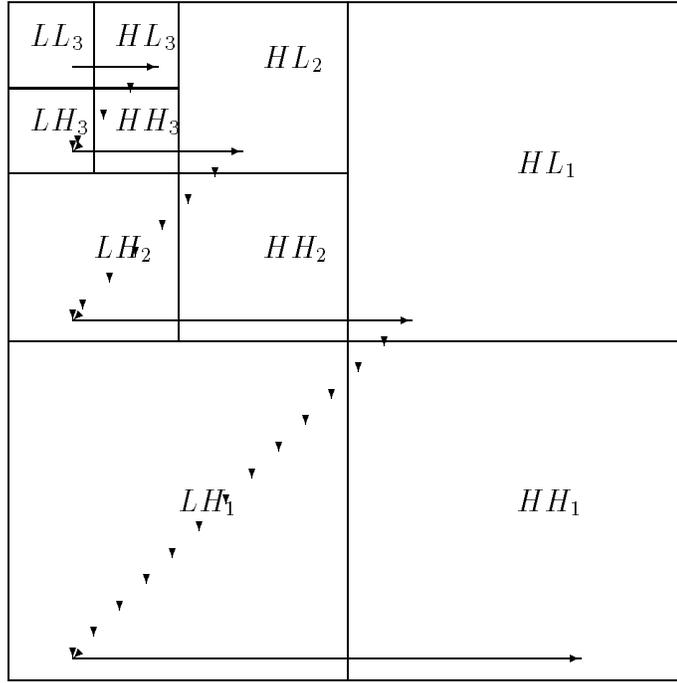


Figure 5.2 Scanning Order of the Wavelet Transform Coefficients with 3 Scales.

Initially **LSP** is empty, and **LIP** and **LIS** are equal to LL_N . In the sorting pass, all x_j in **LIP** will be checked for significance, and the test result will be output. When some x_j is found significant, move it to **LSP** and output its sign right after the test result. Then for x_j in **LIS**, if it is type **A**, i.e., the set of all descendants of x_j , check if x_j has any significant descendants and output the result. If yes, check the children of x_j and output the test result. If some child is significant, move the child to **LSP** and output its sign. If some child is not significant, move the child to **LIP**. If x_j has other descendants than children, i.e., x_j is at the third scale level or even coarser, move x_j to type **B**. (That's where type **B** comes from.) For x_j in **LIS** with type **B**, if it has significant descendants (not including its children), move x_j from **LIS** and move all its children to **LIS** as type **A**. After the sorting pass, we have a refinement pass. In the refinement pass, the refinement values of elements in **LSP** except those included in the sorting pass of this round (since these coefficients' refinement values must be "1"), will be output. Then the threshold is divided by 2, and it goes back to the sorting pass and another around starts.

In SPC, the set partitioning sorting algorithm is very efficient such that the symbol stream can be represented by using only two symbols, “0” and “1”. (Even the positive and negative signs can be represented by “0” and “1”.) Without arithmetic coding, its performance is already superior to EZW.

5.4 Wavelet-Difference-Reduction Algorithm

In this section, we propose a new embedded image coding algorithm. Similar to EZW and SPC, it consists of two parts, a sorting pass and a refinement pass. In the sorting pass, we will show how to use reduced indices of wavelet transform coefficients to encode the positions of significant coefficients directly. As in Section 5.2, a wavelet transform coefficient x is defined as *significant* with respect to a threshold T if $|x| \geq T$, otherwise x is said to be *insignificant*.

5.4.1 Differential Coding

Differential coding [18] takes the difference of values of successive elements. It usually produces a set of smaller values, compared with the original set. For example, if the elements of a set \mathcal{S} are preordered in a monotonically increasing order, then the differential coding result \mathcal{S}' will be a set with smaller values. For example, if

$$\mathcal{S} = \{1, 2, 5, 36, 42\},$$

then

$$\mathcal{S}' = \{1, 1, 3, 31, 6\}.$$

As it is clear, \mathcal{S}' contains relatively smaller values than \mathcal{S} .

We will call \mathcal{S}' the *difference set* of \mathcal{S} . And we make the convention that the value of the first element will not be changed in differential coding. Meanwhile it is straightforward to retrieve back the original set \mathcal{S} from the difference set \mathcal{S}' by taking the partial sum of \mathcal{S}' . So we can say that \mathcal{S} and \mathcal{S}' contain the same information though \mathcal{S}' will be much easier for transmission and storage.

5.4.2 Binary Reduction

Binary reduction [15] is one of the representation of positive integers, with the shortest representation length. By definition, the binary reduction removes all the leading “0” bits and the first “1” in the binary representation of a number, or equivalently,

removes the most significant binary digit. For example, to get the binary reduction of 19, since

$$19_{10} = 10011_2,$$

where the subindex 10 means the decimal representation, and the subindex 2 means the binary representation, we have the binary reduction of 19 is 0011.

When we have a set \mathcal{S} of positive integers, we can apply binary reduction to each element to get the *reduction set* $\mathcal{R}(\mathcal{S})$ of \mathcal{S} . For example, if

$$\mathcal{S} = \{1, 1, 3, 31, 6\},$$

then the reduction set

$$\mathcal{R}(\mathcal{S}) = \{, , 1, 1111, 10\}. \quad (5.1)$$

Note that there are no coded symbols before the first two commas “,” in $\mathcal{R}(\mathcal{S})$. Since all elements in \mathcal{S} are positive integers, the inverse operation of the binary reduction will be adding a “1” as the most significant bit in the binary representation. Applying this inverse operation on the reduction set $\mathcal{R}(\mathcal{S})$ will produce the original set \mathcal{S} . Again, the two sets \mathcal{S} and $\mathcal{R}(\mathcal{S})$ will contain the same information while $\mathcal{R}(\mathcal{S})$ will require less space in digital computers. In practice, one will need some end of message symbol to separate different elements in the reduction set $\mathcal{R}(\mathcal{S})$, like the comma “,” in (5.1).

5.4.3 Coding the Significant Maps

The following problem is considered, how to code the indices of significant wavelet transform coefficient, or the significant maps, in an efficient way? Both J. Shapiro's embedded zerotree wavelet algorithm and A. Said and W. A. Pearlman's codetree algorithm use spatial orientation tree structures to implicitly locate the significant coefficient, here we present a direct approach based on differential coding and binary reduction.

Assume \mathcal{S} is the index set of significant wavelet transform coefficients. We can also assume the elements of \mathcal{S} are ordered in a monotonically increasing order. To code \mathcal{S} , the first step, of course, is the differential coding. After differential coding, we get the difference set \mathcal{S}' . From \mathcal{S}' , we can get back \mathcal{S} . Can the difference set \mathcal{S}' be further coded? Well, since \mathcal{S}' is a set of positive integers (note that \mathcal{S} is monotonically increasing), we can apply the binary reduction on \mathcal{S}' to produce the reduction set

$\mathcal{R}(\mathcal{S}')$. Because $\mathcal{R}(\mathcal{S}')$ and \mathcal{S}' contain the same information, the index set \mathcal{S} can be derived by decoding $\mathcal{R}(\mathcal{S}')$. So we will $\mathcal{R}(\mathcal{S}')$ to represent \mathcal{S} . For example, if

$$\mathcal{S} = \{1, 2, 5, 36, 42\}, \quad (5.2)$$

then we have

$$\mathcal{S}' = \{1, 1, 3, 31, 6\},$$

and

$$\mathcal{R}(\mathcal{S}') = \{., ., 1, 1111, 10\}.$$

The coding result of the index set $\{1, 2, 5, 36, 42\}$ is $\{., ., 1, 1111, 10\}$. It is a very compact representation of the indices.

Actually the concept of combining the differential coding and the binary reduction is a fairly general concept and not specific to the wavelet decomposition domain. For example, it can be applied to the Partition Priority Coding (PPC) [23], and one would expect some possible improvement in the image coding results.

5.4.4 Outline of the Algorithm

With the differential coding and binary reduction, the index j of a significant wavelet transform coefficient can be coded very efficiently. And we call the action of the differential coding and binary reduction on wavelet transform coefficients *Wavelet-Difference-Reduction* (in short *WDR*). We now formulate our *WDR image compression algorithm* using this notion of Wavelet-Difference-Reduction. We will use the language “sorting pass” and “refinement pass” from [43].

After taking the discrete wavelet transform of an image, all wavelet transform coefficients will be ordered from coarser scale to finer scale, exactly the same ordering as in Figure 5.2. In an N scale decomposition, it will be LL_N, HL_N, LH_N, HH_N , then $HL_{N-1}, LH_{N-1}, HH_{N-1}, \dots, HL_1, LH_1$, and HH_1 . This order is based on the hypothesis that more significant wavelet transform coefficients are expected to appear in the coarser scale, and statistically this hypothesis is true in almost all the cases. Also this is the natural order of wavelet transform coefficients in the context of signal processing. Three ordered lists of wavelet transform coefficients are defined, **LSC** (list of significant coefficients), **LTP** (a temporary list, for the significant coefficients found in a given sorting pass round), and **LIC** (list of insignificant coefficients). Initially both **LSC** and **LTP** are empty, and **LIC** contains all the wavelet transform coefficients, with the order as shown in Figure 5.2. And the initial threshold T is

chosen such that $|x_j| < 2T$ for all the wavelet transform coefficients x_j , and for some j_0 , $|x_{j_0}| \geq T$. Output the initial threshold T .

First we have a sorting pass. In the sorting pass, all significant coefficients in **LIC** with respect to T will be moved out and put into **LTP**. Let \mathcal{S} be the indices (in **LIC**) of these significant coefficients. Output the reduction set $\mathcal{R}(\mathcal{S}')$ of the difference set of \mathcal{S} . Instead of using “,” as the end of message symbol to separate different elements in $\mathcal{R}(\mathcal{S}')$, we will take the signs (either “+” or “-”) of these significant coefficients as the end of message symbol. For example, if $\mathcal{S} = \{1, 2, 5, 36, 42\}$ as in (5.2), and the signs of these four significant coefficients are “+ - + + -”, then the encoding output $\mathcal{R}(\mathcal{S}')$ will be “+ - 1 + 1111 + 10-”. Then update the indexing in **LIC**, for example, if x_3 is moved to **LTP**, then all coefficients after x_3 in **LIC** will have their indices subtracted by 1, and so on.

The sorting pass is followed by a refinement pass. In the refinement pass, an additional bit of precision of all the coefficients in **LSC** will be obtained. Or equivalently, the width of the uncertainty interval of coefficients in **LSC** will be cut in half. Before the refinement pass, the uncertainty interval of coefficients in **LSC** is $[0, 2T)$. During the refinement pass, those coefficients in **LSC** with magnitude falling in $[0, T)$ will have the refinement value “0”, and those in **LSC** with magnitude falling in $[T, 2T)$ will have the refinement value “1”. And these refinement values “0” or “1” will be the output. For example, if the magnitude of a coefficient in **LSC** is known to be in $[32, 64)$, then it will be decided at this stage whether it is in $[32, 48)$ or $[48, 64)$. And a “0” symbol will indicate it is in the lower half $[32, 48)$, while a “1” symbol will indicate it is in the upper half $[48, 64)$. Output all these refinement values “0” and “1”. Note that for the first round, there will be no output, since those significant coefficients just found in the sorting pass are all in **LTP**, and **LSC** is still empty.

Then append **LTP** to the end of **LSC**, $\mathbf{LSC} = \mathbf{LSC} \cup \mathbf{LTP}$. Reset **LTP** to the empty set. And T is divided by 2. Another round begins with the sorting pass.

The adaptive arithmetic coding [57] is used on the resulting symbol stream in the sorting pass and refinement pass for each round. When the given rate or distortion metric is met, the encoding stops.

To make the above algorithm more clear, we will use an example to illustrate the steps in the WDR algorithm. Only the symbol stream before adaptive arithmetic coding is shown. A 3-scale wavelet transform of an 8×8 image is borrowed from [45]. The array of values is shown in Figure 5.3. The largest wavelet transform coefficient magnitude is 63, and we will choose $T = 32$ as the initial threshold.

63	-34	49	10	7	13	-12	7
-31	23	14	-13	3	4	6	-1
15	14	3	-12	5	-7	3	9
-9	-7	-14	8	4	-2	3	2
-5	9	-1	47	4	6	-2	2
3	0	-3	2	3	-2	0	4
2	-3	6	-4	3	6	3	6
5	11	5	6	0	3	-4	4

Figure 5.3 Example of 3-scale Wavelet Transform of an 8×8 Image.

In the first round, during the sorting pass, there are four significant coefficients, 63, -34, 49, and 47, with the indices 1, 2, 5, and 36. So $\mathcal{S} = \{1, 2, 5, 36\}$. Now apply the differential coding and binary reduction, we get $\mathcal{R}(\mathcal{S}') = \{, , 1, 1111\}$. The sign of a significant coefficient will be followed by its reduced index, so the resulting symbol stream will be “+ - 1 + 1111+”. There is no “0” or “1” preceding the first “+”, which means its reduced index is nothing, then we can know its index difference will be 1. Also for the first “-”, we can know the index difference for the second significant coefficient is 1. And for the second “+”, taking the inverse of binary reduction operation, it will produce 3 from the coded symbol “1”. Thus the index difference will be 3. And so on. When updating the indices in **LIC**, $x_3 = -31, x_4 = 23$ will have the new indices 1 and 2, since 63 and -34 are all moved out. The wavelet transform coefficients after 49 and before 47 will have their indices subtracted by 3, those after 47 will have their indices subtracted by 4. Now **LTP** = {63, -34, 49, 47}. Since **LSC** is empty for this round, there will be no output for refinement values. And **LSC** = **LSC** \cup **LTP** = {63, -34, 49, 47}, and we reset **LTP** to the empty set. At $T = 32$, all four coefficients in **LSC** have their magnitudes in the interval [32, 64),

and we will take the center of this interval as their reconstruction values. Thus, the reconstruction values of these four significant coefficients will be $\{48, -48, 48, 48\}$. Then T is divided by 2, $T = 32/2 = 16$. The second round begins.

In the second round, there are two significant coefficients, -31 , and 23 , with the indices 1, and 2. Note that these are the updated indices. We have $\mathcal{S} = \{1, 2\}$, and $\mathcal{R}(\mathcal{S}') = \{, , \}$. The symbol stream will be “-+”, with $\mathbf{LTP} = \{-31, 23\}$. Then updating the indices in \mathbf{LIC} , in this case, all indices are subtracted by 2. In the refinement pass $\mathbf{LSC} = \{63, -34, 49, 47\}$. The magnitudes of these four coefficients are all in $[32, 64)$, and they will be refined to an additional bit of precision, i.e., either in $[32, 48)$ or in $[48, 64)$. The uncertainty interval is $[0, 32)$, and those magnitudes falling in $[0, 16)$ will have the refinement values “0”, and those falling in $[16, 32)$ will have the refinement values “1”. Since $63 - 32 = 31 > 16$, the refinement value for 63 is 1; $|-34 - (-32)| = 2 < 16$, the refinement value for -34 is 0; $49 - 32 = 17 > 16$, the refinement value for 49 is 1; $47 - 32 = 15 < 16$, the refinement value for 47 is 0. The symbol stream will be “1010”. Now $\mathbf{LSC} = \mathbf{LSC} \cup \mathbf{LTP} = \{63, -34, 49, 47, -31, 23\}$, and \mathbf{LTP} is reset to empty. The reconstruction values of these six significant coefficients at this stage will be $\{56, -40, 56, 40, -24, 24\}$. And $T = 16/2 = 8$, another round begins...

The encoding will stop when some target rate is met.

For this example, there will be at most six rounds. After the sixth round, the resulting compressed bit stream will be lossless.

5.5 Comparisons of Three Algorithms

The main difference among EZW, SPC and WDR is the way how significant wavelet transform coefficients are located. Both EZW and SPC are “zerotree” type schemes using spatial orientation tree structures to implicitly locate significant wavelet transform coefficients, while WDR is a direct approach to find the positions of these significant coefficients based solely on index coding.

Except for the way to locate significant wavelet transform coefficients, and a minor difference of the order of sorting passes and refinement passes, these three algorithms are essentially using the same idea to encode (and consequently decode) images. They can all be included in the following generic model, which consists of five steps:

1. Take the discrete wavelet transform of the original image.

2. Order the wavelet transform coefficients from coarser scale to finer scale, as in Figure 5.2. Set the initial threshold T .
3. (Sorting Pass) Find the positions of significant coefficients with respect to T , and move these significant coefficients out.
4. (Refinement Pass) Get the refinement values of all significant coefficients, except those just found in the sorting pass of this round.
5. Divide T by 2 and go to step **3**.

The resulting symbol stream in step 3 and 4 will be further encoded by a lossless data compression algorithm.

In the decoding operation, the bit stream will first be decoded by the corresponding lossless data decompression algorithm to retrieve the symbol stream. Then the decoded symbol, both during a sorting pass and a refinement pass, refines and reduces the width of the uncertainty interval in which the true value of the coefficient may occur. The reconstruction value of the coefficient can be anywhere in that uncertainty interval. As suggested by Shapiro, we will simply use the center of the uncertainty interval as the reconstruction value. And the last step is to take the inverse wavelet transform to obtain the reconstructed image.

The above model provides a method of successive approximation of an image. It has several remarkable advantages, as described by Shapiro. First, the bits in the bit stream are generated in order of importance, yielding a fully embedded code. Second, the encoder can terminate the encoding at any point thereby allowing a target rate of target distortion metric to be met exactly. And the performance is achieved with a technique that requires absolutely no training, no pre-stored tables or codebooks, and requires no prior knowledge of the image source.

5.6 Applications of WDR Algorithm

The image compression software ICompress[†] is based on our algorithm Wavelet-Difference-Reduction. ICompress has an input channel, an output channel, and a control panel. The input channel supports lots of difference image formats, from the

[†]ICompress is a registered trademark of the Computational Mathematics Laboratory, Rice University.

lowest common denominator format pnm, raw, to the popular formats jpeg, gif, tiff, and others. The compression ratio will be set in the control panel. It can be any real number greater than or equal to 1. If the ratio is 1, then we will get the original image. Also in the control panel, the compression and decompression CPU time will be shown in the Info area. The decompressed image in tiff format will be sent to the output channel. An illustration of ICompress is in Figure 5.4. In Figure 5.4, the input file is lena.tiff, the 8 bits per pixel (bpp), 512×512 grayscale “Lena” image in tiff format. The compression ratio is set to 40:1. The compression and decompression CPU times are 4.0 seconds and 3.5 seconds, respectively. The original “Lena” image and the decompressed image with ratio 40:1 are shown in Display 1 and 2, respectively.

The Wavelet-Difference-Reduction algorithm can be applied to various types of image data, such as still images, medical images, seismic waves, synthetic aperture radar images, FBI fingerprints, etc. In this section we will use ICompress to code different types of images and compare its performance with other well-known image coding algorithms.

5.6.1 Still Images

Experiments have been done on all the 8 bits per pixel (bpp), grayscale still images, available from <ftp://links.uwaterloo.ca:/pub/BragZone/>, which include “Barbara”, “Goldhill”, “Lena” and others. And we used the biorthogonal Coifman wavelet system of degree 3 (BCW-3) with six scales. The symmetry of biorthogonal Coifman wavelet systems allows the “reflection” extension at the images edges. For our purpose, the compression performance is measured by the peak signal to noise ratio

$$\text{PSNR} = 10 \log_{10} \left(\frac{255^2}{\text{MSE}} \right) \text{ dB},$$

where MSE is the mean square error between the original image and the reconstructed one. Some other criterion might have been more preferable. However, to make a direct comparison with other coders, PSNR is chosen. And the bit rate is calculated from the actual size of the compressed file.

Our experimental results show that the coding performance of current implementation of this new WDR algorithm is between EZW and SPC, which all are much better than JPEG [54]. Here we include the coding results for the 8 bpp, 512×512 grayscale “Lena” image. The PSNR versus bit rate is plotted in Figure 5.5. Some reconstructed images are shown in Figure 5.6, 5.7, 5.8, and 5.9. In Figure 5.5, the

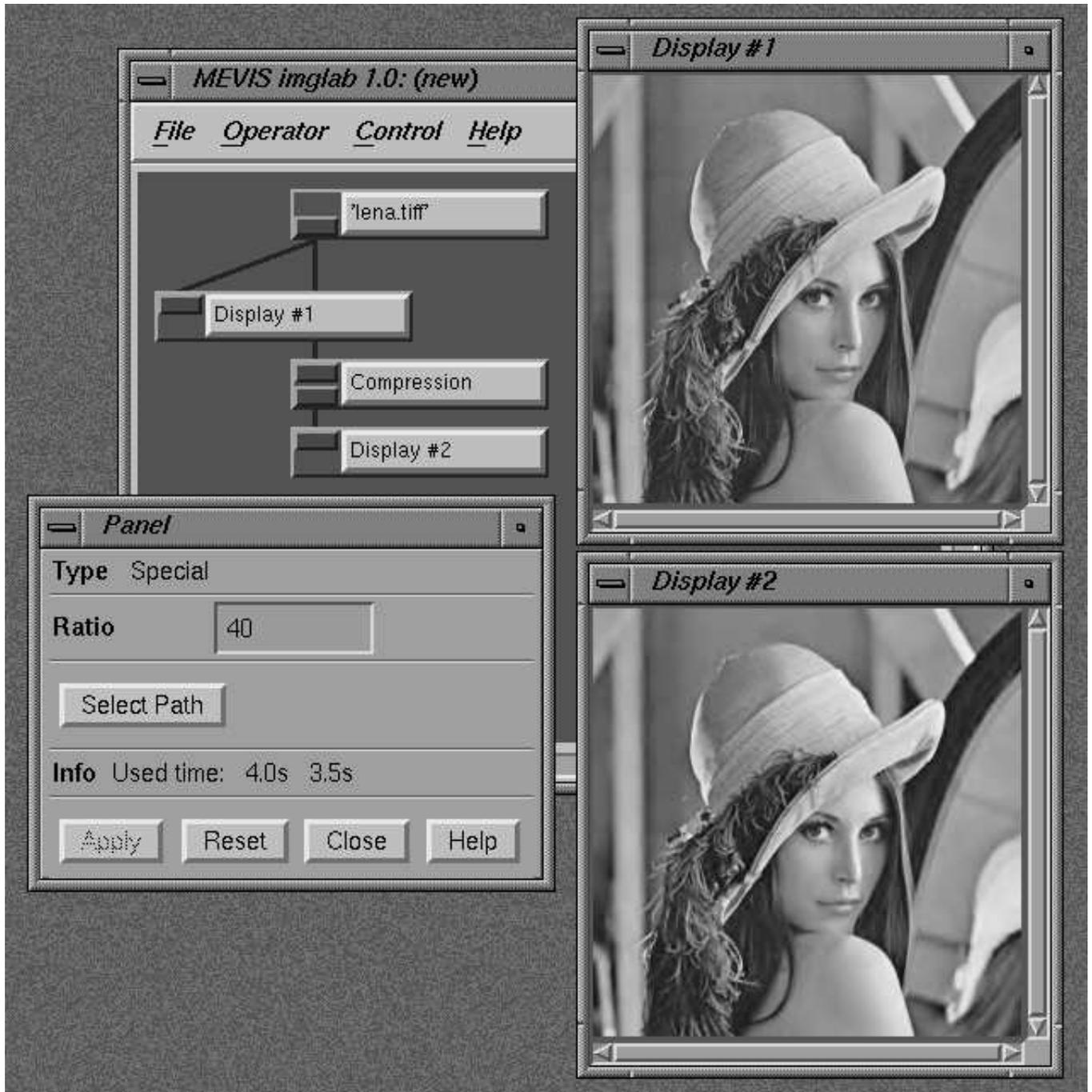


Figure 5.4 Illustration of ICompress

top curve is Said and Pearlman's SPC algorithm. And our new algorithm is about 0.3 dB higher in PSNR than Shapiro's EZW algorithm, 0.6 dB lower in PSNR than Said and Pearlman's SPC algorithm. However, this WDR algorithm is much simpler than both EZW and SPC, hence its encoding and decoding will be faster than both EZW and SPC.

5.6.2 Medical Images

In the automatic detection of microcalcification clusters in digitized mammograms, we compare the performance on the compressed mammograms using the Wavelet-Difference-Reduction algorithm with the performance on the original mammogram images. The surprising result is that at a compression ratio 10:1, the detection performance is improved for small false positive rates, compared with the original. Thus, the Wavelet-Difference-Reduction algorithm not only provides an efficient storage for the mammogram images (which are extremely large in size), but also improves the automatic detection of microcalcification clusters. The reason for the improvement on the compressed data is that at low compression ratios, the compression process is similar to denoising. Lots of the information discarded in the compressed data is the noise. Thus at low compression ratios, the mammogram images quality is improved for the detection of microcalcification clusters. For more details, we refer to our paper [36].

5.6.3 Synthetic Aperture Radar Images

Synthetic aperture radar (SAR) is an active coherent all-weather imaging system that operates in the microwave region of the spectrum. This imagery is well suited to the task of remote ground mapping in many applications, such as surveillance, oceanography, and agriculture. Real-time transmission of SAR data is of great importance for both time critical applications such as military search and destroy missions as well as in scientific survey applications. Furthermore, since post processing of the collected data in either application involves search, classification and tracking of targets, the requirements for a "good" compression algorithm is typically very different from that of lossy image compression algorithms developed for compressing still-images. The definition of targets are application dependent and could be military vehicles, trees in the rain forest, oil spills etc.

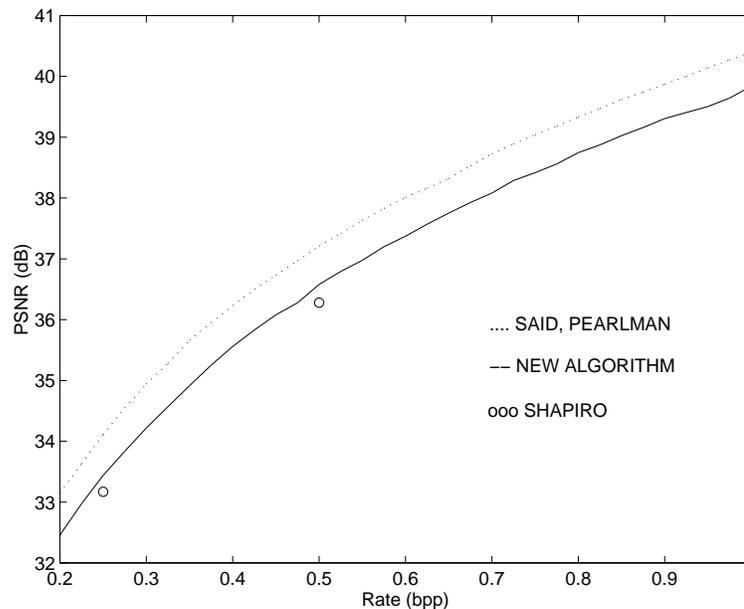


Figure 5.5 Coding Performance of the New Algorithm on “Lena”, Compared with SPC and EZW

If real time transmission of SAR imagery over a $T1$ -carrier (1.544Mb/s) is required, we are faced with having to compress SAR images at a rate of 73:1 (112Mb/s \rightarrow 1.544Mb/s). While a lot of research on lossy still-image compression has taken place over the years, not much attention has been paid to lossy compression of sensor data such as SAR. The main focus of compression of sensor data has been towards lossless compression techniques which at most can achieve a compression of about 2:1 [42], [33]. By applying standard lossy still-image compression algorithms one can achieve good results for automatic target recognition (ATR) at a compression of 16:1 [44]. By using the Wavelet-Difference-Reduction algorithm on the SAR image data which are preprocessed by a technique known as the polarimetric whitening filter (PWF) [37], we found that visually the image quality is still well preserved at the ratio 80:1. The experimental results and more details can be found in our paper [47].

5.7 Image Processing in the Compressed Wavelet Domain

The Wavelet-Difference-Reduction algorithm locates the significant wavelet transform coefficients directly and has a clear geometric structure. In the compressed wavelet



Figure 5.6 Original 8 bpp, 512×512 , Grayscale “Lena”



Figure 5.7 8:1 Compression, PSNR = 39.84 dB



Figure 5.8 16:1 Compression, PSNR = 36.59 dB



Figure 5.9 32:1 Compression, PSNR = 33.44 dB

domain, we know the exact locations of significant coefficients, along with their preceding significant bits values. With such a property, we may process the compressed data directly without applying the decompression first.

5.7.1 Denoising

In his celebrated paper [13], D. Donoho showed that the soft thresholding method is an optimal procedure to recover data from additive Gaussian noise. His method consists of three steps. First take the DWT of the noisy data. Then apply the soft thresholding on the wavelet coefficients. The soft thresholding acts like a shrinkage on real numbers,

$$\mathcal{ST}_t(x) = \begin{cases} x - t & \text{if } x \geq t \\ 0 & \text{if } -t < x < t \\ x + t & \text{if } x \leq -t \end{cases} ,$$

where $t > 0$ is the chosen threshold. The last step is to take the inverse DWT on the resulting numbers.

It is clear that the soft thresholding method only involves the magnitudes of the wavelet transform coefficients. Thus we may denoise the image data in the compressed wavelet domain. In the Wavelet-Difference-Reduction algorithm, the magnitudes of the wavelet transform coefficients are compared with a threshold T and the indices of those significant coefficients will be coded by the differential coding and the binary reduction. The magnitudes of those insignificant coefficients will be compared with the next threshold $T/2$ in the next cycle. Assume a target rate is met exactly at some point, the encoding job is terminated, and the final threshold is T_l . If $T_l \geq t$, where t is the threshold chosen for denoising, then we may simply subtract the magnitudes of significant coefficients by t , and the denoising is done. If $T_l < t$, since in the Wavelet-Difference-Reduction algorithm, the consecutive thresholds in the consecutive cycles are different by a factor of 2, there must exist some cycle with the threshold T_c , such that $T_c < t \leq 2T_c$. To carry out the soft thresholding method in the compressed domain, we just set all magnitudes of those coefficients which are found significant at this cycle (with the threshold T_c) or after this cycle to zero, and subtract by t all magnitudes of those coefficients which are found significant before this cycle.

5.7.2 Speckle Reduction

H. Guo et al. [20] found out that both the soft thresholding and the hard threshold methods are computational efficient and can significantly reduce the speckle while preserving the resolution of the synthetic aperture radar (SAR) images. The hard thresholding behaves like an ideal stop band,

$$\mathcal{HT}_t(x) = \begin{cases} x & \text{if } x \geq t \text{ or } x \leq -t \\ 0 & \text{if } -t < x < t \end{cases},$$

where $t > 0$ is the chosen threshold. The hard thresholding method doesn't require a subtraction operation. Thus it is more easier to be implemented in the compressed wavelet domain. Using the same argument as for the denoising, we may reduce the speckle on the compressed SAR data directly.

5.7.3 Zooming

When the image data size is very large, such as the medical images, seismic waves, image processing will be remarkable slow. Very oftenly we are interested in only a small region of the whole image. Thus it is desirable that we can zoom in the interesting region as quick as possible. Here we propose a novel method to zoom image data in the compressed domain. Assume the image data is stored in a compressed form by the Wavelet-Difference-Reduction algorithm. Since it is an embedded coding algorithm, at the beginning of the bit stream, the embedded code contains all lower rate (here the rate is the bit rate, or the number of bits per pixel) codes. Decompressing these beginning bit stream will give us a reconstruction image at a very low resolution with a relatively small data size. Then it becomes much easier to pick up our interested region in this decompressed image, which has a small data size. After that, as the decoding continues, more and more fine detail in the order of importance will be sent out to produce reconstruction images with higher and higher resolutions. Since we already know the location of our interested region, we also know the locations of significant wavelet transform coefficients, by the nature of the Wavelet-Difference-Reduction algorithm, we can simply ignore those bit stream not related to our interested region and get the reconstruction image of the interested region in a fast fashion.

5.8 Image Coding Evaluation of BCWs

In Chapter 3 we have learned biorthogonal Coifman wavelet systems, their construction and properties. Now it is the time to evaluate their transform coding performance. We will show that biorthogonal Coifman wavelet systems are very suited to image transform coding. Here for simplicity, we choose biorthogonal Coifman wavelet system of degree 3 (BCW-3) and compare it with the most widely used Cohen-Daubechies-Feauveau 9/7-tap filters (CDF-97) [7]. Some image-independent measures as well as some image-dependent measures will be applied to systematically compare these wavelet systems for image transform coding.

5.8.1 Image-Independent Measures

In [53], several image-independent measures based on some properties of the dual filters have been recommended to evaluate wavelet filters for image transform coding. We list the comparison results using these measures in the Table 5.1.

Regularity

It has been well known that the regularity of wavelets is only partially related to the quality of the reconstructed image via wavelet transform coding [51, 53, 41]. However, for short wavelet filters, the regularity is still closely related to the compression performance [41]. We use the algorithm by Rioul [40] to estimate the Hölder regularity of the wavelet filters. Usually, the smoothness of the synthesis scaling function $\tilde{\phi}(x)$ is more important than that of the analysis scaling function $\phi(x)$ in determining the quality of a reconstructed image, and the latter is more relevant to the energy compaction capability than the former for those smooth images. Therefore, there is a tradeoff between these two factors when choosing short wavelet filters. The comparison results in the Table 5.1 indicate that the BCW-3 and CDF-97 are about the same in terms of the distribution of regularity.

Shift-Variant Impulse Response

The impulse response of an L -level combined subband analysis/synthesis system is defined as [53]

$$f(h, \tilde{h}, L; n, n_0) = \mathcal{R}_{h, \tilde{h}, L} \{W_L(n) \mathcal{D}_{h, \tilde{h}, L} \{\delta(n - n_0)\}\}, \quad (5.3)$$

Table 5.1 Comparison Using Image-Independent Measures

	Regularity		MPSR	APSR	MFOSS	AFOSS
	$h_{N,\tilde{N}}$	\tilde{h}_N				
CDF-97	1.00	1.70	11.22	13.93	0.0411	0.0271
BCW-3	0.85	2.00	10.67	15.05	0.0379	0.0263

where h and \tilde{h} are the dual lowpass filters, $\delta(n - n_0)$ is an impulse at $n = n_0$, \mathcal{D} and \mathcal{R} are the decomposition and the reconstruction operators, respectively, and the wavelet domain window function $W_L(n)$ is defined as

$$W_L(n) = \begin{cases} 1, & \text{if } n \in \mathbf{N}, \\ 0, & \text{otherwise,} \end{cases} \quad (5.4)$$

where \mathbf{N} is the set of indices for the subband signal with the lowest resolution. Since the biorthogonal DWT is shift-variant, the impulse response defined above also depends on the location of the impulse. Unlike [53], we here use both the minimum and the average peak-to-sidelobe ratio (MPSR and APSR) in dB among the 2^L possible impulse responses in an L -level decomposition/reconstruction to characterize both the worst-case and the average oscillatory behavior, or ringing effect, in the system response, which usually results in visually annoying artifacts in the reconstructed image. The higher MPSR and APSR correspond to the weaker ringing behavior.

From the Table 5.1 we can see that the APSR of the BCW-3 is better than CDF-97, while the MPSR of the latter is better than the former.

Shift-Variant Step Response

Since the ringing artifact often occurs near the regions of edges in the reconstructed image, it can also be characterized by both the maximum and the average fractional overshoot of the second sidelobe (MFOSS and AFOSS) among the 2^L possible step responses of the combined analysis/synthesis system, which are closely related to the worst-case and the average ringing effect, respectively. A strong overshoot in the step response will lead to significant ringing in the reconstructed image [53].

The comparison results in the Table 5.1 indicate that the BCW-3 has lower AFOSS than the CDF-97, and the BCW-3 is also better than CDF-97 in terms of the MFOSS.

From the above comparison of the impulse responses and the step responses for the four FBs, we can expect that the BCW-3 will exhibit weaker ringing artifact than the CDF-97.

5.8.2 Image-Dependent Measures

The energy compaction is one of the most important metrics in the evaluation of filters used in transform coding schemes. However, to our knowledge, there is no measure for energy compaction that can be used independent of images. We choose six test images: Lena (512×512), Peppers (512×512), Boats (576×720), Building (a synthetic aperture radar image, 800×800), Fingerprint-1 (768×768) and Fingerprint-2 (480×384), which are all 256-gray-level images.

Weighted Subband Coding Gain

For a given image x with size N and a subband decomposition scheme, the energy compaction property of a FB can be characterized by the weighted SBC gain [25]

$$G_{SBC} = C_x \sigma_x^2 \prod_{k=0}^{K-1} \left(\left(\frac{N}{N_k} \right) w_k \sigma_{x_k}^2 \right)^{-N_k/N}, \quad (5.5)$$

where C_x is a constant related to the image x , K is the number of subbands, N_k is the size of the k th subband image, σ_x^2 and $\sigma_{x_k}^2$ are the variances of the image x and the subband image x_k , respectively, and w_k is the weight for the k th subband, which takes into consideration the different energy contribution from different subbands due to the relaxation of the orthogonality. We apply the method in [58] to compute these filter-related weights $\{w_k\}$, and here we also propose a new formula that generalizes the simple cases in [58]. Assume that the channel of the synthesis FB for the k th subband image consists of M filters in the horizontal direction, in the order of $g_{h,1}, g_{h,2}, \dots, g_{h,M}$, and M filters in the vertical direction, in the order of $g_{v,1}, g_{v,2}, \dots, g_{v,M}$ (Here we consider only the two-dimensional separable FBs). The weight w_k is given by

$$w_k = 2^{-2M} \left(\sum_n |g_h(n)|^2 \right) \left(\sum_n |g_v(n)|^2 \right), \quad (5.6)$$

where

$$G_h(z) = \sum_n g_h(n) z^{-n} = \prod_{i=1}^M G_{h,i} \left(z^{2^{M-i}} \right), \quad (5.7)$$

$$G_v(z) = \sum_n g_v(n)z^{-n} = \prod_{i=1}^M G_{v,i}(z^{2^{M-i}}). \quad (5.8)$$

From (5.5) one can see that the SBC gain depends strongly on the image content. It is also partly theoretical, because the definition of the SBC gain includes some assumptions that are not always valid in practice. However, due to the lack of better metrics to measure the energy compaction capability, the SBC gain is still widely used.

The experimental results are given in Table 5.2. We find that the SBC gains of the BCW-3 are slightly better than those of the CDF-97 for the first four images, while for the two fingerprint images the SBC gains of the BCW-3 are much better.

Table 5.2 Comparison Using the Weighted SBC Gain

	Lena	Peppers	Boats	Building	Fingerprint-1	Fingerprint-2
CDF-97	49.29	32.12	37.96	28.36	77.40	46.60
BCW-3	50.34	32.79	40.94	28.74	88.72	53.89

5.9 Conclusions

In this chapter we have presented the Wavelet-Difference-Reduction algorithm. It utilizes the discrete wavelet transform which removes the spatial and spectral redundancies of digital images to a large extent. The combination of the differential coding and the binary reduction represents the positions of significant wavelet transform coefficients very efficiently. This Wavelet-Difference-Reduction algorithm provides a successive approximation of image sources and facilitates progressive image transmission. It requires no training of any kind or prior knowledge of image sources. Since this algorithm doesn't depend on any special statistical model, it can be easily extended to 3-D or even higher image data compression. Also due to its clear geometric structure we may perform image processing directly in the compressed wavelet domain. Considering the large data size nowadays, this may enable us to reach our goal in a fast and efficient way. And with its nice compression performance, fast encoding/decoding speed, this new algorithm looks quite promising in image and video processing.

Chapter 6

Summary

Biorthogonal Coifman wavelet systems have very nice properties both in the theoretical sense and the application sense. The vanishing moments conditions have played an important role in such systems. These conditions not only give growing smoothness of wavelet systems, but also provide fast wavelet sampling approximation. Another attractive feature of biorthogonal Coifman wavelet systems is that all the coefficients are dyadic rational. Thus we can have a very fast multiplication-free discrete wavelet transform implemented on digital computers.

Bibliography

- [1] M. Antonini, M. Barlaud, P. Mathieu, and Ingrid Daubechies. Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1:205–220, April 1992.
- [2] G. Battle. A block spin construction of ondelettes. Part I: Lemarié functions. *Comm. Math. Phys.*, 110:601–615, 1987.
- [3] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183, 1991.
- [4] C. K. Chui. On cardinal spline wavelets. *Ruskai et al.*, pages 419–438, 1992.
- [5] Charles K. Chui. *An Introduction to Wavelets*. Wavelet analysis and its applications. Academic Press, Boston, 1992.
- [6] A. Cohen. Ondelettes, analyses multirésolutions et filtres miroir en quadrature. *Ann. Inst. H. Poincaré, Anal. non linéaire*, 7:439–459, 1990.
- [7] A. Cohen, Ingrid Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, XLV:485–560, 1992.
- [8] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, XLI:909–996, 1988.
- [9] Ingrid Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF regional conference series in applied mathematics. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.
- [10] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets II. variations on a theme. *SIAM J. Math. Anal.*, 24(2):499–519, 1993.

- [11] G. Deslauriers and S. Dubuc. *Interpolation daydique*, pages 44–55. Fractals, dimensions non entières et applications. Masson, Paris, g. cherbit edition, 1987.
- [12] R. A. DeVore, B. Jaweth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Transactions on Information Theory*, 38(2):719–746, March 1992.
- [13] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, 41(3):613–627, May 1995.
- [14] Timo Eirola. Sobolev characterization of solutions of dilation equations. *SIAM J. MATH. ANAL.*, 23(4):1015–1030, July 1992.
- [15] P. Elias. Univeral codeword sets and representations of the integers. *IEEE Trans. Informat. Theory*, 21(2):194–203, March 1975.
- [16] R. Glowinski, W. Lawton, M. Ravachol, and E. Tenenbaum. Wavelet solution of linear and nonlinear elliptic, parabolic and hyperbolic problems in one dimension. In R. Glowinski and A. Lichnewski, editors, *Proceedings of the Ninth International Conference on Computing Methods in Applied Science and Engineering*, 1990.
- [17] R. Gopinath and C. S. Burrus. On the moments of the scaling function ϕ_0 . In *Proc. ISCAS*, volume 2, pages 963–966. IEEE, May 1992.
- [18] D. Gottlieb, S. A. Hagerth, P. G. H. Lehot, and H. S. Rabinowitz. A classification of compression methods and their uesfulness for a large data processing center. In *Nat. Com. Conf.*, pages 453–458, 1975.
- [19] A. Grossmann and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15(4):723–736, July 1984.
- [20] Haitao Guo, Jan E. Odegard, Markus Lang, Ramesh A. Gopinath, Ivan W. Selesnick, and C. Sidney Burrus. Wavelet based speckle reduction with application to SAR based ATD/R. In *Proc. ICIP*, Austin, TX, November 1994.
- [21] A. Haar. Zur theorie der orthogonalen funktionen-systeme. *Math. Ann.*, 69:331–371, 1910.

- [22] Peter N. Heller and Raymond O. Wells, Jr. Spectral theory of multiresolution operators and applications. In C. K. Chui, L. Montefusco, and L. Puccio, editors, *Wavelets: Theory, Algorithms, and Applications*, pages 13–32. Academic Press, San Diego, 1994.
- [23] Y. Huang, H. M. Dreizen, and N. P. Galatsanos. Prioritized DCT for compression and progressive transmission of images. *IEEE Trans. Image Processing*, 1:477–487, October 1992.
- [24] S. Jaffard, P. G. Lemarié, S. Mallat, and Y. Meyer. Multiscale analysis. Unpublished memorandum.
- [25] N. S. Jayant and P. Noll. *Digital Coding of Waveforms*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1984.
- [26] L. Kantorovich. Functional analysis and applied mathematics (Russian). *Uspehi Mat. Nauk*, 3:89–185, 1948.
- [27] L. Kantorovich. On Newton’s method for functional equations (Russian). *Dokl. Akad. Nauk. SSSR*, 59:1237–1240, 1948.
- [28] W. Lawton. Tight frames of compactly supported wavelets. *J. Math. Phys.*, 31:1898–1901, 1990.
- [29] Wayne M. Lawton. Necessary and sufficient conditions for constructing orthonormal wavelet bases. *Journal of Mathematical Physics*, 32(1):57–61, January 1991.
- [30] P. G. Lemarié. Une nouvelle bade d’ondelettes de $L^2(\mathbf{R}^n)$. *J. de Math. Pures et Appl.*, 67:227–236, 1988.
- [31] A. S. Lewis and G. Knowles. Image compression using the 2-D wavelet transform. *IEEE Transactions on Image Processing*, 1:244–250, April 1992.
- [32] Stephane G. Mallat. Multiresolution approximation and wavelet orthonormal bases of $L^2(\mathbf{R})$. *Transactions of the American Mathematical Society*, 315(1):69–87, September 1989.
- [33] N. D. Memon, K. Sayood, and S. S. Magliveras. Lossless compression of multispectral image data. *IEEE Trans. Geoscience and Remote Sensing*, 32(2):282–289, March 1994.

- [34] Y. Meyer. Ondelettes et fonctions splines, December 1986. Séminaire EDP, Ecole Polytechnique, Paris, France.
- [35] Y. Meyer. *Wavelet and Operator*. Cambridge University Press, 1992.
- [36] Thomas Netsch, Jun Tian, Markus Lang, Raymond O. Wells, Jr., Carl Evertsz, and H. Juergens and H. O. Peitgen. Automatic detection of microcalcification clusters on compressed digitized mammograms. Technical Report CML TR 96-04, Computational Mathematics Laboratory, Rice University, April 1996.
- [37] L. M. Novak, M. C. Burl, R. Chaney, and G. J. Owirka. Optimal processing of polarimetric synthetic aperture radar imagery. *Linc. Lab. J.*, 3, 1990.
- [38] Howard L. Resnikoff, Jun Tian, Raymond O. Wells, Jr., and Xiaodong Zhou. Some properties of biorthogonal Coifman wavelet systems. Technical Report CML TR 95-04, Computational Mathematics Laboratory, Rice University, April 1995.
- [39] Howard L. Resnikoff and Raymond O. Wells, Jr. *Wavelet Analysis and the Scalable Structure of Information*. Springer-Verlag, New York, 1996. To appear.
- [40] O. Rioul. Simple regularity criteria for subdivision schemes. *SIAM J. Math. Anal.*, 23(6):1544–1576, November 1992.
- [41] O. Rioul. Regular wavelets: A discrete-time approach. *IEEE Trans. SP*, 41(12):3572–3579, December 1993.
- [42] R. E. Roger and J. F. Arnold. Reversible image compression bounded by noise. *IEEE Trans. Geoscience and Remote Sensing*, 32(1):19–24, January 1994.
- [43] Amir Said and William A. Pearlman. A new fast and efficient image codec based on set partitioning in hierarchical trees. *submitted to IEEE Transactions on Circuits and Systems for Video Technology*, 1995.
- [44] J. F. Schonfeld. Sensor technology office: Review of the university ATR initiative. In *ATR Workshop*, Arlington, VA, August 1994.
- [45] Jerome M. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, December 1993.

- [46] W. Sweldens and R. Piessens. Quadrature formulae and asymptotic error expansions for wavelet approximations of smooth functions. *SIAM J. Numer. Anal.*, 31(4):1240–1264, August 1994.
- [47] Jun Tian, Haitao Guo, Raymond O. Wells, Jr., C. Sidney Burrus, and Jan E. Odegard. Evaluation of a new wavelet based compression algorithm for synthetic aperture radar images. In E. Z. Zelnio and R. J. Douglass, editors, *Algorithms for Synthetic Aperture Radar Imagery III*, volume 2757 of *Proc. SPIE*, 1996.
- [48] Jun Tian and Raymond O. Wells, Jr. Vanishing moments and wavelet approximation. Technical Report CML TR 95-01, Computational Mathematics Laboratory, Rice University, January 1995.
- [49] Jun Tian and Raymond O. Wells, Jr. Image data processing in the compressed wavelet domain. In *Proceedings of the 3rd International Conference on Signal Processing*, Beijing, China, 1996.
- [50] Jun Tian and Raymond O. Wells, Jr. An lossy image codec based on index coding. In *Proceedings of IEEE Data Compression Conference*, Snowbird, Utah, 1996. IEEE Computer Society Press.
- [51] M. Vetterli and C. Herley. Wavelets and filter banks: Theory and design. *IEEE Trans. ASSP*, 40:2207–2232, September 1992.
- [52] Martin Vetterli and Jelena Kovačević. *Wavelets and Subband Coding*. Prentice Hall signal processing series. Prentice Hall PTR, Englewood Cliffs, New Jersey, 1995.
- [53] J. D. Villasenor, D. Delzer, and J. Liao. Wavelet filter evaluation for image compression. *IEEE Trans. on Image Processing*, 4:1053–1060, August 1995.
- [54] G. K. Wallace. The JPEG still picture compression standard. *Comm. ACM*, 34:30–44, April 1991.
- [55] Dong Wei, Jun Tian, Raymond O. Wells, Jr., and C. Sidney Burrus. A new class of biorthogonal wavelet systems for image transform coding. Technical Report CML TR 95-14, Computational Mathematics Laboratory, Rice University, October 1995.

- [56] Raymond O. Wells, Jr. and Xiaodong Zhou. Wavelet interpolation and approximate solutions of elliptic partial differential equations. In R. Wilson and E. A. Tannerr, editors, *Noncompact Lie Groups, Proceedings of NATO Advanced Research Workshop*, 1992.
- [57] I. H. Witten, R. Neal, and J. G. Cleary. Arithmetic coding for data compression. *Comm. ACM*, 30:520–540, 1987.
- [58] J. W. Woods and T. Naveen. A filter based bit allocation scheme for subband compression of HDTV. *IEEE Trans. on Image Processing*, 1:436–440, July 1992.