IFSM, Wavelets and Fractal-Wavelets: Three Methods of Approximation

by

Daniel Piché

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Mathematics
in
Pure Mathematics

Waterloo, Ontario, Canada, 1997

0-612-21538-5

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

IFSM, Wavelets and Fractal-Wavelets:

Three Methods of Approximation

This thesis deals with representations and approximations of functions using iterated function systems (IFS), wavelets and fractal-wavelets.

IFS use self-similarity to approximate a function by contracted and translated copies of itself. Results covered include the Banach Contraction Mapping Principle, the completeness of IFS space and the Collage Theorem. IFS on grey-level maps (IFSM) are defined to generalize IFS to real-valued functions.

Wavelets are discussed, using multiresolution analysis. Stronger convergence results are shown to hold for wavelet expansions than for Fourier expansions. An application of the Mallat algorithm to compression is given.

Fractal-wavelets use the fact that given an orthonormal basis of $L^2(\mathbb{R})$, the mapping which sends a function in $L^2(\mathbb{R})$ to its sequence of basis coefficients is an isometry. An identification is made between IFSM and operators on coefficients. Local IFS on wavelet coefficients are defined and shown to induce IFSM-type operators.

Acknowledgements

I wish to thank my supervisors Edward Vrscay and William Gilbert, for helping me during the writing of this thesis. Ed's constant suggestions of readings and things to try out, as well as his help in planning my thesis were indispensable. Sincere thanks are extended to my readers, Franklin Mendivil and Kathryn Hare, who were able to review the thesis in a very short period of time. I would also wish to acknowledge the members of our Fractalators group and their many informative talks.

I would like to thank my friends and co-graduate students Steve Astels, Matt Becker, Dave Kribs, Karen Meagher, Marni Mishna, Andre Sonnichsen, Nico Spronk, Drew Vandeth, Pete Wood and Chris Wooff, for their company and when I needed to do something other than work.

Je remercie mon professeur d'anglais, Louise Auger, pour m'avoir motivé à travailler fort, et mon professeur de mathématiques en première année d'université, Rolland Gaudet, qui m'a aidé à débuté mon éducation universitaire. Je suis reconnaissant aussi de mon bon ami Christian Dandeneau, avec qui j'ai découvert le monde des ordinateurs.

J'aimerais remercier mes parents, Michelle et René, pour leur encouragement et mon frère Robert, pour son aide à passer le temps durant mon enfance. Je remercie aussi ma bien aimée, Andréanne Bouchard, pour sa tolérance durant le cours de mes études de maîtrise.

There are many others who have been supportive, including Debbie Brown and Carolyn Jackson, from the Department. I extend my deepest thanks to them all.

Dédicace

Cette thèse est dédiée à Zine Guennoun, qui m'a dévoilé l'univers des fractales.

Contents

List of Figures

Introduction

From the beginning, mathematicians have been interested in nature. Indeed, it has been nature in many instances which either inspired or provided ideas for such fields as algebra, geometry, and more recently fractal geometry [9, 10, 33].

Since the discovery of the Cantor set over a century ago, mathematicians have been working with fractals. It is the invention of the computer, in the mid twentieth century, which has enabled us to calculate the fractal objects which would have required years of human computation time previously. However, this same machine has demanded constant input from mathematicians in terms of new theory and algorithms. One such example is the focus of this thesis.

Each day, vast quantities of data are generated by the millions of computers worldwide, data which could never have been generated before the advent of the computer. The need to store this information is critical and has required mathematicians to develop methods of compression. The goal of this thesis is to present three methods of mathematics which have allowed compression of the data representing signals and images. The three topics presented are fractals, wavelets and fractal-wavelets.

Chapter 1 describes iterated function systems (IFS). This method utilizes the inherent self-similarity of an object (set, signal, image) to define maps on it, which in turn allow the reconstruction of the object. For compression, it is the maps that are stored rather than the original object.

The first section of Chapter 1 introduces basic definitions and theorems of metric spaces, including the pivotal Banach Contraction Mapping Principle (BCMP) for complete spaces, upon which rests the entire theory of IFS. Sections 1.2 and 1.3 define the concept of IFS and the space $\mathcal{H}(X)$ where IFS live. This space is shown to be complete, which allows the application of the BCMP. The following section contains examples of some attractors of IFS.

From IFS we move to IFS with grey-level maps (IFSM). This results from the realization that, from a nature perspective, IFS act on black and white images and are inadequate to model the real world. In Section 1.6, IFSM are made more concrete and conditions are given under which they are contractive. Section 1.7 presents a formal solution of the Inverse Problem for IFSM and illustrates an example where IFSM fail to give good approximations. This leads to Section 1.8 where the theory of local IFSM (LIFSM) is presented. It is shown that LIFSM resolve the problems encountered with the initial IFSM method.

The second chapter deals with wavelets. In this thesis, they are considered, for simplicity, to be special types of Hilbert space bases of $L^2(\mathbb{R})$. Basic notions from Hilbert space theory are given in Section 2.1. In Section 2.2 the concept of a multiresolution analysis (MRA) associated with a scaling function is defined, with a couple of example wavelet bases being presented. Sections 2.3 and 2.4 motivate the study of wavelets given the large amount of theory which has been developed in Fourier analysis [43, 55, 56]. It is shown that wavelet series often converge in better ways and much more rapidly than their trigonometric counterparts.

Section 2.5 describes the Mallat algorithm for the decomposition and reconstruction of scaling and wavelet coefficients between levels of a MRA. It is then shown in Section 2.6 how this algorithm can be viewed as a pair of quadrature mirror filters for implementation in digital circuitry. The chapter concludes with some applications of the algorithm in signal compression.

The final chapter combines the ideas of IFS and wavelets, creating fractal-wavelets. Section 3.1 describes the relation between IFSM and wavelets. This leads to local IFS on wavelet coefficients (LIFSW) which are defined in Section 3.2. The following section gives a few examples of LIFSW. The inverse problem for LIFSW is discussed in Section 3.4 with an application to compression being given there.

Appendix A describes a normalized version of the IFSM operator, which is the theoretical generalization of the IFS operator, but which is more difficult to implement in compression methods.

A glossary of notation, list of common abbreviations and index are provided at the end of the thesis. A definition or major reference to a term is indicated by a bold page number. The source code for the applications used in this thesis are available for downloading at

http://links.uwaterloo.ca/~dgpiche

D. G. Piché

dgpiche@links.uwaterloo.ca

Chapter 1

Fractal Transforms

In order to understand the idea behind fractal image compression, we study the subject of iterated function systems.

1.1 Topological Background

We begin with some basic notation and definitions from metric spaces. Other topics not covered here may be found in [28, 54].

Notation 1.1.1 We will use the following notation to denote certain classical sets:

$$\begin{aligned}
\mathbb{N} &= \{0, 1, 2, \ldots\}; \\
\mathbb{N}^+ &= \{1, 2, \ldots\}; \\
\mathbb{Z} &= \text{the integers}; \\
\mathbb{R} &= \text{the set of real numbers}.
\end{aligned}$$

Notation 1.1.2 Throughout the text, (X, d) will denote a metric space where X is the set and d is the metric. Special properties, such as completeness, will be specified as needed. We will denote a sequence in X by $(x_n)_{n \in A}$, where $A \subset \mathbb{N}$. A sequence will be written as (x_n) if the range of the subscripts is clear from the context. If (x_n) converges to x, we write $(x_n) \to x$.

Other notation will be defined as needed.

Definition 1.1.3 *A metric space* (X, d) *is* totally bounded *if for each* $\epsilon > 0$*, there is a finite set* F_ϵ *(called an* ϵ-net*) of* X *such that*

$$X = \bigcup\{N(x;\epsilon) : x \in F_\epsilon\}$$

where $N(x;\epsilon)$ *is the open ball of radius* ϵ *centered at* x*.*

Definition 1.1.4 *A metric space* (X, d) *is* complete *if and only if every Cauchy sequence converges in* X *with respect to the metric* d*.*

Definition 1.1.5 *A function* $f : X \to X$ *is said to be* Lipschitz *if and only if there exists an* $s \in [0, \infty)$ *such that* $\forall x, y \in X$ *we have*

$$d(f(x), f(y)) \leq sd(x, y).$$

We call s *a* Lipschitz constant *of* f*. If there exists such an* $s < 1$*, we say* f *is* contractive *or is a* contraction *and call* s *a* contractivity factor *of* f*. In this case we say that* f *has contractivity at least* s*. We denote the set of all Lipschitz functions on* (X, d) *by* $L(X, d)$ *and write* $Con(X, d)$ *to mean the set of all contractive maps* $f : X \to X$*. If* $X = \mathbb{R}$*, write simply* $Lip(\mathbb{R})$*.*

Proposition 1.1.6 *If* $f \in L(X, d)$*, then* f *is uniformly continuous.*

Proof Let $\epsilon > 0$. We can assume f is not constant, hence let $s > 0$ be a contractivity factor of f. If we let $\delta = \frac{\epsilon}{s}$, then $\forall x, y \in X$,

$$d(x, y) < \delta \implies d(f(x), f(y)) \leq sd(x, y) < \epsilon. \quad \blacksquare$$

Proposition 1.1.7 *Let* $f \in Con(X, d)$*. Define* c_f *by*

$$c_f = \inf\{s : s \textit{ is a contractivity factor of } f\}.$$

Then c_f is a contractivity factor of f.

Proof Let $x, y \in X$ and let S be the set of contractivity factors of f. Then, for each $s \in S$, $d(f(x), f(y)) \leq sd(x, y)$. Hence,

$$\begin{aligned} d(f(x), f(y)) &\leq \inf(S)\, d(x, y) \\ &= c_f d(x, y), \end{aligned}$$

and since $c_f < 1$, $c_f \in S$. ■

Notation 1.1.8 We call c_f the *contractivity* of f. We note that $c_f = 0$ if and only if f is constant.

Example 1.1.9 Consider the function $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \frac{1}{2}x + \frac{1}{2}\ \forall x \in \mathbb{R}$. Then for $x, y \in \mathbb{R}$,

$$\begin{aligned} |f(x) - f(y)| &= \left| \left(\frac{1}{2}x + \frac{1}{2}\right) - \left(\frac{1}{2}y + \frac{1}{2}\right) \right| \\ &= \frac{1}{2}|x - y|. \end{aligned}$$

Therefore, f is contractive with contractivity $\frac{1}{2}$.

Notation 1.1.10 For $x \in X$, we define the *n-fold composition* of a function f at x recursively by

$$\begin{aligned} f^{\circ 1}(x) &= f(x); \\ f^{\circ n+1}(x) &= f(f^{\circ n}(x)). \end{aligned}$$

We call $f^{\circ n}(x)$ the *n-th iterate of f at x.*

Definition 1.1.11 *We say that $y \in X$ is the* attractor *of $f : X \to X$ if and only if*

$$\lim_{n \to \infty} f^{\circ n}(x) = y \quad \forall x \in X.$$

Example 1.1.12 Consider the function f from Example 1.1.9. Then for any $x \in \mathbb{R}$, we have

$$\begin{aligned} f(x) &= \frac{1}{2}x + \frac{1}{2}; \\ f^{\circ 2}(x) &= \frac{1}{2}\left(\frac{1}{2}x + \frac{1}{2}\right) + \frac{1}{2} \\ &= \frac{x}{2^2} + \frac{1}{2} + \frac{1}{4} \end{aligned}$$

and for a general $n > 1$,

$$f^{\circ n}(x) = \frac{x}{2^n} + \sum_{k=1}^{n} \frac{1}{2^i}.$$

Hence, $\lim_{n\to\infty} f^{\circ n}(x) = 1 \ \forall x \in \mathbb{R}$ and $x = 1$ is the attractor of f.

Definition 1.1.13 *Let $f : X \to X$. If for some $x \in X$, $f(x) = x$, we call x a* fixed point *of f.*

Example 1.1.14 Consider the function f from Example 1.1.9. Then

$$\begin{aligned} f(x) = x &\implies \frac{1}{2}x + \frac{1}{2} = x \\ &\implies x = 1. \end{aligned}$$

Hence $x = 1$ is a fixed point of f.

This is not a coincidence, as the next proposition shows.

Proposition 1.1.15 *If a continuous function $f : X \to X$ has an attractor $x \in X$, then x is a fixed point of f.*

Proof Suppose $x \in X$ is the attractor of f. Then, since f is continuous,

$$x = \lim_{n\to\infty} f^{\circ n}(x) = f\left(\lim_{n\to\infty} f^{\circ n-1}(x)\right) = f(x). \qquad \blacksquare$$

We now prove the result upon which the entire theory of iterated function systems is founded. It is the *Banach Contraction Mapping Principle*, or *BCMP* for short [54].

Theorem 1.1.16 (Banach Contraction Mapping Principle) *Suppose (X,d) is a complete metric space and let $f \in Con(X,d)$ with contractivity factor s. Then f has a unique fixed point $\bar{x}_f \in X$. Furthermore, $\bar{x}_f$ is the attractor of f.*

Proof Let $x \in X$ and set $x_n = f^{\circ n}(x)$ for $n \in \mathbb{N}^+$. We will first show (x_n) is a Cauchy sequence. Let $m > n \in \mathbb{N}^+$. Then

$$\begin{aligned} d(x_n, x_m) &= d(f^{\circ n}(x), f^{\circ m}(x)) \\ &\leq sd(f^{\circ n-1}(x), f^{\circ m-1}(x)) \end{aligned}$$

and inductively,

$$d(x_n, x_m) \leq s^n d(x, f^{\circ m-n}(x)). \tag{1.1}$$

Now, for $k \in \mathbb{N}^+$,

$$\begin{aligned} d(x, f^{\circ k}(x)) &\leq d(x, f(x)) + d(f(x), f^{\circ 2}(x)) + \ldots + d(f^{\circ k-1}(x), f^{\circ k}(x)) \\ &\leq d(x, f(x)) + sd(x, f(x)) + \ldots + s^{k-1}d(x, f(x)) \\ &= \sum_{l=0}^{k-1} s^l d(x, f(x)) \\ &= \frac{1-s^k}{1-s} d(x, f(x)) \\ &\leq \frac{1}{1-s} d(x, f(x)). \end{aligned}$$

Thus, by Equation (1.1),

$$d(x_n, x_m) \leq \frac{s^n}{1-s} d(x, f(x)).$$

Since $s < 1$, $d(x_n, x_m) \to 0$ as $n, m \to \infty$. Hence (x_n) is a Cauchy sequence. Therefore, by the completeness of X, let $\bar{x}_f \in X$ with $(x_n) \to \bar{x}_f$. Hence, $\lim_{n\to\infty} f^{\circ n}(x) = \bar{x}_f$.

Now, suppose f has another fixed point $y \in X$. Then

$$d(\bar{x}_f, y) = d(f(\bar{x}_f), f(y)) \le sd(\bar{x}_f, y).$$

However, $s < 1$ hence $d(\bar{x}_f, y) = 0$. Therefore $\bar{x}_f$ is the unique fixed point, and by Proposition 1.1.15, the unique attractor of f. ∎

If f is contractive, we write $\bar{x}_f$ to denote its fixed point.

We will now define a metric on $Con(X, d)$ and show that fixed points vary continuously with respect to contractive maps. The following discussion is a variation of [13].

Proposition 1.1.17 *Define $\bar{d}(f, g)\colon Con(X, d) \to [0, \infty]$ by*

$$\bar{d}(f, g) = \sup_{x\in X} d(f(x), g(x)) \quad \forall f, g \in Con(X, d)$$

and let

$$d_m(f, g) = \min\{\bar{d}(f, g), 1\} \qquad \forall f, g \in Con(X, d).$$

Then $d_m(f, g)$ is a metric on $Con(X, d)$. Furthermore, if (X, d) is compact, $\bar{d}$ is a metric on $Con(X, d)$.

Proof The only problem with $\bar{d}$ is that the distance between certain functions might be infinite, a problem which is eliminated if (X, d) is compact. Let $f, g, h \in Con(X, d)$. Then

i) $0 = d_m(f, g) \iff \bar{d}(f, g) = 0 \iff f = g$.

ii) d_m is symmetric by the symmetry of $\bar{d}$.

iii) To prove the triangle inequality for d_m, it is enough to consider the case when $d_m(f, g) + d_m(g, h) < 1$ since $d_m(f, g)$ is always ≤ 1. Therefore suppose $d_m(f, g) +$

$d_m(g,h) < 1$. Then as $\bar{d}$ satisfies the triangle inequality, we have

$$\bar{d}(f,h) \le \bar{d}(f,g) + \bar{d}(g,h) < 1.$$

Hence,

$$\begin{aligned} d_m(f,h) &= \bar{d}(f,h) \\ &\le \bar{d}(f,g) + \bar{d}(g,h) \\ &= d_m(f,g) + d_m(g,h), \end{aligned}$$

since both $\bar{d}(f,g)$ and $\bar{d}(g,h)$ are less than 1. ∎

Theorem 1.1.18 (Continuity of Fixed Points) *Define F: $Con(X,d) \to X$ by $F(f) = \bar{x}_f$ for each $f \in Con(X,d)$. Then F is continuous with respect to d_m. If X is compact, F is continuous with respect to $\bar{d}$.*

Proof Let $\epsilon > 0$ and let $f, g \in Con(X,d)$. Without loss of generality, we assume $c = c_g \le c_f$. Suppose $d_m(f,g) < \min(\epsilon(1-c), 1)$. Then

$$\begin{aligned} d(\bar{x}_g, \bar{x}_f) &= d(g(\bar{x}_g), f(\bar{x}_f)) \\ &\le d(g(\bar{x}_g), g(\bar{x}_f)) + d(g(\bar{x}_f), f(\bar{x}_f)). \end{aligned} \tag{1.2}$$

By the hypothesis on $d_m(f,g)$, we have $\bar{d}(g,f) < 1$ and $\bar{d}(g,f) < \epsilon(1-c)$. Hence,

$$\begin{aligned} d(g(\bar{x}_f), f(\bar{x}_f)) &\le \bar{d}(g,f) \\ &< \epsilon(1-c). \end{aligned}$$

Therefore, by Equation (1.2),

$$\begin{aligned} & d(\bar{x}_g, \bar{x}_f) < cd(\bar{x}_g, \bar{x}_f) + \epsilon(1-c) \\ \implies & d(\bar{x}_g, \bar{x}_f) < \epsilon. \qquad \blacksquare \end{aligned}$$

Corollary 1.1.19 *If (X,d) is a compact metric space and $f, g \in Con(X,d)$, then*

$$d(\bar{x}_f, \bar{x}_g) < \frac{1}{1-c}\tilde{d}(f,g),$$

where $c = \min(c_f, c_g)$.

Intuitively, if the given maps f and g are close to each other, then their respective fixed points $\bar{x}_f$ and $\bar{x}_g$ are also. This is the fundamental principle behind the fractal-based methods of approximation.

Now, suppose we are given $x \in X$. Is it always possible to construct $f \in Con(X,d)$ such that $x = \bar{x}_f$? In simple cases we might guess at such a function (as in Example 1.1.9). One might indeed think to take the constant function $f(y) = x$ for all $y \in X$. The goal however is to approximate x using a function which is easy to describe, and this constant function would most often necessitate the complete description of x! However, suppose we would be satisfied to find an f with a fixed point close to x? If this is the case, how would we proceed to find f? We can reformulate this problem as follows:

Question 1.1.20 *Given (Y, d_Y) a metric space, $y \in Y$ and $\epsilon > 0$, can we find $f \in Con(Y, d_Y)$ such that $d_Y(y, \bar{y}_f) < \epsilon$?*

This problem is called the *Inverse Problem of Approximation by Fixed Points of Contraction Maps*, or the *Inverse Problem* for short. Detailed discussions can be found in [21] and [49]. Indeed, whether such an f can be constructed, or whether it even exists is uncertain at this stage. The question raised might be "Is $\{\bar{x}_f : f \in Con(Y, d_Y)\}$ dense in Y"? We will attempt to address this question shortly. In practice, Y could be any one of a large number of relevant spaces: compact subsets of $[0,1]^n$; probability measures on $[0,1]$; $L^p(\mathbb{R})$; fuzzy set functions.

For the moment, consider the proof of Theorem 1.1.18 and ask: "Given $y \in Y$, $f \in Con(Y, d_Y)$, how close is y to $\bar{y}_f$"? The following proposition lends an answer:

Proposition 1.1.21 *Let y, Y and f be as above. Then*

$$d_Y(y, \bar{y}_f) \leq \frac{1}{1-c_f} d_Y(y, f(y)).$$

Proof We have

$$\begin{aligned} d_Y(y, \bar{y}_f) &\leq d_Y(y, f(y)) + d_Y(f(y), \bar{y}_f)) \\ &= d_Y(y, f(y)) + d_Y(f(y), f(\bar{y}_f)) \\ &\leq d_Y(y, f(y)) + c_f d_Y(y, \bar{y}_f)) \\ \implies d_Y(y, \bar{y}_f) &\leq \frac{1}{1-c_f} d_Y(y, f(y)). \qquad \blacksquare \end{aligned} \tag{1.3}$$

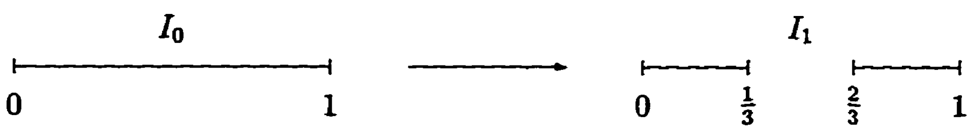
This is often called the *Collage Theorem*. In the light of this new proposition, we see that if $f(y)$ is close to y, then $\bar{y}_f$ is also close to y. Of course, if $c_f \approx 1$, the right hand side of Equation (1.3) might not be very small. This thus gives some insight into finding our desired function. We should find an $f \in Con(Y, d_Y)$ which takes y close to itself. We remember from the BCMP that $\bar{y}_f$ is the attractor of f if Y is complete. Hence we can iterate f to retrieve $\bar{y}_f$ and get the desired approximation to y. Therefore, we can restate the Inverse Problem as

Question 1.1.22 (Inverse Problem) *Let (Y, d_Y) be a complete metric space, and let $y \in Y$. Given $\epsilon > 0$, can we find $f \in Con(Y, d_Y)$ such that $d_Y(y, f(y)) < \epsilon$?*

1.2 Iterated Function Systems: The Idea

The concept we wish to develop in this chapter is that of iterated function systems (IFS) which were first developed by Hutchinson [26]. They were independently discovered by Barnsley and Demko [6] who gave them their name. To motivate their development, we must enter the realm of fractals. We begin with the famous construction of the Cantor "middle-thirds" set [23, pp.114-116].

We construct the Cantor set by induction. Let $I_0 = I = [0,1] \subset \mathbb{R}$. Let $I_1 = I_0 \backslash (\frac{1}{3}, \frac{2}{3})$, that is, the interval $[0,1]$ with the open middle-third removed.



Construct I_2 from I_1 by removing the open middle-thirds from the two remaining closed intervals.

Inductively, construct I_{n+1} by removing the open middle-thirds from the 2^n closed intervals of I_n. We define the *Cantor set* to be $\mathcal{C} = \bigcap_{n=0}^{\infty} I_n$.

Suppose that you were asked to describe $\mathcal{C}$. At this point, it might be difficult without giving the argument for its construction. Returning to the ideas presented at the end of Section 1.1, we would like to find a function f, on some appropriate space, for which $\mathcal{C}$ is the attractor. This function could then be iterated to find $\mathcal{C}$.

To understand what we wish to do in general, let us look at two characteristics of $\mathcal{C}$. By construction, $\mathcal{C}$ is compact. A second characteristic is the self-similarity we find within it. This is one of the reasons $\mathcal{C}$ is called a fractal. A general definition of self-similarity has been given in [30].

Definition 1.2.1 *Let (X, d) be a complete metric space and Λ be a compact topological space. If Λ is finite, it is assumed to have the discrete topology. Suppose that for each $\lambda \in \Lambda$, there is a contraction w_λ on X. Assume that not all the w_λ are constant and that each has contraction factor s. Define the map $w : \Lambda \times X \to X$ by*

$$w(\lambda, x) = w_\lambda(x).$$

Define Ω to be the ordered triple $((X, d), w, \Lambda)$ and $F_\Omega = \{w_\lambda : \lambda \in \Lambda\}$. Then Ω is called a contraction system.

A set A is self-similar under F_Ω *if it is a non-empty compact subset of X such that*

$$A = \bigcup \{w_\lambda(A) : \lambda \in \Lambda\}.$$

Definition 1.2.2 *Let (X, d) be a complete metric space. Then $A \subset X$ is called* self-similar *if there is a contraction system Ω such that A is self-similar under F_Ω.*

To see the self-similarity in $\mathcal{C}$, let $\mathcal{C}_1 = \mathcal{C} \cap [0, \frac{1}{3}]$ and let $\mathcal{C}_2 = \mathcal{C} \cap [\frac{2}{3}, 1]$. Intuitively, if we were to "zoom in" on $\mathcal{C}_1$ or $\mathcal{C}_2$, we could not distinguish either from $\mathcal{C}$. Mathematically, we see that the maps $w_1 : \mathcal{C} \to \mathcal{C}_1$ defined by $x \mapsto \frac{x}{3}$ and $w_2 : \mathcal{C} \to \mathcal{C}_2$ defined by $x \mapsto \frac{x}{3} + \frac{2}{3}$ are metric equivalences under the induced topology of $\mathbb{R}$. Indeed, $\mathcal{C}$ is the disjoint union of two metrically equivalent subsets:

$$\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2.$$

We wish $\mathcal{C}$ to be the fixed point of a certain function. We motivate the following definition by the fact that $\mathcal{C}$ is a subset of I.

Definition 1.2.3 *Let X and Y be sets and $f : X \to Y$. We define the set mapping $\hat{f}: P(X) \to P(Y)$ by*

$$\hat{f}(A) = \{f(a) : a \in A\} \quad \forall A \in P(X),$$

where $P(X)$ denotes the power set of X.

We see that $\mathcal{C}_1 = \hat{w}_1(\mathcal{C})$ and $\mathcal{C}_2 = \hat{w}_2(\mathcal{C})$. Therefore,

$$\mathcal{C} = \hat{w}_1(\mathcal{C}) \cup \hat{w}_2(\mathcal{C}). \tag{1.4}$$

Hence, $\mathcal{C}$ can be written as a union of contracted copies of itself. This is what we wish to do in general. Given a set A, try to write A as a union of contracted copies of itself.

Definition 1.2.4 *Let X and Y be sets and $f_\lambda : X \to Y$, $\lambda \in \mathcal{A}$, where $\mathcal{A}$ is some indexing set. Let $\mathbf{f} = \{f_\lambda\}$. We define $\hat{\mathbf{f}} = \cup_{\lambda \in \mathcal{A}} \hat{f}_\lambda$, that is for $A \subset X$, we have*

$$\hat{\mathbf{f}}(A) = \bigcup_{\lambda \in \mathcal{A}} \hat{f}_\lambda(A).$$

If we now set $\mathbf{w} = \{w_1, w_2\}$, we see by Equation (1.4), that $\mathcal{C}$ is the fixed point of $\hat{\mathbf{w}}$. Now, for this $\mathbf{w}$ to be useful, we would need $\mathcal{C}$ to be its attractor in some appropriate space. Intuitively we might think this is true since $I_{n+1} = \hat{\mathbf{w}}(I_n)$, for each $n \in \mathbb{N}$. Hence, in a way, $\mathbf{w}$ is an exact description of $\mathcal{C}$. In general, the desired approximations would be obtained by iterating maps of the form given in Definition 1.2.4. This is the concept of an iterated function system, or IFS [5].

Definition 1.2.5 *An* iterated function system, *or* IFS, *consists of a complete metric space (X, d) together with a finite set of contraction mappings w_n: $X \to X$ with respective contractivity factors c_n, $n = 1, 2, \ldots, N$.* [1] *Such an IFS, denoted by $\mathbf{w}$, where*

[1] *This definition can be extended to an infinite number of maps [30].*

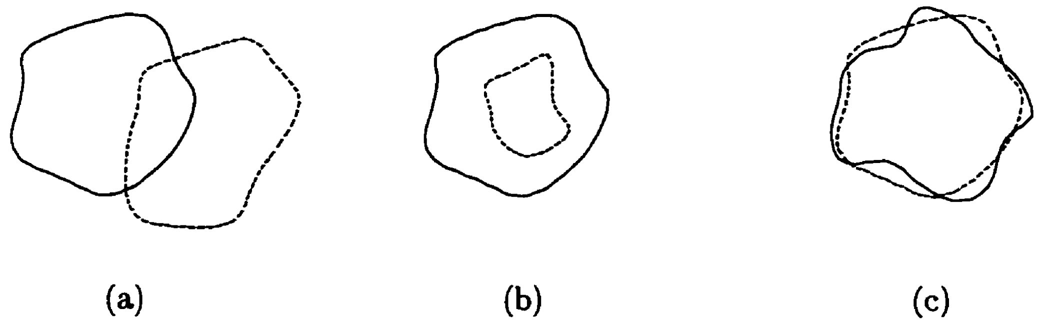


Figure 1.1: Closeness of sets.

w= $\{w_n : n = 1, 2, \dots, N\}$, *is called an* N-map IFS. *The IFS is said to have* contractivity $c = \max\{c_n : n = 1, 2, \dots, N\}$.

The meaning of the contractivity of an IFS will be made clear in the following section. For this it will be necessary to define an appropriate complete space, the elements of which are to be approximated.

1.3 A Complete Space for IFS

We will use the example of $\mathcal{C} \subset \mathbb{R}$ to motivate the search for a space consisting of subsets of a complete metric space. Given a complete space (X, d), the goal is to find a complete space (Y, d_Y) with $Y \subset P(X)$. We will first construct a distance function d_Y on $P(X)$ and use the conditions needed for it to be a metric to help us determine Y. To begin, consider the three pairs of sets in $\mathbb{R}^2$ pictured in Figure 1.1.

Each case consists of two sets, one bounded by the solid line and one by the dashed line. In which case do the two sets seem "closest"? Probably not in (a). In (b), the dashed set certainly seems close to the solid one; it is part of the solid set. However, many points of the solid set seem distant from the dashed set. Case (c) seems intuitively right. Most points of the solid are close to the dashed, and vice-versa. More precisely, each set overlaps the other set rather well. We use these ideas to begin to construct our metric.

Unless otherwise specified, (X, d) will denote a metric space with no other properties.

Notation 1.3.1 Let $x \in X, B \subset X$. Define the distance from x to B by

$$d(x, B) = \inf_{b \in B} d(x, b).$$

Hence, if $x \in B$, $d(x, B) = 0$.

Notation 1.3.2 Let $A, B \subset X$. Define the distance from A to B by

$$d(A, B) = \sup_{a \in A} d(a, B).$$

This seems reasonable since if $A \subset B$, A should be close to B and by this definition we would have $d(A, B) = 0$. Unfortunately, this function is not a metric. For example, $d([0, \frac{1}{2}], [\frac{1}{3}, 1]) = \frac{1}{3}$ but $d([\frac{1}{3}, 1], [0, \frac{1}{2}]) = \frac{1}{2}$. Symmetry is lacking, which motivates the following construction [5, 17, 26]:

Definition 1.3.3 *Let $A, B \subset X$. Define the* Hausdorff distance *between A and B by*

$$h(A, B) = d(A, B) \vee d(B, A),$$

where $x \vee y = \max\{x, y\}$.

This is the function d_Y we seek. It satisfies the intuitive notion of two sets being close, which was found in (c) on page 16. The function h is almost a metric; we use the following table to help rule out certain sets from $P(X)$:

Problem: $h(\emptyset, [0,1]) =?$	Solution: Only consider non-$\emptyset$ sets.
Problem: $h([0,1), [0,1]) = 0$	Solution: Only consider closed sets.
Problem: $h([0,1], [0,\infty)) = \infty$	Solution: The closed sets must be compact.

Notation 1.3.4 Define $\mathcal{H}(X)$ to be the set of all non-empty, compact subsets of X.

Theorem 1.3.5 *Let (X, d) be a metric space. Then $(\mathcal{H}(X), h)$ is a metric space.*

Proof The proof will generally follow the one in [5]. Let $A, B, C \in \mathcal{H}(X)$. As the sets in question are compact, we can change sup to max in the definition of d. We see that

$$h(A, A) = d(A, A) = \max\{d(a, A) : a \in A\} = 0.$$

If $A \neq B$, then without loss of generality let $a \in A \backslash B$. Therefore,

$$\begin{aligned} h(A, B) &\geq d(A, B) && \text{definition of } h(A, B) \\ &\geq d(a, B) && \text{definition of } d(A, B) \\ &> 0 && \text{definition of } d(a, B). \end{aligned}$$

By definition, h is symmetric and we are left to verify the triangle inequality. For $a \in A$ we have

$$\begin{aligned} d(a, B) &= \min\{d(a, b) : b \in B\} \\ &\leq \min\{d(a, c) + d(c, b) : b \in B\} \quad \forall c \in C \\ &= d(a, c) + \min\{d(c, b) : b \in B\} \quad \forall c \in C. \end{aligned}$$

Therefore,

$$\begin{aligned} d(a, B) &\leq \min\{d(a, c) : c \in C\} + \max\{\min\{d(c, b) : b \in B\} : c \in C\} \\ &= d(a, C) + d(C, B). \end{aligned}$$

Hence,

$$\begin{aligned}\max\{d(a,B) : a \in A\} &\leq \max\{d(a,C) : a \in A\} + d(C,B)\\ \Longrightarrow d(A,B) &\leq d(A,C) + d(C,B).\end{aligned}$$

By an argument symmetric in A and B, $d(B,A) \leq d(B,C) + d(C,A)$. Thus we obtain

$$\begin{aligned} h(A,B) &= d(A,B) \vee d(B,A) \\ &\leq [d(A,C) + d(C,B)] \vee [d(B,C) + d(C,A)] \\ &\leq [d(A,C) \vee d(C,A)] + [d(B,C) \vee d(C,B)] \\ &= h(A,C) + h(B,C). \qquad \blacksquare \end{aligned}$$

We are almost at our goal. We have the metric space $(\mathcal{H}(X), h)$; all that remains is to show it is complete. For this, we follow the development in [5].

Notation 1.3.6 Let $S \subset X$ and let $r \geq 0$. Then let $S + r = \{x \in X : d(x,s) \leq r$ for some $s \in S\}$. We call $S + r$ the *dilatation of S by a ball of radius r.*

Lemma 1.3.7 *Let $A, B \in \mathcal{H}(X)$ and let $\epsilon > 0$. Then*

$$h(A,B) \leq \epsilon \iff A \subset B + \epsilon \textit{ and } B \subset A + \epsilon.$$

Proof This is the idea of overlapping as seen in c) on page 16. We will show $d(A,B) \leq \epsilon \iff A \subset B + \epsilon$.

($\Rightarrow$) Suppose $d(A,B) \leq \epsilon$. Then $\max\{d(a,B) : a \in A\} \leq \epsilon$. Therefore, by definition of $B + \epsilon$, $a \in B + \epsilon \ \forall a \in A$. Hence, $A \subset B + \epsilon$.

($\Leftarrow$) Suppose $A \subset B + \epsilon$. For each $a \in A, \exists b \in B$ such that $d(a,b) \leq \epsilon$. Hence, $\forall a \in A$, $d(a,B) \leq \epsilon$ and thus $d(A,B) \leq \epsilon$.

Therefore, by definition of h,

$$\begin{aligned} h(A,B) \leq \epsilon &\iff d(A,B) \leq \epsilon \text{ and } d(B,A) \leq \epsilon \\ &\iff A \subset B+\epsilon \text{ and } B \subset A+\epsilon. \qquad \blacksquare \end{aligned}$$

The goal is to show the completeness of $(\mathcal{H}(X), h)$ when (X, d) is complete. The completeness of X is essential, as the following example demonstrates:

Example 1.3.8 Let $X = [0, 1)$ with the usual Euclidean metric. Then $(\{1 - \frac{1}{n}\}) \to \{1\} \notin \mathcal{H}(X)$. Hence $\mathcal{H}(X)$ is not complete.

It will be necessary to consider the convergence of Cauchy sequences in $(\mathcal{H}(X), h)$. If (A_n) is a Cauchy sequence in $(\mathcal{H}(X), h)$, then by Lemma 1.3.7, given $\epsilon > 0, \exists N$ such that $\forall m, n \geq N, A_m \subset A_n + \epsilon$ and $A_n \subset A_m + \epsilon$. As the completeness of $(\mathcal{H}(X), h)$ relies upon that of (X, d), we need the following lemma which allows the *extension of a Cauchy subsequence* $(x_{n_j} \in A_{n_j})$ to a Cauchy sequence $(x_n \in A_n)$.

Lemma 1.3.9 (The Extension Lemma) *Suppose (A_n) is a Cauchy sequence in $(\mathcal{H}(X), h)$ and let (n_j) be an infinite, strictly increasing, sequence of positive natural numbers. Suppose that $(x_{n_j} \in A_{n_j})$ is a Cauchy sequence in (X, d). Then there exists a Cauchy sequence $(\tilde{x}_n \in A_n)$ such that $\tilde{x}_{n_j} = x_{n_j}$ $\forall j \in \mathbb{N}$.*

Proof Construct the sequence $(\tilde{x}_n \in A_n)$. For each $1 \leq n \leq n_1$, pick $\tilde{x}_n \in \{x \in A_n : d(x_{n_1}, x) = d(x_{n_1}, A_n)\}$. Such a point exists, since each A_n is compact. Proceed in a similar fashion for $n_j + 1 \leq n \leq n_{j+1}$ $\forall j = 1, 2, \ldots$. That is, choose $\tilde{x}_n \in \{x \in A_n : d(x_{n_j}, x) = d(x_{n_j}, A_n)\}$. The claim is that $(\tilde{x}_n)$ is a Cauchy sequence in X.

To see this, let $\epsilon > 0$. Since (x_{n_j}) is a Cauchy sequence, let $M \in \mathbb{N}$ such that $\forall n_k, n_j \geq M$, $d(x_{n_k}, x_{n_j}) \leq \frac{\epsilon}{3}$. Then, choose $N \geq M$ such that $\forall m, n \geq N$, $d(A_m, A_n) \leq \frac{\epsilon}{3}$. Let $m, n \geq N$. Pick $n_{j-1} < m \leq n_j$, and $n_{k-1} < n \leq n_k$. Since $h(A_m, A_{n_j}) < \frac{\epsilon}{3}$, there is a

$y \in A_m \cap (\{x_{n_j}\} + \frac{\epsilon}{3})$, thus $d(\tilde{x}_m, x_{n_j}) \leq \frac{\epsilon}{3}$. Similarly $d(x_{n_k}, \tilde{x}_n) \leq \frac{\epsilon}{3}$. Therefore

$$\begin{aligned} d(\tilde{x}_m, \tilde{x}_n) &\leq d(\tilde{x}_m, x_{n_j}) + d(x_{n_j}, x_{n_k}) + d(x_{n_k}, \tilde{x}_n) \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &\leq \epsilon. \quad \blacksquare \end{aligned}$$

The main result now follows:

Theorem 1.3.10 *If (X, d) is complete, then $(\mathcal{H}(X), h)$ is also. Moreover, if (A_n) is a Cauchy sequence in $\mathcal{H}(X)$, then $A = \lim\limits_{n\to\infty} A_n \in \mathcal{H}(X)$ is given by*

$$A = \{x \in X : \exists \text{ a Cauchy sequence } (x_n \in A_n) \text{ that converges to } x\}.$$

Proof The proof will follow the one given in [5]. Let (A_n) be a Cauchy sequence in $\mathcal{H}(X)$ and let A be as in the statement of the theorem. We break the proof up into five parts:

a) $A \neq \emptyset$;

b) A is closed, hence complete since X is complete;

c) for $\epsilon > 0, \exists N$ such that $\forall n \geq N, A \subset A_n + \epsilon$;

d) A is totally bounded, hence by b) is compact;

e) $\lim\limits_{n\to\infty} A_n = A$.

Proof of a): We use the Extension Lemma to find a Cauchy sequence $(a_i \in A_i)$ in X. Then $\lim a_i = a \in A$ by definition of A. Therefore $A \neq \emptyset$.

Since (A_i) is a Cauchy sequence, get a strictly increasing sequence (N_i) such that $h(A_m, A_n) < \frac{1}{2^i}\ \forall m, n > N_i$. Let $x_{N_1} \in A_{N_1}$. Given that $h(A_{N_1}, A_{N_2}) \leq \frac{1}{2}$, get $x_{N_2} \in A_{N_2}$ such that $d(x_{N_1}, x_{N_2}) \leq \frac{1}{2}$. Suppose $(x_{N_i} \in A_{N_i})_{i=1}^{k}$ is a sequence such that $d(x_{N_{i-1}}, x_{N_i}) \leq \frac{1}{2^{i-1}}$. Then, since $h(A_{N_k}, A_{N_{k+1}}) \leq \frac{1}{2^k}$, choose $x_{N_{k+1}} \in A_{N_{k+1}}$ such that $d(x_{N_k}, x_{N_{k+1}}) \leq \frac{1}{2^k}$.

The claim is that (x_N) is a Cauchy sequence in X. To see this, let $\epsilon > 0$ and choose N such that $\sum_{i=N}^{\infty} \frac{1}{2^i} < \epsilon$. Then for $m > n \geq N$,

$$\begin{aligned} d(x_{N_m}, x_{N_n}) &\leq d(x_{N_m}, x_{N_{m+1}}) + d(x_{N_{m+1}}, x_{N_{m+2}}) + \ldots + d(x_{N_{n-1}}, x_{N_n}) \\ &\leq \sum_{i=N}^{\infty} \frac{1}{2^i} \\ &< \epsilon. \end{aligned}$$

Now, by the Extension Lemma, let $(a_i \in A_i)$ be a Cauchy sequence such that $a_{N_i} = x_{N_i}$. By the completeness of X, $\lim a_i$ exists. Hence $A \neq \emptyset$.

Proof of b): Suppose $(a_i \in A) \to a$. For $i \in \mathbb{N}^+$, get a sequence $(x_{i,n} \in A_n)$ with $\lim_{n\to\infty} x_{i,n} = a_i$, by definition of A. Since $(a_i) \to a$, let (N_i) be an increasing sequence of positive integers such that $d(a_{N_i}, a) < \frac{1}{i}$. Then let $(m_i) \in \mathbb{Z}$ such that $d(x_{N_i,m_i}, a_{N_i}) \leq \frac{1}{i}$. Therefore $d(x_{N_i,m_i}, a) \leq \frac{2}{i}$. Now let $y_{m_i} = x_{N_i,m_i}$. Then by definition of x_{N_i,m_i}, $y_{m_i} \in A_{m_i}$ $\forall i$ and $\lim_{i\to\infty} y_{m_i} = a$. By the Extension Lemma, let $(z_i \in A_i)$ be a sequence such that $z_{m_i} = y_{m_i}$ with $(z_i) \to a$. Hence A is closed, and by the completeness of X, is itself complete.

Proof of c): Let $\epsilon > 0$. Choose N_1 such that $\forall m, n \geq N_1, h(A_n, A_m) < \epsilon$. Let $n \geq N_1$. By Lemma 1.3.7, $\forall m \geq n, A_m \subset A_n + \epsilon$. Now let $a \in A$ and suppose there is a sequence $(a_i \in A_i) \to a$. Choose $N \geq N_1$ such that $\forall m \geq N, d(a_m, a) < \epsilon$. Then $a_m \in A_n + \epsilon$, since $A_m \subset A_n + \epsilon$. As A_n is compact, $A_n + \epsilon$ is closed. Hence as $a_m \in A_n + \epsilon$ $\forall m \geq N$, the limit $a \in A_n + \epsilon$. Since a was arbitrary, $A \subset A_n + \epsilon$.

Proof of d): Suppose A is not totally bounded. Then by definition of total boundedness, for some $\epsilon > 0$, there does not exist a finite ϵ-net of A. Thus, choose $(x_i) \subset A$ such that $d(x_i, x_j) \geq \epsilon$ $\forall i \neq j$. By c), get n such that $A \subset A_n + \frac{\epsilon}{3}$. Then, $\forall x_i$, pick $y_i \in A_n$ such that $d(x_i, y_i) \leq \frac{\epsilon}{3}$. Since A_n is compact, some $(y_{n_i}) \subset (y_i)$ converges. Therefore, let $n_i \neq n_j$ such that $d(y_{n_i}, y_{n_j}) < \frac{\epsilon}{3}$. But then,

$$\begin{aligned} d(x_{n_i}, x_{n_j}) &\leq d(x_{n_i}, y_{n_i}) + d(y_{n_i}, y_{n_j}) + d(y_{n_j}, x_{n_j}) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\ &= \epsilon. \end{aligned}$$

This contradicts the hypothesis on (x_i), hence A is totally bounded, and by b) is compact. Therefore by a), $A \in \mathcal{H}(X)$.

Proof of e): As $A \in \mathcal{H}(X)$, by c) and Lemma 1.3.7, the result will be proven if for $\epsilon > 0, \exists N$ such that $\forall n \geq N, A_n \subset A + \epsilon$.

Let $\epsilon > 0$. Choose N such that $\forall m, n \geq N, h(A_m, A_n) \leq \frac{\epsilon}{2}$. Then, for $m, n \geq N, A_m \subset A_n + \frac{\epsilon}{2}$. Let $l \geq N$. We claim $A_l \subset A + \epsilon$. Let $y \in A_l$ and choose an increasing sequence $(N_i) \subset \mathbb{N}$ such that $l < N_1 < N_2 < \ldots$ and such that $\forall m, n \geq N_j, A_m \subset A_n + \frac{\epsilon}{2^{j+1}}$.

By the choice of l, $A_l \subset A_{N_1} + \frac{\epsilon}{2}$. Since $y \in A_l$, get $x_{N_1} \in A_{N_1}$ such that $d(y, x_{N_1}) \leq \frac{\epsilon}{2}$. As $x_{N_1} \in A_{N_1}$ we get, by compactness of A_{N_2}, an $x_{N_2} \in A_{N_2}$ with $d(x_{N_1}, x_{N_2}) \leq \frac{\epsilon}{2^2}$. By induction, choose $x_{N_j} \in A_{N_j}$ such that $d(x_{N_j}, x_{N_{j+1}}) \leq \frac{\epsilon}{2^{j+1}}$. Hence

$$
\begin{aligned}
d(y, x_{N_j}) &\leq d(y, x_{N_1}) + \sum_{i=1}^{j} d(x_{N_i}, x_{N_{i+1}}) \\
&\leq \frac{\epsilon}{2} + \sum_{i=1}^{j} \frac{\epsilon}{2^{i+1}} \\
&< \epsilon \quad \forall j \in \mathbb{N}^+.
\end{aligned}
$$

For $j \geq i$,

$$
\begin{aligned}
d(x_{N_i}, x_{N_j}) &\leq \sum_{k=i}^{j-1} d(x_{N_k}, x_{N_{k+1}}) \\
&\leq \sum_{k=i}^{j-1} \frac{\epsilon}{2^{k+1}} \\
&< \epsilon.
\end{aligned}
$$

Therefore, (x_{N_i}) is a Cauchy sequence. By construction, $A_{N_j} \subset A_n + \frac{\epsilon}{2}$. Suppose $(x_{N_j}) \to x$. Since $A_n + \frac{\epsilon}{2}$ is closed, this implies $x \in A_n + \frac{\epsilon}{2}$, and since $d(y, x_{N_j}) < \epsilon \; \forall j \in \mathbb{N}^+$, we have $d(y, x) < \epsilon$. Thus $A_n \subset A + \epsilon \; \forall n \geq N$. Combining this with c), $\lim_{n \to \infty} A_n = A$ and hence $(\mathcal{H}(X), h)$ is complete. ■

We will now prove a few properties about the Hausdorff metric which will enable us to justify the contractivity of an iterated function system as seen in Definition 1.2.5. Again, let (X, d) be a metric space.

Notation 1.3.11 Let $Con(X, d, s)$ denote the set of all contractive maps with contractivity at least s.

Lemma 1.3.12 *Let $w \in Con(X, d, s)$, then $\hat{w} \in Con(\mathcal{H}(X), h, s)$.*

Proof Since w is continuous, it takes compact sets to compact sets. Hence $\hat{w} : \mathcal{H}(X) \to \mathcal{H}(X)$. Let $B, C \in \mathcal{H}(X)$. Then

$$\begin{aligned} d(\hat{w}(B), \hat{w}(C)) &= \max\{\min\{d(w(b), w(c)) : c \in C\} : b \in B\} \\ &\leq \max\{\min\{sd(b, c) : c \in C\} : b \in B\} \\ &= sd(B, C). \end{aligned}$$

By a symmetric argument, $d(\hat{w}(C), \hat{w}(B)) \leq sd(C, B)$. Therefore

$$\begin{aligned} h(\hat{w}(B), \hat{w}(C)) &= d(\hat{w}(B), \hat{w}(C)) \vee d(\hat{w}(C), \hat{w}(B)) \\ &\leq (sd(B, C)) \vee (sd(C, B)) \\ &= s(d(B, C) \vee d(C, B)) \\ &= sh(B, C). \quad \blacksquare \end{aligned}$$

The following lemmas have proofs similar to the above.

Lemma 1.3.13 *Let $B, C \in \mathcal{H}(X)$. Then $B \subset C \implies d(x, C) \leq d(x, B)\ \forall x \in X$.*

Lemma 1.3.14 *Let (X, d) be complete. If $A, B, C \in \mathcal{H}(X)$, and $B \subset C$, then*

$$d(A, C) \leq d(A, B).$$

Lemma 1.3.15 *Let A, B, C be as above. Then $d(A \cup B, C) = d(A, C) \vee d(B, C)$.*

Lemma 1.3.16 *Let $A, B, C, D \in \mathcal{H}(X)$. Then $h(A \cup B, C \cup D) \leq h(A, C) \vee h(B, D)$.*

Hence, we may prove the following result:

Proposition 1.3.17 *Let (X, d) be a metric space and let*

$$\mathbf{w} = \{w_n \in Con(X, d, c_n) : n = 1, 2, \ldots, N\}.$$

Then $\hat{\mathbf{w}} \in Con(\mathcal{H}(X), h, c)$, where $c = \max\{c_n : n = 1, 2, \ldots, N\}$.

Proof The proof is by induction on N, the case $N = 1$ having been done in Lemma 1.3.12. Suppose for $2 \leq N \leq k$, $h(\hat{\mathbf{w}}(B), \hat{\mathbf{w}}(C)) \leq sh(B, C)$. Let $\hat{\mathbf{w}}_k = \bigcup\limits_{n=1}^{k} \hat{w}$ and $s = \max\{c_n : n = 1, 2, \ldots, k\}$. Then

$$\begin{aligned}
h(\hat{\mathbf{w}}(B), \hat{\mathbf{w}}(C)) &= h(\hat{\mathbf{w}}_k(B) \cup \hat{w}_{k+1}(B), \hat{\mathbf{w}}_k(C) \cup \hat{w}_{k+1}(C)) && \text{by definition of } \hat{\mathbf{w}}_k \\
&\leq h(\hat{\mathbf{w}}_k(B), \hat{\mathbf{w}}_k(C)) \vee h(\hat{w}_{k+1}(B), \hat{w}_{k+1}(C)) && \text{by Lemma 1.3.16} \\
&\leq sh(B, C) \vee c_{k+1}h(B, C) && \text{by hypothesis on } \hat{\mathbf{w}}_k \\
&\leq ch(B, C). && \blacksquare
\end{aligned}$$

Proposition 1.3.17 implies a crucial result for IFS.

Theorem 1.3.18 (BCMP for IFS) *Let $\mathbf{w}$ be an N-map IFS with contractivity c. Then $\hat{\mathbf{w}} \in Con(\mathcal{H}(X), h, c)$. Furthermore $\hat{\mathbf{w}}$ has a unique fixed point $A_{\hat{\mathbf{w}}} \in \mathcal{H}(X)$ which is also its attractor.*

Proof This follows directly from Proposition 1.3.17 and Theorem 1.1.16. $\blacksquare$

Definition 1.3.19 *The fixed point of $\hat{\mathbf{w}}$ is called the* attractor *of $\hat{\mathbf{w}}$.*

This yields the following version of Proposition 1.1.21 for iterated function systems [5]:

Theorem 1.3.20 (The Collage Theorem) *Let $\mathbf{w}$ be an N-map IFS with contractivity $0 \leq c < 1$. Suppose $L \in \mathcal{H}(X)$ and $\epsilon > 0$ are such that $h(L, \hat{\mathbf{w}}(L)) \leq \epsilon$.* [2] *Then $h(L, A_{\hat{\mathbf{w}}}) \leq \frac{\epsilon}{1-c}$.*

---

[2] *The distance $h(L, \hat{\mathbf{w}}(L))$ is often called the* collage distance.

Proof See the proof of Proposition 1.1.21. ■

The Collage Theorem is important for the Inverse Problem of approximating sets seen in Section 1.1. By the Collage Theorem, one could try to construct an IFS $\mathbf{w}$ which takes L close to itself. The attractor of $\mathbf{w}$ would then be close to L.

It is possible that $c \approx 1$ which, in turn, implies that the constant $\epsilon/(1-c)$ can be large. Thus there is no guarantee that the collage distance is small and the approximation may be quite poor. To make $c \approx 0$, one can use maps with small contractivity factors. However, this might increase the number of maps needed to describe the approximation (hence reducing the compression). This fact is relevant when compression is a principal factor.

In order to calculate fractal images using the theoretical machinery that has been developed, one can use the following algorithm, a consequence of Theorem 1.3.18 [5]:

Corollary 1.3.21 (The Deterministic Algorithm) *Let $\mathbf{w}$ be an N-map IFS with $\mathbf{w} = \{w_j : j = 1, 2, \ldots, N\}$. Let $A_0 \in \mathcal{H}(X)$. Compute $A_n = \hat{\mathbf{w}}^{\circ n}(A)$ by $A_{n+1} = \bigcup_{j=1}^{n} \hat{w}_n(A_n)$ for $n = 1, 2, \ldots$. Then the sequence $(A_n) \subset \mathcal{H}(X)$ converges to the attractor of the IFS in $\mathcal{H}(X)$.*

1.4 Examples of IFS Attractors

In practice, affine IFS contraction maps are used to simplify calculations. Let $M_n(\mathbb{R})$ denote the set of all $n \times n$ matrices on $\mathbb{R}$, where $\mathbb{R}^n$ is the usual Euclidean n-space.

Definition 1.4.1 *Let $X \subset \mathbb{R}^n$, $n \in \mathbb{N}^+$. A map $w : X \to \mathbb{R}^n$ is called an* affine transformation *if $\exists A \in M_n(\mathbb{R})$ and $b \in \mathbb{R}^n$ such that*

$$w(x) = Ax + b \quad \forall x \in X.$$

In general, given vector spaces X and Y, an affine transformation *$f : X \to Y$ is a map of the form*

$$f(x) = Ax + b$$

where A is a linear transformation from X to Y and $b \in Y$.

Example 1.4.2 Let $X = [0, 1]$ and let $w_i(x) = \frac{1}{3}(x + 2i), i = 0, 1$. Then $A_{\hat{\mathbf{w}}} = \mathcal{C}$.

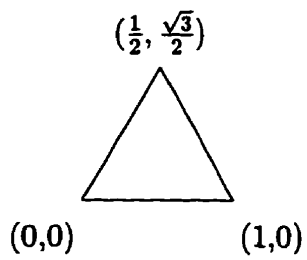
Example 1.4.3 Let $X = [0, 1]^2$. Define the following maps:

$$\begin{aligned} w_1(x, y) &= \left(\frac{x}{2}, \frac{y}{2}\right), \\ w_2(x, y) &= \left(\frac{x}{2} + \frac{1}{2}, \frac{y}{2}\right), \end{aligned}$$

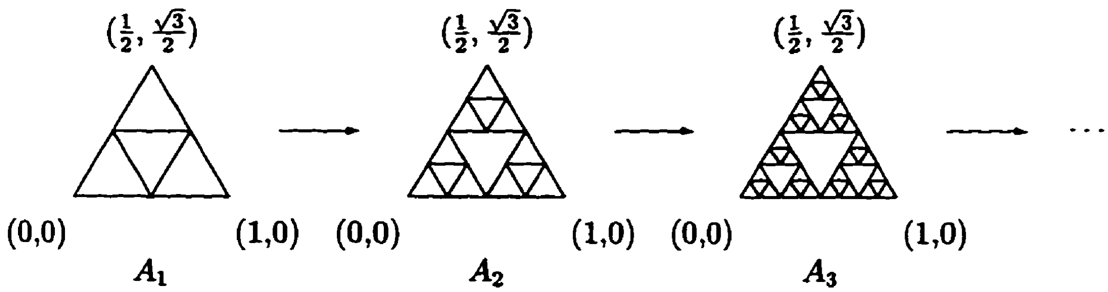
and

$$w_3(x, y) = \left(\frac{x}{2} + \frac{1}{4}, \frac{y}{2} + \frac{\sqrt{3}}{4}\right).$$

To find the attractor of $\hat{\mathbf{w}}$, we use the Deterministic Algorithm. We are allowed to make any choice of A_0. Therefore, let A_0 be the following triangle:



Then, using the algorithm, we obtain the following sequence of sets:



This sequence converges to the Sierpinski gasket [33].

An *affine IFS* $\mathbf{w} = \{w_i\}$ is an IFS where each w_i is affine. Often, affine IFS in $\mathbb{R}^2$ will be written in a table to facilitate their description. Consider an IFS consisting of the maps

$$w_i(x, y) = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e_i \\ f_i \end{pmatrix} \qquad i = 1, 2, \ldots, N.$$

Instead of writing them as above, they are written in a table such as:

$$\begin{array}{cccccc} a_1 & b_1 & c_1 & d_1 & e_1 & f_1 \\ a_2 & b_2 & c_2 & d_2 & e_2 & f_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_N & b_N & c_N & d_N & e_N & f_N \end{array}$$

We now recall the definition of a similitude.

Definition 1.4.4 *A transformation* $w : \mathbb{R}^2 \to \mathbb{R}^2$ *is called a* similitude *if it is an affine transformation of the form*

$$w(x, y) = r \begin{pmatrix} \cos\theta & \pm\sin\theta \\ \sin\theta & \mp\cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}$$

where $(e, f) \in \mathbb{R}^2, r \neq 0, \theta \in [0, 2\pi)$. *The constant* r *is called the* scaling *of* w *or its* scale factor *and* θ *is called its* angle of rotation.

Proposition 1.4.5 *If $w(x) = Ax + b, A \in M_2$, $b, x \in \mathbb{R}^2$ is a similitude in $\mathbb{R}^2$, then its contractivity factor is $|detA|$.*

Proof This is simply a matter of calculating

$$d(w(x_1, y_1), w(x_2, y_2))$$

for the points $(x_1, y_1), (x_2, y_2)$ and using the definition of $detA$ in M_2. ■

Notation 1.4.6 Define the following three sets:

$$\begin{aligned}
Con_1(X,d) &= \{w \in Con(X,d) : w \text{ is 1-1}\};\\
Sim(X,d) &= \{w : w \text{ is a similitude on } X\};\\
Sim_1(X,d) &= Sim(X,d) \cap Con_1(X,d).
\end{aligned}$$

Corollary 1.4.7 *If $w : \mathbb{R}^2 \to \mathbb{R}^2$ is a similitude as above and $|detA| < 1$, then $w \in Con(X, d)$.*

Proof Use Proposition 1.4.5. ■

We now wish to apply this theory to images, i.e. computer images. One can think of an *image* as being a compact subset of $\mathbb{R}^n$. One can model a computer screen by $X = [0, 1]^2$ and define an image on the screen to be a set A in X, with points being screen pixels. If $x \in A$, the associated pixel is plotted white. If $x \notin A$, leave the pixel black. Hence a white screen represents $A = [0, 1]^2$.

Suppose an IFS acts on the screen. When the IFS is iterated, the points of A move about the screen. Looking at $\hat{\mathbf{w}}(A)$, we see that $x \in \hat{\mathbf{w}}(A)$ if $\exists i \in 1, 2, \ldots, N$ such that $x = w_i(y)$ for some $y \in A$. Hence, after one iteration of $\hat{\mathbf{w}}$, a pixel is plotted white if there is a white pixel mapped to it by the IFS. We can therefore think of IFS as mapping black and white images to black and white images.

Unfortunately, as they say, the world is not black and white. What is needed is an IFS-type method which allows for, say, greys?! We might want maps which move pixels around and then scale their grey-levels. These thoughts lead to IFSM [20].

1.5 From IFS to IFSM

The idea of applying IFS methods to grey-level (grey-scale) images was developed by Forte and Vrscay [19, 20, 21]. They formulated an IFS-type method which allows the creation of grey-scale images. Let us consider a compact subset A of $\mathbb{R}^2$ to stimulate some ideas. It is necessary to formulate a definition of A being a grey-scale image. One possible way to do this is to think of the image as a function, rather than a set. What might work is to formulate an IFS method on functions, functions from sets to grey-levels. The question remaining is how?

Following [49], we first formulate this idea for the IFS case. In this case, one finds a simple association to functions. Here, points in images can take on two values: black and white. Therefore, the function associated with a set A is χ_A, the *characteristic function* on A, where

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

The sets considered for IFS are compact, hence it would be natural to consider functions which are characteristic functions of compact sets [49]. Once again, let (X, d) be a metric space.

Notation 1.5.1 Let f be a function from a set A to $\mathbb{R}$. Then inv(f) is defined as

$$\mathrm{inv}(f) = f^{-1}(1).$$

Notation 1.5.2 Let $\mathcal{F}_{BW}(X) = \{f : X \to \{0,1\} \mid \mathrm{inv}(f) \in \mathcal{H}(X)\}$, be the *black and white functions* on X.

We write "BW" to emphasize the fact that we are considering functions which take only two values, 0 (black) and 1 (white).

Recall, for an IFS $\mathbf{w} = \{w_i : i = 1, 2, \ldots, N\}$, the map $\hat{\mathbf{w}} : \mathcal{H}(X) \to \mathcal{H}(X)$ is given by

$$\hat{\mathbf{w}}(S) = \bigcup_{i=1}^{N} \hat{w}_i(S) \quad \forall S \in \mathcal{H}(X).$$

It might therefore be of interest to consider for $A_i \in \mathcal{H}(X), i = 1, 2, \ldots, N$, and $w \in Con(X, d)$,

i) χ_A in terms of $\chi_{A_1}, \ldots, \chi_{A_N}$ where $A = \bigcup_{i=1}^{N} A_i$, and

ii) $\chi_{\hat{w}(A)}$ in terms of χ_A.

Proposition 1.5.3 *Let X be a set and $A_i \subset X$ for $i = 1, 2, \ldots, N$. Then if $A = \bigcup_{i=1}^{N} A_i$,*

$$\chi_A(x) = \max_{i=1,2,\ldots,N} \chi_{A_i}(x).$$

Proof If $x \in A$, choose i such that $x \in A_i$. Hence RHS=LHS. If $x \notin A$, then $\forall i, x \notin A_i$ and $\chi_{A_i}(x) = 0$. This implies that $\max_{i=1,2,\ldots,N} \chi_{A_i}(x) = 0$. Hence RHS=LHS. ∎

Proposition 1.5.4 *Let $A \subset X$ and $w : A \to X$ be 1-1. Then*

$$\chi_{\hat{w}(A)}(x) = \chi_A(w^{-1}(x)) \quad \forall x \in \hat{w}(X).$$

Proof Let $x \in \hat{w}(X)$. Then

$$\chi_{\hat{w}(A)}(x) = \begin{cases} 1 & x \in \hat{w}(A) \\ 0 & x \notin \hat{w}(A). \end{cases}$$

and

$$\chi_A(w^{-1}(x)) = \begin{cases} 1 & w^{-1}(x) \in A \\ 0 & w^{-1}(x) \notin A. \end{cases}$$

However, since w is 1-1, $w^{-1}(x) \in A \iff x \in \hat{w}(A)$. ∎

Proposition 1.5.5 *Let $A \subset X$ and $\mathbf{w} = \{w_1, w_2, \ldots, w_N\}$, with w_i being 1-1 for all $1 \leq i \leq N$. Then $\chi_{\hat{\mathbf{w}}(A)}(x) = \max'_{1\leq i\leq N} \chi_A(w_i^{-1}(x))$ $\forall x \in X$. The notation* max′ *indicates that only subscripts i, where $x \in w_i(X)$, are considered, using the convention* $\max\emptyset = 0$.

Proof The proof follows from Propositions 1.5.3 and 1.5.4. ∎

Therefore, given an IFS $\mathbf{w} = \{w_1, w_2, \ldots, w_N\}$, there is an associated operator $T_{\mathbf{w}}^{BW}$: $\mathcal{F}_{BW}(X) \to \mathcal{F}_{BW}(X)$ defined on $f \in \mathcal{F}_{BW}(X)$ by

$$T_{\mathbf{w}}^{BW} f(x) = \max_{1\leq i\leq N}{}' f(w_i^{-1}(x)) \quad \forall x \in X.$$

The goal is to develop a "black and white" IFS theory on $\mathcal{F}_{BW}(X)$, hence a complete metric must be defined on this space. For insight, consider the next proposition.

Proposition 1.5.6 *Let (X, d) be complete, $w_1, w_2, \ldots, w_N \in Con_1(X, d)$ and $u \in \mathcal{F}_{BW}(X)$. Then*

$$\mathrm{inv}(T_{\mathbf{w}}^{BW} u) = \hat{\mathbf{w}}(\mathrm{inv}(u)),$$

where $\hat{\mathbf{w}}(A) = \bigcup_{i=1}^{N} \hat{w}_i(A)$ for $A \in \mathcal{H}(X)$.

Proof Given $u \in \mathcal{F}_{BW}(X)$,

$$\begin{aligned} x \in \mathrm{inv}(T_{\mathbf{w}}^{BW}u) &\iff T_{\mathbf{w}}^{BW}u(x) = 1 \\ &\iff \exists i \quad u \circ w_i^{-1}(x) = 1 \\ &\iff \exists i \quad w_i^{-1}(x) \in \mathrm{inv}(u) \\ &\iff \exists i \quad x \in \hat{w}_i(\mathrm{inv}(u)) \\ &\iff x \in \hat{\mathbf{w}}(\mathrm{inv}(u)). \quad \blacksquare \end{aligned}$$

It is therefore natural to define the following metric on $\mathcal{F}_{BW}(X)$:

Definition 1.5.7 *Let* $d_{BW}(u,v)= h(\mathrm{inv}(u), \mathrm{inv}(v))$ $\forall u, v \in \mathcal{F}_{BW}(X)$.

Theorem 1.5.8 *If* (X, d) *is complete, then* $(\mathcal{F}_{BW}(X), d_{BW})$ *is also.*

Proof We first show d_{BW} is a metric. Let $u, v \in \mathcal{F}_{BW}(X)$. Then

$$\begin{aligned} d_{BW}(u,v) = 0 &\iff \mathrm{inv}\, u = \mathrm{inv}\, v \\ &\iff (\forall x, u(x) = 1 \iff v(x) = 1). \end{aligned}$$

Since u and v only take on values of 0 or 1, this happens if and only if $u = v$. The symmetry property and the triangle inequality follow since h is a metric. Therefore d_{BW} is a metric.

Now, let (u_n) be a Cauchy sequence in $\mathcal{F}_{BW}(X)$. By definition of d_{BW}, $(\mathrm{inv}(u_n))$ is a Cauchy sequence in $\mathcal{H}(X)$. Hence by completeness of $\mathcal{H}(X)$, $A = \lim_{n\to\infty} \mathrm{inv}(u_n) \in \mathcal{H}(X)$. Let $u = \chi_A$. Then $u \in \mathcal{F}_{BW}(X)$ since A is compact. Given $\epsilon > 0$, choose N such that $\forall n \geq N$, $h(\mathrm{inv}(u), \mathrm{inv}(u_n)) < \epsilon$. Thus $d_{BW}(u, u_n) < \epsilon$, which implies $(u_n) \to u$ in d_{BW}. Hence $(\mathcal{F}_{BW}(X), d_{BW})$ is complete. $\blacksquare$

This leads to the next theorem.

Theorem 1.5.9 *Let* $w_i \in Con_1(X, d)$ *for* $i = 1, 2, \ldots, N$. *Then* $T_{\mathbf{w}}^{BW}$ *is contractive on* $(\mathcal{F}_{BW}(X), d_{BW})$ *and* $c_{T_{\mathbf{w}}^{BW}} = c_{\hat{\mathbf{w}}}$.

Proof For $u, v \in \mathcal{F}_{BW}(X)$,

$$\begin{aligned} d_{BW}(T_{\mathbf{w}}^{BW} u, T_{\mathbf{w}}^{BW} v) &= h(\operatorname{inv}(T_{\mathbf{w}}^{BW} u), \operatorname{inv}(T_{\mathbf{w}}^{BW} v)) \\ &= h(\hat{\mathbf{w}}(\operatorname{inv} u), \hat{\mathbf{w}}(\operatorname{inv} v)) \end{aligned}$$

by Proposition 1.5.6. Therefore, as u and v were arbitrary, $T_{\mathbf{w}}^{BW}$ is contractive and, by the last equality, $c_{T_{\mathbf{w}}^{BW}} = c_{\hat{\mathbf{w}}}$. ■

Corollary 1.5.10 *Let $w_i \in Con_1(X, d)$ for $i = 1, 2, \ldots, N$. Then $T_{\mathbf{w}}^{BW}$ has a unique, attracting, fixed point $\bar{u}_{T_{\mathbf{w}}^{BW}} \in \mathcal{F}_{BW}(X)$. Furthermore $\hat{\mathbf{w}}(\operatorname{inv}(\bar{u}_{T_{\mathbf{w}}^{BW}})) = \operatorname{inv}(\bar{u}_{T_{\mathbf{w}}^{BW}})$.*

We are now in a position to extend this work to *grey-level maps*; that is, functions $u : X \to \mathbb{R}$. We write $\mathcal{F}(X)$ for the set of grey-level maps on X. Hence, letting $u \in \mathcal{F}(X)$, and using our previous work, we could define an operator T by

$$Tu(x) = \max_{1 \leq k \leq N}{}' u(w_k^{-1}(x)) \quad \forall x \in X.$$

This was fine in the IFS case since u only took on two values. In a sense, an IFS changes u only in physical space. We wish to allow this new operator to modify the grey-level values when the function is displaced physically. For this, a grey-level component is added [20].

Definition 1.5.11 *Let (X, d) be a metric space. Let $\mathbf{w} = \{w_k\}_{k=1}^N$ where $w_k \in Con_1(X, d)$ for $k = 1, 2, \ldots, N$. Then, let $\Phi = \{\phi_k\}_{k=1}^N$ where $\phi_k : \mathbb{R} \to \mathbb{R}$ for $k = 1, 2, \ldots, N$. The pair $(\mathbf{w}, \Phi)$ will be called an* iterated function system with grey-level maps, *or* IFSM *for short.*

Define the IFSM operator $T_{(\mathbf{w},\Phi)}^{\max} : \mathcal{F}(X) \to \mathcal{F}(X)$, on $u \in \mathcal{F}(X)$, by

$$T_{(\mathbf{w},\Phi)}^{\max} u(x) = \max_{1 \leq k \leq N}{}' \phi_k(u(w_k^{-1}(x))) \quad \forall x \in X.$$

Now, when u is displaced in space, its grey-level values are also modified. When viewing IFS as IFSM, this definition reduces to the case when $\phi_i = id_{\mathbb{R}}$ $\forall i = 1, 2, \ldots, N$.

Since the grey-level function u will be allowed to assume values between 0 and 1, we note the following: In the IFS case, if $x \in w_i(X) \cap w_j(X)$ for some $1 \leq i \neq j \leq N$, i.e. when there is *overlapping*, then $T^{\max}_{(\mathbf{w},\Phi)}u(x) = 1$, since both $\phi_i(u(w_i^{-1}(x))) = 1$ and $\phi_j(u(w_j^{-1}(x))) = 1$. This was fine since the function u assumed only values of 0 or 1. In the IFSM case, where $T^{\max}_{(\mathbf{w},\Phi)}u(x)$ can assume values between 0 and 1, the grey-level mappings ϕ_i and ϕ_j could be more general.

Suitable operators for both non-overlapping and overlapping cases have been studied [20, 21]. We will focus our attention here on the more general and probable situation where the sets $w_i(X)$ do overlap. One way to accommodate the problem of overlap is to consider taking a linear combination of the $\phi_k \circ u \circ w_k^{-1}$. We therefore define the operator $T_{(\mathbf{w},\Phi)}$ on $u \in \mathcal{F}(X)$ by:

$$T_{(\mathbf{w},\Phi)}u(x) = \sum_{k=1}^{N}{}' \phi_k(u(w_k^{-1}(x))) \quad \forall x \in X, \tag{1.5}$$

where $\sum'$ indicates that the sum runs over the indices k with $x \in w_k(X)$. We use the convention that an empty sum has a value of 0.

It should be noted that $T_{(\mathbf{w},\Phi)}$ is not an exact generalization of $T^{BW}_{\mathbf{w}}$. It has been chosen in this way since it will allow us to find a nice solution the Inverse Problem. We refer the reader to Appendix A for a discussion on the generalization of $T^{BW}_{\mathbf{w}}$ for the IFSM case. The following section will focus on $T_{(\mathbf{w},\Phi)}$ as defined in Equation (1.5). Further generalizations may be found in [22].

1.6 IFSM on $L^p(X, \mu)$

Let $(\mathbf{w}, \Phi)$ be an IFSM on a complete metric space (X, d) where $\mathbf{w} = \{w_1, w_2, \dots, w_N\}$, $w_k \in Con_1(X)$ and $\Phi = \{\phi_1, \phi_2, \dots, \phi_N\}$, $\phi_k : \mathbb{R} \to \mathbb{R}$. When it is understood that a specific IFSM is being considered, write T for $T_{(\mathbf{w},\Phi)}$, to denote the associated IFSM operator.

The theory of IFSM was developed for $L^p(X, \mu)$ by Forte and Vrscay in [20]. We present a few of their results here.

Proposition 1.6.1 *Let $(\mathbf{w}, \Phi)$ be an N-map IFSM and let T be the associated IFSM operator. Suppose:*

i) $\forall u \in L^p(X, \mu), u \circ w_k^{-1} \in L^p(X, \mu), 1 \le k \le N$ and

ii) $\phi_k \in Lip(\mathbb{R}), 1 \le k \le N$.

Then for $1 \le p \le \infty, T : L^p(X, \mu) \to L^p(X, \mu)$.

Proof Let $1 \le p \le \infty$ and $u \in L^p(X, \mu)$. Let $1 \le k \le N$. By *i)*, $u \circ w_k^{-1} \in L^p(X, \mu)$, and hence by *ii)* $\phi_k \circ u \circ w_k^{-1} \in L^p(X, \mu)$. Therefore $T_{(\mathbf{w},\Phi)} u = \sum_{k=1}^{N}{}' \phi_k \circ u \circ w_k^{-1} \in L^p(X, \mu)$. ∎

We now show contractivity of $T_{(\mathbf{w},\Phi)}$ under certain conditions. Let $M(X)$ denote the set of finite measures on $B(X)$, the Borel sets of X.

Proposition 1.6.2 *Let $(\mathbf{w}, \Phi)$ be an N-map IFSM such that $\phi_k(t) = a_k \in \mathbb{R}$ for all $t \in \mathbb{R}$ and $1 \le k \le N$. Then $\forall p \in [1, \infty)$ and $\mu \in M(X)$, the associated IFSM operator T is contractive on $L^p(X, \mu)$, with contractivity factor $c_T = 0$. Furthermore, its fixed point $\bar{u}_T$ is*

$$\bar{u}_T = \sum_{k=1}^{N} a_k \chi_{\bar{w}_k(X)}.$$

Proof Let $u, v \in L^p(X, \mu)$. Then

$$\begin{aligned} Tu &= \sum_{k=1}^{N}{}' \phi_k \circ u \circ w_k^{-1} \\ &= \sum_{k=1}^{N} a_k \chi_{\hat{w}_k(X)}. \end{aligned}$$

Therefore,

$$\begin{aligned} \|Tu - Tv\|_p &= \left[\int_X \left| \sum_{k=1}^{N}{}' \phi_k(u(w_k^{-1}(x))) - \phi_k(v(w_k^{-1}(x))) \right|^p d\mu(x) \right]^{\frac{1}{p}} \\ &\leq \sum_{k=1}^{N} \left[\int_{X_k} |\phi_k(u(w_k^{-1}(x))) - \phi_k(v(w_k^{-1}(x)))|^p \, d\mu(x) \right]^{\frac{1}{p}}, \quad X_k = \hat{w}_k(X) \\ &= \sum_{k=1}^{N} \int_{X_k} |a_k - a_k|^p d\mu(x) \\ &= 0. \quad \blacksquare \end{aligned}$$

Proposition 1.6.3 *Let $X \subset \mathbb{R}^D, D \in \mathbb{N}^+$, and let $\mu = m^{(D)}$ be the Lebesgue measure on $\mathbb{R}^D$ and d be the usual Euclidean metric. Let $(\mathbf{w}, \Phi)$ be an N-map IFSM such that, for $1 \leq k \leq N$,*

i) $w_k \in Sim_1(X, d)$ with contractivity factor c_k and

ii) $\phi_k \in Lip(\mathbb{R})$, with Lipschitz constant K_k.

Then for $p \in [1, \infty)$ and $u, v \in L^p(X, \mu)$, we have

$$\|Tu - Tv\|_p \leq C(D, p)\|u - v\|_p,$$

where $C(D, p) = \sum_{k=1}^{N} c_k^{D/p} K_k$.

Proof Let $u, v \in L^p(X, \mu)$. Then

$$\begin{aligned}
||Tu - Tv||_p &= \left[\int_X \left|\sum_{k=1}^{N}{}' \phi_k(u(w_k^{-1}(x))) - \phi_k(v(w_k^{-1}(x)))\right|^p dx\right]^{\frac{1}{p}} \\
&\leq \sum_{k=1}^{N} \left[\int_{X_k} |\phi_k(u(w_k^{-1}(x))) - \phi_k(v(w_k^{-1}(x)))|^p dx\right]^{\frac{1}{p}} \\
&= \sum_{k=1}^{N} c_k^{D/p} \left[\int_X |\phi_k(u(y)) - \phi_k(v(y))|^p dy\right]^{\frac{1}{p}} \\
&\leq \sum_{k=1}^{N} c_k^{D/p} K_k \left[\int_X |u(y) - v(y)|^p dy\right]^{\frac{1}{p}} \\
&= C(D,p)||u - v||_p. \qquad \blacksquare
\end{aligned}$$

Hence, if $C(D,p) < 1$, T is contractive on $L^p(X,\mu)$ and has a unique, attracting fixed point. It is not necessary that all IFS maps be contractive (in the base space X) for T to be contractive. The contractivity of the ϕ_k (in the grey-level range) can contribute in this aspect [20].

Example 1.6.4 Let $X = [0,1]$ and μ be the Lebesgue measure on X. Let $w_i(x) = \frac{1}{3}(x + i - 1), i = 1, 2, 3$. Let $\phi_1(t) = \frac{1}{2}t, \phi_2(t) = \frac{1}{2}, \phi_3(t) = \frac{1}{2}t + \frac{1}{2}$, for $t \in \mathbb{R}$. The fixed point of this IFSM is the *Devil's staircase*, which is continuous almost everywhere on X and differentiable on $X \backslash \mathcal{C}$. The attractor $\bar{u}$ is shown in Figure 1.2.

Given two N-map IFSM $(\mathbf{w}, \Phi_i)$, $i = 1, 2$, where $\Phi_i = \{\phi_{i_1}, \phi_{i_2}, \dots, \phi_{i_N}\}$, define the distance [3] between the grey-level components by

$$d_{\Phi}^{N}(\Phi_1, \Phi_2) = \sup_{1 \leq k \leq N} \sup_{t \in \mathbb{R}} |\phi_{1k}(t) - \phi_{2k}(t)|.$$

[3] This function will be a metric when X is compact.

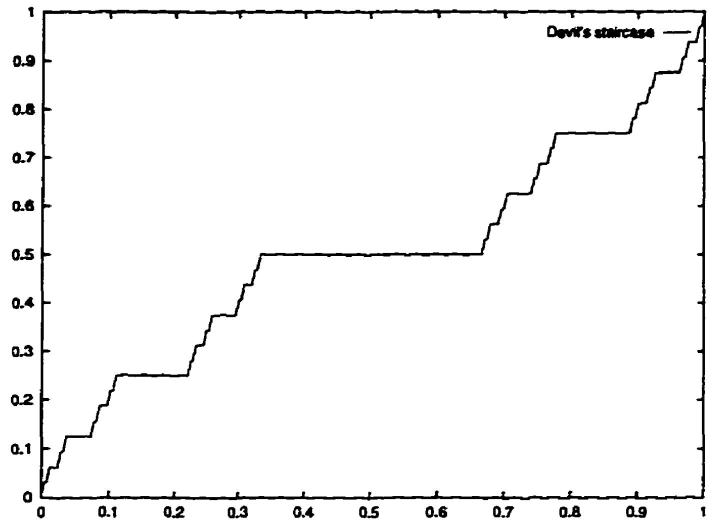


Figure 1.2: The Devil's staircase. This is also the distribution function $F(x) = \int_0^x d\mu$ of the Cantor-Lebesgue measure μ.

The following result from [20] establishes the continuity of fixed points for IFSM (c.f. Theorem 1.1.18).

Proposition 1.6.5 *Let $(\mathbf{w}, \Phi_1)$ be an N-map IFSM with fixed point $\bar{u}_1 \in L^p(X, \mu)$. Then given $\epsilon > 0$, $\exists \delta > 0$ such that for all N-map IFSM $(\mathbf{w}, \Phi_2)$ with $d_\Phi^N(\Phi_1, \Phi_2) < \delta$, then $||\bar{u}_1 - \bar{u}_2||_p < \epsilon$, where $\bar{u}_2$ is the fixed point of $(\mathbf{w}, \Phi_2)$.*

Proof Let $Y = L^p(X, \mu)$ and T_i be the IFSM operators of $(\mathbf{w}, \Phi_i)$, $i = 1, 2$ with contractivity factors c_i. Then

$$\begin{aligned}
\bar{d}(T_1, T_2) &= \sup_{u \in Y} ||T_1 u - T_2 u||_p \\
&= \sup_{u \in Y} \left(\int_X \left| \sum_{k=1}^{N}{}' \phi_{1k}(u(w_k^{-1}(x))) - \phi_{2k}(u(w_k^{-1}(x))) \right|^p dx \right)^{1/p}
\end{aligned}$$

$$\begin{aligned}
&\leq \sup_{u\in Y}\sum_{k=1}^{N}\left(\int_{X_k}\left|\phi_{1k}(u(w_k^{-1}(x))) - \phi_{2k}(u(w_k^{-1}(x)))\right|^p dx\right)^{1/p}, \quad X_k = w_k^{-1}(x)\\
&\leq \sum_{k=1}^{N}\mu(X_k)^{1/p} d_{\Phi}^{N}(\Phi_1, \Phi_2)\\
&= M.
\end{aligned}$$

The result then follows by setting $\delta = \epsilon(1-c)/M$, where $c = \min\{c_1, c_2\}$ and using Corollary 1.1.19 on page 11. ■

1.7 Inverse Problem Using IFSM

In this section we present a formal solution to the Inverse Problem for IFSM. Consider the following formulation:

Question 1.7.1 *For $v \in L^p(X,\mu)$ and $\epsilon > 0$, can we find an IFSM $(\mathbf{w}, \Phi)$ with associated operator T such that $||v - Tv||_p < \epsilon$?*

A formal solution was obtained in [20] by constructing sequences of N-map IFSM $(\mathbf{w}^N, \Phi^N)$, $N = 1, 2, 3, \ldots$ where $\mathbf{w}^N$ is chosen from a fixed set $\mathcal{W}$ of contraction maps.

Definition 1.7.2 *Let $\mathcal{W} = \{w_1, w_2, \ldots\}$ be an infinite set of contraction maps on X. Then $\mathcal{W}$ generates a μ-dense and non-overlapping, or μ-d-n, family of subsets of X if $\forall\epsilon > 0$ and $\forall B \subset X$, there exists a finite set of integers $i_k \geq 1$, $1 \leq k \leq N$ such that*

i) $A = \cup_{k=1}^{N} w_{i_k}(X) \subset B$;

ii) $\mu(B\backslash A) < \epsilon$ *and*

iii) $\mu(w_{i_k}(X) \cap w_{i_l}(X)) = 0$ *whenever* $k \neq l$.

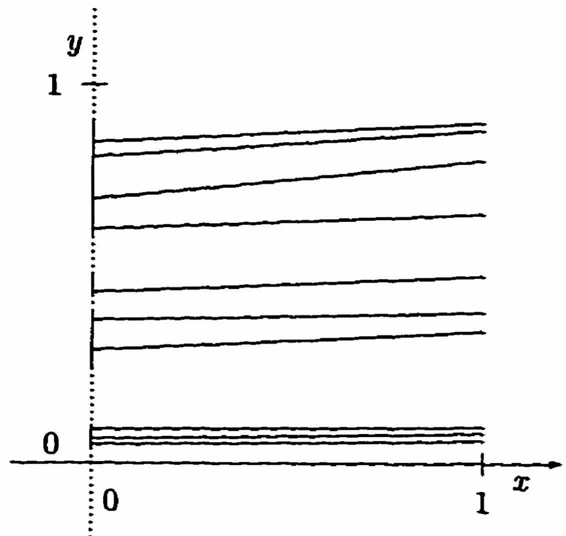


Figure 1.3: The set B is the union of the solid lines on the vertical axis. The set A is the union of the lines (projected onto the vertical axis).

Example 1.7.3 Let $X = [0,1]$ with Lebesgue measure. Let $w_{ij}(x) = 2^{-i}(x + j - 1)$, $i = 1, 2, \ldots$, $1 \le j \le 2^i$. For each $i \ge 1$, the set of maps $\{w_{ij}, 1 \le j \le 2^i\}$ is a set of 2^i contractions of $[0,1]$ which tile $[0,1]$. Then $\mathcal{W} = \{w_{ij}\}$ is μ-d-n. Figure 1.3 illustrates the idea.

Now, suppose $\mathcal{W} = \{w_i\}$, with $w_i \in Con_1(X,d)$, generates a μ-d-n family of subsets of X. Let

$$\mathbf{w}^N = \{w_1, w_2, \ldots, w_N\} \quad N = 1, 2, \ldots,$$

denote the N-map truncations of $\mathcal{W}$. Assume that for each $k \in \mathbb{N}^+$, $\phi_k \in Lip(\mathbb{R})$ is the associated grey-level map of w_k and let

$$\mathbf{\Phi}^N = \{\phi_1, \phi_2, \ldots, \phi_N\}.$$

Let $T^N : L^p(X,\mu) \to L^p(X,\mu)$ be the associated IFSM operator of $(\mathbf{w}^N, \Phi^N)$. Then the following result holds:

Theorem 1.7.4 *Let $v \in L^p(X,\mu)$, $1 \leq p < \infty$ and $\mathcal{W}$ as above. Then*

$$\lim_{N\to\infty} \inf ||v - T^N v||_p = 0.$$

Proof A proof can be found in [20]. ■

Using this result and Example 1.7.3, we are now in a position to develop an algorithm for the construction of IFSM approximations of target functions $v \in L^p(X,\mu)$. Then, given an N-map IFSM $(\mathbf{w}, \Phi)$ on (X,d) with associated operator T, we have the squared L^2 distance

$$\begin{aligned}\Delta^2 &= ||v - Tv||_2^2 \\ &= \int_X \left(\sum_{k=1}^{N}{}' \phi_k(v(w_k^{-1}(x))) - v(x) \right)^2 d\mu(x). \end{aligned} \tag{1.6}$$

With the formal solution in mind, we assume the IFS maps w_k are fixed and search for grey-level maps ϕ_k which minimize Δ^2 for the given target v. This is the key idea for IFSM [20].

For computational simplicity, assume the maps w_k and ϕ_k are affine. The pair $(\mathbf{w}, \Phi)$ will be called an *affine IFSM*. Assuming that $\phi_k(t) = \alpha_k t + \beta_k$ $\forall t \in \mathbb{R}, k = 1, 2, \ldots, N$, then

$$Tu(x) = \sum_{k=1}^{N}{}' \left[\alpha_k u(w_k^{-1}(x)) + \beta_k \chi_{w_k(X)}(x) \right]. \tag{1.7}$$

If $X \subset \mathbb{R}^D$, then by Proposition 1.6.3 on page 37, $\forall u, v \in L^p(X,\mu)$,

$$||Tu - Tv||_p \leq C(D,p)||u-v||_p$$

with $C(D,p) = \sum_{k=1}^{N} c_k^{D/p}\alpha_k$. Hence, if $C(D,p) < 1$, T is contractive on $L^p(X,\mu)$ and has a unique fixed point $\bar{u}_T$.

Example 1.7.5 If $\beta_k = 0$ for $1 \leq k \leq N$, then $\bar{u}_T \equiv 0$.

Example 1.7.6 If $X = [0,1]$, $w_k(x) = a_k x + b_k, 1 \leq k \leq N$, and T is contractive with fixed point $\bar{u}_T$, then by Equation (1.7),

$$\begin{aligned}\bar{u}_T(x) &= \sum_{k=1}^{N}{}' \alpha_k \bar{u}_T\left(\frac{x-b_k}{a_k}\right) + \beta_k \chi_{w_k(X)}(x)\\ &= \sum_{k=1}^{N}{}' \alpha_k \psi_k(x) + \beta_k \phi_k(x).\end{aligned}$$

Therefore, $\bar{u}_T$ is a linear combination of piecewise constant functions ϕ_k, and functions ψ_k, which are dilations and translations of $\bar{u}_T$. [4] This idea is reminiscent of the wavelets relations and will be discussed in Chapter 2.

By the following theorem, it is sufficient in practical situations to study the subclass of affine IFSM [20].

Theorem 1.7.7 *Let $X = \mathbb{R}^D$ and let $\mu \in M(X)$. Given $p \geq 1$, let $L^p_A(X,\mu) \subset L^p(X,\mu)$ be the set of fixed points of contractive N-map affine IFSM on X. Then $L^p_A(X,\mu)$ is dense in $L^p(X,\mu)$.*

Proof For simplicity, we prove the result for $D = 1$. Let S be the set of step functions on X. Then, given $\phi \in S$, get N, with $1 \leq N < \infty$, a set of numbers $\xi_k \in \mathbb{R}$ and intervals

[4] If $f : \mathbb{R} \to \mathbb{R}$, then given $a \in \mathbb{R}$, $f(a\cdot)$ is called a *dilation* of f and $f(\cdot - a)$ is called a *translation* of f.

$J_k = [a_k, b_k] \subset [0,1]$, $k = 1, 2, \ldots, N$ such that

$$\phi = \sum_{k=1}^{N} \xi_k \chi_{J_k}.$$

Then, ϕ is the attractor of the N-map affine IFSM $(\mathbf{w}, \Phi)$ with

$$\begin{aligned} w_k(x) &= (b_k - a_k)x + a_k \\ \phi_k(t) &= \xi_k, \quad 1 \leq k \leq N. \end{aligned}$$

Thus, $S \subset L^p_A(X, \mu)$. But, since S is dense in $L^p(X, \mu)$, the result follows. The argument in higher dimensions follows in a similar manner, by replacing the intervals J_k by appropriate rectangles in $\mathbb{R}^D$. ■

Now, suppose $(\mathbf{w}, \Phi)$ is an N-map affine IFSM with

i) $w_k \in Con_1(X)$ with contractivity factors $c_k > 0$ for $1 \leq k \leq N$;

ii) $\cup_{k=1}^{N} w_k(X) = X$, and

iii) $\phi_k : \mathbb{R} \to \mathbb{R}$, where $\phi_k(t) = \alpha_k t + \beta_k$, $t \in \mathbb{R}$, $1 \leq k \leq N$.

Then, going back to Equation (1.6),

$$\begin{aligned} \Delta^2 &= < v - Tv, v - Tv > \\ &= \sum_{k=1}^{N}\sum_{l=1}^{N}(< \psi_k, \psi_l > \alpha_k\alpha_l + 2 < \psi_k, \chi_k > \alpha_k\beta_l + < \chi_k, \chi_l > \beta_k\beta_l) \\ &\quad - 2\sum_{k=1}^{N}(< v, \psi_k > \alpha_k + < v, \chi_k > \beta_k) + < v, v >, \end{aligned}$$

where $\psi_k = v \circ w_k^{-1}$ and $\chi_k = \chi_{w_k(X)}$. Then Δ^2 can be written as a *quadratic form* in the parameters α_k and β_k as

$$\Delta^2 = \mathbf{x}^T A \mathbf{x} + \mathbf{b}^T \mathbf{x} + c,$$

where $\mathbf{x}^T = (\alpha_1, \ldots, \alpha_N, \beta_1, \ldots, \beta_N) \in \mathbb{R}^{2N}$. Minimizing Δ^2 is a *quadratic programming* (QP) problem in the α_k and β_k. A detailed discussion is given in [20, 49].

We end this section with a few examples of approximations which demonstrate the application of IFSM. The second example reveals a problem with this method.

Example 1.7.8 Let $u(x) = \sin(x)$ and $X = [0,1]$. The approximations of u are given in Figure 1.4. The maps w_k map X to evenly divided subintervals. For example, in the case of 2 maps, $w_1(x) = x/2$ and $w_2(x) = x/2 + 1/2$. The following table gives the L^2 distance between u and the approximations.

Number of maps	Distance	File size (bytes)	Computation time (sec.)
u	0.0	30878	n.a.
2	0.0199362	42	1.17
4	0.0191687	82	1.17
16	0.0188445	320	1.17

In this case, the results are quite nice. This happens since the parts of the function on the subintervals are similar to the entire function.

Example 1.7.9 Let $u(x) = \sin(\pi x)$ and $X = [0,1]$. Consider the approximations in Figures 1.5. The following table gives the L^2 distance between u and the approximations.

Number of maps	Distance	File size (bytes)	Computation time (sec)
u	0.0	30768	n.a.
2	0.307759	39	1.13
4	0.158104	82	1.13
16	0.0400456	339	1.17

Here, it is difficult to get a good approximation since the best fits are given by piecewise constant functions.

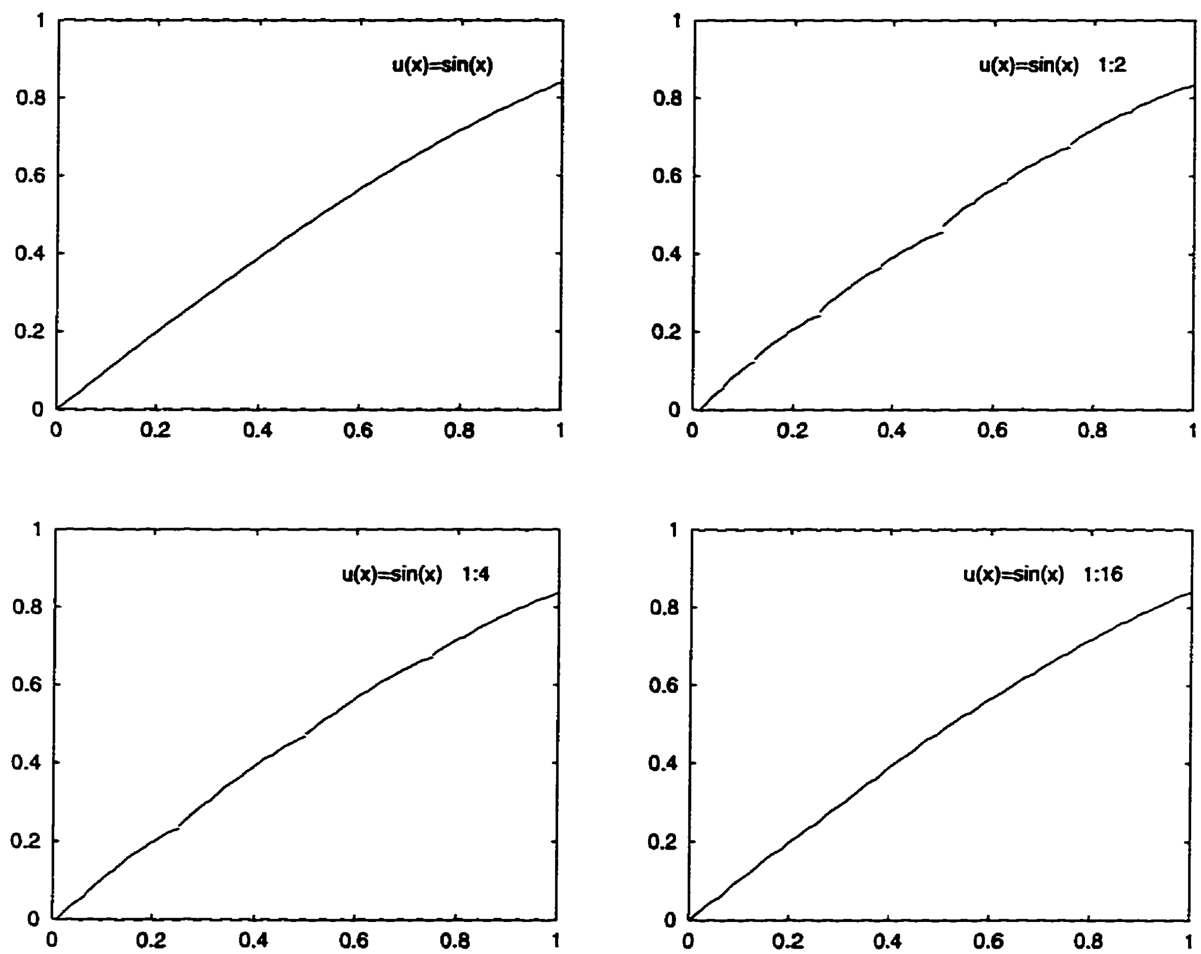


Figure 1.4: IFSM approximations of $u(x) = \sin(x)$.

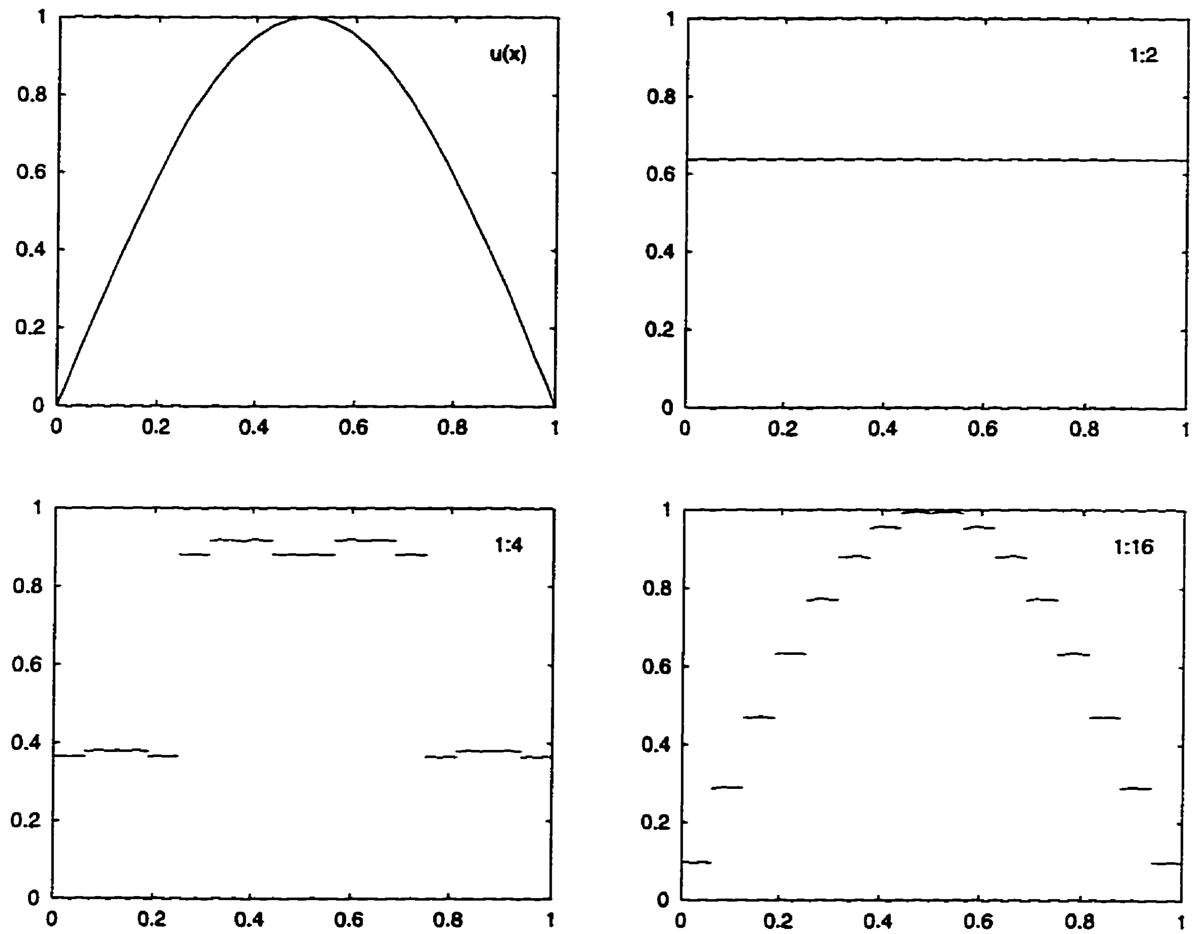


Figure 1.5: IFSM approximation of $u(x) = \sin(\pi x)$ with 2, 4 and 16 range blocks.

The problem described above arises in most situations, when the smaller portions of an image are not similar to the entire image. However, this situation can be remedied by considering local IFSM.

1.8 LIFSM

A method, which in general yields better approximations than IFSM, is the method of local IFSM (LIFSM) [20].

Definition 1.8.1 *Let $X \subset \mathbb{R}^D$ and $\mu = m^{(D)}$. Let $J_k \subset X$, $k = 1, 2, \ldots, N$, such that*

i) $\cup_{k=1}^{N} J_k = X$ (covering condition) and

ii) $\mu(J_j \cap J_k) = 0$ when $j \neq k$ (μ-non-overlapping condition).

Suppose also that $\forall J_k$, $\exists I_{j(k)} \subset X$ with an associated map $w_{j(k),k} \in Con(X, d)$ with contractivity factor $c_{j(k),k}$ such that

$$w_{j(k),k}(I_{j(k)}) = J_k.$$

The set J_k is called the range block *of the* domain block *$I_{j(k)}$.* [5] *For each $w_{j(k),k}$, let $\phi_k : \mathbb{R} \to \mathbb{R}$ be an associated grey-level map. Then define*

$$\mathbf{w}_{loc} = \{w_{i(1),1}, \ldots, w_{i(N),N}\} \quad \text{and} \quad \Phi = \{\phi_1, \ldots, \phi_N\}.$$

The pair $(\mathbf{w}_{loc}, \Phi)$ is called an N-map local IFSM, *or* LIFSM. *The associated operator $T^{loc}_{(\mathbf{w},\Phi)} : \mathcal{F}(X) \to \mathcal{F}(X)$ is defined by*

$$T^{loc}_{(\mathbf{w},\Phi)} u(x) = \begin{cases} \phi_k(u(w^{-1}_{i(k),k})) & x \in J_k \backslash \cup_{l \neq k}^{N} J_l(X), \\ 0 & x \text{ otherwise.} \end{cases}$$

[5] *Often, one calls the domain blocks "parent" blocks and the range blocks "child" blocks.*

A result similar to Proposition 1.6.3 can then be obtained:

Proposition 1.8.2 *Let $X \subset \mathbb{R}^D$ and $\mu = m^{(D)}$. Let $(\mathbf{w}_{loc}, \Phi)$ be a LIFSM as above with $\phi_k \in Lip(\mathbb{R})$ for $1 \leq k \leq N$ and let T^{loc} be the associated LIFSM operator. Then, for $u, v \in L^p(X, \mu)$,*

$$||T^{loc}u - T^{loc}||_p \leq C_{loc}(D,p)||u-v||_p,$$

where $C_{loc}(D,p) = \left(\sum_{k=1}^{N} c_{j(k),k} K_k^p\right)^{1/p}$. Thus, if $C_{loc}(D,p) < 1$, T^{loc} is contractive on $L^p(X,\mu)$ and has a unique fixed point.

Proof The proof is similar to that of Proposition 1.6.3 on page 37 and is omitted. ∎

Now, suppose $X = [0,1]^D$, $\mu = m^{(D)}$ and $v \in L^2(X, \mu)$. Then, given an N-map LIFSM as above, the squared collage distance is given by

$$\begin{aligned}
\Delta^2 &= ||T^{loc}v - v||_2^2 \\
&= \sum_{k=1}^{N} \int_{J_k} [\phi_k(v(w_{j(k),k}^{-1}(x))) - v(x)]^2 dx \\
&= \sum_{k=1}^{N} \Delta_{j(k),k}^2.
\end{aligned}$$

It is therefore sufficient to minimize the $\Delta_{j(k),k}^2$ individually for each range block J_k. In the case where the maps ϕ_k are affine, this becomes a QP problem [20].

To apply this idea to the Inverse Problem, consider the following:

i) $X \subset \mathbb{R}^D$, $\mu = m^{(D)}$, d usual Euclidean metric;

ii) $w_k \in Sim_1(X, d)$, with $X = \cup_{k=1}^{N} X_k$, where $X_k = w_k(X)$ (*covering condition*);

iii) $\mu(X_i \cap X_j) = 0$ when $i \neq j$ (μ-*non-overlapping condition*) and

iv) $\phi_k : \mathbb{R} \to \mathbb{R}$ are affine with $\phi_k(t) = \alpha_k t + \beta_k$, $t \in \mathbb{R}$.

Then

$$\begin{aligned}\Delta_k^2 &\equiv \int_{X_k} [\alpha_k v(w_k^{-1}(x)) + \beta_k - v(x)]^2 d\mu \\ &= c_k^D \int_X [\alpha_k v(x) + \beta_k - v(w_k^{-1}(x))]^2 d\mu.\end{aligned}$$

As before, with the formal solution of the Inverse Problem in mind, assume that the w_k are fixed, and hence for each k, Δ_k can be viewed as a quadratic form in the parameters α_k and β_k:

$$\begin{aligned}c_k^{-D}\Delta_k^2 = ||v||_2^2\alpha_k^2 + 2\alpha_k\beta_k||v||_1 + \beta_k^2 - 2 < v, v \circ w_k > \alpha_k \\ - 2||v \circ w_k||_1\beta_k + ||v \circ w_k||_2^2.\end{aligned}$$

The problem can be viewed as a least squares minimization of Δ_k with respect to α_k and β_k. Set

$$\frac{\partial \Delta_k^2}{\partial \alpha_k} = \frac{\partial \Delta_k^2}{\partial \beta_k} = 0, \quad k = 1, 2, \ldots, N.$$

Then

$$||v||_2^2\alpha_k + ||v||_1\beta_k =< v \circ w_k, v >,$$

$$||v||_1\alpha_k + \beta_k = ||v \circ w_k||_1,$$

for $k = 1, 2, \ldots, N$.

Then, if $D_v \equiv ||v||_1^2 - ||v||_2^2 \neq 0$, the solutions are given by

$$\alpha_k = D_v^{-1}(< v \circ w_k, v > - ||v \circ w_k||_1 \, ||v||_1)$$
$$\beta_k = D_v^{-1}(||v||_2^2 \, ||v \circ w_k||_1 - ||u||_1 < v \circ w_k, v >),$$

for $1 \leq k \leq N$.

When considering images, the condition that $\phi_k : \mathbb{R}^+ \to \mathbb{R}^+$ would be needed. This forces $\alpha_k, \beta_k \geq 0$. It is not guaranteed that the α_k and β_k given by the above method will be nonnegative. However, if we consider an image to be a function defined on a compact subset A of $\mathbb{R}$, the condition on the α_k and β_k could be relaxed, with $\phi_k(v(x))$ still being nonnegative on A.

Hence, given $v \in L^2(X, \mu)$, fix N_J range blocks J_k, $1 \leq k \leq N_J$, and N_I domain blocks I_j, $1 \leq j \leq N_I$. For each range block J_k, minimize the distance $\Delta^2_{j,k}$, for each domain block I_j, $1 \leq j \leq N_I$. Then, let $I_{j(k)}$ be the domain block for which $\Delta_{j(k),k}$ is minimized over the domains. The values of $I_{j(k)}$, and the associated parameters α_k and β_k, are then stored, for $1 \leq k \leq N_J$. These values are called an IFSM approximation of v.

Example 1.8.3 Consider the function $u(x) = \sin(\pi x)$ for $x \in X = [0, 1]$. Some approximations to u using the LIFSM method are shown in Figure 1.6. The following table gives the L^2 distance between u and the approximations.

Domains	Ranges	Distance	File size (bytes)	Computation time (sec)
u	n.a.	0.0	30768	n.a.
2	4	0.0266135	82	1.16
2	8	0.0144324	162	1.18
2	16	0.00762873	322	1.18
4	16	0.00131272	324	1.24

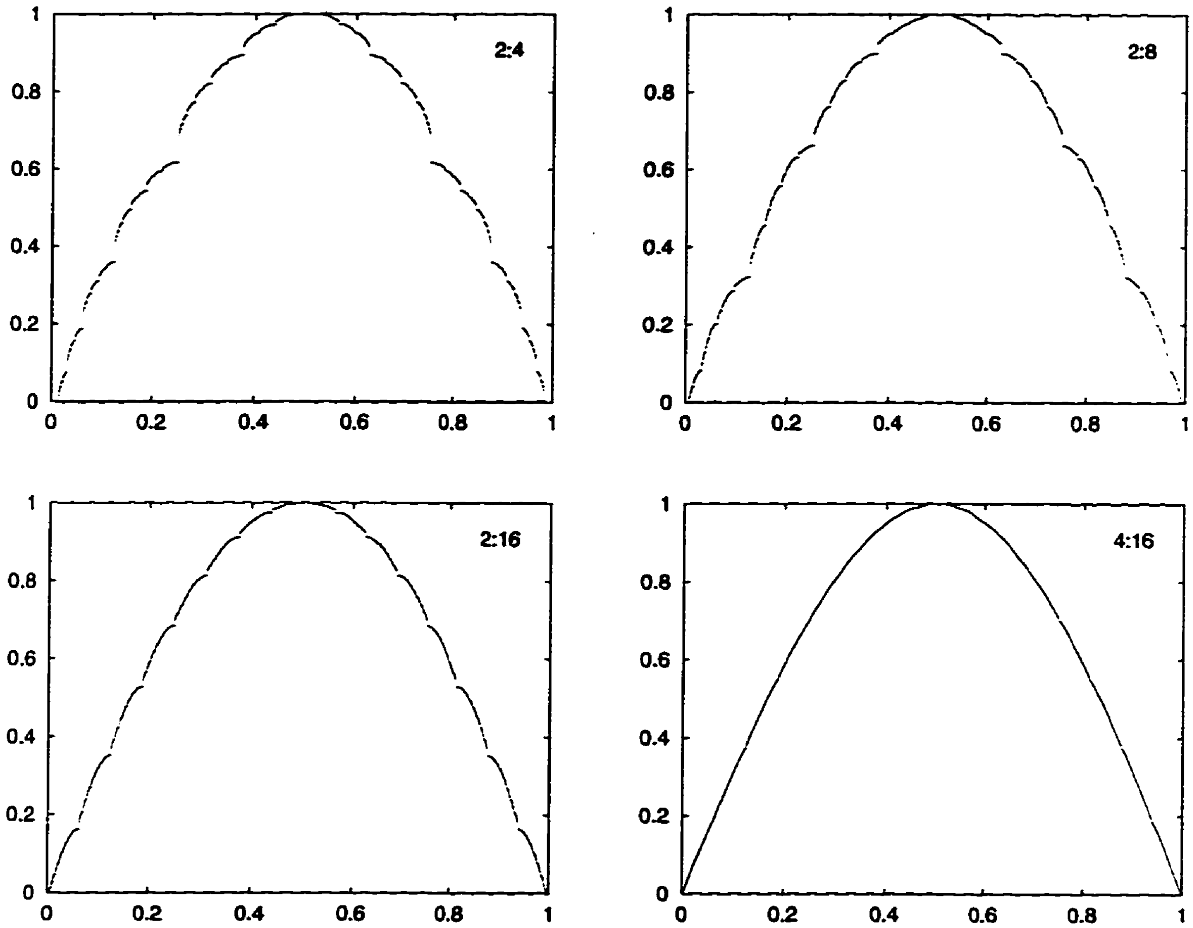


Figure 1.6: LIFSM approximation of $\sin(\pi x)$ with block ratio (D:R) from left to right, top to bottom, 2:4, 2:8, 2:16 and 4:16.

Comparing these results with the IFSM case shown in Example 1.7.9, the strength of LIFSM is revealed.

Chapter 2

Wavelets

This chapter will deal with function approximation in a different way, using wavelets (c.f. Example 1.7.6). Our general goal is the representation or approximation of arbitrary target functions by functions which we know. An often used method is through bases.

2.1 Hilbert Space Background

Notation 2.1.1 Let $(H, <\cdot,\cdot>)$ denote a Hilbert space over $\mathbb{R}$ with inner product $<\cdot,\cdot>$. The *norm* of $f \in H$ is $||f|| = \sqrt{<f,f>}$. The distance between $f, g \in H$ is $d(f,g) = ||f-g||$.

For simplicity, the Hilbert space of focus in this thesis will be $L^2(\mathbb{R})$, the square integrable functions on $\mathbb{R}$, with the usual inner product.

Definition 2.1.2 *Two elements $f, g \in H$ are* orthogonal *if $<f,g>= 0$. Write $f \perp g$ to denote this fact. An element is called* normalized *if it has norm 1. A set $\{h_\alpha\} \subset H$ is* orthogonal *if all pairs of distinct elements are orthogonal. It will be called* orthonormal *if it is orthogonal and all its elements are normalized.*

Example 2.1.3 Let $H = L^2(0,\pi)$. For $n \in \mathbb{N}$, let $h_n(x) = \sin(nx)$ for $x \in (0,\pi)$. Then $\{h_n\}$ is an orthogonal set.

Example 2.1.4 Let $H = L^2[0, \infty)$. For $n \in \mathbb{N}$, let $h_n = \chi_{[n,n+1)}$. Then $\{h_n\}$ is an orthonormal set.

Definition 2.1.5 *If M is a subspace of H, define $M^\perp$, the* orthogonal complement *of M, as*

$$M^\perp = \{f \in H : <f, m>= 0 \quad \forall m \in M\}.$$

Proposition 2.1.6 *If M is a subspace of H, then $M^\perp$ is a subspace of H and if M is closed, $H = M \oplus M^\perp$, the* direct sum *of M and $M^\perp$.*

Recall the definition of a projection:

Definition 2.1.7 *Let $\{h_\alpha\}$ be an orthonormal set in H and let $M = \overline{\langle h_\alpha \rangle}$, where $\langle h_\alpha \rangle$ denotes the linear span of the set $\{h_\alpha\}$ and $\overline{A}$ denotes the closure of the set A. Then, the function P_M: $H \to M$ defined by*

$$P_M f = \sum_\alpha < f, h_\alpha > h_\alpha$$

is called the orthogonal projection *of H onto M.*

The following proposition lists a few basic facts about P_M:

Proposition 2.1.8 *The function P_M is well-defined, linear, continuous and idempotent with respect to composition. In addition, $P_M f = f \iff f \in M$ and $P_M f = 0 \iff f \in M^\perp$.*

Recall the following important theorems from Hilbert space theory. Standard proofs can be found in [4, 8, 44].

Theorem 2.1.9 (Pythagorean Theorem) *If $f, g \in H$ and $f \perp g$, then*

$$||f + g||^2 = ||f||^2 + ||g||^2.$$

Theorem 2.1.10 (Bessel's inequality) *Suppose (h_n) is an orthonormal sequence in H and $f \in H$. Then $||f||^2 \geq \sum_{n=1}^{\infty} | < f, h_n > |^2$.*

Corollary 2.1.11 *If $\{h_\alpha\}$ is orthonormal and $f \in H$, then $\sum_\alpha < f, h_\alpha > h_\alpha$ converges. In addition, if $f = \sum_\alpha c_\alpha h_\alpha$, then $c_\alpha = < f, h_\alpha >$.*

Definition 2.1.12 *A* basis *of H is a maximal orthonormal set in H. That is, $\{h_\alpha\} \subset H$ is a basis if no element $f \in H$, $f \neq 0$, is orthogonal to each of the h_α. A basis is also called a* complete orthonormal set.

Theorem 2.1.13 (Parseval's equality) *Suppose $\{h_n\}$ is a complete orthonormal set and $f, g \in H$. Then*

$$< f, g > = \sum_{n=0}^{\infty} c_n \bar{d}_n,$$

where $c_n = < f, h_n >$ and $d_n = < g, h_n >$. Therefore,

$$||f||^2 = \sum_{n=0}^{\infty} |c_n|^2.$$

Example 2.1.14 The trigonometric system $\{\frac{1}{2\pi} e^{inx}\}_{n \in \mathbb{Z}}$ is a complete orthonormal set on $L^2(-\pi, \pi)$.

Theorem 2.1.15 *Every Hilbert space has a basis $\{h_\alpha\}$ and $H = \overline{\langle h_\alpha \rangle}$.*

Definition 2.1.16 *If $\{h_\alpha\}$ is a set such that*

$$H = \overline{\langle h_\alpha \rangle},$$

but is not orthonormal then $\{h_\alpha\}$ is called complete.

The next example will be important in motivating the definition of a large class of bases for $L^2(\mathbb{R})$. This example is the *Haar basis* [51].

Example 2.1.17 (The Haar Wavelets) Let $H = L^2(\mathbb{R})$ and let $\phi = \chi_{[0,1)}$. We wish to use ϕ to construct a basis of $L^2(\mathbb{R})$. For each nonzero $n \in \mathbb{Z}$, $\phi(t-n) \perp \phi(t)$. This is trivial since the supports of the two functions are disjoint. The set $\{\phi(t-n)\}$ is not a basis of $L^2(\mathbb{R})$ since the set

$$V_0 = \overline{\langle \phi(t-n) : n \in \mathbb{Z} \rangle}$$

consists of piecewise constant functions with jumps only on $\mathbb{Z}$.

Consider the *dilated* and *translated* versions of $\phi(t)$

$$\phi(2^m t - n), \quad m, n \in \mathbb{Z}.$$

Given $m \in \mathbb{Z}$, the set $\{2^{m/2}\phi(2^m t - n) : n \in \mathbb{Z}\}$ is orthonormal, since the supports of any two distinct functions in it are disjoint. For $m \in \mathbb{Z}$, let V_m

$$V_m = \overline{\langle 2^{m/2}\phi(2^m t - n) : n \in \mathbb{Z} \rangle}.$$

Then, the space $V = \cup_{m \in \mathbb{Z}} V_m$ consists of piecewise constant functions with jumps at dyadic rationals. As these functions are dense in $L^2(\mathbb{R})$, we have that $\overline{V} = L^2(\mathbb{R})$.

Define $\phi_{m,n}(t) = 2^{m/2}\phi(2^m t - n)$ for $m, n \in \mathbb{Z}$. Then $\{\phi_{m,n}\}$ is complete in $L^2(\mathbb{R})$. However, $<\varphi_{00}, \varphi_{10}>= \frac{1}{\sqrt{2}}$, therefore, $\{\phi_{m,n}\}$ is not orthonormal. To solve this problem, let $\psi(t) = \phi(2t) - \phi(2t-1)$. Then $\{\psi(t-n)\}$ is orthonormal and $\psi(2t-k) \perp \psi(t-n)$ $\forall n, k \in \mathbb{Z}$. Therefore, we obtain the following theorem:

Theorem 2.1.18 *Let $\psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n)$ for $m, n \in \mathbb{Z}$. Then $\{\psi_{m,n}\}$ is a complete orthonormal system in $L^2(\mathbb{R})$.*

Proof The proof for general functions of this type will be given later (see Proposition 2.2.11). ∎

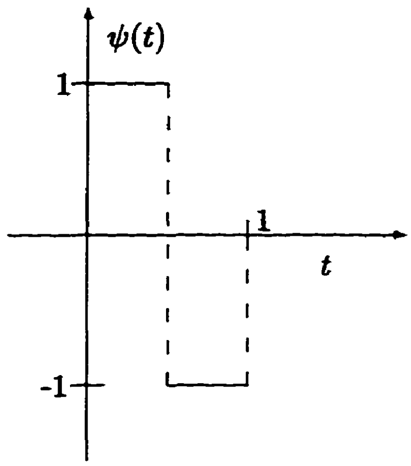


Figure 2.1: The mother wavelet $\psi(t)$ of the Haar system.

The set $\{\psi_{m,n}\}$ is called the set of *Haar wavelets*. The function ψ is called the *mother wavelet* and is shown in Figure 2.1. We see that $V_m = \overline{\langle \psi_{k,n} : k, n \in \mathbb{Z}, k \leq m-1 \rangle}$. The standard approximation of a function $f \in L^2(\mathbb{R})$ is

$$f_m = \sum_{k=-\infty}^{m-1} \sum_{n=-\infty}^{\infty} < f, \psi_{k,n} > \psi_{k,n}.$$

Therefore, $f_m \in V_m$. By Parseval's equality, we have

$$\begin{aligned} < f_m, \phi_{m,n} > &= \sum_{k=-\infty}^{m-1} \sum_{j=-\infty}^{\infty} < f, \psi_{k,j} >< \psi_{k,j}, \phi_{m,n} > \\ &= < f, \phi_{m,n} > . \end{aligned}$$

Hence, $f_m = P_{V_m} f$, that is

$$f_m = \sum_{n=-\infty}^{\infty} < f, \phi_{m,n} > \phi_{m,n}.$$

We therefore obtain a strong result on the convergence of the approximations.

Proposition 2.1.19 *Let f be continuous on $\mathbb{R}$ with compact support; then $f_m \to f$ uniformly.*

Proof Since f has compact support, it is uniformly continuous. Therefore, $\forall \epsilon > 0$ we can find an m such that

$$|f(x) - f(y)| < \epsilon \text{ when } |x - y| \leq 2^{-m}.$$

Now, for each $n \in \mathbb{Z}$ and $x \in [n2^{-m}, (n+1)2^{-m})$ we have

$$f_m(x) = 2^{m/2} \int_{2^{-m}n}^{2^{-m}(n+1)} f(t) 2^{m/2} \phi(2^m x - n) dt$$

by definition of $\phi_{m,n}$. Then, by the Mean Value Theorem,

$$\begin{aligned} f_m(x) &= 2^{m/2}(f(c)2^{-m})2^{m/2}\phi(2^m x - n) \\ &= f(c), \end{aligned}$$

for some $c \in [n2^{-m}, (n+1)2^{-m})$. Since $|x - c| \leq 2^{-m}$, $|f_m(x) - f(x)| < \epsilon$, as required. ∎

The Haar system is an example of a wavelet basis of $L^2(\mathbb{R})$.

2.2 Multiresolution Analysis

A *wavelet* has been defined by Meyer [38] as an integrable function ψ whose integral is zero such that

$$\int_0^\infty \frac{|\hat{\psi}(\xi t)|^2}{t} dt = 1$$

for all $\xi \neq 0$ where $\hat{\psi}$ is the Fourier transform of ψ. Recall the following definitions:

Definition 2.2.1 *The* (infinite) Fourier transform *of the function* $f \in L^1(\mathbb{R})$ *is*

$$\hat{f}(\omega) = \int_{\mathbb{R}} f(t)e^{-i\omega t}dt, \quad \omega \in \mathbb{R}.$$

If the transform is in $L^1(\mathbb{R})$, *then the inverse is given by*

$$f(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(\omega)e^{i\omega t}d\omega, \quad t \in \mathbb{R}.$$

For $f \in L^2(\mathbb{R})$, Parseval's equality yields the following two relations:

$$||f||^2 = \frac{1}{2\pi}||\hat{f}||^2 \quad \text{and} \quad < f, g >= \frac{1}{2\pi} < \hat{f}, \hat{g} > .$$

Example 2.2.2 The Fourier transform of the Haar mother wavelet given in Section 2.1 is

$$\hat{\psi}(\omega) = \frac{i}{w}\left[2e^{-i\omega/2} - 1 - e^{-i\omega}\right].$$

A complex parametric plot is given in Figure 2.2.

One way to obtain wavelet bases like the Haar basis is through multiresolution analysis. Begin by considering regular functions [51].

Definition 2.2.3 *Let* S *be the space of all* $C^\infty(\mathbb{R})$ *functions* θ *such that*

$$|\theta^{(k)}(t)| \leq C_{p,k}(1+|t|)^{-p} \quad p, k \in \mathbb{N}, t \in \mathbb{R}, \tag{2.1}$$

where $\theta^{(k)}$ *denotes the k-th derivative of* θ, *with convergence given by the semi-norms*

$$\gamma_{p,k} = \sup_t (1+|t|)^p|\theta^{(k)}(t)|.$$

The space S *is called the space of* rapidly decreasing *or* regular $C^\infty(\mathbb{R})$ *functions on* $\mathbb{R}$. [1]

[1] *This is sometimes called the* Schwartz class.

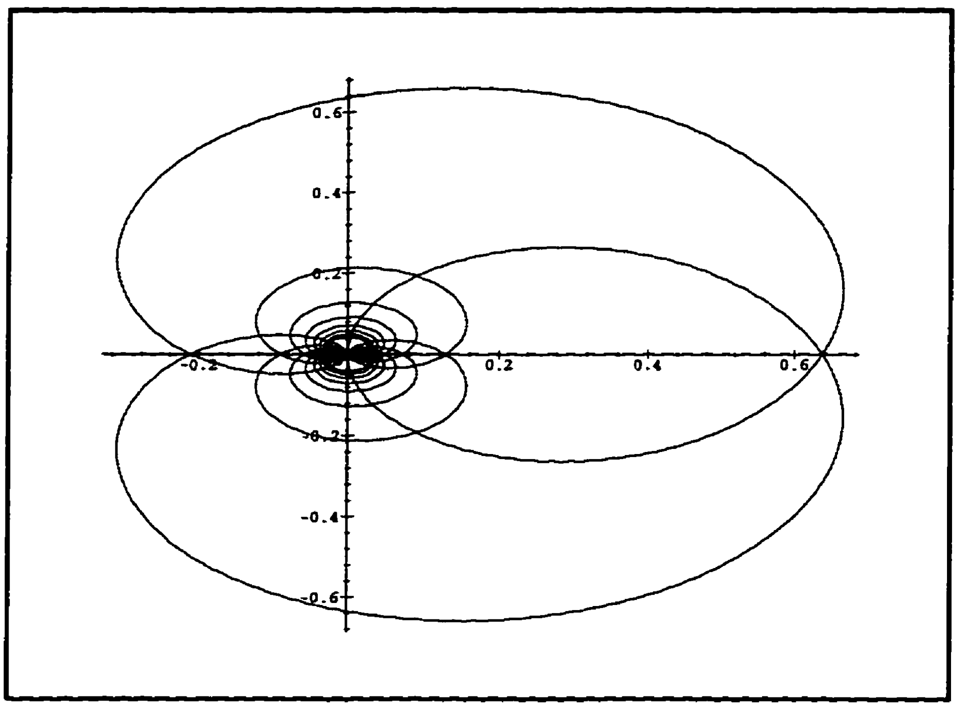


Figure 2.2: The Fourier transform of ψ, the mother wavelet of the Haar system.

In S, $\theta_\nu \to \theta$ whenever

$$(1+|t|)^p D^k(\theta_\nu(t) - \theta(t)) \to 0$$

uniformly in t $\forall p, k \in \mathbb{N}$ as $\nu \to \infty$. Here, D is the derivative operator.

Example 2.2.4 *Hermite functions* on $\mathbb{R}$, defined by

$$h_0(x) = \frac{e^{-x^2/2}}{\sqrt{\pi}} \quad x \in \mathbb{R}$$

and $(x - D)h_n(x) = \sqrt{2n+2}h_{n+1}(x), n \in \mathbb{N}, x \in \mathbb{R}$, are in S. Therefore, since these functions form an orthonormal basis in $L^2(\mathbb{R})$, $\bar{S} = L^2(\mathbb{R})$ [4].

Example 2.2.5 All $C^\infty(\mathbb{R})$ functions of compact support are regular.

The Haar scaling function does not satisfy Equation (2.1), but satisfies a less restrictive condition.

Definition 2.2.6 *For $r \in \mathbb{N}$, let S_r be the space of all $\theta \in C^r(\mathbb{R})$ satisfying Equation (2.1) for all $k \leq r$ and for all $p \in \mathbb{N}$, with the topology restricted by $k \leq r$. Functions in S_r are called* r-regular.

Example 2.2.7 The function ϕ of the Haar system is in S_0.

We now define the concept of a multiresolution analysis, or MRA of $L^2(\mathbb{R})$. This will allow us to construct more wavelet bases. More general definitions can be found in [7, 32, 35, 38].

Definition 2.2.8 *Let $\phi \in S_r$. The function ϕ defines a* multiresolution analysis, *or MRA, of $L^2(\mathbb{R})$, and is called a* scaling function, *if there is a nested sequence of closed subspaces $\{V_m\}_{m\in\mathbb{Z}}$ satisfying the following conditions:*

i) $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ is an orthonormal basis of V_0;

ii) $\ldots \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots \subset L^2(\mathbb{R})$;

iii) $f \in V_m \iff f(2\cdot) \in V_{m+1} \quad \forall m \in \mathbb{Z}$;

iv) and $\overline{\bigcup_{m\in\mathbb{Z}} V_m} = L^2(\mathbb{R})$.

The map $T : V_0 \to V_1$ defined by $f \mapsto \sqrt{2}f(2\cdot)$ is an isometric isomorphism from V_0 to V_1, hence $\{\sqrt{2}\phi(2t-n)\}$ is a orthonormal basis for V_1. Therefore, since $\phi \in V_1$,

$$\phi(t) = \sum_{k\in\mathbb{Z}} h_k \sqrt{2}\phi(2t-k) \quad \forall t \in \mathbb{R}, \tag{2.2}$$

where $\{h_k\} \in \ell^2(\mathbb{Z})$. Equation (2.2) is called the *dilation equation* and the coefficients $\{h_k\}$ are called the *dilation coefficients* of ϕ.

Usually, the condition that

$$\bigcap_{m\in\mathbb{Z}} V_m = \{0\}$$

is included in the definition. It was shown in [31] that this property follows from the definition given above.

By what was shown earlier, the Haar system satisfies this definition for $r = 0$. Another example is the *Shannon system*.

Example 2.2.9 (The Shannon Wavelets) Let ϕ be the Fourier transform of a function resembling the scaling function of the Haar system. That is

$$\hat{\phi}(\omega) = \begin{cases} 1 & if -\pi \le \omega \le \pi \\ 0 & otherwise. \end{cases}$$

Then

$$\begin{aligned}
\phi(t) &= \frac{1}{2\pi}\int_{\mathbb{R}} \hat{\phi}(\omega)e^{i\omega t}d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t}d\omega \\
&= \frac{\sin \pi t}{\pi t}.
\end{aligned}$$

Then for $0 \neq n \in \mathbb{Z}$,

$$\begin{aligned}
\int_{\mathbb{R}} \phi(t)\overline{\phi(t-n)}dt &= \frac{1}{2\pi}\int_{\mathbb{R}} \hat{\phi}(\omega)\overline{\hat{\phi}(\omega)}e^{i\omega n}d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega n}d\omega \\
&= \frac{\sin \pi n}{\pi n} \\
&= 0.
\end{aligned}$$

Notation 2.2.10 Let f be a function from a set A to $\mathbb{R}$. Then the *support* of f, $\operatorname{supp}(f)$, is defined as

$$\operatorname{supp}(f) = \overline{\{a \in A : |f(a)| > 0\}}.$$

Let $f \in L^2(\mathbb{R})$ with $\operatorname{supp} \hat{f} \subset [-\pi, \pi]$. Then

$$\hat{f}(\omega) = \sum_{n\in\mathbb{Z}} c_n e^{i\omega n} \quad |\omega| \leq \pi,$$

where $c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{f}(\omega)e^{-i\omega n}d\omega = f(-n)$ by the Fourier integral theorem. By the same theorem,

$$\begin{aligned}
f(t) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{f}(\omega)e^{i\omega t}d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{n\in\mathbb{Z}} f(-n)e^{i\omega n}e^{i\omega t}d\omega \\
&= \sum_{n\in\mathbb{Z}} f(-n)\frac{\sin\pi(t+n)}{\pi(t+n)}.
\end{aligned}$$

The last equality is often referred to as the Shannon Sampling Theorem. Hence $V_0 \equiv \overline{\langle\phi(t-n) : n \in \mathbb{Z}\rangle}$ is the set of all such functions. Therefore V_0 is a closed subspace of $L^2(\mathbb{R})$ and *i)* of Definition 2.2.8 is satisfied.

Now consider $g(x) = f(2x)$, where $2x = t$. Let V_1 be the space of all functions g such that $f \in V_0$. These functions have Fourier transforms vanishing outside of $[-2\pi, 2\pi]$. In this manner, construct an increasing sequence of spaces $V_m = \{\sqrt{2}f(2\cdot) : f \in V_{m-1}\}$. Functions $f \in V_m$ have Fourier transforms which vanish outside $[-2m\pi, 2m\pi]$. Similarly, by letting $x = t/2$, we can construct a decreasing sequence of spaces V_m for $m < 0$. These spaces will contain functions with Fourier transforms vanishing outside $[-\pi/(2m), \pi/(2m)]$. Hence properties *ii)* and *iii)* of Definition 2.2.8 are satisfied. We obtain condition *iv)* since the supports of the Fourier transforms expand to $\mathbb{R}$ as $m \to \infty$.

In a similar fashion to the Haar system, we can construct the function ψ in V_1 which is orthogonal to $\phi(\cdot - n)$ for all $n \in \mathbb{Z}$ by letting $\psi(t) = 2\phi(2t) - \phi(t)$. Figure 2.3 shows the graphs of ϕ and ψ. These form the *Shannon system.* [2]

Three general approaches have been used to construct MRA [51]:

---

[2] A smoothed version of the Shannon wavelets are the *Meyer wavelets* and are given in [38, p.66].

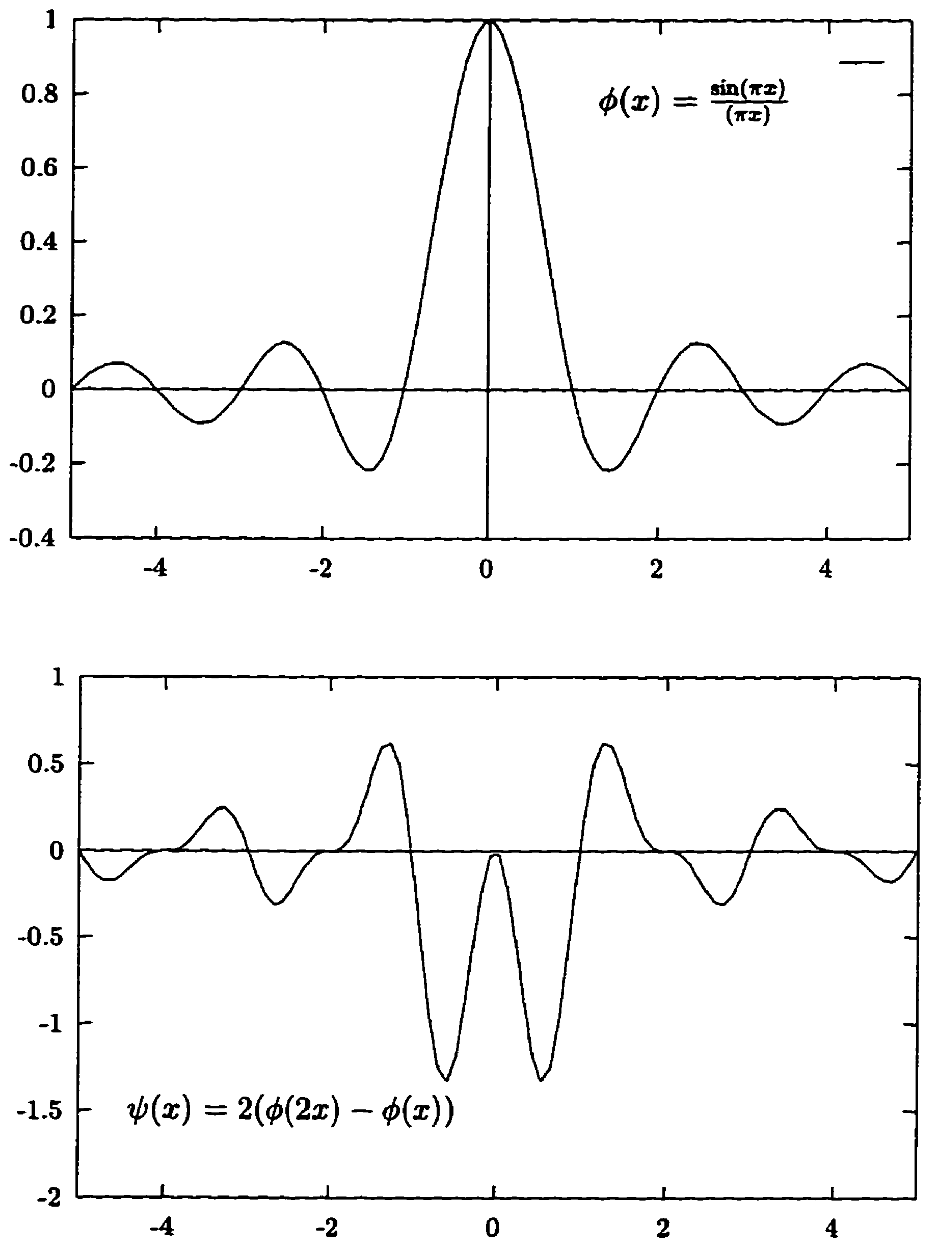


Figure 2.3: The Shannon scaling function ϕ and a mother wavelet ψ.

i) Begin with an existing MRA $\{V_m\}_{m\in\mathbb{Z}}$ then try to find an orthonormal basis. For example, let V_0 be defined by a Riesz basis of translates of a fixed function $\{\theta(t-n)\}$. Then use the orthogonalization procedure of Lemarié and Meyer to find an orthogonal system $\{\phi(t-n)\}$ [29].

ii) Choose dilation coefficients (h_k) such that all the requirements of a MRA are satisfied [51, pp.32-33].

iii) Choose the Fourier transform of $\phi(t)$ such that it has compact support and the transformed versions of *i)-iv)* and of the dilation equation are satisfied [51, pp.30-32].

Once a scaling function $\phi(t)$ has been found, we wish to use it to construct a *mother wavelet*, $\psi(t)$. We want $\psi(t)$ to satisfy the property that $\{\psi(t-n)\}$ is an orthonormal basis of $W_0 = V_0^{\perp_{V_1}}$, where $A^{\perp_B}$ denotes the orthogonal complement of A in B. Then $V_1 = V_0 \oplus W_0$, hence we want the functions

$$\psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n)$$

to form an orthonormal basis of $W_m = V_m^{\perp_{V_{m+1}}}$. The following property is therefore satisfied:

Proposition 2.2.11 $\bigoplus_{m\in\mathbb{Z}} W_m = L^2(\mathbb{R})$.

Proof We have for $m \in \mathbb{Z}$,

$$\begin{aligned} V_{m+1} &= V_m \oplus W_m \\ &= V_0 \oplus W_0 \oplus W_1 \oplus \ldots \oplus W_m. \end{aligned}$$

Since $\underset{m\in\mathbf{Z}}{\cup} V_m$ is dense in $L^2(\mathbb{R})$, then

$$V_0 \oplus (\overset{\infty}{\underset{m=0}{\oplus}} W_m) = L^2(\mathbb{R}).$$

Similarly

$$\begin{aligned} V_0 &= V_{-1} \oplus W_{-1} \\ &= V_{-k} \oplus W_{-k} \oplus \ldots \oplus W_{-1}. \end{aligned}$$

Taking the limit as $k \to \infty$, an recalling that $\underset{m\in\mathbf{Z}}{\cap} V_m = \{0\}$, $V_{-k} \to \{0\}$. Therefore $\bigoplus_{m\in\mathbf{Z}} W_m = L^2(\mathbb{R})$. ∎

Hence the following corollary is obtained:

Corollary 2.2.12 *The set $\{\psi_{m,n}\}_{m,n\in\mathbf{Z}}$ is an orthonormal basis of $L^2(\mathbb{R})$.*

As in the case of $\phi(t)$, there are two methods for constructing a mother wavelet $\psi(t)$. The first is similar to the construction of the Franklin scaling function [51, p.34-35]. The second is to note that

$$\psi(t) = \sqrt{2}\sum h_{1-k}(-1)^k\phi(2t-k) \tag{2.3}$$

satisfies the necessary orthogonality conditions [15, p.135].

2.3 Convergence of Wavelet Expansions

Before going further in our study of wavelets, we must see whether they are indeed worth our attention. In view of our goal of approximating functions, we would like to have nice results for the convergence of wavelet expansions. Recall the large amount of work necessary to get nice convergence results for trigonometric series. Even then, one finds examples such as the following:

i) There is a continuous function $f : [-\pi, \pi] \to \mathbb{R}$ such that its *Fourier series* $S(f) = \sum_{n=-\infty}^{\infty} c_n e^{inx}$ is pointwise divergent at a dense set of points [55, p.300].

ii) Even with summability methods, convergence for smooth functions is not much more rapid than for other continuous functions [55, p.122].

Proposition 2.1.19 on page 58 suggests that wavelet expansions might have much nicer convergence properties than those of trigonometric series. We will demonstrate some of these in this section. Much of classical approximation theory is based on delta sequences [51, p.116]. For convenience, we make the following definition:

Definition 2.3.1 *A* tempered distribution *is an element of S^*, the dual of S.*

Example 2.3.2 Let f be a locally integrable function of polynomial growth. Then we can view $f \in S^*$ by defining, for $\theta \in S$,

$$(f, \theta) = \int_{\mathbb{R}} f(t)\theta(t)dt.$$

Another example is the delta function.

Example 2.3.3 The *delta function*, or more correctly *delta distribution*, δ_α, $\alpha \in \mathbb{R}$ is the element of S^* satisfying

$$(\delta_\alpha, f) = f(\alpha) \quad \forall f \in S.$$

Denote δ_0 by δ.

Definition 2.3.4 *A* delta sequence *is a sequence $(\delta_m(\cdot, y)) \subset S^*$ such that $\delta_m(\cdot, y) \to \delta_y$ in S^*.*

Example 2.3.5 The Dirichlet kernel $D_m(\cdot, y)$:

$$D_m(x,y) = \frac{\sin\left[\left(m+\frac{1}{2}\right)(x-y)\right]}{2\pi \sin\left(\frac{x-y}{2}\right)} \chi_{[-\pi,\pi]}(x-y).$$

We consider a subclass of such sequences, namely quasi-positive delta sequences, and derive some interesting convergence results.

Definition 2.3.6 *A quasi-positive delta sequence, or* QPDS, *is a sequence* $(\delta_m(\cdot, y)) \subset L^1(\mathbb{R}), y \in \mathbb{R}$ *such that*

i) $\exists C > 0$ *such that*

$$\int_{\mathbb{R}} |\delta_m(x,y)| dx \leq C \quad \forall y \in \mathbb{R}, m \in \mathbb{N};$$

ii) $\exists c > 0$ *such that*

$$\int_{y-c}^{y+c} \delta_m(x,y) dx \to 1$$

uniformly on compact subsets of $\mathbb{R}$ *as* $m \to \infty$.

iii) $\forall \gamma > 0$,

$$\sup_{|x-y| \geq \gamma} |\delta_m(x,y)| \to 0$$

as $m \to \infty$.

Example 2.3.7 Let $F_m(\cdot, y)$ be the *Fejér kernel*. That is

$$F_m(x,y) = \frac{\sin^2\left[\left(\frac{m+1}{2}\right)(x-y)\right]}{2(m+1)\pi \sin^2\left(\frac{x-y}{2}\right)} \chi_{[-\pi,\pi]}(x-y).$$

In fact, the Fejér kernel satisfies the stronger conditions of a positive delta sequence [51, p.132]. It will be shown later that the Dirichlet kernel is not a QPDS.

The following result indicates that QPDS may be useful for approximations:

Proposition 2.3.8 *Let $(\delta_m(\cdot, y))$ be a QPDS and $f \in L^1(\mathbb{R})$ be continuous on (a, b). Then*

$$f_m(y) = \int_{\mathbb{R}} \delta_m(x, y) f(x) dx \to f(y)$$

as $m \to \infty$ uniformly on compact subsets of (a, b).

Proof A proof may be found in [51, p.118-119]. ∎

To see the relevance of these sequences for wavelets, consider for $m \in \mathbb{Z}$ the approximation f_m to $f \in L^2(\mathbb{R})$:

$$\begin{aligned} f_m(y) &= \sum_n < f, \phi_{m,n} > \phi_{m,n}(y) \\ &= \sum_n \left[\int_{\mathbb{R}} f(x) \overline{2^{m/2} \phi(2^m x - n)} dx \right] 2^{m/2} \phi(2^m y - n) \\ &= \int_{\mathbb{R}} \left[2^m \sum_n \overline{\phi(2^m x - n)} \phi(2^m y - n) \right] f(x) dx. \end{aligned} \tag{2.4}$$

Let $q_m(x, y) = 2^m \sum_n \overline{\phi(2^m x - n)} \phi(2^m y - n)$. If $f \in V_m$ then $f_m = f$. This implies that each of the spaces V_m is a *reproducing kernel Hilbert space*. For a more detailed discussion see [2, 52]. The *reproducing kernel*, or *RK*, of V_0 is

$$q(x, t) = \sum_n \overline{\phi(x - n)} \phi(t - n).$$

By Equation (2.4), the RK of V_m is

$$q_m(x, t) = 2^m q(2^m x, 2^m t).$$

Similarly W_m has RK

$$r_m(x,t) = 2^m \sum_n \overline{\psi(2^m x - n)}\psi(2^m t - n).$$

By the regularity of ϕ, the series defining $q(x,t)$, and its derivatives with respect to t of order $\leq r$, converge uniformly for $x \in \mathbb{R}$.

Example 2.3.9 For the Haar system, $\phi = \chi_{[0,1)}$, so

$$\begin{aligned} q(x,t) &= \sum_n \chi_{[0,1)}(x-n)\chi_{[0,1)}(t-n) \\ &= \begin{cases} 0 & \text{if } x,t \notin [n,n+1) \text{ for some } n \in \mathbb{Z} \\ 1 & \text{otherwise, that is } x - [t] \in [0,1), \end{cases} \end{aligned}$$

where $[t]$ is the greatest integer $\leq t$. Therefore $q(x,t) = \phi(x - [t])$ for $x,t \in \mathbb{R}$.

It will be shown that $(q_m(\cdot,y))$ is a QPDS. In this aim, consider the following result [51]:

Theorem 2.3.10 *Let $\phi \in S_r$ generate a MRA $\{V_m\}$ and $q_m(x,t)$ be the reproducing kernel of V_m. Let δ_{mn} denote the Kronecker delta. If $\hat{\phi}(0) \geq 0$, then*

i) $\int_{\mathbb{R}} \phi(y)dy = 1$;

ii) $\hat{\phi}(2\pi k) = \delta_{0k}, k \in \mathbb{Z}$;

iii) $\sum_n \phi(x-n) = 1, x \in \mathbb{R}$;

iv) $\int_{\mathbb{R}} q(x,y)dy = 1, x \in \mathbb{R}$.

Proof Proof of *i)*: Let $m \in \mathbb{N}$ and $\hat{f}(\omega) = \chi_{[0,1]}(\omega)$. Then

$$\begin{aligned}
f_m(x) &= \int_{\mathbb{R}} q_m(x,t) f(t) dt \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \hat{q}_m(x,\omega) \hat{f}(\omega) d\omega \\
&= \frac{1}{2\pi} \int_0^1 2^m \sum_n \phi(2^m x - n) \left[\int_{\mathbb{R}} \phi(2^m t - n) e^{-i\omega t} dt \right] d\omega \\
&= \frac{1}{2\pi} \int_0^1 \sum_n \phi(2^m x - n) \left[\int_{\mathbb{R}} \phi(y) e^{-i\omega(2^{-m} y + 2^{-m} n)} dy \right] d\omega \\
&= \frac{1}{2\pi} \int_0^1 \sum_n \phi(2^m x - n) \left[\int_{\mathbb{R}} \phi(y) e^{-i\omega 2^{-m} y} e^{-i\omega 2^{-m} n} dy \right] d\omega \\
&= \frac{1}{2\pi} \int_0^1 \sum_n \phi(2^m x - n) \hat{\phi}(\omega 2^{-m}) e^{-i\omega 2^{-m} n} d\omega.
\end{aligned}$$

Let

$$\begin{aligned}
a_{k,l} &= < f_m, \phi_{k,l} > \\
&= \int_{\mathbb{R}} \left(\frac{1}{2\pi} \int_0^1 \sum_n \hat{\phi}(\omega 2^{-m}) e^{-i\omega 2^{-m} n} \phi(2^m x - n) d\omega \right) \phi_{k,l}(x) dx \\
&= \frac{1}{2\pi} \int_0^1 2^{-m/2} \hat{\phi}(\omega 2^{-m}) e^{-in2^{-m}\omega} d\omega \sum_n \int_{\mathbb{R}} 2^{m/2} \phi(2^m x - n) 2^{k/2} \phi(2^k x - l) dx \\
&= \left[\frac{1}{2\pi} \int_0^1 2^{-m/2} \hat{\phi}(\omega 2^{-m}) e^{-in2^{-m}\omega} d\omega \right] \delta_{ln} \delta_{km}.
\end{aligned}$$

Hence, by Parseval's equality,

$$
\begin{aligned}
||f_m||^2 &= \sum_{k,l} |a_{k,l}|^2 \\
&= \sum_n |a_{m,n}|^2 \\
&= \sum_n \left(\frac{1}{2\pi}\right)^2 \left(\int_0^1 \hat{\phi}(\omega 2^{-m}) e^{-in2^{-m}\omega} d\omega\right)^2 2^{-m} \\
&= \frac{1}{2\pi} \sum_n \left| \int_{-2^m\pi}^{2^m\pi} \hat{\phi}(\omega 2^{-m}) \hat{f}(\omega) \frac{e^{-in2^{-m}\omega}}{\sqrt{2^{m+1}\pi}} d\omega \right|^2
\end{aligned}
$$

However, since $\{e^{-in2^{-m}\omega}/\sqrt{2^{m+1}\pi}\}$ is orthonormal on $[-2^m\pi, 2^m\pi]$, we use Parseval's equality to get

$$
\begin{aligned}
||f_m||^2 &= \frac{1}{2\pi} \int_{-2^m\pi}^{2^m\pi} \left| \hat{\phi}(\omega 2^{-m}) \hat{f}(\omega) \right|^2 d\omega \\
&= \frac{1}{2\pi} \int_0^1 \left| \hat{\phi}(\omega 2^{-m}) \right|^2 d\omega.
\end{aligned}
$$

Then, since $\hat{\phi}$ is bounded and continuous,

$$
\begin{aligned}
\frac{1}{2\pi} \int_0^1 \left| \hat{\phi}(0) \right|^2 d\omega &= \lim_{m\to\infty} ||f_m||^2 \\
&= ||f||^2 \\
&= \frac{1}{2\pi} \int_0^1 |\hat{f}(\omega)|^2 d\omega \\
&= \frac{1}{2\pi}.
\end{aligned}
$$

Hence, as $\hat{\phi}(0) \geq 0$, $\hat{\phi}(0) = 1$, and

$$
\int_{\mathbb{R}} \phi(y) dy = \hat{\phi}(0) = 1.
$$

Proof of *ii)*: By the orthonormality of $\{\phi(t-n)\}$, we have

$$
\begin{aligned}
\delta_{0n} &= \int_{\mathbf{R}} \phi(t-n)\phi(t)dt \\
&= \frac{1}{2\pi}\int_{\mathbf{R}} \hat{\phi}(\omega)e^{-i\omega n}\hat{\phi}(\omega)d\omega \\
&= \frac{1}{2\pi}\sum_k \int_0^{2\pi} |\hat{\phi}(\omega+2\pi k)|^2 e^{-i\omega n}d\omega \\
&= \frac{1}{2\pi}\int_0^{2\pi} \sum_k |\hat{\phi}(\omega+2\pi k)|^2 e^{-i\omega n}d\omega. \qquad (2.5)
\end{aligned}
$$

Let $|\hat{\phi}_2^*(\omega)|^2 = \sum_k |\hat{\phi}(\omega+2\pi k)|^2$. Then Equation (2.5) gives the Fourier coefficients of $|\hat{\phi}_2^*(\omega)|^2$. Hence

$$|\hat{\phi}_2^*(\omega)|^2 = \sum_n c_n e^{i\omega n} = 1,$$

and therefore

$$\sum_k |\hat{\phi}(\omega+2\pi k)|^2 = 1.$$

Since $\hat{\phi}(0)=1$, then $\hat{\phi}(2\pi k)=0\ \forall k\neq 0$. Therefore *ii)* is proved.

Proof of *iii)*: Consider the Fourier series of

$$\sum_n \phi(x-n) = \sum_k d_k e^{i2\pi kx}.$$

Then

$$\begin{aligned} d_k &= \int_0^1 \sum_n \phi(x-n)e^{-2\pi ikx}dx \\ &= \int_{\mathbb{R}} \phi(x)e^{-2\pi ikx}dx \\ &= \hat{\phi}(2\pi k) \\ &= \delta_{0k}, \end{aligned}$$

whence $\sum_n \phi(x-n) = 1$.

Proof of *iv)*: This follows immediately from *i)* and *iii)* since

$$\begin{aligned} \int_{\mathbb{R}} q(x,y)dy &= \int_{\mathbb{R}} \sum_n \phi(x-n)\phi(y-n)dy \\ &= \sum_n \phi(x-n) \int_{\mathbb{R}} \phi(y-n)dy \\ &= \sum_n \phi(x-n) \int_{\mathbb{R}} \phi(t)dt \\ &= \sum_n \phi(x-n) \\ &= 1. \qquad \blacksquare \end{aligned}$$

Lemma 2.3.11 *Let $\phi \in S_r$, then $\forall m \in \mathbb{N}, \exists C_m \in \mathbb{R}$ such that*

$$|\partial_x^\alpha \partial_y^\beta q(x,y)| \le C_m(1+|x-y|)^{-m} \quad \forall 0 \le \alpha, \beta \le r.$$

Proof Let $\alpha, \beta, m \in \mathbb{N}$ with $0 \leq \alpha, \beta \leq r$. Then

$$\begin{aligned}
|\partial_x^\alpha \partial_y^\beta q(x,y)| &\leq \sum_n |\partial_x^\alpha \phi(x-n)||\partial_y^\beta \phi(y-n)| \\
&\leq \sum_n \frac{C_{m+2,\alpha}}{(1+|x-n|)^{m+2}} \frac{C_{m+2,\beta}}{(1+|y-n|)^{m+2}} \\
&\leq \frac{C_{m+2,\alpha} C_{m+2,\beta}}{(1+|x-y|)^m} \sum_n \frac{1}{(1+|x-n|)^2} \frac{1}{(1+|y-n|)^2},
\end{aligned}$$

where ∂_x^α denotes the α-th partial derivative with respect to x. The last sum is uniformly bounded hence we can get $C_m > 0$ such that

$$|\partial_x^\alpha \partial_y^\beta q(x,y)| \leq C_m (1+|x-y|)^{-m}. \quad \blacksquare$$

It now follows that $(q_m(\cdot, y))$ is a QPDS:

Proposition 2.3.12 *Suppose $\phi \in S_r$ generates a MRA and let q_m be the RK of V_m. Then $(q_m(\cdot, y))$ is a QPDS.*

Proof Use the previous lemma and Theorem 2.3.10.

i) We have

$$\begin{aligned}
\int_{\mathbb{R}} |q_m(x,y)| dx &= 2^m \int_{\mathbb{R}} |q(2^m x, 2^m y) dx \\
&= \int_{\mathbb{R}} |q(x, 2^m y)| dx \\
&\leq C_2 \int_{\mathbb{R}} (1+|x-2^m y|)^2 dx \\
&= C_2 \int_{\mathbb{R}} (1+|x|)^2 dx \\
&= C.
\end{aligned}$$

ii) Let $c > 0$, then for $y \in \mathbb{R}$,

$$\begin{aligned}
\int_{y-c}^{y+c} q_m(x,y)dx &= \int_{2^m(y-c)}^{2^m(y+c)} q(x, 2^m y)dx \\
&= \int_{t-2^m c}^{t+2^m c} q(x,t)dx \\
&= 1 - \int_{t+2^m c}^{\infty} q(x,t)dx - \int_{-\infty}^{t-2^m c} q(x,t)dx.
\end{aligned}$$

Now

$$\begin{aligned}
\left| \int_{t+2^m c}^{\infty} q(x,t)dx \right| &\le C_2 \int_{t+2^m c}^{\infty} \frac{1}{1+(t-x)^2} dx \\
&= C_2 \int_{2^m c}^{\infty} \frac{1}{1+x^2} dx \to 0
\end{aligned}$$

as $m \to \infty$. A similar argument shows that $\int_{-\infty}^{t-2^m c} q(x,t)dx \to 0$ as $m \to \infty$.

iii) Let $\gamma > 0$, then for $x, y \in \mathbb{R}$ with $|x-y| \ge \gamma$,

$$\begin{aligned}
|q_m(x,y)| &= 2^m |q(2^m x, 2^m y)| \\
&\le \frac{2^m C_2}{(1+|2^m x - 2^m y|)^2} \\
&= \frac{C_2}{2^{2m}(2^{-m}+|x-y|)^2} \\
&\le \frac{C_2}{2^{2m}(2^{-m}+\gamma)^2} \to 0
\end{aligned}$$

as $m \to \infty$. ∎

An immediate corollary follows:

Corollary 2.3.13 *Let $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ be continuous on (a,b) and let $f_m = P_{V_m} f$. Then*

$$f_m \to f,$$

as $m \to \infty$, uniformly on compact subsets of (a, b).

Proof It was shown on page 70 that

$$f_m(y) = \int_{\mathbb{R}} q_m(x, y) f(y) dy.$$

The result now follows immediately from Propositions 2.3.8 and 2.3.12. ∎

It is of interest to compare this result with the following from Fourier analysis:

Theorem 2.3.14 *Let f satisfy a uniform Lipschitz condition of order $\alpha > 0$ in (a, b). Then the Fourier series $S_m \to f$ uniformly in any subinterval $[c, d] \subset (a, b)$.*

Proof See [51, p.53]. ∎

Recall that the partial sums of the Fourier series of f are given by

$$S_m(x) = \int_{\mathbb{R}} D_m(x, y) f(y) dy,$$

where $D_m(x, y)$ is the Dirichlet kernel given in Example 2.3.5 and that f satisfies a *Lipschitz condition of order* α, $\alpha > 0$ at x if there exists $C > 0$ such that

$$|f(y) - f(x)| \leq C|y - x|^{\alpha}$$

in some neighbourhood of x. The function f satisfies a *uniform Lipschitz condition* if the condition holds with the same C for all x. [3]

The Lipschitz condition cannot be relaxed since there are many continuous functions with Fourier series converging everywhere, but not uniformly, and also some with divergent Fourier series [55, p.298]. This leads to the investigation of summability methods like

[3]The name Lipschitz is often replaced by Hölder.

Cesàro summability and Abel summability. These methods yield kernels which are QPDS. and even positive delta sequences, which is a stronger condition [12, 51]. For the moment then, it still seems that Fourier analysis yields the same type of convergence results as does wavelet analysis. Our next goal will be to study the rate of convergence of the expansions f_m. It is here that we will see a major advantage of wavelets.

2.4 Rate of Convergence

We will need to introduce the Zak transform [25] and the concept of Sobolev spaces in order to study the rate of convergence of the wavelet approximations.

Definition 2.4.1 *Let $\phi \in S_r$. The* Zak transform *$Z\phi$ of ϕ is defined by*

$$Z\phi(t,\omega) = \sum_n e^{-i\omega n}\phi(t-n),$$

for $t, \omega \in \mathbb{R}$.

Given $\phi \in S_r$, it follows that $Z\phi(t,\cdot) \in C^\infty(\mathbb{R})$ $\forall t \in \mathbb{R}$. By part *iii)* of Theorem 2.3.10, $Z\phi(t,0) = 1$, thus

$$e^{i\omega t} Z\phi(t,\omega) = 1 + O(|\omega|).$$

In some cases, $\hat{\phi}(\omega) = 1$ in a neighbourhood of $\omega = 0$, and $\hat{\phi}(\omega) = 1 + O(|\omega|^\lambda)$ for λ arbitrarily large. The same holds for $Z\phi$. Thus for some $\lambda \geq 1$, the following definition is satisfied:

Definition 2.4.2 *Let $\phi \in S_r$ be a scaling function. Then ϕ* satisfies property Z_λ *if*

i) $\hat{\phi}(\omega) = 1 + O(|\omega|^\lambda)$ *as* $\omega \to 0$;

ii) $Z\phi(t,\omega) = e^{-i\omega t}(1 + O(|\omega|^{\lambda})$ *uniformly as* $\omega \to 0$.

In fact, it can be shown that if $\phi \in S_r$ then ϕ satisfies Z_λ if $r \geq \lambda - 1$ and $\hat{\phi}^{(k)}(0) = 0$, $k = 1, 2, \ldots, \lambda - 1$ [51].

The rate of convergence of the expansions f_m to f will be studied using Sobolev norms.

Definition 2.4.3 *Let* $\alpha \in \mathbb{R}$. *The* Sobolev space H^α *consists of all functions* $f \in S^*$ *such that*

$$\int_{\mathbb{R}} |\hat{f}(\omega)|^2(\omega^2 + 1)^\alpha d\omega < \infty.$$

Example 2.4.4 For $\alpha = 0$, $H^0 = L^2(\mathbb{R})$.

Example 2.4.5 Let $\alpha \in \mathbb{N}^+$, then H^α consists of functions in $L^2(\mathbb{R})$ which are $(\alpha - 1)$ times differentiable and whose α-th derivative is in $L^2(\mathbb{R})$ [1].

The inner product of $f, g \in H^\alpha$ is defined by

$$< f, g >_\alpha = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(\omega)\overline{\hat{g}(\omega)}(\omega^2 + 1)^\alpha d\omega.$$

The space H^α is complete with respect to this inner product, and is therefore a Hilbert Space. We write $||\cdot||_\alpha$ to denote the Sobolev norm on H^α. The dual of H^α is $H^{-\alpha}$. For a more detailed discussion of Sobolev Spaces, see [44, 45].

We now come to our first convergence result [51].

Theorem 2.4.6 *Let* $\phi \in S_r$ *be a scaling function satisfying property* Z_λ *for some* $\lambda > 0$. *If* $q_m(x,t)$ *is the reproducing kernel of* V_m, *then*

$$||q_m(\cdot, y) - \delta(\cdot - y)||_{-\alpha} = O(2^{-m\lambda})$$

uniformly for $y \in \mathbb{R}$, *when* $\alpha > \lambda + \frac{1}{2}$.

Proof The proof uses the following two lemmas [51]:

Lemma 2.4.7 *Let $\psi \in S_r$, with $\{\psi_{m,n}(x) = 2^{m/2}\psi(2^m x - n)\}$ an orthonormal system in $L^2(\mathbb{R})$. Then the k-th* moment [4] *of ψ,*

$$\int_{\mathbb{R}} x^k \psi(x)dx = 0, \quad 0 \leq k \leq r.$$

Proof Proceed by induction on k. For $k = 0$, let N be a dyadic rational such that $\psi(N) \neq 0$. This is possible since ψ is continuous and since dyadics are dense in $\mathbb{R}$. Choose $m > 1$ sufficiently large such that $2^m N \in \mathbb{Z}$. Then, by the orthogonality of the $\psi_{m,n}$,

$$\begin{aligned} 0 &= 2^m \int_{\mathbb{R}} \psi(x)\psi(2^m x - 2^m N)dx \\ &= \int_{\mathbb{R}} \psi(2^{-m}t + N)\psi(t)dt. \end{aligned} \tag{2.6}$$

Since $\psi \in S_r$, $\psi(2^{-m}t + N)$ is uniformly bounded, hence the integrand is dominated by a multiple of $|\psi(t)|$. Therefore, by the Lebesgue Dominated Convergence Theorem,

$$\begin{aligned} 0 &= \lim_{m \to \infty} \int_{\mathbb{R}} \psi(2^{-m}t + N)\psi(t)dt \\ &= \int_{\mathbb{R}} \lim_{m \to \infty} \psi(2^{-m}t + N)\psi(t)dt \\ &= \psi(N) \int_{\mathbb{R}} \psi(t)dt. \end{aligned}$$

Thus the case $k = 0$ is proved.

Assume the theorem holds for $k < n \leq r$. Choose N such that $\psi^{(n)}(N) \neq 0$. Then, by Taylor's Theorem,

$$\psi(x) = \sum_{k=0}^{n} \psi^{(k)}(N)\frac{(x-N)^k}{k!} + r_n(x)\frac{(x-N)^n}{n!}$$

[4] *Moments are used in many fields including branches of Fractal theory [24, 48].*

where $r_n(x)$ is uniformly bounded and $\lim_{x\to N} r_n(x)(x-N)^n = 0$. Then by substituting into Equation (2.6), we have

$$\begin{aligned}
0 &= \int_{\mathbb{R}} \left(\sum_{k=0}^{n} \psi^{(k)}(N) \frac{(2^{-m}t)^k}{k!} + r_n(2^{-m}t+N)\frac{(2^{-m}t)^n}{k!} \right) \psi(t)dt \\
&= \psi^{(n)}(N) \int_{\mathbb{R}} 2^{-mn} \frac{t^n}{n!} \psi(t)dt + \int_{\mathbb{R}} r_n(2^{-m}t+N) \frac{2^{-mn}t^n}{n!} \psi(t)dt.
\end{aligned}$$

Now, multiplying both sides by $2^{mn}n!$ and letting $m \to \infty$,

$$\begin{aligned}
0 &= \psi^{(n)}(N) \int_{\mathbb{R}} t^n \psi(t)dt + \int_{\mathbb{R}} \lim_{m\to\infty} r_n(2^{-m}t+N)t^n\psi(t)dt \\
&= \psi^{(n)}(N) \int_{\mathbb{R}} t^n \psi(t)dt + 0.
\end{aligned}$$

As $\psi^{(n)}(N) \neq 0$, $\int_{\mathbb{R}} t^n \psi(t)dt = 0$. ■

Lemma 2.4.8 *For* $n = 0, 1, \ldots, r$,

$$\int_{\mathbb{R}} q_m(x,y)y^n dy = x^n \quad \forall m \in \mathbb{Z}.$$

Proof By Lemma 2.4.7, for $k \in \mathbb{Z}$ and $0 \leq n \leq r$,

$$\int_{\mathbb{R}} \psi(y-k)y^n dy = 0.$$

Therefore

$$\int_{\mathbb{R}} r(x,y)y^n dy = \sum_k \psi(x-k) \int_{\mathbb{R}} \psi(y-k)y^n dy = 0,$$

where $r(x,y)$ is the reproducing kernel of ψ. Since $V_1 = V_0 \oplus W_0$, the RK of V_1 is also given by

$$q_1(x,y) = q(x,y) + r(x,y).$$

Therefore,

$$\begin{aligned}\int_{\mathbb{R}} q_1(x,y)y^n dy &= \int_{\mathbb{R}} q(x,y)y^n dy + \int_{\mathbb{R}} r(x,y)y^n dy \\ &= \int_{\mathbb{R}} q(x,y)y^n dy, \end{aligned} \tag{2.7}$$

hence $\int_{\mathbb{R}} q_m(x,y)y^n dy$ does not change with m.

Now, consider $x \in \mathbb{R}$ with $|x| \leq 1$. Let $\theta \in S_r$ with $0 \leq \theta(y) \leq 1$ for all $y \in \mathbb{R}$ and $\theta(y) = 1$ when $|y| \leq 2$. Then

$$\int_{\mathbb{R}} q_m(x,y)y^n dy = \int_{\mathbb{R}} q_m(x,y)y^n(\theta(y) + (1-\theta(y))dy.$$

By definition of θ, $y^n\theta(y) \in S_r$, hence by Proposition 2.3.8,

$$\int_{\mathbb{R}} q_m(x,y)y^n\theta(y)dy \to x^n\theta(x) = x^n.$$

The remaining integral becomes

$$\int_{\mathbb{R}} q_m(x,y)y^n(1-\theta(y))dy = \left(\int_{-\infty}^{-2} + \int_2^{\infty}\right) q_m(x,y)y^n(1-\theta(y))dy.$$

Examining the second part gives

$$\begin{aligned}\int_2^{\infty} |q_m(x,y)y^n(1-\theta(y))|dy &\leq \int_2^{\infty} \frac{C_{n+2}}{(1+2^m|x-y|)^{n+2}} y^n dy \\ &\leq C_{n+2} \max_{\substack{y\geq 2 \\ |x|\leq 1}} \frac{y^n}{(1+2^m(y-x))^n} \max_{|x|\leq 1} \int_2^{\infty} \frac{dy}{(1+2^m(y-x))^2} \\ &= C_{n+2}\frac{2^n}{(1+2^m(2-1))^n} \int_2^{\infty} \frac{dy}{(1+2^m(y-1))^2}. \end{aligned} \tag{2.8}$$

Therefore,

$$\int_2^{\infty} q_m(x,y)y^n(1-\theta(y))dy \to 0$$

uniformly for $|x| \leq 1$. The other integral can be bounded similarly. Thus, by changing the scale, the condition $|x| \leq 1$ can be removed, and hence

$$\int_{\mathbb{R}} q_m(x,y)y^n dy \to x^n,$$

as $m \to \infty$, for all $x \in \mathbb{R}$ and $0 \neq n \leq r$. By Equation (2.7), since the integral does not change with m, the result follows. ■

Proof of Theorem 2.4.6: Note that

$$\begin{aligned} 2^m \int_{\mathbb{R}} f(x)\delta(2^m x - 2^m y)dx &= \int_{\mathbb{R}} f(2^{-m}z + y)\delta(z)dz \\ &= f(y). \end{aligned}$$

Therefore

$$2^m \delta(2^m x - 2^m y) = \delta(x - y). \tag{2.9}$$

Now, define $\epsilon(x,y) = q(x,y) - \delta(x-y)$ and let

$$\epsilon_m(x,y) = 2^m \epsilon(2^m x, 2^m y).$$

By Equation (2.9), $\epsilon_m(x,y) = q_m(x,y) - \delta(x-y)$. Therefore,

$$||q_m(\cdot,y) - \delta(\cdot - y)||_{-\alpha} = ||\epsilon_m(\cdot,y)||_{-\alpha}.$$

Now, consider $||\epsilon_m(\cdot, y)||_{-\alpha}^2$:

$$
\begin{aligned}
||\epsilon_m(\cdot, y)||_{-\alpha}^2 &= \int_{\mathbf{R}} (1+\omega^2)^{-\alpha} |\hat{\epsilon}_m(\omega, y)|^2 d\omega \\
&= 2^m \int_{\mathbf{R}} (1+2^{2m}\xi^2)^{-\alpha} |\hat{\epsilon}(\xi, 2^m y)|^2 d\xi \\
&= 2^m \left\{ \int_{|\xi|\le 1} + \int_{|\xi|>1} \right\} (1+2^{2m}\xi^2)^{-\alpha} |\hat{\epsilon}(\xi, 2^m y)|^2 d\xi \\
&= I_1 + I_2.
\end{aligned}
$$

Since

$$
\begin{aligned}
\hat{q}(\omega, y) &= \sum_n \hat{\phi}(\omega) e^{-i\omega n} \phi(y-n) \\
&= \hat{\phi}(\omega) Z\phi(y, \omega)
\end{aligned}
$$

and since ϕ satisfies Z_λ,

$$
\hat{q}(\omega, y) = e^{-i\omega y}(1 + O(|\omega|^\lambda)),
$$

then

$$
\begin{aligned}
\hat{\epsilon}(\omega, y) &= e^{-i\omega y}(1 + O(|\omega|^\lambda)) - e^{-i\omega y} \\
&= O(|\omega|^\lambda).
\end{aligned}
$$

Hence

$$
\begin{aligned}
|I_1| &\le 2 \cdot 2^m \int_0^1 (1+2^{2m}\xi^2)^{-\alpha} (C|\xi|^\lambda)^2 d\xi \\
&= 2C^2 \int_0^{2^m} (1+\omega^2)^{-\alpha} |2^{-m}\omega|^{2\lambda} d\omega \\
&= 2^{-2m\lambda} C_1 \int_0^{2^m} (1+\omega^2)^{-\alpha} \omega^{2\lambda} d\omega \\
&= O(2^{-2m\lambda}).
\end{aligned}
$$

For the second integral, we have

$$\begin{aligned}
|I_2| &\le 2^m \int_{|\xi|>1} (1+2^{2m}\xi^2)^{-\alpha} |\hat{q}(\xi, 2^m y) - e^{-i\xi 2^m y}|^2 d\xi \\
&\le 2^m \int_{|\xi|>1} (2^{2m}\xi^2)^{-\alpha} \cdot 2(|\hat{q}(\xi, 2^m y)|^2 + 1) d\xi \\
&= 2^{m(1-2\alpha)+1} \int_{|\xi|>1} |\xi|^{-2\alpha} (|\hat{q}(\xi, 2^m y)|^2 + 1) d\xi \\
&= 2^{m(1-2\alpha)+1} \int_{|\xi|>1} |\xi|^{-2\alpha} (|\hat{\phi}(\xi)|^2 |Z\phi(2^m y, \xi)|^2 + 1) d\xi \\
&\le 2^{m(1-2\alpha)+1} \left[||\phi||_1 \int_{|\xi|>1} |\xi|^{-2\alpha} |Z\phi(2^m y, \xi)|^2 d\xi + \int_{|\xi|>1} |\xi|^{-2\alpha} d\xi \right] \\
&\le 2^{m(1-2\alpha)+1} \left[||\phi||_1 \cdot 2 \sum_{k=0}^{\infty} \int_0^{2\pi} \frac{|Z\phi(2^m y, \omega+1)|^2}{|\omega + 1 + 2\pi k|^{2\alpha}} d\omega + 2 \cdot \int_1^{\infty} \xi^{-2\alpha} d\xi \right] \\
&\le 2^{m(1-2\alpha)+2} \left[||\phi||_1 \sum_{k=0}^{\infty} \int_0^{2\pi} \frac{D_1}{|1+2\pi k|^{2\alpha}} d\omega + \frac{1}{2\alpha - 1} \right] \\
&\le 2^{m(1-2\alpha)+2} \left[D_2 \sum_{k=0}^{\infty} |1+2\pi k|^{-2\alpha} + \frac{1}{2\alpha-1} \right] \\
&= O\left(2^{-m(2\alpha-1)}\right).
\end{aligned}$$

Since $\alpha > \lambda + \frac{1}{2}$,

$$2^{-m(2\alpha-1)} < 2^{-m\left(2\left(\lambda+\frac{1}{2}\right)-1\right)}$$

thus $|I_2| = O\left(2^{-m(2\alpha-1)}\right)$. Therefore,

$$\begin{aligned}
&||\epsilon_m(\cdot, y)||_{-\alpha}^2 = O(2^{-2\lambda m}) \\
\Longrightarrow\ &||\epsilon_m(\cdot, y)||_{-\alpha} = O(2^{-\lambda m}). \qquad \blacksquare
\end{aligned}$$

The following corollary is immediate:

Corollary 2.4.9 *Let $f \in H^\alpha$, $\phi \in S_r$ satisfy Z_λ for some $\lambda > 0$, and let $\alpha > \lambda + \frac{1}{2}$. Then the projections f_m of f onto V_m satisfy*

$$||f - f_m||_\infty = O(2^{-m\lambda}).$$

Proof By the Sobolev inequality, if $f \in H^\alpha$ and $g \in H^{-\alpha}$ then

$$| < g, f > | \leq ||g||_{-\alpha}||f||_\alpha.$$

Since $H^\alpha \subset H^\beta$ for $\beta \leq \alpha$, $f_m \in H^{-\alpha}$. Therefore, given $y \in \mathbb{R}$,

$$\begin{aligned} |f(y) - f_m(y)| &= | < f, \delta(\cdot - y) - q_m(\cdot, y) > | \\ &\leq ||f||_\alpha ||\delta(\cdot - y) - q_m(\cdot, y)||_{-\alpha} \\ &= O(2^{-m\lambda}). \quad \blacksquare \end{aligned}$$

We can compare this to a similar result for Fourier series. Let $H^\beta_{2\pi}$ be the space of all periodic $f \in S^*$ such that

$$\sum_{n=-\infty}^{\infty} |c_n|^2(n^2 + 1)^\alpha < \infty$$

where c_n are the Fourier coefficients of f.

Proposition 2.4.10 *Let $f \in H^\beta_{2\pi}$, $\beta > \frac{1}{2}$, then the Fourier series of f converges to f uniformly at a rate of $O(n^{-\beta+\frac{1}{2}})$.*

Proof See [51, p.54]. $\blacksquare$

Even with summability methods (using the Abel or Fejér kernels), this rate of convergence cannot be improved in general, even if f is smooth [55, p.122].

2.5 The Mallat Algorithm

We now construct an algorithm, developed by Mallat [32], which relates coefficients at different scales in a MRA. Let $\phi \in S_r$ generate a MRA $\{V_m\}$. Then $f \in V_1$ can be written in two forms

$$\begin{aligned} f(x) &= \sum_n a_n^1 \sqrt{2}\phi(2x-n) \\ &= \sum_n a_n^0 \phi(x-n) + \sum_n b_n^0 \psi(x-n) \end{aligned}$$

since $V_1 = V_0 \oplus W_0$.

By the dilation equations

$$\phi(x-l) = \sqrt{2}\sum_k h_k \phi(2x-2l-k)$$

and

$$\psi(x-l) = \sqrt{2}\sum_k g_k \phi(2x-2l-k).$$

Making the choice $g_k = (-1)^k h_{1-k}$ from Equation (2.3) on page 67,

$$\begin{aligned} f(x) &= \sum_l a_l^0 \sqrt{2}\sum_k h_k \phi(2x-2l-k) \\ &\quad + \sum_l b_l^0 \sqrt{2} \sum_k (-1)^k h_{1-k}\phi(2x-2l-k) \\ &= \sqrt{2}\sum_n \sum_j h_{n-2j} a_j^0 \phi(2x-n) \\ &\quad + \sqrt{2}\sum_n \sum_j (-1)^n h_{1-n+2j} b_j^0 \phi(2x-n). \end{aligned}$$

By the orthogonality conditions, for $n \in \mathbb{Z}$,

$$a_n^1 = \sum_j h_{n-2j} a_j^0 + \sum_j (-1)^n h_{1-n+2j} b_j^0.$$

This relation can be derived at each scale in an analogous way, hence

$$a_n^m = \sum_j h_{n-2j} a_j^{m-1} + \sum_j (-1)^n h_{1-n+2j} b_j^{m-1}.$$

The coefficients a_n^0 and b_n^0 can be found in terms of the a_n^1 as follows:

$$\begin{aligned}
a_n^0 &= \int_{\mathbb{R}} f(x)\phi(x-n)dx \\
&= \int_{\mathbb{R}} f(x)\sqrt{2}\sum_k h_k \phi(2x-2n-k) \\
&= \sum_k h_k \int_{\mathbb{R}} f(x)\sqrt{2}\phi(2x-2n-k) \\
&= \sum_k h_k a_{2n+k}^1 \\
&= \sum_k h_{k-2n} a_k^1,
\end{aligned}$$

and similarly

$$\begin{aligned}
b_n^0 &= \int_{\mathbb{R}} f(x)\psi(x-n) \\
&= \sum_k g_{k-2n} a_k^1 \\
&= \sum_k (-1)^k h_{1-k+2n} a_k^1.
\end{aligned}$$

Therefore,

$$a_n^{m-1} = \sum_k a_k^m h_{k-2n}$$

and

$$b_n^{m-1} = \sum_k a_k^m (-1)^k h_{1-k+2n}.$$

The above algorithm is called the *Mallat algorithm*. We can interpret these results by viewing the MRA as a sequence of varying resolutions of $L^2(\mathbb{R})$. Given $f \in V_1$, $f = f_1 = f_0 + e_0$, where e_0 is the projection of f onto W_0. One can think of f_0 as a coarser version of f, and e_0 as the error in the approximation; that is, $\{V_m\}$ contains approximations of $L^2(\mathbb{R})$ functions and $\{W_m\}$ contains the error in their approximations. Therefore, the first part of the algorithm consists of the decomposition of f into its scaling coefficients (a_n^m) and its wavelet coefficients (b_n^m) at a selected level $m < 1$. The function can then be reconstructed by applying the second part. Figure 2.4 shows the two parts of the algorithm. This is a fast algorithm in practical applications, being of complexity $O(N)$ as opposed to $O(N \log N)$ for the Fast Fourier Transform [7, pp.80-81].

2.6 Filters

The Mallat algorithm is a useful tool for approximating discrete-time signals. It will be shown that it acts like a pair of quadrature mirror filters (QMF). Recall from Section 2.2

$$\begin{tikzcd}
 & & b_n^{m-1} & b_n^{m-2} & b_n^{m-3} & \\
\cdots \arrow[r] & a_n^m \arrow[r] \arrow[ur] & a_n^{m-1} \arrow[r] \arrow[ur] & a_n^{m-2} \arrow[r] \arrow[ur] & a_n^{m-3} \arrow[r] \arrow[ur] & \cdots
\end{tikzcd}$$

$$\begin{tikzcd}
 & & b_n^{m-1} \arrow[dl] & b_n^{m-2} \arrow[dl] & b_n^{m-3} \arrow[dl] & {} \arrow[dl] \\
\cdots & a_n^m \arrow[l] & a_n^{m-1} \arrow[l] & a_n^{m-2} \arrow[l] & a_n^{m-3} \arrow[l] & \cdots \arrow[l]
\end{tikzcd}$$

Figure 2.4: Above: decomposition algorithm. Below: reconstruction algorithm.

on page 64 that for $f \in L^2(\mathbb{R})$ with $\operatorname{supp} \hat{f} \subset [-\pi, \pi]$, we have

$$f(t) = \sum_n f(n) \frac{\sin \pi(t-n)}{\pi(t-n)}.$$

Such a function f is called a *bandlimited* function [15, p.20]. We can therefore consider a sequence $(a_n) \in \ell^2(\mathbb{Z})$ as a sequence of sampled values of a function $f \in L^2(\mathbb{R})$ with

$$f(t) = \sum_n a_n \frac{\sin \pi(t-n)}{\pi(t-n)};$$
$$\hat{f}(\omega) = \sum_n a_n e^{-in\omega}.$$

Consider a function g defined by

$$g = h * f,$$

where $\hat{h}(\omega) = H(\omega) = \sum_n h_n e^{-in\omega}$ is 2π-periodic. Then

$$\begin{aligned}\hat{g}(\omega) &= \hat{h}(\omega)\hat{f}(\omega)\\ &= \sum_k h_k e^{-ik\omega} \sum_n a_n e^{-in\omega}\\ &= \sum_m e^{-im\omega} \sum_n h_{m-n} a_n\end{aligned}$$

and

$$g(t) = \sum_m \sum_n h_{m-n} a_n \frac{\sin \pi(x-m)}{\pi(x-m)}.$$

In the case of a discrete signal $(x_n) \in \ell^2(\mathbb{Z})$, the convolution becomes

$$y_n = \sum_k h_{n-k} x_k,$$

hence the Fourier series of (y_n) is

$$\begin{aligned}Y(\omega) &= \sum_n y_n e^{i\omega n}\\ &= \sum_n \sum_k h_{n-k} x_k e^{i\omega n}\\ &= \sum_l \sum_k h_l x_k e^{i\omega l} e^{i\omega k}\\ &= \sum_l h_l e^{i\omega l} \sum_k x_k e^{i\omega k}\\ &= H(\omega) X(\omega).\end{aligned}$$

Definition 2.6.1 *The above operation is called a* continuous linear system. *The function $H(\omega)$ is called the* system transfer function *and $h(t)$ is called the* impulse response. *If $H(\omega) = 0, |\omega| \geq \omega_0$ then H is called a* low-pass filter. *If $H(\omega) = 0, |\omega| \leq \omega_1$ then H is* *called a* high-pass filter. *If* $H(\omega) = 0, \omega_0 \leq |\omega| \leq \omega_1$ *then* H *is called a* band-pass filter.

Consider the Mallat algorithm. For the decomposition algorithm we have

$$\begin{aligned} a_n^0 &= \sum_k h_{k-2n} a_k^1, \\ b_n^0 &= \sum_k g_{k-2n} a_k^1 \\ &= \sum_k (-1)^k h_{1-k+2n} a_k^1. \end{aligned}$$

This can be decomposed into a filter

$$e_n^0 = \sum_k h_{k-n} a_k^1$$

followed by decimation

$$a_n^0 = e_{2n}^0.$$

For the wavelet coefficients, let

$$\begin{aligned} f_n^0 &= \sum_k g_{k-n} a_k^1 \\ &= \sum_k (-1)^{k-n} h_{1-k+n} a_k^1 \end{aligned}$$

with decimation

$$b_n^0 = f_{2n}^0.$$

Let $E(\omega) = \sum_n e_n^0 e^{i\omega n}$. Then

$$\begin{aligned}E(\omega) &= \sum_n \sum_k h_{k-n} a_k^1 e^{i\omega n} \\ &= \sum_l h_{-l} e^{i\omega l} \sum_k a_k^1 e^{i\omega k} \\ &= H(\omega) A(\omega),\end{aligned}$$

hence the filter has an impulse response (h_{-n}) with system transfer function $H(\omega) = \sum_n h_{-n} e^{i\omega n}$ [51].

The dilation equation for ϕ may be written in terms of Fourier transforms as

$$\begin{aligned}\hat{\phi}(\omega) &= \frac{1}{\sqrt{2}} \sum_n h_n \hat{\phi}\left(\frac{\omega}{2}\right) e^{-i\omega n/2} \\ &= m_0\left(\frac{\omega}{2}\right) \hat{\phi}\left(\frac{\omega}{2}\right),\end{aligned}$$

thus $H(\omega) = m_0(\omega)/\sqrt{2}$.

For the Meyer wavelets [38, p.66], $m_0\left(\frac{\omega}{2}\right) = \sum_k \hat{\phi}(\omega + 4\pi k)$. On the interval $\omega \in [-\pi, \pi]$, the support of $H(\omega)$ is $|\omega| < \frac{\pi+\epsilon}{2}$, for some $\epsilon > 0$. That is, $H(\omega)$ is a low-pass filter [51].

For the b_k^1, let

$$\begin{aligned}F(\omega) &= \sum_n f_n^0 e^{i\omega n} \\ &= \sum_n \sum_k (-1)^{k-n} h_{1-k+n} a_k^1 e^{i\omega n} \\ &= \sum_l \sum_k (-1)^l h_{1+l} a_k^1 e^{i\omega(l+k)} \\ &= \sum_l (-1)^l h_{1+l} e^{i\omega l} \sum_k a_k^1 e^{i\omega k} \\ &= G(\omega) A(\omega).\end{aligned}$$

Then

$$\begin{aligned} G(\omega) &= \sum_l (-1)^l h_{1+l} e^{i\omega l} \\ &= \sum_l h_{1+l} e^{i(\omega+\pi)l} \\ &= \sum_n h_{-n} e^{-i(\omega+\pi)(n+1)} \\ &= e^{-i(\omega+\pi)} \sum_n h_{-n} e^{-i(\omega+\pi)n} \\ &= e^{-i(\omega+\pi)} H(-(\omega+\pi)). \end{aligned}$$

Therefore, for $\omega \in [-\pi, \pi]$, $G(\omega) = 0$, for $|\omega| < \frac{\pi-\epsilon}{2}$ and some $\epsilon > 0$. Hence $(g_k) = ((-1)^k h_{1-k})$ is the impulse response of a high-pass filter.

In a similar way, we can consider the reconstruction algorithm as a pair of filters. Recall

$$a_n^1 = \sum_k h_{n-2k} a_k^0 + \sum_k g_{n-2k} b_k^0.$$

If (c_n^0) is a sequence, define $(c_n^{0,0})$ to be the sequence obtained from (c_n^0) by interlacing zeros; that is, let

$$c_n^{0,0} = \begin{cases} c_{\frac{n}{2}}^0 & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd.} \end{cases}$$

Then, let

$$e_n^{0*} = \sum_k h_{n-k} a_k^{0,0}$$

and

$$\begin{aligned} f_n^{0*} &= \sum_k g_{n-k} b_k^{0,0} \\ &= \sum_k (-1)^{n-k} h_{1-n+k} b_k^{0,0}. \end{aligned}$$

Hence, $b_n^1 = e_n^{0*} + f_n^{0*}$. The Fourier series of (e_n^{0*}) and (f_n^{0*}) are, respectively,

$$\begin{aligned} E^*(\omega) &= \sum_n e_n^{0*} e^{i\omega n} \\ &= \sum_n \sum_k h_{n-k} a_k^{0,0} e^{i\omega n} \\ &= \sum_l h_l e^{i\omega l} \sum_k a_k^{0,0} e^{i\omega k} \\ &= H^*(\omega) A^*(\omega) \end{aligned}$$

and

$$\begin{aligned} F^*(\omega) &= \sum_n f_n^{0*} e^{i\omega n} \\ &= \sum_n \sum_k (-1)^{n-k} h_{1-n+k} b_k^{0,0} e^{i\omega n} \\ &= \sum_l (-1)^l h_{1-l} e^{i\omega l} \sum_k b_k^{0,0} e^{i\omega k} \\ &= G^*(\omega) B^*(\omega). \end{aligned}$$

Therefore, $H^*(\omega)$ is a low-pass filter, and

$$\begin{aligned}
G^*(\omega) &= \sum_l (-1)^l h_{1-l} e^{i\omega l} \\
&= \sum_l h_{1-l} e^{i(\omega+\pi)l} \\
&= \sum_n h_n e^{i(\omega+\pi)(1-n)} \\
&= e^{i(\omega+\pi)} \sum_n h_n e^{-i(\omega+\pi)n} \\
&= e^{i(\omega+\pi)} H^*(-(\omega+\pi)),
\end{aligned}$$

hence is a high-pass filter. The filters H^* and G^* are called the conjugate filters of H and G respectively. The Mallat algorithm can be represented schematically by Figure 2.5.

2.7 Applications

The goal now is to use the Mallat algorithm to approximate functions. First, consider the following result of Daubechies [15, pp.202-204]:

Theorem 2.7.1 *Suppose ϕ has compact support. If f is continuous on $\mathbb{R}$ then $\forall x \in \mathbb{R}$,*

$$\lim_{m\to\infty} \int_{\mathbb{R}} f(x+y) 2^{m/2} \overline{\phi_{m,n}(y)} dy = f(x).$$

If f is uniformly continuous then the convergence is uniform. If f is Lipschitz continuous with exponent α, then

$$\left| f(x) - \int_{\mathbb{R}} f(x+y) 2^{m/2} \overline{\phi_{m,n}(y)} dy \right| = O(2^{-m\alpha}). \tag{2.10}$$

$$
\begin{tikzcd}
 & e_n^{m-1} & a_n^{m-1} & a_n^{m-1,0} & e_n^{m-1,0} & \\
 & H \arrow[u] \arrow[r] & 2\searrow \arrow[u] \arrow[r] & 2\nearrow \arrow[u] \arrow[r] & H^* \arrow[u] \arrow[dr] & \\
a_n^m \arrow[ur] \arrow[dr] & & & & & a_n^m \\
 & G \arrow[d] \arrow[r] & 2\searrow \arrow[d] \arrow[r] & 2\nearrow \arrow[d] \arrow[r] & G^* \arrow[d] \arrow[ur] & \\
 & f_n^{m-1} & b_n^{m-1} & b_n^{m-1,0} & f_n^{m-1,0} &
\end{tikzcd}
$$

Figure 2.5: The Mallat algorithm. The left half denotes the decomposition and the right denotes the reconstruction. The symbols $2\searrow$ and $2\nearrow$ represent decimation and interleaving by zeros respectively.

Proof Let $m \in \mathbb{N}$, then

$$\begin{aligned}\left|f(x) - \int_{\mathbb{R}} f(x+y)2^{m/2}\overline{\phi_{m,n}(y)}dy\right| &= \left|\int_{\mathbb{R}}(f(x) - f(x+2^{-m}(t+n)))\overline{\phi(t)}dt\right| \\ &\leq ||\phi||_1 \sup_{|t|\leq K} |(f(x) - f(x+2^{-m}(t+n)))|\end{aligned}$$

where $\operatorname{supp}\phi \subset [-K, K]$. Since f is continuous, we can find M such that $\forall m \geq M$, the right hand side is arbitrarily small. If f is uniformly continuous, then the choice of M can be made independent of x, hence the convergence is uniform. If f is Lipschitz continuous, Equation (2.10) follows immediately. ■

Hence, we have a way of computing ϕ. Suppose ϕ is continuous, or Lipschitz continuous with exponent α. Let $x = 2^{-M}N$, $M, N \in \mathbb{N}$. Then, by Theorem 2.7.1,

$$\begin{aligned}\phi(x) &= \lim_{m\to\infty} 2^m \int_{\mathbb{R}} \phi\left(2^{-M}N + y\right) \overline{\phi(2^m y)}dy \\ &= \lim_{m\to\infty} 2^{m/2} \int_{\mathbb{R}} \phi(t)\overline{\phi_{m,2^{m-M}N}(t)}dt \\ &= \lim_{m\to\infty} 2^{m/2} < \phi, \phi_{m,2^{m-M}N} > .\end{aligned}$$

In addition, we can find M_0 such that $\forall m \geq M_0$,

$$\left|\phi(2^{-M}N) - 2^{m/2} < \phi, \phi_{m,2^{m-M}N} >\right| \leq C2^{-m\alpha}, \tag{2.11}$$

where C and M_0 depend on M or N.

Assuming $(\phi_{0,n})$ are orthonormal, then ϕ is the unique function f satisfying

$$< f, \phi_{0,n} > = \delta_{0n},$$

$$< f, \psi_{m,n} > = 0 \quad \forall m > 0, n \in \mathbb{Z}.$$

We can use this, along with the filtering scheme in Section 2.6, to compute ϕ given a set of filter coefficients (h_n). For each $n \in \mathbb{Z}$, let $a_n^0 = \delta_{0n}$ be the low-pass sequence and at each level m, let $b_n^m = 0$ be the high-pass sequence. Then

$$\begin{aligned} a_n^1 &= \sum_k h_{n-2k} a_n^0 \\ &= h_n. \end{aligned}$$

Therefore, for each m, $a_n^m = < \phi, \phi_{m,n} >$. By Equation (2.11) the algorithm converges to the values of ϕ at the dyadics. Defining $f_k(2^{-m}n) = 2^{m/2} < \phi, \phi_{m,n} >$, with f_k being piecewise constant on $[2^{-m}(n-1/2), 2^{-m}(n+1/2))$, for $n \in \mathbb{Z}$, we have the following:

Proposition 2.7.2 *If ϕ is Lipschitz continuous with exponent α, then there exists a $C > 0$ and $M_0 \in \mathbb{N}$ such that $\forall m \geq M_0$,*

$$||\phi - f_m||_\infty \leq C2^{-\alpha m}.$$

Proof A proof can be found in [15, p.205]. ■

Therefore, to compute approximate values of $\phi(x)$ we have the *cascade algorithm* [15, p.205].

i) Start with a sequence $(f_0(n))$ with $f_0(n) = \delta_{0n}$.

ii) Compute $f_m(2^{-m}n), n \in \mathbb{Z}$ using $a_n^m = \sum_k h_{n-2k} a_k^{m-1}$. At every step, the number of values doubles: values at "even points", $2^{-m}(2n)$ by

$$f_m(2^{-m}2n) = \sum_k h_{2(n-k)} f_{m-1}(2^{-m}k)$$

and "odd points", $2^{-m}(2n+1)$ by

$$f_m(2^{-m}(2n+1)) = \sum_k h_{2(n-k)+1} f_{m-1}(2^{-m}k).$$

iii) Interpolate the $f_m(2^{-m}n)$ to get $f_m(x)$ for non-dyadic x.

Similarly, we could calculate ψ by starting with the low-pass sequence $a_n^0 = 0$ and high-pass sequence $b_n^0 = \delta_{0n}$ or moreover, calculate $\phi_{m,n}$ or $\psi_{m,n}$, by choosing the appropriate initial sequences. Figure 2.6 shows the Daubechies-4 scaling function and mother wavelet. They are given by the filter coefficients

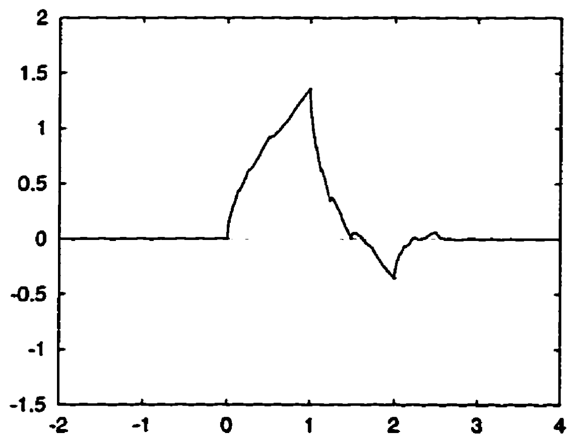
n	h_n
0	0.4829629131445341
1	0.8365163037378079
2	0.2241438680420134
3	-0.1294095225512604

Finally, the Mallat algorithm can be used for compression. Start with a function f which, by Theorem 2.7.1, can be approximated by assuming, for large enough m, that

$$\begin{aligned} f(2^{-m}n) &= 2^{m/2} < f, \phi_{m,n} > \\ &= 2^{m/2} a_n^m. \end{aligned}$$

Then, use the decomposition algorithm to compute the wavelet coefficients of f.

For finite sequences, suppose we start with a signal f consisting of 2^M samples on $[0,1]$ at the dyadics $x = 2^{-M}n, 0 \le n < 2^M$. By Theorem 2.7.1 we assume these values are the



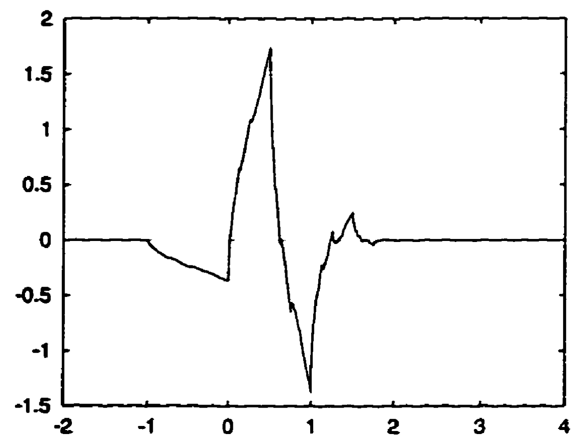


Figure 2.6: Daubechies-4 scaling function and mother wavelet.

scaling coefficients of f, that is

$$f(2^{-M}n) = 2^{m/2}a_n^m.$$

We then periodize f by assuming it has period one. This means that we assume $f = f^*$, where

$$f^*(x) = \sum_{n\in\mathbf{Z}} f(x-n).$$

For simplicity in coding, we adopt a new notation for the coefficients, letting

$$a_{m,n} = a_n^m \text{ and } b_{m,n} = b_n^m.$$

By the first application of the decomposition algorithm, we obtain $M/2$ wavelet coefficients $(b_{m-1,n})$ and $M/2$ scaling coefficients $(a_{m-1,n})$. Continuing in this manner, we obtain $2^M - 1$ wavelet coefficients $(b_{m,n}), m = 0, 1, \ldots M, 0 \leq n \leq 2^m - 1$ and one scaling coefficient $a_{0,0}$. The set of wavelet coefficients and the scaling coefficient is called the *Fast*

<table>
<tr><td colspan="8">$a_{0,0}$</td></tr>
<tr><td colspan="8">$b_{0,0}$</td></tr>
<tr><td colspan="4">$b_{1,0}$</td><td colspan="4">$b_{1,1}$</td></tr>
<tr><td colspan="2">$b_{2,0}$</td><td colspan="2">$b_{2,1}$</td><td colspan="2">$b_{2,2}$</td><td colspan="2">$b_{2,3}$</td></tr>
<tr><td>$b_{3,0}$</td><td>$b_{3,1}$</td><td>$b_{3,2}$</td><td>$b_{3,3}$</td><td>$b_{3,4}$</td><td>$b_{3,5}$</td><td>$b_{3,6}$</td><td>$b_{3,7}$</td></tr>
<tr><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td></tr>
</table>

Figure 2.7: The wavelet tree of a function. The horizontal axis indicates a displacement in time, or location, whereas the vertical axis is a change in frequency.

Wavelet Transform of f. The coefficients can be arranged in a meaningful structure, called the *wavelet tree*, as shown in Figure 2.7.

This technique can be used to plot both periodized wavelets and scaling functions. Figure 2.8 shows the Daubechies-4 periodized scaling functions. Figure 2.9 shows some of the periodized Daubechies-4 wavelets. Recall from Theorem 2.3.10 that

$$\phi^*(x) = \sum_n \phi(x-n) = 1, \quad x \in \mathbb{R}.$$

Hence, the periodized scaling function $\phi^*(t) = 1$, $\forall t \in \mathbb{R}$, for any given MRA.

One method of compression consists of pruning branches of the tree by assuming that if a wavelet coefficient has absolute value below some threshold, the coefficients below it can be pruned (set to zero). Figure 2.10 shows a sequence of approximations of $u(x) = \sin(x)$ with varying threshold values for the wavelet coefficients. The Coifman-6 wavelets are determined by the following filter coefficients:

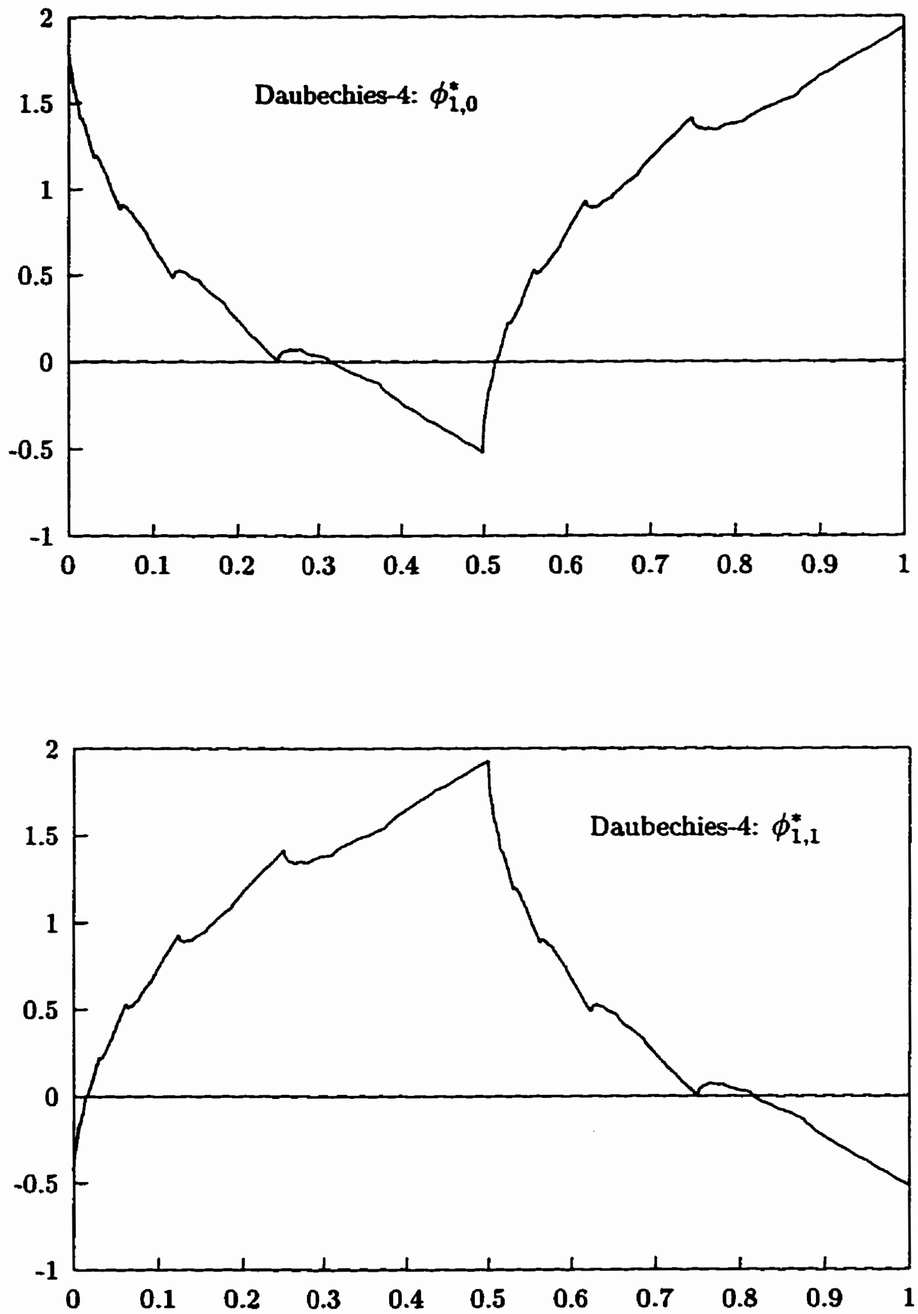


Figure 2.8: Daubechies-4 periodized scaling functions. The functions $\phi^*_{1,0}$ and $\phi^*_{1,1}$ form an orthonormal basis of V^*_1 [51, p.106].

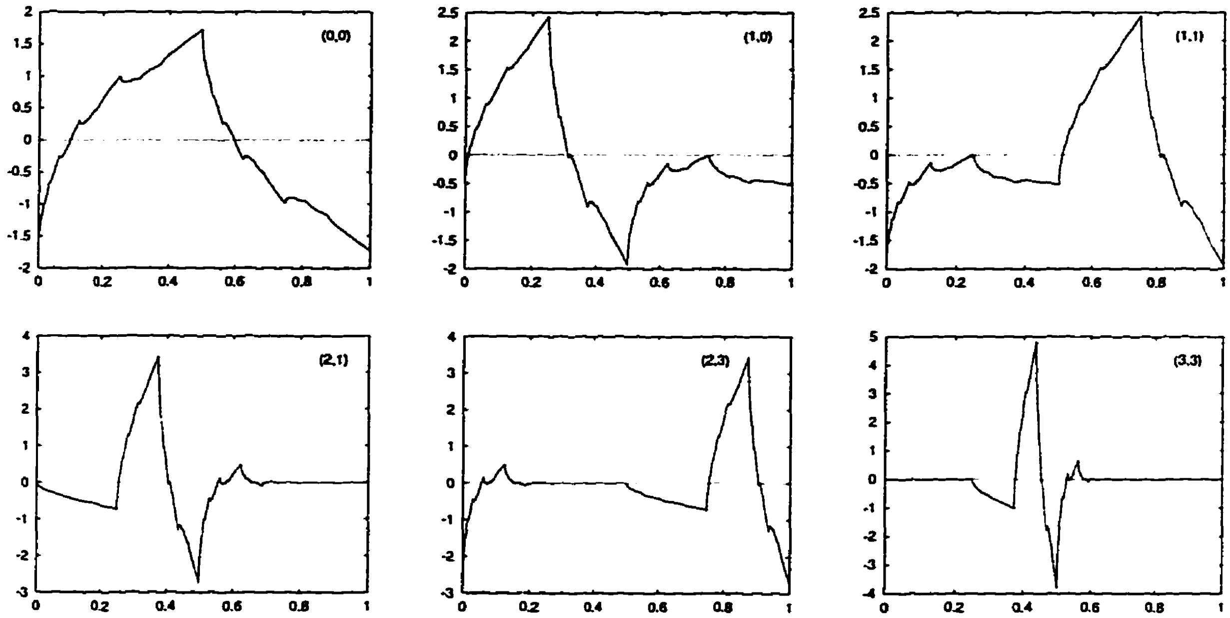


Figure 2.9: Daubechies-4 periodized wavelets, $\psi^*_{m,n}$.

n	h_n
0	0.22658426510
1	0.74568755880
2	0.60749164120
3	-0.07716155548
4	-0.12696912540
5	0.03858077774

From Figure 2.10, it is evident that the majority of the information about u is contained in the lower frequency coefficients. This method, called the *zero-tree* method, is similar to the method of JPEG compression [7, 27, 41]. The second method consists of using the fractal methods developed in Chapter 1.

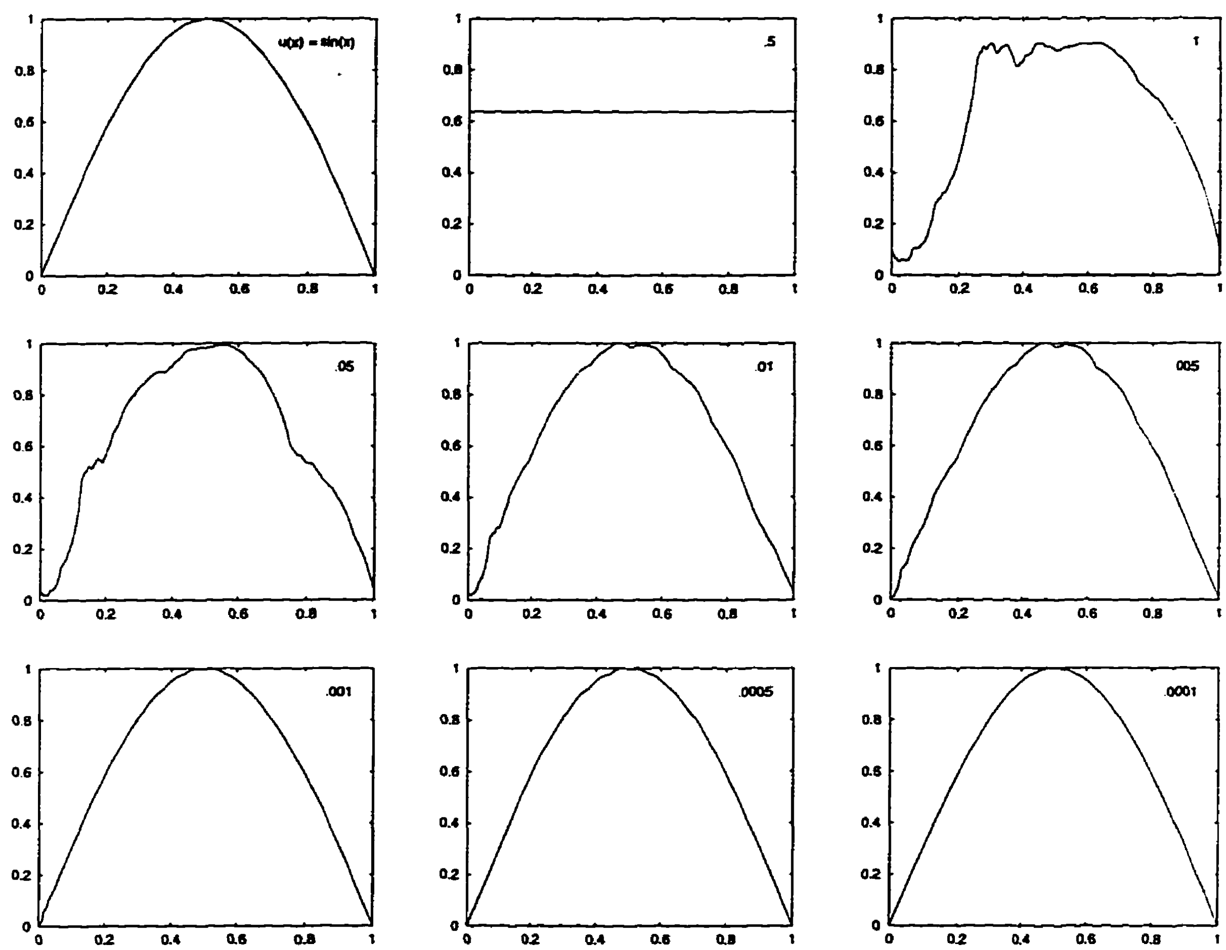


Figure 2.10: Fast Wavelet Transform approximation of $u(x) = \sin(\pi x)$ (top-left) using Coifman-6 wavelets with thresholds from .5 to .0001. The original function is at the top left.

Chapter 3

Fractal Wavelet Compression

3.1 Relations

We have seen in Chapter 1 the method of IFSM which allows the construction of functions through an iterative process. Given an N-map IFSM $(\mathbf{w}, \Phi)$, the associated operator T was defined by

$$Tu = \sum_{k=1}^{N}{}' \phi_k \circ u \circ w_k^{-1},$$

for all $u \in L^2(\mathbb{R})$ (see Equation (1.5) on page 35). Under certain conditions, T was contractive and had a unique fixed point $\bar{u}_T$, which was also the attractor of T. In the light of the results of Chapter 2, one might consider an operator M, associated to T, which acts on wavelet coefficients of functions [19, 37].

Let (q_n) be an orthonormal basis of $L^2(\mathbb{R})$. Then a function $u \in L^2(\mathbb{R})$ can be written as

$$u = \sum_n u_n q_n \tag{3.1}$$

where $u_n = < u, q_n >$.

Let $F : L^2(\mathbb{R}) \to \ell^2(\mathbb{Z})$ be the transform $Fu = (u_n)$. By Plancherel's Theorem, F is an isometry. Its inverse is $F^{-1}(c_n) = \sum_n c_n q_n$, $(c_n) \in \ell^2(\mathbb{Z})$. In general, write $\mathbf{u} = Fu$. Now define the operator M given by the commutative diagram

$$\begin{tikzcd} L^2(\mathbb{R}) \arrow[r, "F", "\cong"'] \arrow[d, "T"'] & \ell^2(\mathbb{Z}) \arrow[d, "M"] \\ L^2(\mathbb{R}) \arrow[r, "F", "\cong"'] & \ell^2(\mathbb{Z}) \end{tikzcd}$$

Theorem 3.1.1 *Let (q_n) be an orthonormal basis of $L^2(\mathbb{R})$ and F be its associated transform as given above. Then T is a contractive operator on $L^2(\mathbb{R})$ with fixed point $\bar{u}_T$ if and only if the operator $M = F \circ T \circ F^{-1}$ is contractive on $\ell^2(\mathbb{Z})$ with fixed point $\bar{\mathbf{u}}_M$, where $\bar{\mathbf{u}}_M = F\bar{u}_T$.*

Proof Let $u, v \in L^2(\mathbb{R})$ with basis coefficient sequences $\mathbf{u}$ and $\mathbf{v}$ respectively. By Parseval's equality, $||u|| = ||\mathbf{u}||$, hence

$$\begin{aligned} &||Tu - Tv|| \leq c||u - v|| \\ \iff &||M\mathbf{u} - M\mathbf{v}|| \leq c||\mathbf{u} - \mathbf{v}||. \end{aligned}$$

Hence T is contractive $\iff$ M is contractive. Furthermore,

$$M(F\bar{u}_T) = F(T\bar{u}_T) = F\bar{u}_T,$$

therefore by the BCMP, $F\bar{u}_T = \bar{\mathbf{u}}_M$. ■

Consider the case when T is the associated IFSM operator of an N-map affine IFSM on $X = \mathbb{R}$. Then given $u \in L^2(\mathbb{R})$, if $v = Tu$,

$$v = \sum_m v_m q_m$$

where

$$\begin{aligned} v_m &=< v, q_m > \\ &=< Tu, q_m > \\ &= \sum_{k=1}^{N} \alpha_k < u \circ w_k^{-1}, q_m > + \sum_{k=1}^{N} \beta_k < \chi_{w_k(X)}, q_m > . \end{aligned}$$

Therefore, by Equation (3.1) on page 107,

$$\begin{aligned} v_m &= \sum_{k=1}^{N} \alpha_k \sum_n u_n < q_n \circ w_k^{-1}, q_m > + \sum_{k=1}^{N} \beta_k < \chi_{w_k(X)}, q_m > \\ &= \sum_n a_{mn} u_n + e_m \end{aligned} \tag{3.2}$$

where $a_{mn} = \sum_{k=1}^{N} \alpha_k < q_n \circ w_k^{-1}, q_m >$ and $e_m = \sum_{k=1}^{N} \beta_k < \chi_{w_k(X)}, q_m >$.

By Equation (3.2), we get the following result [19]:

Proposition 3.1.2 *Let (q_n) be an orthonormal basis of $L^2(\mathbb{R})$ with associated transform F. If T is an affine IFSM on $L^2(\mathbb{R})$, then $M = F \circ T \circ F^{-1}$ is an affine IFS on coefficients (IFSC) on $\ell^2(\mathbb{Z})$ and has the form $M\mathbf{u} = A\mathbf{u} + \mathbf{e}$, where $A = (a_{mn})$ and $\mathbf{e} = (e_m)$, $m, n \in \mathbb{Z}$ are given above.*

In general, the matrix A is not sparse, for example with the Discrete Cosine Transform [37]. However, due to the localization properties of wavelets, many of these elements will vanish.

Example 3.1.3 Let $X = [0,1]$ and T be the operator defined by

$$Tu(x) = \frac{1}{2}u(2x) + \frac{1}{2}u(2x-1) \quad x \in [0,1].$$

This is the IFSM operator of the affine IFSM given by $w_1(x) = \frac{1}{2}x$, $w_2(x) = \frac{1}{2}(x+1)$ and $\phi_1(t) = \phi_2(t) = \frac{1}{2}t$. The fixed point of T is $\bar{u}_T \equiv 0$. Consider the operator M given when (q_n) is chosen to be the Haar basis on $L^2[0,1]$. We assign the following ordering to the basis elements:

$$\begin{aligned} q_{-1} &= \phi; \\ q_{2^i+j-1} &= \psi_{i,j}, \quad i \in \mathbb{N}, 0 \le j \le 2^i - 1. \end{aligned}$$

Then the operator A is

$$A = \frac{1}{2\sqrt{2}} \begin{pmatrix} \sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & 0 & \ddots \end{pmatrix}$$

For the moment at least, the above proposition only works one way. The question remaining is, given an affine IFSC M on $\ell^2(\mathbb{Z})$, is the operator T an affine IFSM. [1] The question is therefore:

Question 3.1.4 *Given an affine IFSC M on $\ell^2(\mathbb{Z})$, defined by $M\mathbf{u} = A\mathbf{u} + \mathbf{e}$ with $A = (a_{mn})$ and $\mathbf{e} = (e_k)$, does there exist an N-map IFSM $(\mathbf{w}, \Phi)$, for some orthonormal basis (q_n) of $L^2(\mathbb{R})$, such that*

$$a_{mn} = \sum_{k=1}^{N} \alpha_k < q_n \circ w_k^{-1}, q_m >$$

and

$$e_m = \sum_{k=1}^{N} \beta_k < \chi_{w_k(X)}, q_m >$$

for $m, n \in \mathbb{Z}$?

A case where this question has been solved is for LIFSW.

3.2 LIFSW

We present the general method of the 1-dimensional case of local IFS on wavelet coefficients (LIFSW) presented in [19, 37]. The 2-dimensional extension can be found in [47].

Let ϕ be a scaling function with MRA (V_m) and ψ be the standard associated mother wavelet (see Equation (2.3) on page 67). We focus our attention on functions $f \in L^2(\mathbb{R})$

[1] By an IFSC, we mean an operator which acts on sequences in some "IFS"-type manner.

which have expansions

$$f = a_{00}\phi + \sum_{i=0}^{\infty}\sum_{j=0}^{2^i-1} b_{i,j}\psi_{i,j}$$

where $a_{00} =< f, \phi >$ and $b_{i,j} =< f, \psi_{i,j} >$. Assume ψ has compact support on $\mathbb{R}$. The expansion coefficients can be written in a meaningful way in the form

<table>
<tr><td colspan="4" align="center">$a_{0,0}$</td></tr>
<tr><td colspan="4" align="center">$b_{0,0}$</td></tr>
<tr><td colspan="2" align="center">$b_{1,0}$</td><td colspan="2" align="center">$b_{1,1}$</td></tr>
<tr><td colspan="4" align="center">$\vdots$</td></tr>
<tr><td>$B_{k,0}$</td><td>$B_{k,1}$</td><td>$\ldots$</td><td>$B_{k,2^k-1}$</td></tr>
</table>

where $B_{i,j}$ represents the branch of coefficients with node $b_{i,j}$, and is called the *block* $B_{i,j}$. We say that the coefficients $b_{i,j}$ are on level i, and $a_{0,0}$ is at level -1. The above diagram is called the wavelet (coefficient) tree of f and is denoted by B^f.

Definition 3.2.1 *Consider the operator W defined on wavelet trees as follows:* [2] *Suppose there is a $k \geq 0, k^* > k, \alpha_j, |\alpha_j| < 2^{(k-k^*)/2}, 0 \leq j \leq 2^{k^*} - 1$ such that given a wavelet tree $B_{0,0}$, $W(B_{0,0}) = B^*_{0,0}$, where the coefficients of $B^*_{0,0}$ are given by*

$$\begin{aligned}
a^*_{0,0} &= a_{0,0}, & \\
b^*_{i,j} &= b_{i,j} & 0 \leq i \leq k^* - 1, 0 \leq j \leq 2^i - 1, \\
B^*_{k^*,j} &= \alpha_j B_{k,l(j)} & 0 \leq j \leq 2^{k^*} - 1, l(j) \in \{0, 1, \ldots, 2^k - 1\}.
\end{aligned}$$

Then W will be called a local IFS on wavelet coefficients, *or* LIFSW. *The blocks $B_{k,j}, 0 \leq j \leq 2^k - 1$ are called the* domain blocks. *The blocks $B^*_{k^*,j}, 0 \leq j \leq 2^{k^*} - 1$ are called the* range blocks. *The parameters α_i are called the* scaling factors *of W.*

---

[2] *The condition on the α_j guarantees that $W : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$. These α_j will correspond to grey-level maps $\phi_j(t) = 2^{(k^*-k)/2}\alpha_j(t)$.*

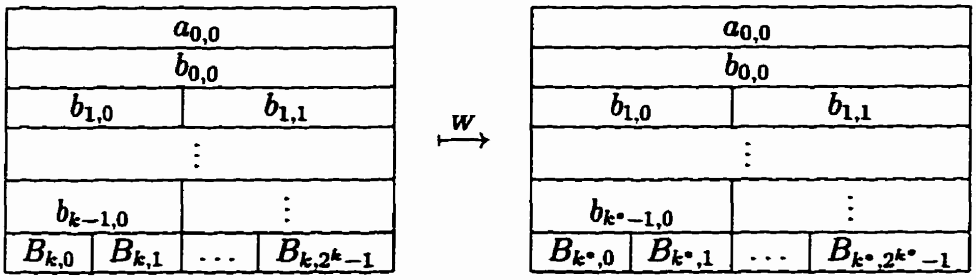


Figure 3.1: Action of W on a wavelet tree.

Definition 3.2.2 *Given $f \in L^2(\mathbb{R})$, with wavelet expansion*

$$f = a_{0,0}\phi_{0,0} + \sum_{i=0}^{\infty}\sum_{j=0}^{2^i-1} b_{i,j}\psi_{i,j},$$

define the function $f_{k,p}$, for $p \geq 0$, by

$$f_{k,p} = \sum_{i=0}^{\infty}\sum_{j=0}^{2^i-1} b_{k+i,2^ip+j}\psi_{k+i,2^ip+j}.$$

The action of W on the tree B^f is given in Figure 3.1. Consider the function $v_l^* = (Tf)_{k^*,l}$. By the definition of W, all its wavelet coefficients are equal to 0 except that $B^{v_l^*}_{k^*,l} = \alpha_l B^f_{k,j(l)}$. Therefore,

$$v_l^* = \sum_{i=0}^{\infty}\sum_{n=0}^{2^i-1} b_{k^*+i,2^ij(l)+n}\psi_{k^*+i,2^il+n}.$$

However, notice that

$$f_{k,j(l)} = \sum_{i=0}^{\infty}\sum_{n=0}^{2^i-1} b_{k+i,2^ij(l)+n}\psi_{k+i,2^ij(l)+n},$$

and that

$$b^*_{k^*+k',2^{k'}l+l'} = \alpha_l b_{k+k',2^{k'}j(l)+l'} \quad k' \geq 0, 0 \leq l' \leq 2^{k'} - 1. \tag{3.3}$$

We can thus use the scaling and dilation relations between the $\psi_{i,j}$ to write v_l^* as a multiple of $f_{k,j(l)} \circ w_l^{-1}$ for some appropriate function w_l. The function w_l can be calculated as follows:

$$\begin{aligned}
\psi_{k,j(l)}(x) &= 2^{k/2}\psi(2^k x - j(l)), \\
\psi_{k^*,l}(x) &= 2^{k^*/2}\psi(2^{k^*}x - l) \\
&= 2^{(k^*-k)/2}2^{k/2}\psi(2^k w_l^{-1}(x) - j(l)).
\end{aligned}$$

Equating the arguments of ψ we have

$$w_l^{-1}(x) = 2^{k^*-k}x + \frac{j(l) - l}{2^k}, \tag{3.4}$$

and hence by Equations (3.3) and (3.4),

$$v_l^*(x) = 2^{(k^*-k)/2} f_{k,j(l)}\left(2^{k^*-k}x + \frac{j(l) - l}{2^k}\right).$$

Therefore T is a *recurrent vector IFSM with condensation* (c.f. [11]). By this we mean that T acts between orthogonal components of the wavelet tree and has condensation function

$$\sum_{i=k}^{k^*-1}\sum_{j=0}^{2^i-1} b_{i,j}\psi_{i,j}.$$

A useful space, over which T is contractive, was constructed in [19]. Let $u \in L^2(\mathbb{R})$ and let W be as above. Let

$$C_w(u, k^*) = \{a_{0,0}, b_{i,j}, i \geq 0, 0 \leq j \leq 2^i - 1 : \sum |b_{i,j}|^2 < \infty$$
$$\text{with } b_{i,j} =< u, \psi_{i,j} >, 0 \leq i \leq k^* - 1, 0 \leq j \leq 2^i - 1\}.$$

Consider the metric d_w on C_w by

$$d_w(c, d) = \max_{0 \leq l \leq 2^{k^*}-1} \Delta_l^2,$$

where

$$\Delta_l^2 = \sum_{k'=0}^{\infty} \sum_{l'=0}^{2^{k'}-1} (b^c_{k+k', 2^{k'}l+l'} - b^d_{k+k', 2^{k'}l+l'})^2,$$

where b^c and b^d refer to the wavelet coefficients of c and d respectively. Note that since $k \geq 0$, $d_w(c, d)$ is always independent of $a_{0,0}$. By the completeness of $\ell^2(\mathbb{Z})$ it follows that

Proposition 3.2.3 *The metric space $(C_w(u, k^*), d_w)$ is complete.*

In addition

Proposition 3.2.4 *For $\mathbf{c}, \mathbf{d} \in C_w(u, k^*)$,*

$$d_w(W\mathbf{c}, W\mathbf{d}) \leq c_w d_w(\mathbf{c}, \mathbf{d}) \quad c_w = \max_{0 \leq l \leq 2^{k^*}-1} |\alpha_l|.$$

Therefore, the BCMP yields the following result:

Corollary 3.2.5 *If $c_w < 1$, there exists a unique $\bar{\mathbf{u}} \in C_w(u, k^*)$ such that $W\bar{\mathbf{u}} = \bar{\mathbf{u}}$.*

Corollary 3.2.6 *Let $\epsilon > 0$ and $\mathbf{c} \in C_w(u, k^*)$. Suppose there exists an LIFSW, with* *associated transformation W, such that $d_w(\mathbf{c}, W\mathbf{c}) < \epsilon$. Then*

$$d_w(\mathbf{c}, \bar{\mathbf{u}}) < \frac{\epsilon}{1 - c_w},$$

where $W\bar{\mathbf{u}} = \bar{\mathbf{u}}$.

Proof The result follows directly from Proposition 1.1.21 on page 12. ∎

3.3 Examples of LIFSW

We present here a few examples of LIFSW and their attractors [37, 47].

Example 3.3.1

$$W : B_{00} \mapsto \begin{array}{|c|c|} \hline \multicolumn{2}{|c|}{b_{00}} \\ \hline \alpha_0 B_{00} & \alpha_1 B_{00} \\ \hline \end{array}$$

where $|\alpha_i| < \frac{1}{\sqrt{2}}$. We have $k = 0, k^* = 1, j(0) = j(1) = 0$. Therefore

$$\begin{aligned} w_0^{-1}(x) &= 2x + \frac{0-0}{2^0} = 2x \\ w_1^{-1}(x) &= 2x + \frac{0-1}{2^0} = 2x - 1. \end{aligned}$$

Also, $f = f_0$, hence

$$\begin{aligned} Tf(x) &= b_{0,0}\psi_{0,0}(x) + \sqrt{2}\alpha_0 f_{j(0)} \circ w_0^{-1}(x) + \sqrt{2}\alpha_1 f_{j(1)} \circ w_1^{-1}(x) \\ &= b_{0,0}\psi_{0,0}(x) + \sqrt{2}\alpha_0 f(2x) + \sqrt{2}\alpha_1 f(2x-1). \end{aligned}$$

In general, for $k = 0$, T is an *IFSM with condensation* (see [5] for a discussion on IFS with condensation). The attractors of T, using the Coifman-6, Daubechies-4 and Haar wavelets, are shown in Figure 3.2, where $a_{0,0} = 0$, $b_{0,0} = 1$, $\alpha_1 = 0.2$ and $\alpha_2 = 0.3$. Note the dependence of the attractor on the basis chosen. However, the attractor of W is basis independent.

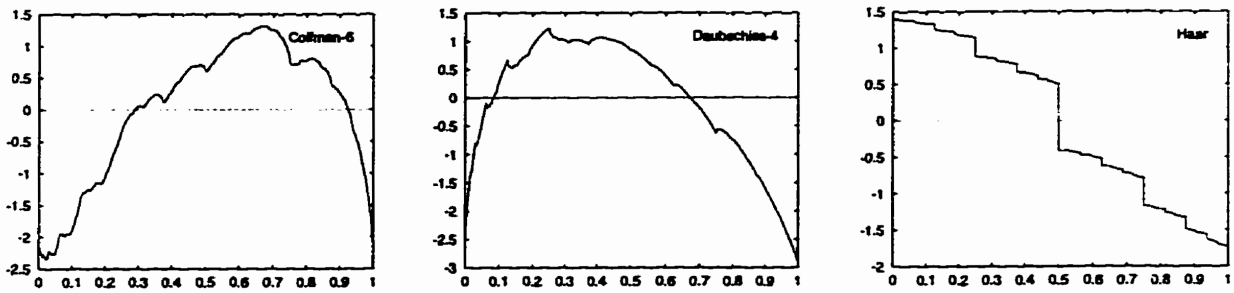


Figure 3.2: The LIFSW attractors of T in Example 3.3.1 using Coifman-6, Daubechies-4 and Haar wavelets.

Example 3.3.2

$W : B_{00} \mapsto$

<table>
<tr><td colspan="4">b_{00}</td></tr>
<tr><td colspan="2">b_{10}</td><td colspan="2">b_{11}</td></tr>
<tr><td>$\alpha_0 B_{10}$</td><td>$\alpha_1 B_{11}$</td><td>$\alpha_2 B_{10}$</td><td>$\alpha_3 B_{11}$</td></tr>
</table>

We have $k = 1, k^* = 2, j(0) = j(2) = 0, j(1) = j(3) = 1$. Therefore

$$Tf(x) = b_{00}\psi_{00}(x) + v_0 + v_1$$

where

$$v_0(x) = b_{10}\psi_{10}(x) + \sqrt{2}\alpha_0 f_{0,0}(2x) + \sqrt{2}\alpha_1 f_{0,1}(2x)$$

and

$$v_1(x) = b_{11}\psi_{11}(x) + \sqrt{2}\alpha_2 f_{0,1}(2x - 1/2) + \sqrt{2}\alpha_3 f_{0,0}(2x - 3/2).$$

The attractors of T are given in Figure 3.3, where $a_{0,0} = 0$, $b_{0,0} = b_{1,0} = 1$, $b_{0,1} = 0.1$ and $\alpha_i = 0.5$, for $i = 0, 1, 2, 3$.

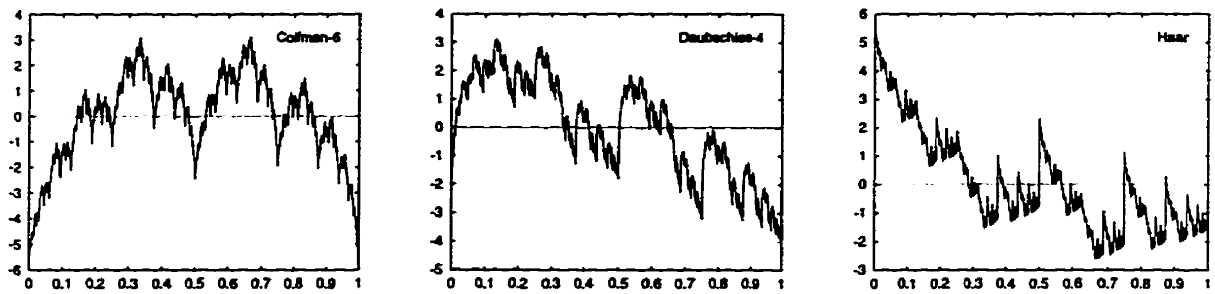


Figure 3.3: The attractors of T in Example 3.3.2 using, from left to right, Coifman-6, Daubechies-4 and Haar wavelets.

Example 3.3.3

$W : B_{00} \mapsto$

<table>
<tr><td colspan="4">b_{00}</td></tr>
<tr><td colspan="2">b_{10}</td><td colspan="2">b_{11}</td></tr>
<tr><td>$\alpha_0 B_{10}$</td><td>$\alpha_1 B_{11}$</td><td>$\alpha_2 B_{11}$</td><td>$\alpha_3 B_{10}$</td></tr>
</table>

In this case

$$Tf = b_{00}\psi_{00}(x) + v_0 + v_1$$

where

$$v_0(x) = b_{10}\psi_{10}(x) + \sqrt{2}\alpha_0 f_{0,0}(2x) + \sqrt{2}\alpha_1 f_{0,1}(2x)$$

and

$$v_1(x) = b_{11}\psi_{11}(x) + \sqrt{2}\alpha_2 f_{0,0}(2x-1) + \sqrt{2}\alpha_3 f_{0,1}(2x-1).$$

The LIFSW attractors of T are given in Figure 3.4 with the same parameters as in Example 3.3.2 above.

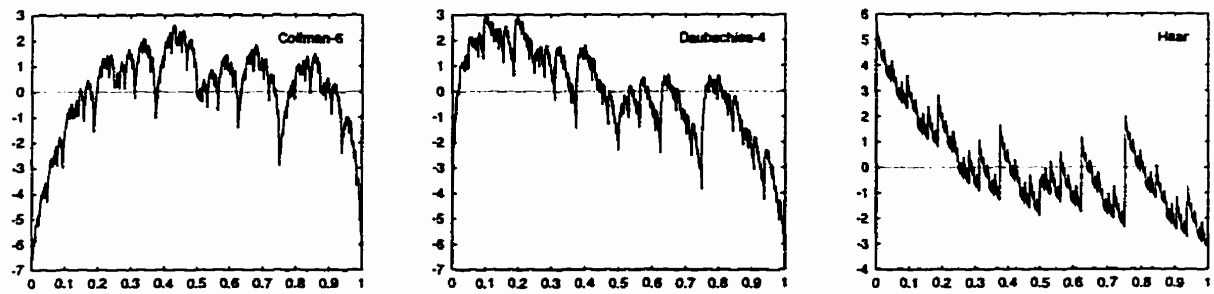


Figure 3.4: The attractors of T using, from left to right, Coifman-6, Daubechies-4 and Haar wavelets.

3.4 Inverse Problem and Compression

Given a target function $v \in L^2(\mathbb{R})$, we can use Corollary 3.2.6 to construct an LIFSW on its coefficient tree [19]. The squared L^2 distance associated with each range block $B^*_{k^*,l}$ and domain block $B_{k,j}$ is given by

$$\Delta^2_{l,j} = \sum_{k'=0}^{\infty} \sum_{l'=0}^{2^{k'}-1} \left(c_{k^*+k',2^{k'}l+l'} - \alpha_l c_{k+k',2^{k'}j+l'}\right)^2.$$

The optimal scaling factor $\bar{\alpha}_{l,j}$ given by the least square minimization is

$$\bar{\alpha}_{l,j} = \frac{S_{k^*,l,k,j}}{S_{k,j,k,j}}$$

where

$$S_{a,b,c,d} = \sum_{k'=0}^{\infty} \sum_{l'=0}^{2^{k'}-1} c_{a+k',2^{k'}b+l'} c_{c+k',2^{k'}d+l'}.$$

The minimized collage distance is then

$$\Delta^{\min}_{l,j} = [S_{k^*,l,k^*,l} - \bar{\alpha}_{l,j} S_{k^*,l,k,j}]^{1/2}.$$

Thus, as with LIFSM, for each range block $B^*_{k^*,l}$, choose the domain block $B_{k,j(l)}$ for which $\Delta^{\min}_{l,j(l)}$ is minimized. Then, iterate the associated operator W on any initial $c \in C_w(v, k^*)$. For simplicity, one can let c be the sequence with $c_{i,j} = 0$ for all $i \geq k^*$. The function $\bar{u}$ associated to the fixed point $\bar{\mathbf{u}}$ of W is then given by

$$\bar{u} = \bar{a}_{0,0}\phi + \sum_{i=0}^{\infty}\sum_{j=0}^{2^i-1} \bar{b}_{i,j}\psi_{i,j}.$$

To apply this method to compression, assume we are given a discrete signal f consisting of 2^M samples on $[0, 1]$ at the dyadics $x = 2^{-M}n, 0 \leq n < 2^M$. By Theorem 2.7.1 we assume these values are the scaling coefficients of f, that is

$$f(2^{-M}n) = 2^{m/2}a_{m,n}.$$

Then assume f has period one as in Section 2.7, page 102 and generate the wavelet coefficients $(b_{m,n})$ for $m = 0, 1, \ldots, M, 0 \leq n \leq 2^m - 1$. Choose a level k in the tree for the domain blocks, and k^* for the range blocks. For each range block, calculate the distances $\Delta^{\min}_{l,j}$ for each $0 \leq j \leq 2^k - 1$. Choose $j(l)$ to be the index of the domain block for which $\Delta^{\min}_{l,j}$ is minimized over $0 \leq j \leq 2^k - 1$. The LIFSW approximation to the target will then consist of the set of coefficients $\{a_{0,0}, b_{i,j} |\ 0 \leq i \leq k^* - 1, 0 \leq j \leq 2^i - 1\}$, and the set of pairs $\{(\bar{\alpha}_l, j(l)) |\ 0 \leq l \leq 2^{k^*} - 1\}$. These two sets are called the *Fractal Wavelet Transform* of f.

To obtain the approximation to the original signal, iterate W on any initial tree, such as c given above. Use the reconstruction algorithm of the Mallat algorithm to construct the scaling coefficients $(a^*_{M,n})$, $0 \leq n \leq 2^M - 1$. Finally, use these coefficients as the values of the approximation f^* by

$$f^*(2^{-M}n) = 2^{m/2}a^*_{M,n}.$$

Example 3.4.1 Let $u(x) = \sin(\pi x)$ on $X = [0,1]$. Figure 3.5 shows successive approximations of u using the LIFSW method with $M = 10$, hence 1024 samples. The following table gives the L^2 error in the approximations. In each case, the computation time involved was approximately 1 second. The original file for u was 18335 bytes.

Domain level (k)	Range level (k^*)	Error	File size (bytes)
0	1	0.07836540	42
0	2	0.03548420	88
0	3	0.01179280	188
0	4	0.00401203	396
0	5	0.00140071	845
1	2	0.00972692	84
1	4	0.00102580	374
2	5	0.00024186	788

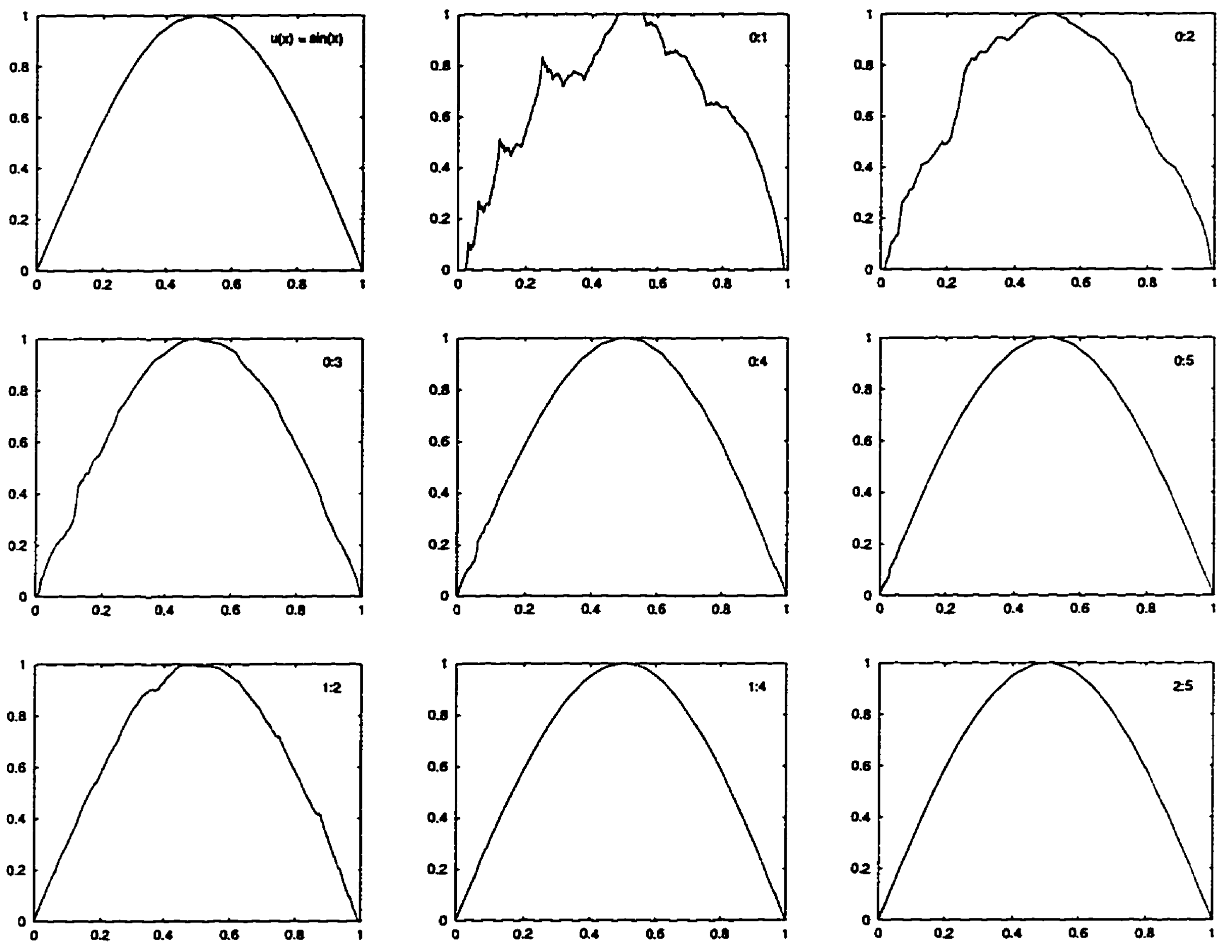


Figure 3.5: LIFSW approximation of $u(x) = \sin(\pi x)$ (top-left) using Coifman-6 wavelets going between levels (k, k^*).

Appendix A

A.1 Generalization of $T_{\mathbf{w}}^{BW}$

Given an N-map IFSM $(\mathbf{w}, \Phi)$, the IFSM operator $T_{(\mathbf{w},\Phi)}$ was defined by

$$T_{(\mathbf{w},\Phi)}u(x) = \sum_{k=1}^{N}{}' \phi_k(u(w_k^{-1}(x))) \quad x \in X.$$

This is a useful form since it permits a simple formulation and solution to the inverse problem. Unfortunately, this is not a generalization of $T_{\mathbf{w}}^{BW}$. This can be seen by noting that, given an N-map IFS $\mathbf{w}$, we may obtain an IFSM $(\mathbf{w}', \Phi')$ by defining $w_i' = w_i$ $\forall i = 1, 2, \ldots, N$ and setting $\phi_i = id_{\mathbb{R}}$ $\forall i = 1, 2, \ldots, N$. Then, for $u \in \mathcal{F}_{BW}(X)$, $\forall i = 1, 2, \ldots, N$, $\phi_i \circ u \circ w_i'^{-1} = u \circ w_i^{-1} \in \mathcal{F}_{BW}(X)$. However $T_{(\mathbf{w}',\Phi')}(u)$ and $T_{\mathbf{w}}^{BW}(u)$ are not necessarily the same. Indeed, $T_{(\mathbf{w}',\Phi')}(u)$ may take on values other than 0 or 1. It would therefore be reasonable to introduce a normalized version of $T_{(\mathbf{w},\Phi)}$, $T_{(\mathbf{w},\Phi)}^{nor}$, such that $T_{(\mathbf{w},\Phi)}^{nor} \restriction \mathcal{F}_{BW}(X) = T_{\mathbf{w}}^{BW}$. With this in mind, we return to the introduction of $T_{\mathbf{w}}^{BW}$. The reasoning behind its construction was that

$$\chi_{A\cup B}(x) = \max\{\chi_A(x), \chi_B(x)\}.$$

Instead, consider $\chi_{A\cup B}(x)$ with $T_{(\mathbf{w},\Phi)}^{nor}$ in mind, as a weighted sum.

Proposition A.1.1 *Let X be a set and $A_i \subset X, i = 1, 2, \ldots, N$. Let $A = \bigcup_{i=1}^{N} A_i$. Define $\sigma_A(x) = \sum_{i=1}^{N} \chi_{A_i}(x)$. Then,*

$$\chi_A(x) = \frac{\sum_{i=1}^{N} \chi_{A_i}(x)}{\sigma_A(x)} \quad \forall x \in X,$$

using the convention that $0 \cdot \infty = 0$.

Given that the numerator and denominator are equal, this result is rather trivial. We are simply dividing the sum by $\sigma_A(x)$, the number of sets A_i which contain x. This number is precisely the numerator. However, we can also interpret the numerator as a sum of grey-level values.

Recalling Proposition 1.5.3 and Proposition 1.5.4, we obtain the following result:

Proposition A.1.2 *Let (X, d) be a metric space and $w_i \in Con_1(X, d), i = 1, 2, \ldots, N$. Then letting $\sigma(x) = \sigma_{\hat{\mathbf{w}}(A)}(x)$, for $x \in X$, we have*

$$\chi_{\hat{\mathbf{w}}(A)}(x) = \frac{\sum_{i=1}^{N} \chi_A(w_i^{-1}(x))}{\sigma(x)} \quad \forall x \in X.$$

Proof The proof follows directly from Proposition 1.5.3 and Proposition 1.5.4. ∎

Hence, we can associate with the IFS $\mathbf{w}$, defined by $\hat{\mathbf{w}}(A) = \bigcup_{i=1}^{N} \hat{w}_i(A)$, the operator $T_{\mathbf{w}}^{BW} : \mathcal{F}_{BW}(X) \to \mathcal{F}_{BW}(X)$ by

$$T_{\mathbf{w}}^{BW} f(x) = \frac{\sum_{i=1}^{N} f(w_i^{-1}(x))}{\sigma(x)} \quad \forall x \in X,$$

where $f(x) = \chi_A(x)$.

This new operator is the same as the one previously defined on $\mathcal{F}_{BW}(X)$, simply written in a different form. As such, it can be directly extended to an IFSM operator as follows:

Definition A.1.3 *Let (X,d) be complete, $w_k \in Con_1(X,d)$ and $\phi_k : \mathbb{R} \to \mathbb{R}$ for $k = 1,2,\dots,N$. Let $(\mathbf{w},\Phi)$ be the associated IFSM. We define the operator $T^{nor}_{(\mathbf{w},\Phi)}$ on a function $u : X \to \mathbb{R}$ by*

$$T^{nor}_{(\mathbf{w},\Phi)}u(x) = \frac{\sum_{k=1}^{N}{}' \phi_k(u(w_k^{-1}(x)))}{\sigma(x)} \quad x \in X.$$

where the $\sum'$ indicates that the sum is taken over indices k for which $w_k^{-1}(x)$ exists. The convention is that an empty sum has value zero and that $0 \cdot \infty = 0$.

We call $T^{nor}_{(\mathbf{w},\Phi)}$ the IFSM operator associated with $(\mathbf{w},\Phi)$.

A.2 IFSM on $L^p(X,\mu)$

We will show that $T^{nor}_{(\mathbf{w},\Phi)}$ has the same properties as the operator $T_{(\mathbf{w},\Phi)}$ defined in Section 1.5 on page 35. Note that $T^{nor}_{(\mathbf{w},\Phi)} = T_{(\mathbf{w},\Phi)}$ when the set X_i are non-overlapping.

Let $(\mathbf{w},\Phi)$ denote the IFSM, on the complete metric space (X,d), associated to $\mathbf{w} = \{w_1, w_2, \dots, w_N\}$, $w_i \in Con_1(X,d)$ and $\Phi = \{\phi_1, \phi_2, \dots, \phi_N\}$, $\phi_i : \mathbb{R} \to \mathbb{R}$. The important step is to show its associated operator $T^{nor}_{(\mathbf{w},\Phi)} : L^p(X,\mu) \to L^p(X,\mu)$. Hence, for $u \in L^p(X,\mu)$, we must show $T^{nor}_{(\mathbf{w},\Phi)}u$ is still measurable and integrable. To accomplish this, we make the following definition:

Definition A.2.1 *Given an IFS $\mathbf{w} = \{w_1, w_2, \dots, w_N\}$, let*

$$A_1 = X \backslash w_1(X) \text{ and let } A_2 = w_1(X).$$

Recursively, for $2 \leq n \leq N-1$, *let*

$$A_{i_1 i_2 \dots i_n 1} = A_{i_1 i_2 \dots i_n} \setminus w_{n+1}(X)$$

and let

$$A_{i_1 i_2 \dots i_n 2} = A_{i_1 i_2 \dots i_n} \cap w_{n+1}(X),$$

for $i_j = 1, 2$ *and* $j = 1, 2, \dots, n$. *We define*

$$\mathcal{A}_n = \{A_{i_1 i_2 \dots i_n} : i_j = 1, 2; j = 1, 2, \dots, n\}.$$

Finally, set $\mathcal{A}_{\mathbf{w}} = \mathcal{A}_N$. *We call* $\mathcal{A}_{\mathbf{w}}$ *the* **w**-cover *of* X.

Proposition A.2.2 $\mathcal{A}_{\mathbf{w}}$ *is a collection of disjoint sets.*

Proof We proceed by induction on n, with $n=1$ being clear. For $n > 1$, assume $\{A_{i_1 i_2 \dots i_n} : i_j = 1, 2; j = 1, 2, \dots, n\}$ is a disjoint collection. Let $A = A_{i_1 i_2 \dots i_n i_{n+1}}$ and let $B = A_{j_1 j_2 \dots j_n j_{n+1}} \in \mathcal{A}_{n+1}$. By definition, $A_{k_1 k_2 \dots k_n k_{n+1}} \subset A_{k_1 k_2 \dots k_n}$, hence if for some $1 \leq m \leq n, i_m \neq j_m$, then by the induction hypothesis, $A_{i_1 i_2 \dots i_n} \cap A_{j_1 j_2 \dots j_n} = \emptyset$, and $A \cap B = \emptyset$. Otherwise, $i_{n+1} \neq j_{n+1}$ and A and B are disjoint by construction. Therefore, by induction, $\mathcal{A}_n$ is a disjoint collection and the result follows. ■

We note certain characteristics of $\mathcal{A}_{\mathbf{w}}$.

a) Some members of $\mathcal{A}_{\mathbf{w}}$ may be empty.

b) $\sigma_{\hat{\mathbf{w}}(X)}$ is constant on each $A_{i_1 i_2 \dots i_N}$. Define $\sigma_{\hat{\mathbf{w}}(X)}(x) \equiv c_{i_1 i_2 \dots i_N}$ for $x \in A_{i_1 i_2 \dots i_N}$.

c) If (X, μ) is a measure space with σ-algebra generated by open sets, and the w_k are bi-continuous, then each member of $\mathcal{A}_{\mathbf{w}}$ is measurable.

d) $\int_X f d\mu = \sum_{A \in \mathcal{A}_{\mathbf{w}}} \int_A f d\mu$.

We therefore obtain a version of Proposition 1.6.1.

Proposition A.2.3 *Let $(\mathbf{w}, \Phi)$ be an N-map IFSM on (X, d) and let T be its associated normalized IFSM operator. Suppose:*

i) $\forall u \in L^p(X, \mu), u \circ w_k^{-1} \in L^p(X, \mu), 1 \leq k \leq N$ *and*

ii) $\phi_k \in Lip(\mathbb{R}), 1 \leq k \leq N$.

Then for $1 \leq p \leq \infty, T : L^p(X, \mu) \to L^p(X, \mu)$.

Proof Let $1 \leq p \leq \infty$ and let $u \in L^p(X, \mu)$. Let $1 \leq k \leq N$. By *i)*, $u \circ w_k^{-1} \in L^p(X, \mu)$, hence by *ii)*, $\phi_k \circ u \circ w_k^{-1} \in L^p(X, \mu)$. Therefore $\sum_{k=1}^{N}{}' \phi_k \circ u \circ w_k^{-1} \in L^p(X, \mu)$. By b) and c) above, $Tu|_{A_{i_1 i_2 \dots i_N}}$ is measurable for each $i_j = 1, 2$ and $j = 1, 2, \dots, N$. Hence, Tu is measurable.

Suppose $1 \leq p < \infty$. Then, by c) and d),

$$\int_X |Tu(x)|^p \, d\mu(x) = \sum_{i_1, i_2, \dots, i_N = 1}^{2} \frac{1}{C_{i_1 i_2 \dots i_N}^p} \int_{A_{i_1 i_2 \dots i_N}} \left| \sum_{k=1}^{N}{}' \phi_k(u(w_k^{-1}(x))) \right|^p d\mu(x).$$

The last integrals exist since $\sum_{k=1}^{N}{}' \phi_k \circ u \circ w_k^{-1} \in L^p(X, \mu)$ and since each $A_{i_1 i_2 \dots i_N}$ is measurable. Hence, $Tu \in L^p(X, \mu)$.

Suppose $p = \infty$. By *ii)*, $\sum_{k=1}^{N}{}' \phi_k \circ u \circ w_k^{-1} \in L^\infty(X, \mu)$, $1 \leq k \leq N$. Hence by b), $Tu|_{A_{i_1 i_2 \dots i_N}} \in L^\infty(X, \mu)$. Therefore, since Tu is a finite sum of $L^\infty(X, \mu)$ functions, $Tu \in L^\infty(X, \mu)$. ∎

Proposition A.2.4 *Let $(\mathbf{w}, \Phi)$ be an N-map IFSM such that $\phi_k(t) = c_k \in \mathbb{R}, 1 \leq k \leq N$. Then $\forall p \in [1, \infty)$ and $\mu \in M(X)$, the associated operator T is contractive on $L^p(X, \mu)$, with contractivity factor $c_T = 0$. Furthermore, its fixed point $\bar{u}_T$ is*

$$\bar{u}_T(x) = \frac{\sum_{k=1}^{N} c_k \chi_{\tilde{w}_k(X)}(x)}{\sigma(x)} \quad \forall x \in X.$$

Proof Let $u, v \in L^p(X,\mu)$. Then

$$\begin{aligned}
||Tu - Tv||_p &= \left[\int_X \left| \left[\sum_{k=1}^{N}{}' \phi_k(u(w_k^{-1}(x))) - \phi_k(v(w_k^{-1}(x))) \right] \sigma(x)^{-1} \right|^p d\mu(x) \right]^{1/p} \\
&\leq \left[\int_X \left| \sum_{k=1}^{N}{}' \phi_k(u(w_k^{-1}(x))) - \phi_k(v(w_k^{-1}(x))) \right|^p d\mu(x) \right]^{1/p} \\
&= 0 \qquad \text{as in Proposition 1.6.2.}
\end{aligned}$$

Also, $\forall u \in L^p(X,\mu)$ and $x \in X$,

$$\begin{aligned}
Tu(x) &= \frac{\sum_{k=1}^{N}{}' \phi_k(u(w_k^{-1}(x)))}{\sigma(x)} \\
&= \frac{\sum_{k=1}^{N} c_k \chi_{\hat{w}_k(X)}(x)}{\sigma(x)}. \qquad \blacksquare
\end{aligned}$$

Proposition A.2.5 *Let $X \subset \mathbb{R}^D, D \in \mathbb{N}^+$, and let $\mu = m^{(D)}$. Suppose $(\mathbf{w}, \Phi)$ is an N-map IFSM such that*

i) $w_k \in Sim_1(X, d)$ with contractivity factors c_k, and

ii) $\phi_k \in Lip(\mathbb{R})$, with Lipschitz constants K_k, for $1 \leq k \leq N$.

Then, for $p \in [1,\infty)$ and $u, v \in L^p(X,\mu)$, we have

$$||Tu - Tv||_p \leq C(D,p)||u - v||_p,$$

where $C(D,p) = \sum_{k=1}^{N} c_k^{D/p} K_k$.

Proof Let $u, v \in L^p(X,\mu)$. Use the fact that $\sigma(x) \geq 1$ and the proof follows as in Proposition 1.6.3 on page 37. $\blacksquare$

Hence, if $C(D, p) < 1$, T is contractive on $L^p(X, \mu)$ and has a unique, attracting fixed point. In addition, if there is a lot of overlapping in the w_i, for example if each $x \in \hat{\mathbf{w}}(X)$ is in at least two $w_k(X)$, then $\sigma(x) > 1$. In this case, the new operator T will be more contractive than the previous one, leading to faster convergence to its fixed point.

Bibliography

[1] R. A. Adams. *Sobolev Spaces.* Pure and applied mathematics. Academic Press, 1975.

[2] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[3] R. B. Ash. *Measure, Integration, and Functional Analysis.* Academic Press, New York, 1972.

[4] G. Bachman and L. Narici. *Functional Analysis.* Academic Press, 1966.

[5] M. F. Barnsley. *Fractals Everywhere.* Academic Press, Boston, 2nd edition, 1993.

[6] M. F. Barnsley and S. Demko. Iterated function systems and the global construction of fractals. *Proc. R. Soc. Lond. A*, 399:243–275, 1985.

[7] J. J. Benedetto and M. W. Frazier, editors. *Wavelets: Mathematics and Applications.* Studies in Advanced Mathematics. CRC Press, 1994.

[8] B. Bollobás. *Linear Analysis: An Introductory Course.* Cambridge University Press, 1992.

[9] J. Burke. *Connections.* Macmillan, London, 1978.

[10] J. Burke. *The axemaker's gift: a double-edged history of human culture.* G.P. Putnam's Sons, 1995.

[11] C. Cabrelli, U. Molter and E. R. Vrscay. Recurrent iterated function systems: Invariant measures, a collage theorem and moment relations. In H.-O. Peitgen, J. M. Henriques and L. F. Penedo, editors, *Fractals in the Fundamental and Applied Sciences*, pages 71–80, North-Holland, Amsterdam, 1991.

[12] C. Canuto and A. Quarteroni. Approximation results for orthogonal polynomials in Sobolev spaces. *Math. Comp.*, 38:67–86, 1982.

[13] P. M. Centore and E. R. Vrscay. Continuity of attractors and invariant measures for iterated function systems. *Canadian Mathematical Bulletin*, 37(3):315–329, 1994.

[14] J. B. Conway. *A Course in Functional Analysis.* Springer-Verlag, New York, 1985.

[15] I. Daubechies. *Ten Lectures on Wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, 1992.

[16] N. Dunford and J. T. Schwartz. *Linear Operators Part I: General Theory.* John Wiley and Sons, New York, 3rd edition, 1996.

[17] K. J. Falconer. *The Geometry of Fractal Sets.* Cambridge University Press, Cambridge, U. K., 1985.

[18] Y. Fisher, editor. *Fractal Image Compression.* Springer-Verlag, 1995.

[19] B. Forte and E. R. Vrscay. Inverse problem methods for generalized fractal transforms. In Y. Fisher, editor, *Fractal Image Encoding and Analysis*, Proceedings of the NATO Advanced Study Institute, Trodheim, Norway, July 8-17, 1995.

[20] B. Forte and E. R. Vrscay. Solving the inverse problem for function and image approximation using iterated function systems. *Dynamics of Continuous and Impulsive Systems*, 1:177–231, 1995.

[21] B. Forte and E. R. Vrscay. Theory of generalized fractal transforms. In Y. Fisher, editor, *Fractal Image Encoding and Analysis*, Proceedings of the NATO Advanced Study Institute, Trodheim, Norway, July 8-17, 1995.

[22] M. Ghazel. *Fractal, Function and Image Approximation Using Iterated Transformation Systems.* PhD thesis, Department of Applied Mathematics, University of Waterloo, 1994.

[23] D. Gulick. *Encounters with Chaos.* McGraw Hill, 1992.

[24] C. Handy and G. Mantica. Inverse problems in fractal construction: Moment method solution. *Physica*, D43:17–36, 1990.

[25] C. E. Heil and D. F. Walnut. Continuous and discrete wavelet transforms. *SIAM Review*, 31:628–666, 1989.

[26] J. E. Hutchinson. Fractals and self-similarity. *Indiana University Journal of Mathematics*, 30:713–747, 1981.

[27] J. J. Hwang and K. R. Rao. *Techniques and Standards for Image, Video and Audio Coding.* Prentice Hall, 1996.

[28] R. H. Kasriel. *Undergraduate Topology*. Saunders, 1986.

[29] G. Lemarié and Y. Meyer. Ondelettes et bases Hilbertiennes. *Revista Matematica Iberoamericana*, 2:1–18, 1986.

[30] G. B. Lewellen. Self-similarity. *Rocky Mountain Journal of Mathematics*, 23(3):1023–1040, Summer 1993.

[31] W. R. Madych. Multiresolution analyses, tiles, and scaling functions. In *Probabilistic and Stochastic Methods in Analysis, with Applications*, pages 233–243. Kluwer Academic Publishers, 1992.

[32] S. Mallat. Multiresolution approximation and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. AMS*, 315:69–88, 1989.

[33] B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Company, New York, 1977.

[34] J. T. Marti. *Introduction to Sobolev Spaces and Finite Element Solution of Elliptic Boundary Value Problems*. Computational mathematics and applications. Academic Press, 1986.

[35] P. R. Massopust. *Fractal Functions, Fractal Surfaces, and Wavelets*. Academic Press, San Diego, 1994.

[36] F. Mendivil. A generalization of IFS with probabilities to infinitely many maps. To appear in Rocky Mountain Journal of Mathematics, 1996.

[37] F. Mendivil and E. R. Vrscay. Correspondence between fractal-wavelet transforms and iterated functions systems with grey level maps. To appear in the Proceedings of Fractals in Engineering Conference, Arcachon, France, June 25-27, 1997.

[38] Y. Meyer. *Ondelettes et opérateurs I.* Herman, Paris, 1990.

[39] A. Mukherjea and K. Pothoven. *Real and Functional Analysis Part A: Real Analysis.* Plenum Press, New York, 2nd edition, 1984.

[40] A. Mukherjea and K. Pothoven. *Real and Functional Analysis Part B: Functional Analysis.* Plenum Press, New York, 2nd edition, 1984.

[41] W. B. Pennebaker and J. L. Mitchell. *JPEG Still Image Data Compression Standard.* Van Nostrand Reinhold, New York, 1993.

[42] H. L. Royden. *Real Analysis.* Macmillan, 3rd edition, 1988.

[43] W. Rudin. *Fourier Analysis on Groups.* Interscience Publishers, New York, 1962.

[44] W. Rudin. *Functional Analysis.* McGraw-Hill, 1973.

[45] R. Strichartz. *A Guide to Distribution Theory and Fourier Transforms.* Studies in Advanced Mathematics. CRC Press, 1994.

[46] H. E. Taylor and T. L. Wade. *University Calculus.* John Wiley & Sons, 1962.

[47] E. R. Vrscay. A new class of fractal-wavelet transforms for image representation and compression. To appear in the Proceedings of ImageTech '97 Conference on Multimedia Imaging Technology and Applications, Atlanta, GA, April 13-16, 1997.

[48] E. R. Vrscay. Moment and collage methods for the inverse problem of fractal construction with iterated function systems. In H.-O. Peitgen, J. M. Henriques and L. F. Penedo, editors, *Fractals in the Fundamental and Applied Sciences*, pages 443–461. North-Holland, Amsterdam, 1991.

[49] E. R. Vrscay. Iterated function systems: Theory, applications and the inverse problem, Spring 1995. A course in Fractal Image Compression, University of Waterloo.

[50] E. R. Vrscay. Mathematical theory of generalized fractal transforms and associated inverse problems. In *ImageTech Conference on the Mathematics of Imaging and its Applications*, March 1996.

[51] G. G. Walter. *Wavelets and Other Orthogonal Systems with Applications*. CRC Press, 1994.

[52] H. L. Weinert, editor. *Reproducing Kernel Hilbert Spaces*. Benchmark papers in electrical engineering and computer science. Hutchinson Ross, 1982.

[53] R. L. Wheeden and A. Zygmund. *Measure and Integral: An Introduction to Real Analysis*. Marcel Dekker, Inc., 1977.

[54] S. Willard. *General Topology*. Addison-Wesley, 1970.

[55] A. Zygmund. *Trigonometric Series*, volume I. Cambridge University Press, 1959.

[56] A. Zygmund. *Trigonometric Series*, volume II. Cambridge University Press, 1959.

Glossary

$((X,d),w,\Lambda)$, **14**
$(C_w(u,k^*),d_w)$, **115**
$(\mathcal{H}(X),h)$, **18**
$(H,<\cdot,\cdot>)$, **53**
(X,d), **4**
$(\mathbf{w},\Phi)$, **34**
$(\mathbf{w}_{loc},\Phi)$, **48**
$(\mathcal{F}_{BW}(X),d_{BW})$, **33**
$(x_n)\to x$, **4**
$<\cdot,\cdot>$, **53**
$2\nearrow$, **98**
$2\searrow$, **98**

$\mathcal{A}$, **15**
$\mathcal{A}_{\mathbf{w}}$, **126**
$\overline{\mathcal{A}}$, **54**
$A_{\tilde{\mathbf{w}}}$, **25**
$a_{m,n}$, **102**
$A^{\perp_B}$, **66**

$B(X)$, **36**
b^c, **115**
B^f, **112**
$B_{i,j}$, **112**
$b_{m,n}$, **102**

$\mathcal{C}$, **13**
$C(D,p)$, **38**, **128**
c_f, **5**
χ_A, **30**
$C^\infty(\mathbb{R})$, **59**
$C_{loc}(D,p)$, **49**
$c_n^{0,0}$, **95**
$Con(\mathcal{H}(X),h,s)$, 24
$Con(X,d)$, **5**
$Con(X,d,s)$, **24**
$Con_1(X,d)$, **29**
$C^r(\mathbb{R})$, 61
$C_w(u,k^*)$, **115**

D, **61**
$\bar{d}(f,g)$, **9**
d, **4**
$d(A,B)$, **17**
$d_\Phi^N(\Phi_1,\Phi_2)$, **38**
$d(x,B)$, **17**
δ_α, **68**
$d_{BW}(u,v)$, **33**
Δ^2, **42**
Δ_l^2, **115**
$\Delta_{l,j}^2$, **119**
$\Delta_{l,j}^{\min}$, **119**
$\delta_m(\cdot,y)$, **68**
δ_{mn}, **71**
$detA$, 29
∂_x^α, **76**
$D_m(\cdot,y)$, **69**
$d_m(f,g)$, **9**
D_v, **51**
$d_w(c,d)$, **115**

e_n^{0*}, **95**

F, **10**

$T_{(\mathbf{w},\Phi)}$, **35**, 123
$T_{\mathbf{w}}^{BW}$, **32**, 123, **124**
$\theta^{(k)}$, **59**
$T_{(\mathbf{w},\Phi)}^{loc}$, **48**
$T_{(\mathbf{w},\Phi)}^{\max}$, **34**
$T_{(\mathbf{w},\Phi)}^{nor}$, **125**

$\mathbf{u}$, **108**
$\bar{\mathbf{u}}_M$, **108**
$\bar{u}_T$, 107

V_m, **56**, **61**, 71

W, **112**
$\mathcal{W}$, **40**
$w(\lambda, x)$, **14**
$\mathbf{w}$, **16**
$\hat{\mathbf{w}}$, **25**
$\hat{w}$, 24
$\mathbf{w}_{loc}$, **48**
w_n, **15**

X, **4**
$\bar{x}_f$, **9**
(x_n), **4**
$(x_n)_{n\in A}$, **4**
$x \vee y$, **17**

$\mathbb{Z}$, **4**
Z_λ, 79
$Z\phi$, **79**

Abbreviations

Index

Bold page numbers indicate a definition or major reference.