

# COMPARISON OF DISCRETE AND CONTINUOUS WAVELET TRANSFORMS

PALLE E.T. JORGENSEN AND MYUNG-SIN SONG

## ARTICLE OUTLINE

Glossary	1
1. Definition	8
2. Introduction	9
3. The discrete vs continuous wavelet Algorithms	9
3.1. The Discrete Wavelet Transform	9
3.2. The Continuous Wavelet Transform	13
3.3. Some background on Hilbert space	13
3.4. Connections to group theory	17
4. List of names and discoveries	19
5. History	21
6. Tools from Mathematics	21
7. A Transfer Operator	22
8. Future Directions	23
9. Literature	24
10. Bibliography	24

## GLOSSARY

This *glossary* consists of a list of terms used inside the paper: In mathematics, in probability, in engineering, and on occasion in physics. To clarify the seemingly confusing use of up to four different names for the same idea or concept, we have further added informal explanations spelling out the reasons behind the differences in current terminology from neighboring fields.

DISCLAIMER: This glossary has the structure of four columns. A number of terms are listed line by line, and each line is followed by explanation. Some “terms” have up to four separate (yet commonly accepted) names.

MATHEMATICS	PROBABILITY	ENGINEERING	PHYSICS
<b>function</b> (measurable)	<b>random variable</b>	<b>signal</b>	<b>state</b>
Mathematically, functions may map between any two sets, say, from $X$ to $Y$ ; but if $X$ is a probability space (typically called $\Omega$ ), it comes with a $\sigma$ -algebra $\mathcal{B}$ of measurable sets, and probability measure $P$ . Elements $E$ in $\mathcal{B}$ are called events, and $P(E)$			

---

Work supported in part by the U.S. National Science Foundation.

---

MATHEMATICS      PROBABILITY      ENGINEERING      PHYSICS

---

the probability of  $E$ . Corresponding measurable functions with values in a vector space are called random variables, a terminology which suggests a stochastic viewpoint. The function values of a random variable may represent the outcomes of an experiment, for example “throwing of a die.”

Yet, function theory is widely used in engineering where functions are typically thought of as signal. In this case,  $X$  may be the real line for time, or  $\mathbb{R}^d$ . Engineers visualize functions as signals. A particular signal may have a stochastic component, and this feature simply introduces an extra stochastic variable into the “signal,” for example noise.

Turning to physics, in our present application, the physical functions will be typically be in some  $L^2$ -space, and  $L^2$ -functions with unit norm represent quantum mechanical “states.”

**sequence (incl. random walk      time-series      measurement vector-valued)**

Mathematically, a sequence is a function defined on the integers  $\mathbb{Z}$  or on subsets of  $\mathbb{Z}$ , for example the natural numbers  $\mathbb{N}$ . Hence, if time is discrete, this to the engineer represents a time series, such as a speech signal, or any measurement which depends on time. But we will also allow functions on lattices such as  $\mathbb{Z}^d$ .

In the case  $d = 2$ , we may be considering the grayscale numbers which represent exposure in a digital camera. In this case, the function (grayscale) is defined on a subset of  $\mathbb{Z}^2$ , and is then simply a matrix.

A random walk on  $\mathbb{Z}^d$  is an assignment of a sequential and random motion as a function of time. The randomness presupposes assigned probabilities. But we will use the term “random walk” also in connection with random walks on combinatorial trees.

**nested refinement      multiresolution      scales of visual subspaces      resolutions**

While finite or infinite families of nested subspaces are ubiquitous in mathematics, and have been popular in Hilbert space theory for generations (at least since the 1930s), this idea was revived in a different guise in 1986 by Stéphane Mallat, then an engineering graduate student. In its adaptation to wavelets, the idea is now referred to as the multiresolution method.

What made the idea especially popular in the wavelet community was that it offered a skeleton on which various discrete algorithms in applied mathematics could be attached and turned into wavelet constructions in harmonic analysis. In fact what we now call multiresolutions have come to signify a crucial link between the world of discrete wavelet algorithms, which are popular in computational mathematics and in engineering (signal/image

processing, data mining, etc.) on the one side, and on the other side continuous wavelet bases in function spaces, especially in  $L^2(\mathbb{R}^d)$ . Further, the multiresolution idea closely mimics how fractals are analyzed with the use of finite function systems.

But in mathematics, or more precisely in operator theory, the underlying idea dates back to work of John von Neumann, Norbert Wiener, and Herman Wold, where nested and closed subspaces in Hilbert space were used extensively in an axiomatic approach to stationary processes, especially for time series. Wold proved that any (stationary) time series can be decomposed into two different parts: The first (deterministic) part can be exactly described by a linear combination of its own past, while the second part is the opposite extreme; it is *unitary*, in the language of von Neumann.

von Neumann's version of the same theorem is a pillar in operator theory. It states that every isometry in a Hilbert space  $\mathcal{H}$  is the unique sum of a shift isometry and a unitary operator, i.e., the initial Hilbert space  $\mathcal{H}$  splits canonically as an orthogonal sum of two subspaces  $\mathcal{H}_s$  and  $\mathcal{H}_u$  in  $\mathcal{H}$ , one which carries the shift operator, and the other  $\mathcal{H}_u$  the unitary part. The shift isometry is defined from a nested scale of closed spaces  $V_n$ , such that the intersection of these spaces is  $\mathcal{H}_u$ . Specifically,

$$\begin{aligned} \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_n \subset V_{n+1} \subset \cdots \\ \bigwedge_n V_n = \mathcal{H}_u, \text{ and } \bigvee_n V_n = \mathcal{H}. \end{aligned}$$

However, Stéphane Mallat was motivated instead by the notion of scales of resolutions in the sense of optics. This in turn is based on a certain "artificial-intelligence" approach to vision and optics, developed earlier by David Marr at MIT, an approach which imitates the mechanism of vision in the human eye.

The connection from these developments in the 1980s back to von Neumann is this: Each of the closed subspaces  $V_n$  corresponds to a level of resolution in such a way that a larger subspace represents a finer resolution. Resolutions are relative, not absolute! In this view, the relative complement of the smaller (or coarser) subspace in larger space then represents the visual detail which is added in passing from a blurred image to a finer one, i.e., to a finer visual resolution.

This view became an instant hit in the wavelet community, as it offered a repository for the fundamental father and the mother functions, also called the scaling function  $\varphi$ , and the wavelet function  $\psi$ . Via a system of translation and scaling operators, these functions then generate nested subspaces, and we recover the scaling identities which initialize the appropriate algorithms. What results is now called the family of pyramid algorithms in wavelet analysis. The approach itself is called the multiresolution approach (MRA) to wavelets. And in the meantime various generalizations (GMRA) have emerged.

MATHEMATICS      PROBABILITY      ENGINEERING      PHYSICS

---

In all of this, there was a second “accident” at play: As it turned out, pyramid algorithms in wavelet analysis now lend themselves via multiresolutions, or nested scales of closed subspaces, to an analysis based on frequency bands. Here we refer to bands of frequencies as they have already been used for a long time in signal processing.

One reason for the success in varied disciplines of the same geometric idea is perhaps that it is closely modeled on how we historically have represented numbers in the positional number system. Analogies to the Euclidean algorithm seem especially compelling.

<b>operator</b>	<b>process</b>	<b>black box</b>	<b>observable (if selfadjoint)</b>
-----------------	----------------	------------------	--

In linear algebra students are familiar with the distinctions between (linear) transformations  $T$  (here called “operators”) and matrices. For a fixed operator  $T: V \rightarrow W$ , there is a variety of matrices, one for each choice of basis in  $V$  and in  $W$ . In many engineering applications, the transformations are not restricted to be linear, but instead represent some experiment (“black box,” in Norbert Wiener’s terminology), one with an input and an output, usually functions of time. The input could be an external voltage function, the black box an electric circuit, and the output the resulting voltage in the circuit. (The output is a solution to a differential equation.)

This context is somewhat different from that of quantum mechanical (QM) operators  $T: V \rightarrow V$  where  $V$  is a Hilbert space. In QM, selfadjoint operators represent observables such as position  $Q$  and momentum  $P$ , or time and energy.

<b>Fourier dual pair</b>	<b>generating function</b>	<b>time/frequency</b>	$P/Q$
------------------------------	--------------------------------	-----------------------	-------

The following dual pairs position  $Q$ /momentum  $P$ , and time/energy may be computed with the use of Fourier series or Fourier transforms; and in this sense they are examples of Fourier dual pairs. If for example time is discrete, then frequency may be represented by numbers in the interval  $[0, 2\pi)$ ; or in  $[0, 1)$  if we enter the number  $2\pi$  into the Fourier exponential. Functions of the frequency are then periodic, so the two endpoints are identified. In the case of the interval  $[0, 1)$ , 0 on the left is identified with 1 on the right. So a low frequency band is an interval centered at 0, while a high frequency band is an interval centered at  $1/2$ . Let a function  $W$  on  $[0, 1)$  represent a probability assignment. Such functions  $W$  are thought of as “filters” in signal processing. We say that  $W$  is low-pass if it is 1 at 0, or if it is near 1 for frequencies near 0.

---

MATHEMATICS      PROBABILITY      ENGINEERING      PHYSICS

---

Low-pass filters pass signals with low frequencies, and block the others.

If instead some filter  $W$  is 1 at  $1/2$ , or takes values near 1 for frequencies near  $1/2$ , then we say that  $W$  is high-pass; it passes signals with high frequency.

**convolution**      —      **filter**      **smearing**

Pointwise multiplication of functions of frequencies corresponds in the Fourier dual time-domain to the operation of convolution (or of Cauchy product if the time-scale is discrete.) The process of modifying a signal with a fixed convolution is called a linear filter in signal processing. The corresponding Fourier dual frequency function is then referred to as “frequency response” or the “frequency response function.”

More generally, in the continuous case, since convolution tends to improve smoothness of functions, physicists call it “smearing.”

**decomposition**      —      **analysis**      **frequency components**  
(e.g., Fourier coefficients in a Fourier expansion)

Calculating the Fourier coefficients is “analysis,” and adding up the pure frequencies (i.e., summing the Fourier series) is called synthesis. But this view carries over more generally to engineering where there are more operations involved on the two sides, e.g., breaking up a signal into its frequency bands, transforming further, and then adding up the “banded” functions in the end. If the signal out is the same as the signal in, we say that the analysis/synthesis yields perfect reconstruction.

**integrate**      **reconstruct**      **synthesis**      **superposition**  
(e.g., inverse Fourier transform)

Here the terms related to “synthesis” refer to the second half of the kind of signal-processing design outlined in the previous paragraph.

**subspace**      —      **resolution**      **(signals in a frequency band)**

For a space of functions (signals), the selection of certain frequencies serves as a way of selecting special signals. When the process of scaling is introduced into optics of a digital camera, we note

MATHEMATICS	PROBABILITY	ENGINEERING	PHYSICS
that a nested family of subspaces corresponds to a grading of visual resolutions.			
<b>Cuntz relations</b>	—	<b>perfect reconstruction from subbands</b>	<b>subband decomposition</b>

$$\sum_{i=0}^{N-1} S_i S_i^* = \mathbf{1}, \text{ and } S_i^* S_j = \delta_{i,j} \mathbf{1}.$$

<b>inner product</b>	<b>correlation</b>	<b>transition probability</b>	<b>probability of transition from one state to another</b>
----------------------	--------------------	-------------------------------	--

In many applications, a vector space with inner product captures perfectly the geometric and probabilistic features of the situation. This can be axiomatized in the language of Hilbert space; and the inner product is the most crucial ingredient in the familiar axiom system for Hilbert space.

$f_{\text{out}} = T f_{\text{in}}$	—	<b>input/output</b>	<b>transformation of states</b>
------------------------------------	---	---------------------	---------------------------------

Systems theory language for operators  $T: V \rightarrow W$ . Then vectors in  $V$  are input, and in the range of  $T$  output.

<b>fractal</b>	—	—	—
----------------	---	---	---

Intuitively, think of a fractal as reflecting similarity of scales such as is seen in fern-like images that look “roughly” the same at small and at large scales. Fractals are produced from an infinite iteration of a finite set of maps, and this algorithm is perfectly suited to the kind of subdivision which is a cornerstone of the discrete wavelet algorithm. Self-similarity could refer alternately to space, and to time. And further versatility is added, in that flexibility is allowed into the definition of “similar.”

—	—	<b>data mining</b>	—
---	---	--------------------	---

The problem of how to handle and make use of large volumes of data is a corollary of the digital revolution. As a result, the subject of data mining itself changes rapidly. Digitized information (data) is now easy to capture automatically and to store electronically. In science, in commerce, and in industry, data represents collected observations and information: In business, there is data on markets, competitors, and customers. In manufacturing, there is data for optimizing production opportunities, and for improving processes. A tremendous potential for data mining exists in

medicine, genetics, and energy. But raw data is not always directly usable, as is evident by inspection. A key to advances is our ability to *extract information and knowledge* from the data (hence “data mining”), and to understand the phenomena governing data sources. Data mining is now taught in a variety of forms in engineering departments, as well as in statistics and computer science departments.

One of the structures often hidden in data sets is some degree of *scale*. The goal is to detect and identify one or more natural global and local scales in the data. Once this is done, it is often possible to detect associated similarities of scale, much like the familiar scale-similarity from multidimensional wavelets, and from fractals. Indeed, various adaptations of wavelet-like algorithms have been shown to be useful. These algorithms themselves are useful in *detecting* scale-similarities, and are applicable to other types of pattern recognition. Hence, in this context, generalized multiresolutions offer another tool for discovering structures in large data sets, such as those stored in the resources of the Internet. Because of the sheer volume of data involved, a strictly manual analysis is out of the question. Instead, sophisticated query processors based on statistical and mathematical techniques are used in generating insights and extracting conclusions from data sets.

*Multiresolutions.* Haar’s work in 1909–1910 had implicitly the key idea which got wavelet mathematics started on a roll 75 years later with Yves Meyer, Ingrid Daubechies, Stéphane Mallat, and others—namely the idea of a multiresolution. In that respect Haar was ahead of his time. See Figures 1 and 2 for details.

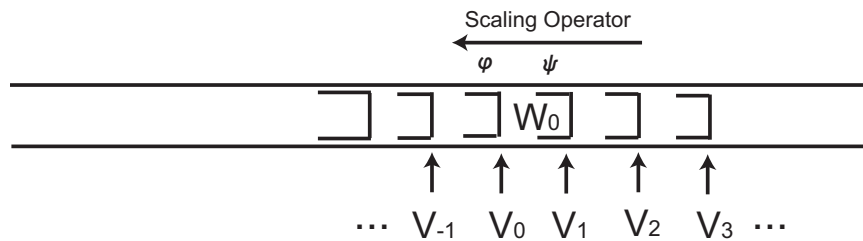


FIGURE 1. Multiresolution.  $L^2(\mathbb{R}^d)$ -version (continuous);  $\varphi \in V_0$ ,  $\psi \in W_0$ .

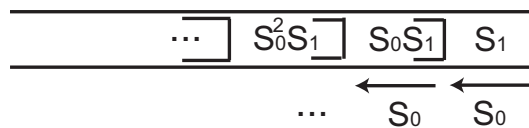


FIGURE 2. Multiresolution.  $l^2(\mathbb{Z})$ -version (discrete);  $\varphi \in V_0$ ,  $\psi \in W_0$ .

$$\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots, V_0 + W_0 = V_1$$

The word “multiresolution” suggests a connection to optics from physics. So that should have been a hint to mathematicians to take a closer look at trends in signal and image processing! Moreover, even staying within mathematics, it turns out that as a general notion this same idea of a “multiresolution” has long roots in mathematics, even in such modern and pure areas as operator theory and Hilbert-space geometry. Looking even closer at these interconnections, we can now recognize scales of subspaces (so-called multiresolutions) in classical algorithmic construction of orthogonal bases in inner-product spaces, now taught in lots of mathematics courses under the name of the Gram–Schmidt algorithm. Indeed, a closer look at good old Gram–Schmidt reveals that it is a matrix algorithm, Hence new mathematical tools involving non-commutativity!

If the signal to be analyzed is an image, then why not select a fixed but suitable *resolution* (or a subspace of signals corresponding to a selected resolution), and then do the computations there? The selection of a fixed “resolution” is dictated by practical concerns. That idea was key in turning computation of wavelet coefficients into iterated matrix algorithms. As the matrix operations get large, the computation is carried out in a variety of paths arising from big matrix products. The dichotomy, continuous vs. discrete, is quite familiar to engineers. The industrial engineers typically work with huge volumes of numbers.

Numbers! — So why wavelets? Well, what matters to the industrial engineer is not really the wavelets, but the fact that special wavelet functions serve as an efficient way to encode large data sets—I mean encode for computations. And the wavelet algorithms are computational. They work on numbers. Encoding numbers into pictures, images, or graphs of functions comes later, perhaps at the very end of the computation. But without the graphics, I doubt that we would understand any of this half as well as we do now. The same can be said for the many issues that relate to the crucial mathematical concept of self-similarity, as we know it from fractals, and more generally from recursive algorithms.

## 1. DEFINITION

In this paper we outline several points of view on the interplay between discrete and continuous wavelet transforms; stressing both pure and applied aspects of both. We outline some new links between the two transform technologies based on the theory of representations of generators and relations. By this we mean a finite system of generators which are represented by operators in Hilbert space. We further outline how these representations yield sub-band filter banks for signal and image processing algorithms.

The word “wavelet transform” (WT) means different things to different people: Pure and applied mathematicians typically give different answers the questions “What is the WT?” And engineers in turn have their own preferred quite different approach to WTs. Still there are two main trends in how WTs are used, the *continuous* WT on one side, and the *discrete* WT on the other. Here we offer a userfriendly outline of both, but with a slant toward geometric methods from the theory of operators in Hilbert space.

Our paper is organized as follows: For the benefit of diverse reader groups, we begin with Glossary (section ). This is a substantial part of our account, and it reflects the multiplicity of how the subject is used.

The concept of multiresolutions or multiresolution analysis (MRA) serves as a link between the discrete and continuous theory.

In section 4, we summarize how different mathematicians and scientists have contributed to and shaped the subject over the years.

The next two sections then offer a technical overview of both discrete and the continuous WTs. This includes basic tools from Fourier analysis and from operators in Hilbert space. In sections 6 and 7 we outline the connections between the separate parts of mathematics and their applications to WTs.

## 2. INTRODUCTION

While applied problems such as time series, signals and processing of digital images come from engineering and from the sciences, they have in the past two decades taken a life of their own as an exciting new area of applied mathematics. While searches in Google on these keywords typically yield sites numbered in the millions, the diversity of applications is wide, and it seems reasonable here to narrow our focus to some of the approaches that are both more mathematical and more recent. For references, see for example [1, 6, 23, 31]. In addition, our own interests (e.g., [20, 21, 27, 28]) have colored the presentation below. Each of the two areas, the discrete side, and the continuous theory is huge as measured by recent journal publications. A leading theme in our article is the independent interest in a multitude of interconnections between the discrete algorithm and their uses in the more mathematical analysis of function spaces (continuous wavelet transforms). The mathematics involved in the study and the applications of this interaction we feel is of benefit to both mathematicians and to engineers. See also [20]. An early paper [9] by Daubechies and Lagarias was especially influential in connecting the two worlds, discrete and continuous.

## 3. THE DISCRETE VS CONTINUOUS WAVELET ALGORITHMS

**3.1. The Discrete Wavelet Transform.** If one stays with function spaces, it is then popular to pick the  $d$ -dimensional Lebesgue measure on  $\mathbb{R}^d$ ,  $d = 1, 2, \dots$ , and pass to the Hilbert space  $L^2(\mathbb{R}^d)$  of all square integrable functions on  $\mathbb{R}^d$ , referring to  $d$ -dimensional Lebesgue measure. A wavelet basis refers to a family of basis functions for  $L^2(\mathbb{R}^d)$  generated from a finite set of normalized functions  $\psi_i$ , the index  $i$  chosen from a fixed and finite index set  $I$ , and from two operations, one called scaling, and the other translation. The scaling is typically specified by a  $d$  by  $d$  matrix over the integers  $\mathbb{Z}$  such that all the eigenvalues in modulus are bigger than one, lie outside the closed unit disk in the complex plane. The  $d$ -lattice is denoted  $\mathbb{Z}^d$ , and the translations will be by vectors selected from  $\mathbb{Z}^d$ . We say that we have a wavelet basis if the triple indexed family  $\psi_{i,j,k}(x) := |\det A|^{j/2} \psi(A^j x + k)$  forms an orthonormal basis (ONB) for  $L^2(\mathbb{R}^d)$  as  $i$  varies in  $I$ ,  $j \in \mathbb{Z}$ , and  $k \in \mathbb{R}^d$ . The word “orthonormal” for a family  $F$  of vectors in a Hilbert space  $\mathcal{H}$  refers to the norm and the inner product in  $\mathcal{H}$ : The vectors in an orthonormal family  $F$  are assumed to have norm one, and to be mutually orthogonal. If the family is also total (i.e., the vectors in  $F$  span a subspace which is dense in  $\mathcal{H}$ ), we say that  $F$  is an orthonormal basis (ONB.)

While there are other popular wavelet bases, for example frame bases, and dual bases (see e.g., [2, 14] and the papers cited there), the ONBs are the most agreeable at least from the mathematical point of view.

That there are bases of this kind is not at all clear, and the subject of wavelets in this continuous context has gained much from its connections to the discrete world of signal- and image processing.

Here we shall outline some of these connections with an emphasis on the mathematical context. So we will be stressing the theory of Hilbert space, and bounded linear operators acting in Hilbert space  $\mathcal{H}$ , both individual operators, and families of operators which form algebras.

As was noticed recently the operators which specify particular subband algorithms from the discrete world of signal- processing turn out to satisfy relations that were found (or rediscovered independently) in the theory of operator algebras, and which go under the name of Cuntz algebras, denoted  $\mathcal{O}_N$  if  $n$  is the number of bands. For additional details, see e.g., [21].

In symbols the  $C^*$ -algebra has generators  $(S_i)_{i=0}^{N-1}$ , and the relations are

$$(3.1) \quad \sum_{i=0}^{N-1} S_i S_i^* = \mathbf{1}$$

(where  $\mathbf{1}$  is the identity element in  $\mathcal{O}_N$ ) and

$$(3.2) \quad \sum_{i=0}^{N-1} S_i S_i^* = \mathbf{1}, \text{ and } S_i^* S_j = \delta_{i,j} \mathbf{1}.$$

In a representation on a Hilbert space, say  $\mathcal{H}$ , the symbols  $S_i$  turn into bounded operators, also denoted  $S_i$ , and the identity element  $\mathbf{1}$  turns into the identity operator  $I$  in  $\mathcal{H}$ , i.e., the operator  $I : h \rightarrow h$ , for  $h \in \mathcal{H}$ . In operator language, the two formulas 3.1 and 3.2 state that each  $S_i$  is an isometry in  $\mathcal{H}$ , and that the respective ranges  $S_i \mathcal{H}$  are mutually orthogonal, i.e.,  $S_i \mathcal{H} \perp S_j \mathcal{H}$  for  $i \neq j$ . Introducing the projections  $P_i = S_i S_i^*$ , we get  $P_i P_j = \delta_{i,j} P_i$ , and

$$\sum_{i=0}^{N-1} P_i = I$$

In the engineering literature this takes the form of programming diagrams:

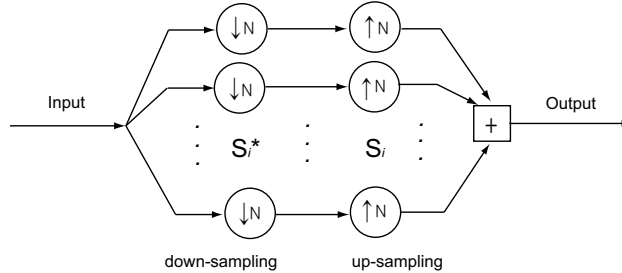
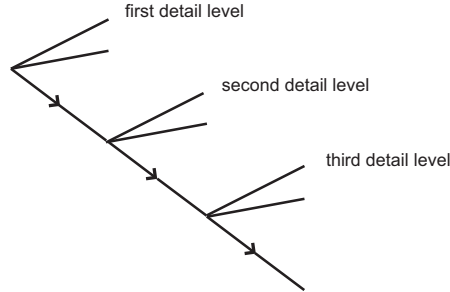


FIGURE 3. Perfect reconstruction in a subband filtering as used in signal and image processing.

If the process of Figure 3 is repeated, we arrive at the discrete wavelet transform



or stated in the form of images ( $n = 5$ )

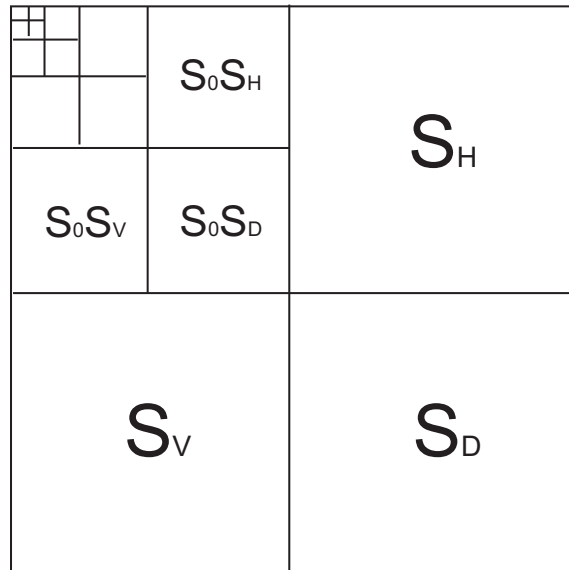


FIGURE 4. The subdivided squares represent the use of the pyramid subdivision algorithm to image processing, as it is used on pixel squares. At each subdivision step the top left-hand square represents averages of nearby pixel numbers, averages taken with respect to the chosen low-pass filter; while the three directions, horizontal, vertical, and diagonal represent detail differences, with the three represented by separate bands and filters. So in this model, there are four bands, and they may be realized by a tensor product construction applied to dyadic filters in the separate x- and the y-directions in the plane. For the discrete WT used in image-processing, we use iteration of four isometries  $S_0, S_H, S_V, and S_D$  with mutually orthogonal ranges, and satisfying the following sum-rule  $S_0 S_0^* + S_H S_H^* + S_V S_V^* + S_D S_D^* = I$ , with  $I$  denoting the identity operator in an appropriate  $l^2$ -space.

FIGURE 5.  $n = 2$  Jorgensen

Selecting a resolution subspace  $V_0 = \text{closure span}\{\varphi(\cdot - k) | k \in \mathbb{Z}\}$ , we arrive at a wavelet subdivision  $\{\psi_{j,k} | j \geq 0, k \in \mathbb{Z}\}$ , where  $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ , and the continuous expansion  $f = \sum_{j,k} \langle \psi_{j,k} | f \rangle \psi_{j,k}$  or the discrete analogue derived from the isometries,  $i = 1, 2, \dots, N-1$ ,  $S_0^k S_i$  for  $k = 0, 1, 2, \dots$ ; called the discrete wavelet transform.

*Notational convention.* In algorithms, the letter  $N$  is popular, and often used for counting more than one thing.

In the present context of the Discrete Wavelet Algorithm (DWA) or DWT, we count two things, “the number of times a picture is decomposed via subdivision”. We have used  $n$  for this. The other related but different number  $N$  is the number of subbands,  $N = 2$  for the dyadic DWT, and  $N = 4$  for the image DWT. The image-processing WT in our present context is the tensor product of the 1-D dyadic WT, so  $2 \times 2 = 4$ . Caution: Not all DWAs arise as tensor products of  $N = 2$  models. The wavelets coming from tensor products are called separable. When a particular image-processing scheme is used for generating continuous wavelets it is not transparent if we are looking at a separable or inseparable wavelet!

To clarify the distinction, it is helpful to look at the representations of the Cuntz relations by operators in Hilbert space. We are dealing with representations of the

two distinct algebras  $\mathcal{O}_2$ , and  $\mathcal{O}_4$ ; two frequency subbands vs 4 subbands. Note that the Cuntz  $\mathcal{O}_2$ , and  $\mathcal{O}_4$  are given axiomatic, or purely symbolically. It is only when subband filters are chosen that we get representations. This also means that the choice of  $N$  is made initially; and the same  $N$  is used in different runs of the programs. In contrast, the number of times a picture is decomposed varies from one experiment to the next!

**Summary:**  $N = 2$  for the dyadic DWT: The operators in the representation are  $S_0$ ,  $S_1$ . One average operator, and one detail operator. The detail operator  $S_1$  “counts” local detail variations.

Image-processing. Then  $N = 4$  is fixed as we run different images in the DWT: The operators are now:  $S_0$ ,  $S_H$ ,  $S_V$ ,  $S_D$ . One average operator, and three detail operator for local detail variations in the three directions in the plane.

**3.2. The Continuous Wavelet Transform.** Consider functions  $f$  on the real line  $\mathbb{R}$ . We select the Hilbert space of functions to be  $L^2(\mathbb{R})$  To start a continuous WT, we must select a function  $\psi \in L^2(\mathbb{R})$  and  $r, s \in \mathbb{R}$  such that the following family of functions

$$\psi_{r,s}(x) = r^{-1/2}\psi\left(\frac{x-s}{r}\right)$$

creates an over-complete basis for  $L^2(\mathbb{R})$ . An over-complete family of vectors in a Hilbert space is often called a coherent decomposition. This terminology comes from quantum optics. What is needed for a continuous WT in the simplest case is the following representation valid for all  $f \in L^2(\mathbb{R})$ :

$$f(x) = C_\psi^{-1} \int \int_{\mathbb{R}^2} \langle \psi_{r,s} | f \rangle \psi_{r,s}(x) \frac{dr ds}{r^2}$$

where  $C_\psi := \int_{\mathbb{R}} |\hat{\psi}(\omega)|^2 \frac{d\omega}{\omega}$  and where  $\langle \psi_{r,s} | f \rangle = \int_{\mathbb{R}} \overline{\psi_{r,s}(y)} f(y) dy$ . The refinements and implications of this are spelled out in tables in section 3.4

**3.3. Some background on Hilbert space.** Wavelet theory is the art of finding a special kind of basis in Hilbert space. Let  $\mathcal{H}$  be a Hilbert space over  $\mathbb{C}$  and denote the inner product  $\langle \cdot | \cdot \rangle$ . For us, it is assumed linear in the second variable. If  $\mathcal{H} = L^2(\mathbb{R})$ , then

$$\langle f | g \rangle := \int_{\mathbb{R}} \overline{f(x)} g(x) dx.$$

If  $\mathcal{H} = \ell^2(\mathbb{Z})$ , then

$$\langle \xi | \eta \rangle := \sum_{n \in \mathbb{Z}} \bar{\xi}_n \eta_n.$$

Let  $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ . If  $\mathcal{H} = L^2(\mathbb{T})$ , then

$$\langle f | g \rangle := \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{f(\theta)} g(\theta) d\theta.$$

Functions  $f \in L^2(\mathbb{T})$  have Fourier series: Setting  $e_n(\theta) = e^{in\theta}$ ,

$$\hat{f}(n) := \langle e_n | f \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-in\theta} f(\theta) d\theta,$$

and

$$\|f\|_{L^2(\mathbb{T})}^2 = \sum_{n \in \mathbb{Z}} |\hat{f}(n)|^2.$$

Similarly if  $f \in L^2(\mathbb{R})$ , then

$$\hat{f}(t) := \int_{\mathbb{R}} e^{-ixt} f(x) dx,$$

and

$$\|f\|_{L^2(\mathbb{R})}^2 = \frac{1}{2\pi} \int_{\mathbb{R}} |\hat{f}(t)|^2 dt.$$

Let  $J$  be an index set. We shall only need to consider the case when  $J$  is countable. Let  $\{\psi_\alpha\}_{\alpha \in J}$  be a family of nonzero vectors in a Hilbert space  $\mathcal{H}$ . We say it is an *orthonormal basis* (ONB) if

$$(3.3) \quad \langle \psi_\alpha | \psi_\beta \rangle = \delta_{\alpha,\beta} \quad (\text{Kronecker delta})$$

and if

$$(3.4) \quad \sum_{\alpha \in J} |\langle \psi_\alpha | f \rangle|^2 = \|f\|^2 \quad \text{holds for all } f \in \mathcal{H}.$$

If only (3.4) is assumed, but not (3.3), we say that  $\{\psi_\alpha\}_{\alpha \in J}$  is a (normalized) *tight frame*. We say that it is a *frame* with *frame constants*  $0 < A \leq B < \infty$  if

$$A \|f\|^2 \leq \sum_{\alpha \in J} |\langle \psi_\alpha | f \rangle|^2 \leq B \|f\|^2 \quad \text{holds for all } f \in \mathcal{H}.$$

Introducing the rank-one operators  $Q_\alpha := |\psi_\alpha\rangle \langle \psi_\alpha|$  of Dirac's terminology, see [3], we see that  $\{\psi_\alpha\}_{\alpha \in J}$  is an ONB if and only if the  $Q_\alpha$ 's are projections, and

$$(3.5) \quad \sum_{\alpha \in J} Q_\alpha = I \quad (= \text{the identity operator in } \mathcal{H}).$$

It is a (normalized) tight frame if and only if (3.5) holds but with no further restriction on the rank-one operators  $Q_\alpha$ . It is a frame with frame constants  $A$  and  $B$  if the operator

$$S := \sum_{\alpha \in J} Q_\alpha$$

satisfies

$$AI \leq S \leq BI$$

in the order of hermitian operators. (We say that operators  $H_i = H_i^*$ ,  $i = 1, 2$ , satisfy  $H_1 \leq H_2$  if  $\langle f | H_1 f \rangle \leq \langle f | H_2 f \rangle$  holds for all  $f \in \mathcal{H}$ ). If  $h, k$  are vectors in a Hilbert space  $\mathcal{H}$ , then the operator  $A = |h\rangle \langle k|$  is defined by the identity  $\langle u | Av \rangle = \langle u | h \rangle \langle k | v \rangle$  for all  $u, v \in \mathcal{H}$ .

Wavelets in  $L^2(\mathbb{R})$  are generated by simple operations on one or more functions  $\psi$  in  $L^2(\mathbb{R})$ , the operations come in pairs, say scaling and translation, or phase-modulation and translations. If  $N \in \{2, 3, \dots\}$  we set

$$\psi_{j,k}(x) := N^{j/2} \psi(N^j x - k) \quad \text{for } j, k \in \mathbb{Z}.$$

**3.3.1. Increasing the dimension.** In wavelet theory, [7] there is a tradition for reserving  $\varphi$  for the father function and  $\psi$  for the mother function. A 1-level wavelet transform of an  $N \times M$  image can be represented as

$$(3.6) \quad \mathbf{f} \mapsto \begin{pmatrix} \mathbf{a}^1 & | & \mathbf{h}^1 \\ \hline \mathbf{v}^1 & | & \mathbf{d}^1 \end{pmatrix}$$

where the subimages  $\mathbf{h}^1, \mathbf{d}^1, \mathbf{a}^1$  and  $\mathbf{v}^1$  each have the dimension of  $N/2$  by  $M/2$ .

$$\begin{aligned}
 (3.7) \quad \mathbf{a}^1 &= V_m^1 \otimes V_n^1 : \varphi^A(x, y) = \varphi(x)\varphi(y) = \sum_i \sum_j h_i h_j \varphi(2x - i)\varphi(2y - j) \\
 \mathbf{h}^1 &= V_m^1 \otimes W_n^1 : \psi^H(x, y) = \psi(x)\varphi(y) = \sum_i \sum_j g_i h_j \varphi(2x - i)\varphi(2y - j) \\
 \mathbf{v}^1 &= W_m^1 \otimes V_n^1 : \psi^V(x, y) = \varphi(x)\psi(y) = \sum_i \sum_j h_i g_j \varphi(2x - i)\varphi(2y - j) \\
 \mathbf{d}^1 &= W_m^1 \otimes W_n^1 : \psi^D(x, y) = \psi(x)\psi(y) = \sum_i \sum_j g_i g_j \varphi(2x - i)\varphi(2y - j)
 \end{aligned}$$

where  $\varphi$  is the father function and  $\psi$  is the mother function in sense of wavelet,  $V$  space denotes the average space and the  $W$  spaces are the difference space from multiresolution analysis (MRA) [7].

In the formulas, we have the following two indexed number systems  $\mathbf{a} := (h_i)$  and  $\mathbf{d} := (g_i)$ ,  $\mathbf{a}$  is for averages, and  $\mathbf{d}$  is for local differences. They are really the input for the DWT. But they also are the key link between the two transforms, the discrete and continuous. The link is made up of the following scaling identities:

$$\begin{aligned}
 \varphi(x) &= 2 \sum_{i \in \mathbb{Z}} h_i \varphi(2x - i); \\
 \psi(x) &= 2 \sum_{i \in \mathbb{Z}} g_i \varphi(2x - i);
 \end{aligned}$$

and (low-pass normalization)  $\sum_{i \in \mathbb{Z}} h_i = 1$ . The scalars  $(h_i)$  may be real or complex; they may be finite or infinite in number. If there are four of them, it is called the “four tap”, etc. The finite case is best for computations since it corresponds to compactly supported functions. This means that the two functions  $\varphi$  and  $\psi$  will vanish outside some finite interval on a real line.

The two number systems are further subjected to orthogonality relations, of which

$$(3.8) \quad \sum_{i \in \mathbb{Z}} \bar{h}_i h_{i+2k} = \frac{1}{2} \delta_{0,k}$$

is the best known.

The systems  $h$  and  $g$  are both low-pass and high-pass filter coefficients. In 3.7,  $\mathbf{a}^1$  denotes the first averaged image, which consists of average intensity values of the original image. Note that only  $\varphi$  function,  $V$  space and  $h$  coefficients are used here. Similarly,  $\mathbf{h}^1$  denotes the first detail image of horizontal components, which consists of intensity difference along the vertical axis of the original image. Note that  $\varphi$  function is used on  $y$  and  $\psi$  function on  $x$ ,  $W$  space for  $x$  values and  $V$  space for  $y$  values; and both  $h$  and  $g$  coefficients are used accordingly. The data  $\mathbf{v}^1$  denotes the first detail image of vertical components, which consists of intensity difference along the horizontal axis of the original image. Note that  $\varphi$  function is used on  $x$  and  $\psi$  function on  $y$ ,  $W$  space for  $y$  values and  $V$  space for  $x$  values; and both  $h$  and  $g$  coefficients are used accordingly. Finally,  $\mathbf{d}^1$  denotes the first detail image of diagonal components, which consists of intensity difference along the diagonal axis of the original image. The original image is reconstructed from the decomposed image by taking the sum of the averaged image and the detail images and scaling by a scaling factor. It could be noted that only  $\psi$  function,  $W$  space and  $g$  coefficients are used here. See [33, 28].

This decomposition not only limits to one step but it can be done again and again on the averaged detail depending on the size of the image. Once it stops at certain level, quantization (see [26, 32]) is done on the image. This quantization step may be lossy or lossless. Then the lossless entropy encoding is done on the decomposed and quantized image.

The relevance of the system of identities (3.8) may be summarized as follows. Set

$$m_0(z) := \frac{1}{2} \sum_{k \in \mathbb{Z}} h_k z^k \text{ for all } z \in \mathbb{T};$$

$$g_k := (-1)^k \bar{h}_{1-k} \text{ for all } k \in \mathbb{Z};$$

$$m_1(z) := \frac{1}{2} \sum_{k \in \mathbb{Z}} g_k z^k; \text{ and}$$

$$(S_j f)(z) = \sqrt{2} m_j(z) f(z^2), \text{ for } j = 0, 1, f \in L^2(\mathbb{T}), z \in \mathbb{T}.$$

Then the following conditions are equivalent:

- (a) The system of equations (3.8) is satisfied.
- (b) The operators  $S_0$  and  $S_1$  satisfy the Cuntz relations.
- (c) We have perfect reconstruction in the subband system of Figure 3.

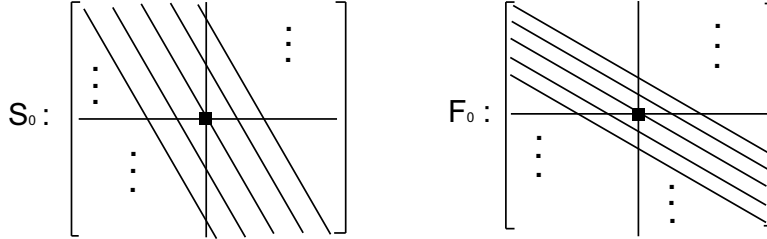
Note that the two operators  $S_0$  and  $S_1$  have equivalent matrix representations. Recall that by Parseval's formula we have  $L^2(\mathbb{T}) \simeq l^2(\mathbb{Z})$ . So representing  $S_0$  instead as an  $\infty \times \infty$  matrix acting on column vectors  $x = (x_j)_{j \in \mathbb{Z}}$  we get

$$(S_0 x)_i = \sqrt{2} \sum_{j \in \mathbb{Z}} h_{i-2j} x_j$$

and for the adjoint operator  $F_0 := S_0^*$ , we get the matrix representation

$$(F_0 x)_i = \frac{1}{\sqrt{2}} \sum_{j \in \mathbb{Z}} \bar{h}_{j-2i} x_j$$

with the overbar signifying complex conjugation. This is computationally significant to the two matrix representations, both the matrix for  $S_0$ , and for  $F_0 := S_0^*$ , is slanted. However, the slanting of one is the mirror-image of the other, i.e.,



*Significance of slanting.* The slanted matrix representations refers to the corresponding operators in  $L^2$ . In general operators in Hilbert function spaces have many matrix representations, one for each orthonormal basis (ONB), but here we are concerned with the ONB consisting of the Fourier frequencies  $z^j$ ,  $j \in \mathbb{Z}$ . So in our matrix representations for the  $S$  operators and their adjoints we will be acting on column vectors, each infinite column representing a vector in the sequence space  $l^2$ . A vector in  $l^2$  is said to be of finite size if it has only a finite set of non-zero entries.

It is the matrix  $F_0$  that is effective for iterated matrix computation. Reason: When a column vector  $x$  of a fixed size, say  $2s$  is multiplied, or acted on by  $F_0$ , the result is a vector  $y$  of half the size, i.e., of size  $s$ . So  $y = F_0 x$ . If we use  $F_0$  and  $F_1$  together on  $x$ , then we get two vectors, each of size  $s$ , the other one  $z = F_1 x$ ,

and we can form the combined column vector of  $y$  and  $z$ ; stacking  $y$  on top of  $z$ . In our application,  $y$  represents averages, while  $z$  represents local differences: Hence the wavelet algorithm.

$$\begin{bmatrix} \vdots \\ y_{-1} \\ y_0 \\ y_1 \\ \vdots \\ \text{---} \\ \vdots \\ z_{-1} \\ z_0 \\ z_1 \\ \vdots \end{bmatrix} = \begin{bmatrix} F_0 \\ \text{---} \\ F_1 \end{bmatrix} \begin{bmatrix} \vdots \\ x_{-2} \\ x_{-1} \\ x_0 \\ x_1 \\ x_2 \\ \vdots \end{bmatrix}$$

$$y = F_0 x$$

$$z = F_1 x$$

**3.4. Connections to group theory.** The first line in the two tables below is the continuous wavelet transform. It comes from what in physics is called *coherent vector decompositions*. Both transforms applies to vectors in Hilbert space  $\mathcal{H}$ , and  $\mathcal{H}$  may vary from case to case. Common to all transforms is vector input and output. If the input agrees with output we say that the combined process yields the identity operator image.  $\mathbf{1} : \mathcal{H} \rightarrow \mathcal{H}$  or written  $\mathbf{1}_{\mathcal{H}}$ . So for example if  $(S_i)_{i=0}^{N-1}$  is a finite operator system, and input/output operator example may take the form

$$\sum_{i=0}^{N-1} S_i S_i^* = \mathbf{1}_{\mathcal{H}}.$$

Summary of and variations on the resolution of the identity operator  $\mathbf{1}$  in  $L^2$  or in  $\ell^2$ , for  $\psi$  and  $\tilde{\psi}$  where  $\psi_{r,s}(x) = r^{-\frac{1}{2}} \psi\left(\frac{x-s}{r}\right)$ ,

$$C_\psi = \int_{\mathbb{R}} \frac{d\omega}{|\omega|} |\hat{\psi}(\omega)|^2 < \infty,$$

similarly for  $\tilde{\psi}$  and  $C_{\psi, \tilde{\psi}} = \int_{\mathbb{R}} \frac{d\omega}{|\omega|} \overline{\hat{\psi}(\omega)} \hat{\tilde{\psi}}(\omega)$ :

$N = 2$	Overcomplete Basis	Dual Bases
continuous resolution	$C_\psi^{-1} \iint_{\mathbb{R}^2} \frac{dr ds}{r^2}  \psi_{r,s}\rangle \langle \psi_{r,s} $ $= \mathbf{1}$	$C_{\psi, \tilde{\psi}}^{-1} \iint_{\mathbb{R}^2} \frac{dr ds}{r^2}  \psi_{r,s}\rangle \langle \tilde{\psi}_{r,s} $ $= \mathbf{1}$
discrete resolution	$\sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}}  \psi_{j,k}\rangle \langle \psi_{j,k}  = \mathbf{1}$ , $\psi_{j,k}$ corresponding to $r = 2^{-j}$ , $s = k2^{-j}$	$\sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}}  \psi_{j,k}\rangle \langle \tilde{\psi}_{j,k}  = \mathbf{1}$
$N \geq 2$	Isometries in $\ell^2$	Dual Operator System in $\ell^2$
sequence spaces	$\sum_{i=0}^{N-1} S_i S_i^* = \mathbf{1}$ , where $S_0, \dots, S_{N-1}$ are adjoints to the quadrature mirror filter operators $F_i$ , i.e., $S_i = F_i^*$	$\sum_{i=0}^{N-1} S_i \tilde{S}_i^* = \mathbf{1}$ , for a dual operator system $S_0, \dots, S_{N-1}$ , $\tilde{S}_0, \dots, \tilde{S}_{N-1}$

Then the assertions in the first table amount to:

$C_\psi^{-1} \iint_{\mathbb{R}^2} \frac{dr ds}{r^2}  \langle \psi_{r,s}   f \rangle ^2$ $= \ f\ _{L^2}^2 \quad \forall f \in L^2(\mathbb{R})$	$C_{\psi, \tilde{\psi}}^{-1} \iint_{\mathbb{R}^2} \frac{dr ds}{r^2} \langle f   \psi_{r,s} \rangle \langle \tilde{\psi}_{r,s}   g \rangle$ $= \langle f   g \rangle \quad \forall f, g \in L^2(\mathbb{R})$
$\sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}}  \langle \psi_{j,k}   f \rangle ^2$ $= \ f\ _{L^2}^2 \quad \forall f \in L^2(\mathbb{R})$	$\sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \langle f   \psi_{j,k} \rangle \langle \tilde{\psi}_{j,k}   g \rangle$ $= \langle f   g \rangle \quad \forall f, g \in L^2(\mathbb{R})$
$\sum_{i=0}^{N-1} \ S_i^* c\ ^2 = \ c\ ^2 \quad \forall c \in \ell^2$	$\sum_{i=0}^{N-1} \langle S_i^* c   \tilde{S}_i^* d \rangle = \langle c   d \rangle \quad \forall c, d \in \ell^2$

A function  $\psi$  satisfying the resolution identity is called a *coherent vector* in mathematical physics. The representation theory for the  $(ax+b)$ -group, i.e., the matrix group  $G = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \mid a \in \mathbb{R}_+, b \in \mathbb{R} \right\}$ , serves as its underpinning. Then the tables above illustrate how the  $\{\psi_{j,k}\}$  wavelet system arises from a discretization of the following unitary representation of  $G$ :

$$\left( U_{\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}} f \right) (x) = a^{-\frac{1}{2}} f \left( \frac{x-b}{a} \right)$$

acting on  $L^2(\mathbb{R})$ . This unitary representation also explains

the discretization step in passing from the first line to the second in the tables above. The functions  $\{\psi_{j,k} \mid j, k \in \mathbb{Z}\}$  which make up a wavelet system result from

the choice of a suitable coherent vector  $\psi \in L^2(\mathbb{R})$ , and then setting

$$\psi_{j,k}(x) = \left( U_{\begin{pmatrix} 2^{-j} & k \cdot 2^{-j} \\ 0 & 1 \end{pmatrix}} \psi \right) (x) = 2^{\frac{j}{2}} \psi(2^j x - k).$$

Even though this representation lies at the historical origin of the subject of wavelets, the  $(ax + b)$ -group seems to be now largely forgotten in the next generation of the wavelet community. But Chapters 1–3 of [7] still serve as a beautiful presentation of this (now much ignored) side of the subject. It also serves as a link to mathematical physics and to classical analysis.

#### 4. LIST OF NAMES AND DISCOVERIES

Many of the main discoveries summarized below are now lore.

- |  |  |
|--|--|
| <p><b>1807</b><br/>Jean Baptiste Joseph<br/>Fourier<br/>mathematics, physics<br/>(heat conduction)</p> | <p>Expressing functions as sums of sine and cosine waves of frequencies in arithmetic progression (now called Fourier series).</p>   |
| <p><b>1909</b><br/>Alfred Haar<br/>mathematics</p>   | <p>Discovered, while a student of David Hilbert, an orthonormal basis consisting of step functions, applicable both to functions on an interval, and functions on the whole real line. While it was not realized at the time, Haar's construction was a precursor of what is now known as the Mallat subdivision, and multiresolution method, as well as the subdivision wavelet algorithms.</p> |
| <p><b>1946</b><br/>Denes Gabor<br/>(Nobel Prize): physics<br/>(optics, holography)</p>                 | <p>Discovered basis expansions for what might now be called time-frequency wavelets, as opposed to time-scale wavelets.</p>  |
| <p><b>1948</b><br/>Claude Elwood Shannon<br/>mathematics, engineering<br/>(information theory)</p>     | <p>A rigorous formula used by the phone company for sampling speech signals. Quantizing information, entropy, founder of what is now called the mathematical theory of communication.</p>  |
| <p><b>1976</b><br/>Claude Garland, Daniel<br/>Esteban (both)<br/>signal processing</p>                 | <p>Discovered subband coding of digital transmission of speech signals over the telephone.</p>   |

- 1981**  
Jean Morlet  
petroleum engineer
- Suggested the term “ondelettes.” J.M. decomposed reflected seismic signals into sums of “wavelets (Fr.: ondelettes) of constant shape,” i.e., a decomposition of signals into wavelet shapes (now called wavelet series). Received somewhat late recognition for his work. Due to contributions by A. Grossman and Y. Meyer, Morlet’s discoveries have now come to play a central role in the theory.
- 1985**  
Yves Meyer  
mathematics,  
applications
- Mentor for A. Cohen, S. Mallat, and other of the wavelet pioneers, Y.M. discovered infinitely often differentiable wavelets.
- 1989**  
Albert Cohen  
mathematics (orthogonality relations),  
numerical analysis
- Discovered the use of wavelet filters in the analysis of wavelets—the so-called Cohen condition for orthogonality.
- 1986**  
Stéphane Mallat  
mathematics, signal  
and image processing
- Discovered what is now known as the subdivision, and multiresolution method, as well as the subdivision wavelet algorithms. This allowed the effective use of operators in the Hilbert space  $L^2(\mathbb{R})$ , and of the parallel computational use of recursive matrix algorithms.
- 1987**  
Ingrid Daubechies  
mathematics, physics,  
and communications
- Discovered differentiable wavelets, with the number of derivatives roughly half the length of the support interval. Further found polynomial algorithmic for their construction (with coauthor Jeff Lagarias; joint spectral radius formulas).
- 1991**  
Wayne Lawton  
mathematics  
(the wavelet  
transfer operator)
- Discovered the use of a transfer operator in the analysis of wavelets: orthogonality and smoothness.
- 1992**  
The FBI  
using wavelet algorithms in digitizing and compressing fingerprints
- C. Brislawn and his group at Los Alamos created the theory and the codes which allowed the compression of the enormous FBI fingerprint file, creating A/D, a new database of fingerprints.

<b>2000</b> The International Standards Organization	A wavelet-based picture compression standard, called JPEG 2000, for digital encoding of images.
<b>1994</b> David Donoho statistics, mathematics	Pioneered the use of wavelet bases and tools from statistics to “denoise” images and signals.

## 5. HISTORY

While wavelets as they have appeared in the mathematics literature (e.g., [7]) for a long time, starting with Haar in 1909, involve function spaces, the connections to a host of discrete problems from engineering is more subtle. Moreover the deeper connections between the discrete algorithms and the function spaces of mathematical analysis are of a more recent vintage, see e.g., [31] and [21].

Here we begin with the function spaces. This part of wavelet theory refers to continuous wavelet transforms (details below). It dominated the wavelet literature in the 1980s, and is beautifully treated in the first four chapters in [7] and in [8]. The word “continuous” refers to the continuum of the real line  $\mathbb{R}$ . Here we consider spaces of functions in one or more real dimensions, i.e., functions on the line  $\mathbb{R}$  (signals), the plane  $\mathbb{R}^2$  (images), or in higher dimensions  $\mathbb{R}^d$ , functions of  $d$  real variables.

## 6. TOOLS FROM MATHEMATICS

In our presentation, we will rely on tools from at least three separate areas of mathematics, and we will outline how they interact to form a coherent theory, and how they come together to form a link between what is now called the discrete and the continuous wavelet transform. It is the discrete case that is popular with engineers ([1, 23, 29, 30]), while the continuous case has come to play a central role in the part of mathematics referred to as harmonic analysis, [8]. The three areas are, operator algebras, dynamical systems, and basis constructions:

- a. Operator algebras. The theory of operator algebras in turn breaks up in two parts: One the study of “the algebras themselves” as they emerge from the axioms of von Neumann (von Neumann algebras), and Gelfand, Kadison and Segal ( $C^*$ -algebras.) The other has a more applied slant: It involves “the representations” of the algebras. By this we refer to the following: The algebras will typically be specified by generators and by relations, and by a certain norm-completion, in any case by a system of axioms. This holds both for the norm-closed algebras, the so called  $C^*$ -algebras, and for the weakly closed algebras, the von Neumann algebras. In fact there is a close connection between the two parts of the theory: For example, representations of  $C^*$ -algebras generate von Neumann algebras.

To talk about representations of a fixed algebra say  $A$  we must specify a Hilbert space, and a homomorphism  $\rho$  from  $A$  into the algebra  $\mathcal{B}(H)$  of all bounded operators on  $\mathcal{H}$ . We require that  $\rho$  sends the identity element in

$A$  into the identity operator acting on  $\mathcal{H}$ , and that  $\rho(a^*) = (\rho(a))^*$  where the last star now refers to the adjoint operator.

It was realized in the last ten years (see for example [3, 21, 22]) that a family of representations that wavelets which are basis constructions in harmonic analysis, in signal/image analysis, and in computational mathematics may be built up from representations of an especially important family of simple  $C^*$ -algebras, the Cuntz algebras. The Cuntz algebras are denoted  $\mathcal{O}_2, \mathcal{O}_3, \dots$ , including  $\mathcal{O}_\infty$ .

- b. Dynamical systems. The connection between the Cuntz algebras  $\mathcal{O}_N$  for  $N = 2, 3$ , are relevant to the kind of dynamical systems which are built on branching-laws, the case of  $\mathcal{O}_N$  representing  $N$ -fold branching. The reason for this is that if  $N$  is fixed,  $\mathcal{O}_N$  includes in its definition an iterated subdivision, but within the context of Hilbert space. For more details, see e.g., [15, 13, 16, 17, 18, 12, 22].
- c. Analysis of bases in function spaces. The connection to basis constructions using wavelets is this: The context for wavelets is a Hilbert space  $\mathcal{H}$ , where  $\mathcal{H}$  may be  $L^2(\mathbb{R}^d)$  where  $d$  is a dimension,  $d = 1$  for the line (signals),  $d = 2$  for the plane (images), etc. The more successful bases in Hilbert space are the orthonormal bases ONBs, but until the mid 1980s, there were no ONBs in  $L^2(\mathbb{R}^d)$  which were entirely algorithmic and effective for computations. One reason for this is that the tools that had been used for 200 years since Fourier involved basis functions (Fourier wave functions) which are not localized. Moreover these existing Fourier tools were not friendly to algorithmic computations.

## 7. A TRANSFER OPERATOR

A popular tool for deciding if a candidate for a wavelet basis is in fact an ONB uses a certain transfer operator. Variants of this operator is used in diverse areas of applied mathematics. It is an operator which involves a weighted average over a finite set of possibilities. Hence it is natural for understanding random walk algorithms. As remarked in for example [20, 21, 22, 15], it was also studied in physics, for example by David Ruelle who used to prove results on phase transition for infinite spin systems in quantum statistical mechanics. In fact the transfer operator has many incarnations (many of them known as Ruelle operators), and all of them based on  $N$ -fold branching laws.

In our wavelet application, the Ruelle operator weights in input over the  $N$  branch possibilities, and the weighting is assigned by a chosen scalar function  $W$ . the and the  $W$ -Ruelle operator is denoted  $R_W$ . In the wavelet setting there is in addition a low-pass filter function  $m_0$  which in its frequency response formulation is a function on the  $d$ -torus  $\mathbf{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ .

Since the scaling matrix  $A$  has integer entries  $A$  passes to the quotient  $\mathbb{R}^d/\mathbb{Z}^d$ , and the induced transformation  $r_A : \mathbb{T}^d \rightarrow \mathbb{T}^d$  is an  $N$ -fold cover, where  $N = |\det A|$ , i.e., for every  $x$  in  $\mathbb{T}^d$  there are  $N$  distinct points  $y$  in  $\mathbb{T}^d$  solving  $r_A(y) = x$ .

In the wavelet case, the weight function  $W$  is  $W = |m_0|^2$ . Then with this choice of  $W$ , the ONB problem for a candidate for a wavelet basis in the Hilbert space  $L^2(\mathbb{R}^d)$  as it turns out may be decided by the dimension of a distinguished eigenspace for  $R_W$ , by the so called Perron-Frobenius problem.

This has worked well for years for the wavelets which have an especially simple algorithm, the wavelets that are initialized by a single function, called the scaling function. These are called the multiresolution analysis (MRA) wavelets, or for short the MRA-wavelets. But there are instances, for example if a problem must be localized in frequency domain, when the MRA-wavelets do not suffice, where it will by necessity include more than one scaling function. And we are then back to trying to decide if the output from the discrete algorithm, and the  $\mathcal{O}_N$  representation is an ONB, or if it has some stability property which will serve the same purpose, in case where asking for an ONB is not feasible.

### 8. FUTURE DIRECTIONS

The idea of a scientific analysis by subdividing a fixed picture or object into its finer parts is not unique to wavelets. It works best for structures with an inherent self-similarity; this self-similarity can arise from numerical scaling of distances. But there are more subtle non-linear self-similarities. The Julia sets in the complex plane are a case in point [4, 5, 10, 11, 24, 25]. The simplest Julia sets come from a one parameter family of quadratic polynomials  $\varphi_c(z) = z^2 + c$ , where  $z$  is a complex variable and where  $c$  is a fixed parameter. The corresponding Julia sets  $J_c$  have a surprisingly rich structure. A simple way to understand them is the following: Consider the two branches of the inverse  $\beta_{\pm} = z \mapsto \pm\sqrt{z-c}$ . Then  $J_c$  is the unique minimal non-empty compact subset of  $\mathbb{C}$ , which is invariant under  $\{\beta_{\pm}\}$ . (There are alternative ways of presenting  $J_c$  but this one fits our purpose. The Julia set  $J$  of a holomorphic function, in this case  $z \mapsto z^2 + c$ , informally consists of those points whose long-time behavior under repeated iteration, or rather iteration of substitutions, can change drastically under arbitrarily small perturbations.) Here “long-time” refers to large  $n$ , where  $\varphi^{(n+1)}(z) = \varphi(\varphi^{(n)}(z))$ ,  $n = 0, 1, \dots$ , and  $\varphi^{(0)}(z) = z$ .

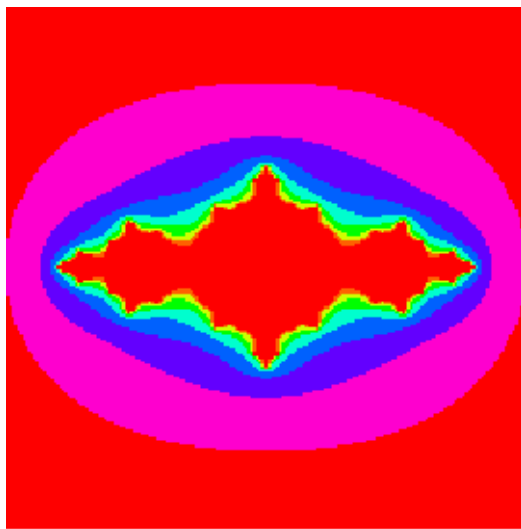


FIGURE 6. Julia set with  $c = -1$

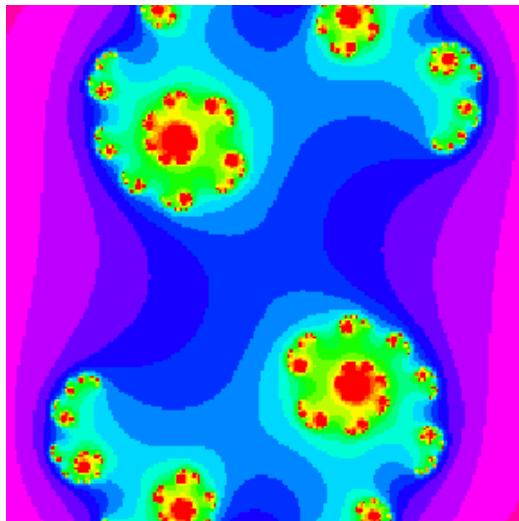


FIGURE 7. Julia set with  $c = 0.45 - 0.1428i$

It would be interesting to adapt and modify the Haar wavelet, and the other wavelet algorithms to the Julia sets. The two papers [16, 17] initiate such a development.

## 9. LITERATURE

As evidenced by a simple Google check, the mathematical wavelet literature is gigantic in size, and the manifold applications spread over a vast number of engineering journals. While we cannot do justice to this volumest literature, we instead offer a collection of the classics [19] edited recently by C. Heil et.al.

*Acknowledgements.* We thank Professors Dorin Dutkay and Judy Packer for helpful discussions.

## 10. BIBLIOGRAPHY

- [1] Gilles Aubert and Pierre Kornprobst. *Mathematical problems in image processing*. 2006.
- [2] Lawrence Baggett, Palle Jorgensen, Kathy Merrill, and Judith Packer. A non-MRA  $C^r$  frame wavelet with rapid decay. *Acta Appl. Math.*, 2005.
- [3] O. Bratelli and P. Jorgensen. *Wavelets Through a Looking Glass: The World of the Spectrum*. Birkhäuser, 2002.
- [4] M. Braverman and M. Yampolsky. Non-computable Julia sets. *J. Amer. Math. Soc.*, 19(3):551–578 (electronic), 2006.
- [5] Mark Braverman. Parabolic Julia sets are polynomial time computable. *Nonlinearity*, 19(6):1383–1401, 2006.
- [6] K. Bredies, D. A. Lorenz, and P. Maass. *An optimal control problem in medical image processing*. 2006.
- [7] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. 1992.
- [8] Ingrid Daubechies. *Wavelet transforms and orthonormal wavelet bases*. Proc. Sympos. Appl. Math. 1993.

- [9] Ingrid Daubechies and Jeffrey C. Lagarias. Two-scale difference equations. II. Local regularity, infinite products of matrices and fractals. *SIAM J. Math. Anal.*, 1992.
- [10] Robert L. Devaney and Daniel M. Look. A criterion for Sierpinski curve Julia sets. *Topology Proc.*, 30(1):163–179, 2006. Spring Topology and Dynamical Systems Conference.
- [11] Robert L. Devaney, Mónica Moreno Rocha, and Stefan Siegmund. Rational maps with generalized Sierpinski gasket Julia sets. *Topology Appl.*, 154(1):11–27, 2007.
- [12] Dorin E. Dutkay and Palle E. T. Jorgensen. Wavelets on fractals. *Rev. Mat. Iberoamericana*, 22, 2006.
- [13] Dorin E. Dutkay and Kjetil Roysland. The algebra of harmonic functions for a matrix-valued transfer operator. *arXiv:math/0611539*, 2007.
- [14] Dorin E. Dutkay and Kjetil Roysland. Covariant representations for matrix-valued transfer operators. *arXiv:math/0701453*, 2007.
- [15] Dorin Ervin Dutkay. The spectrum of the wavelet Galerkin operator. *Integral Equations Operator Theory*, 2004.
- [16] Dorin Ervin Dutkay and Palle E. T. Jorgensen. Wavelet constructions in non-linear dynamics. *Electron. Res. Announc. Amer. Math. Soc.*, 2005.
- [17] Dorin Ervin Dutkay and Palle E. T. Jorgensen. Hilbert spaces built on a similarity and on dynamical renormalization. *J. Math. Phys.*, 2006.
- [18] Dorin Ervin Dutkay and Palle E. T. Jorgensen. Iterated function systems, Ruelle operators, and invariant projective measures. *Math. Comp.*, 2006.
- [19] Christopher Heil and David F. Walnut, editors. *Fundamental papers in wavelet theory*. Princeton University Press, Princeton, NJ, 2006.
- [20] Palle E. T. Jorgensen. Matrix factorizations, algorithms, wavelets. *Notices Amer. Math. Soc.*, 50, 2003.
- [21] Palle E. T. Jorgensen. *Analysis and probability: wavelets, signals, fractals*, volume 234 of *Graduate Texts in Mathematics*. Springer, New York, 2006.
- [22] Palle E. T. Jorgensen. Certain representations of the Cuntz relations, and a question on wavelets decompositions. 414:165–188, 2006.
- [23] Feng Liu. Diffusion filtering in image processing based on wavelet transform. *Sci. China Ser. F*, 49, 2006.
- [24] John Milnor. Pasting together Julia sets: a worked out example of mating. *Experiment. Math.*, 13(1):55–92, 2004.
- [25] C. L. Petersen and S. Zakeri. On the Julia set of a typical quadratic polynomial with a Siegel disk. *Ann. of Math. (2)*, 159(1):1–52, 2004.
- [26] A. Skodras, C. Christopoulos, and T. Ebrahimi. Jpeg 2000 still image compression standard” iee signal processing magazine. *IEEE Signal processing Magazine*, 18:36–58, Sept. 2001.
- [27] Myung-Sin Song. *Wavelet image compression*. Ph.D. thesis. The University of Iowa, 2006.
- [28] Myung-Sin Song. Wavelet image compression. In *Operator theory, operator algebras, and applications*, volume 414 of *Contemp. Math.*, pages 41–73. Amer. Math. Soc., Providence, RI, 2006.
- [29] Gilbert Strang. *Wavelets from filter banks*. Springer, 1997.
- [30] Gilbert Strang. *Signal processing for everyone*, volume 1739 of *Lecture Notes in Math*. 2000.
- [31] Gilbert Strang and Truong Nguyen. *Wavelets and filter banks*. Wellesley-Cambridge Press, Wellesley, MA, 1996.
- [32] B. E. Usevitch. A tutorial on modern lossy wavelet image compression: Foundations of jpeg 2000. *IEEE Signal processing Magazine*, 18:22–35, Sept. 2001.
- [33] J. S Walker. *A Primer on Wavelets and their Scientific Applications*. Chapman & Hall, CRC, 1999.

(Palle E.T. Jorgensen) DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF IOWA, 14 MACLEAN HALL, IOWA CITY, IA 52242

*E-mail address:* `jorgen@math.uiowa.edu`

(Myung-Sin Song) DEPARTMENT OF MATHEMATICS AND STATISTICS, SOUTHERN ILLINOIS UNIVERSITY EDWARDSVILLE, BOX 1653, SCIENCE BUILDING, EDWARDSVILLE, IL 62026

*E-mail address:* `msong@siue.edu`