

Matrix Factorizations, Algorithms, Wavelets

Palle E. T. Jorgensen

with illustrations by Brian Treadway

Amath story: When I was little, growing up in Denmark, before kindergarten—only there wasn't any then—so before I heard of *The Tinder Box* or *The Ugly Duckling* from the books of Hans Christian Andersen, my father told me about low-pass and high-pass filters. He was a telephone engineer and worked on the filters used in signals transmitted over long cables, just after World War II. The “high” and “low” part of the story refers to frequency bands of the sound signals. Not that this meant much to me at the time. Rather, I was fascinated by the pictures in the electrical engineering journals that were stacked up on the floor next to me, and I spent hours looking at them (that was all there was, there on the floor!), so these pictures¹ of filter design, some in color, occupied me on long Sundays while my dad was building instruments in the living room. Nothing else for me to do! Then, after going to school I forgot all about my dad's explanation of quadrature mirror filters (no wonder!) and they were out of mind for a very long time. I never had any particular reason to think much about them at all, I mean the low-pass frequency bands and all that, but I am sure they in some strange way created a lasting visual impres-

Palle E. T. Jorgensen is professor of mathematics at the University of Iowa. His email address is jorgen@math.uiowa.edu; <http://www.math.uiowa.edu/~jorgen/>.

We gratefully acknowledge illustrations, graphics production, Mathematica programming, suggestions, and typesetting by Brian Treadway. Encouragement and suggestions from David Larson, Yves Meyer, Gil Strang, R. S. Strichartz, Bill Casselman, and Harold Boas inspired several improvements to the original version of the article. Work supported by the U.S. National Science Foundation under grant DMS-0139473.

¹*These very same pictures are now popular in wavelet books; see, e.g., [Mal99], [Dau92, pp. 262 and 315], and [StNg96]. The back cover of [StNg96] offers a variant of the diagrams, and the idea they represent is now getting belated recognition in math circles.*

sion on me. If I heard them mentioned later, I might have been slightly amused but no more than that. Perhaps not before wavelet math, or rather my interests in wavelets in the late 1980s, did all of that stuff about frequency bands gradually resurface from out of a mist recalled from the back reaches of my mind.

It was only after I grew up and matured that I realized how these subband filters define operators in Hilbert space which satisfy all kinds of abstract relations, now known as Cuntz and Cuntz-Krieger relations and thought in some circles to have been invented in 1977 by Cuntz or in 1965 by Dixmier. These are tools from math that I had gotten involved with in the late 1970s for completely different reasons. My impression is that the operator relations that are called the Cuntz relations in math go way back, probably back to before I was born, and they are and have been used every day, and twice on Sundays, ever since by signal processing engineers and others I probably don't even know about—in addition to their extensive use in several areas of math! Our matrix functions from math are actually called polyphase matrices by engineers, and they are scattering matrices in other circles, and quantum gates in physics. In fact a lot of the things we do in operator theory are known and used in other fields but are known under different names and in different ways. And important for different reasons! In any case, they *are* really important and for all kinds of good reasons, not least of which is their rediscovery in operator algebras and in wavelets.

Discussions between engineers and mathematicians, both pure and applied, have always seemed to me somewhat surrealistic. Independent lines of thinking fortuitously converge on a common ground of some algorithms and some matrix tricks, but the strands might well come from vastly different vantage points in their approach to a particular subject, reflected in the multiplicity of terminology that is used, i.e., the same mathematical term being assigned different names by the

different groups, and yet the discussion is about the same fundamental underlying idea. See an abbreviated dictionary at the end of this paper. The diverse approaches share a common central mathematical concept. I think few of us so far really have made a conscious effort to articulate this in a way that rings bells with all the diverse communities: (1) the telephone engineer who is wondering about signals with discrete time and trying to model the transmission of sound in optical cables and working on filtering out noise in the receiving end; (2) the graphics designer from computer science or from computational geometry who is doing subdivisions; (3) the engineer who is building chips for data compression in his/her attempts at digitizing archives of fingerprints stored in analog form or chips for JPEG 2000 implementations; and, finally, (4) the mathematician, perhaps the last one in the bunch, who is searching for wavelets with the best smoothness or wavelets with a shorter pyramid algorithm. Each one of them, the practitioner from one or the other of the communities, invents his or her tools quite independently of the others; and the connections, those that are understood up to now, are only realized with hindsight, often only much later, after the fact. But we have to first look for these connections, and this present attempt is the tip of the iceberg—a start, at least.

What Is a Wavelet? In its simplest form, it is a function ψ on the real line \mathbb{R} such that the doubly indexed family $\{2^{n/2}\psi(2^n x - k)\}_{n,k \in \mathbb{Z}}$ provides a basis for all the functions in a suitable space such as $L^2(\mathbb{R})$. Since $L^2(\mathbb{R})$ comes with a norm and inner product, it is natural to ask that the basis functions be normalized and mutually orthogonal (but many useful wavelets are not orthogonal). The analog-to-digital problem from signal processing concerns the correspondence

$$(1) \quad f(x) \longmapsto c_{n,k}$$

for the representation

$$(2) \quad f(x) = \sum_{n \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} c_{n,k} 2^{n/2} \psi(2^n x - k).$$

We will be working primarily with the Hilbert space $L^2(\mathbb{R})$, and we allow complex-valued functions.

Hence the inner product $\langle f | g \rangle = \int \overline{f(x)} g(x) dx$ has a complex conjugate on the first factor in the product under the integral sign. If f represents a signal in analog form, the wavelet coefficients $c_{n,k}$ offer a digital representation of the signal, and the correspondence between the two sides in (1) is a new form of the analysis/synthesis problem, quite analogous to Fourier's analysis/synthesis problem of classical mathematics. One reason for the success of wavelets is the fact that the algorithms for the problem (1) are faster than the classical ones in the context of Fourier.

Other efficient algorithms include the fast Fourier transform [Wic94] applied to digitized signals. It is based on dyadic scaling, a feature it shares with the hierarchical wavelet algorithms. But the latter have further advantages related to localization (see, e.g., [Mal99] and [Mey00, §11]): At discontinuities or sharp spikes, the edge effects for wavelet algorithms in the analog-to-digital problem are moderate, and they do not build up as in the Gibbs phenomenon with Fourier-based tools. An intrinsic feature of the subdivision scheme of wavelets is that the edge effects are concentrated in a few large wavelet coefficients, allowing us to neglect the rest; the asymptotic gain compared to Fourier series is exponential. A color picture is usually made up of a few homogeneous chunks of different colors or shades, one separated from the others by edges. It is for this reason that the wavelet algorithm is so much more efficient in digitizing graphics than are the alternative Fourier-based algorithms. Similarly, for signals, wavelet analysis breaks up $f(x)$ into its frequency components, with each component in a resolution in time that is matched to its scale.

Resolution and Detail

Mathematically, a *multiresolution* is a telescoping family of closed subspaces in some Hilbert space, doubly infinite but generated by a single operator U and a single subspace \mathcal{V}_0 . The operator U represents some scaling, and \mathcal{V}_0 some fixed resolution, for example, step functions of step size one; the larger spaces in the family represent a finer resolution, and the smaller spaces coarser resolutions. When applied to signal processing or to optics, each resolution space is assigned a band of frequencies.

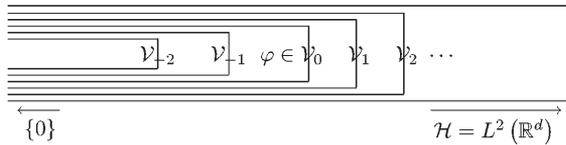
Matrix factorizations have a long history as a tool for designing fast algorithms. Factorizations have been used from the beginning in classical algorithms of signal processing and, more recently, in wavelet subdivision schemes. Since algorithms for quantum information are also based on factorization of unitary matrices, it is not surprising that the subdivision-wavelet algorithms have proved to adapt especially well to the realm of qubits in quantum theory: see, for example, [PBK03] in the case of the Grover search algorithm.

The wavelet algorithms can be cast geometrically in terms of subspaces in Hilbert space which describe a scale of resolutions of some signal or some picture. They are tailor-made for an algorithmic approach that is based upon unitary matrices or upon functions with values in the unitary matrices. Wavelet analysis takes place in some Hilbert space \mathcal{H} of functions on \mathbb{R}^d , for example, $\mathcal{H} = L^2(\mathbb{R}^d)$. An indexed family of closed subspaces $\{\mathcal{V}_n\}_{-\infty < n < \infty}$ such that

$$(3) \quad \mathcal{V}_n \subset \mathcal{V}_{n+1}, \quad \bigcap_{n \in \mathbb{Z}} \mathcal{V}_n = \{0\}, \quad \text{and} \\ \bigvee_{n \in \mathbb{Z}} \mathcal{V}_n = L^2(\mathbb{R}^d)$$

is said to offer a *resolution*. (To stress the *variety* of spaces in this telescoping family, we often use the word *multiresolution*.) Here the symbol \bigvee denotes the closed linear span. In pictures, the configuration of subspaces looks like:

The subspaces of a resolution

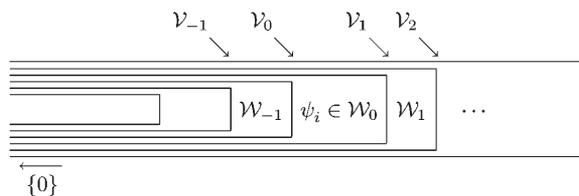


Remember when shopping for a new digital camera: just as important as the resolutions themselves (as given here by the scale of closed subspaces \mathcal{V}_n) are the associated spaces of *detail*. As expected, the details of a signal represent the relative complements between the two resolutions, a *coarser* one and a more *refined* one. **Starting with the Hilbert-space approach to signals**, we are led to the following closed subspaces (relative orthogonal complements):

$$(4) \quad \mathcal{W}_n := \mathcal{V}_{n+1} \ominus \mathcal{V}_n \\ = \{f \in \mathcal{V}_{n+1} : \langle f | h \rangle = 0, h \in \mathcal{V}_n\},$$

and the signals in these intermediate spaces \mathcal{W}_n then constitute the amount of detail which must be added to the resolution \mathcal{V}_n in order to arrive at the next refinement \mathcal{V}_{n+1} . In the diagram below, the intermediate spaces \mathcal{W}_n of (4) represent incremental details in the resolution.

Incremental Detail



The simplest instance of this is the one which Haar discovered in 1910 [Haa10] for $L^2(\mathbb{R})$. There, for each $n \in \mathbb{Z}$, \mathcal{V}_n represents the space of all step functions with step size 2^{-n} , i.e., the functions f on \mathbb{R} which are constant in each of the dyadic intervals $j2^{-n} \leq x < (j+1)2^{-n}$, $j = 0, \dots, 2^n - 1$, and their integral translates, and which satisfy $\|f\|^2 = \int_{-\infty}^{\infty} |f(x)|^2 dx < \infty$. The inner product for Haar is the familiar one,

$$(5) \quad \langle f | h \rangle = \int_{-\infty}^{\infty} \overline{f(x)} h(x) dx,$$

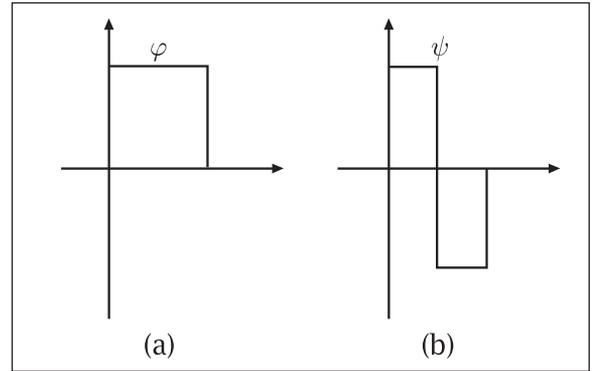


Figure 1. Haar wavelet functions.

and similarly for our present $L^2(\mathbb{R}^d)$, with the modification that the integration is now over \mathbb{R}^d .

An operator U in a Hilbert space is *unitary* if it is onto and preserves the norm or, equivalently, the inner product. Unitary operators are invertible, and $U^{-1} = U^*$ where the $*$ refers to the adjoint. Similarly, the orthogonality property for a projection P in a Hilbert space may be stated purely algebraically as $P = P^2 = P^*$. The adjoint $*$ is also familiar from matrix theory, where $(A^*)_{i,j} = \overline{A_{j,i}}$: in words, the $*$ refers to the operation of transposing and taking the complex conjugate. In the matrix case, the norm on \mathbb{C}^n is $(\sum_k |x_k|^2)^{1/2}$. In infinite dimensions, there are isometries which map the Hilbert space into a proper subspace of itself.

For Haar's case we can scale between the resolutions using $f(x) \mapsto f(x/2)$, which represents a dyadic scaling.

To make it unitary, take

$$(6) \quad U = U_2: f \mapsto 2^{-1/2} f\left(\frac{x}{2}\right),$$

which maps each space \mathcal{V}_n onto the next coarser subspace \mathcal{V}_{n-1} , and $\|Uf\| = \|f\|$, $f \in L^2(\mathbb{R})$. This can be stated geometrically, using the respective *orthogonal projections* P_n onto the resolution spaces \mathcal{V}_n , as the identity

$$(7) \quad UP_nU^{-1} = P_{n-1}.$$

And (7) is a basic geometric reflection of a self-similarity feature of the cascades of wavelet approximations. It is made intuitively clear in Haar's simple but illuminating example, included below. The important fact is that this geometric self-similarity, in the form of (7), holds completely generally. Moreover, it serves as a tool for generating new wavelets and for analyzing them. A crucial observation Haar made in his 1910 paper was that the box-wavelet of Figure 1 is actually singly generated. Without making it explicit, Haar also further noticed a special case of what is now called *multiresolution analysis* (MRA). Haar considered the two functions

$$(8) \quad \varphi = \chi_{[0,1)} \quad \text{and} \quad \psi = \chi_{[0,1/2)} - \chi_{[1/2,1)}$$

shown in Figures 1(a) and 1(b), where we use χ to denote the indicator function. With these two functions, φ and ψ , it is clear that

$$\mathcal{V}_0 = \bigvee \{\varphi(\cdot - k) : k \in \mathbb{Z}\} \quad \text{and} \\ \mathcal{W}_0 = \bigvee \{\psi(\cdot - k) : k \in \mathbb{Z}\},$$

where \bigvee denotes the closed linear span of the functions inside $\{ \}$. Similarly, using the translation operators

$$(T_y f)(x) := f(x - y), \quad f \in L^2(\mathbb{R}), \quad y \in \mathbb{R},$$

and the relation

$$(9) \quad UT_y U^{-1} = T_{2y}, \quad y \in \mathbb{R},$$

we get

$$(10) \quad \begin{aligned} \varphi(x) &= \varphi(2x) + \varphi(2x - 1), \\ \psi(x) &= \varphi(2x) - \varphi(2x - 1), \end{aligned}$$

and, for each $n \in \mathbb{Z}$,

$$(11) \quad \mathcal{V}_n = \bigvee_{k \in \mathbb{Z}} \varphi(2^n x - k),$$

representing the closed subspace generated by all the \mathbb{Z} -translates as specified, and

$$(12) \quad \mathcal{W}_n = \bigvee_{k \in \mathbb{Z}} \psi(2^n x - k).$$

From Haar to Daubechies

While the identities (10) and the properties sketched in (11) are in fact clear from inspection of the shapes in Figure 1, the first surprise in wavelet theory is that smooth wavelet shapes, represented by differentiable functions φ and ψ of compact support, are also possible and with the functions satisfying the exact same resolution properties which were first noticed in a very special (nonsmooth) case by Alfred Haar. We now turn to this crucial issue.

The issue of differentiability of wavelets is a rather large subject. We will only be able to touch on it here. Our viewpoint is that when the support size is specified, then we are able to display a corresponding variety of wavelets. The next step then is to identify the most differentiable specimens in the variety. This is in fact an area with current and exciting research, much of it dictated by applications, but to get started we first need easy matrix formulations which facilitate computations. The tools are somewhat technical. Here is a sample of them, identified by their technical names: (i) the vanishing moment method (based on polynomial factorization), (ii) the joint spectral radius method (a clear-cut test, but difficult to apply), (iii) the dominant eigenvalue test (sketched in Figure 5 below), (iv) the spin vector test. This last one amounts to writing the so-called polyphase matrix function $z \mapsto G(z)$ as a product of matrix functions $zp + p^\perp$ where p is a rank-one projection in some \mathbb{C}^N ; i.e., p is the projection onto some $\mathbf{v} \in \mathbb{C}^N$, $\|\mathbf{v}\| = 1$. By

adjusting the configuration of vectors \mathbf{v} contributing to the product factorization for $G(z)$, the more differentiable wavelets can be identified with a search algorithm. The graphics around Figures 5 and 6 may help the reader to visualize method (iv). Figure 6 picks out a particular sample of configurations of two spin vectors. We explain below how variations of a single unit vector \mathbf{v} in \mathbb{C}^2 describe a variety of dyadic wavelets supported from 0 to 3 on the x -axis, while two independently moving spin vectors, i.e., unit vectors \mathbf{v}_1 and \mathbf{v}_2 in \mathbb{C}^2 , describe the variety of wavelets supported in $[0, 5]$.

Hermitian projections, especially finite-dimensional and one-dimensional ones, along with issues arising from their interactions and compositions, form a chapter of lore in complex geometry and in complex Hilbert space as it is used in quantum theory. This framework describes the wavelet varieties perfectly: First recall that the complex n -dimensional subspaces in \mathbb{C}^N are viewed as, and are by definition, points in the Grassmannian $G(n, N)$. And, starting with Wolfgang Pauli, $n = 1$, $N = 2$, it is popular to identify points in three pairwise isomorphic manifolds, (A)–(C), described as follows (for details see, for example, [BrJo02]): (A) $G(1, 2)$; (B) the two-sphere S^2 , which goes under the name “the Bloch sphere” in physics circles (to Pauli, a point in S^2 represents the state of an electron or of some spin-1/2 particle, and the points in the open ball inside S^2 represent mixed states); and, finally, (C) equivalence classes of unit-vectors in \mathbb{C}^2 , where equivalence of vectors \mathbf{u} and \mathbf{v} is defined by $\mathbf{u} = c\mathbf{v}$ with $c \in \mathbb{C}$, $|c| = 1$. With this viewpoint, a one-dimensional projection p in \mathbb{C}^N is identified with the equivalence class defined from a basis vector, say \mathbf{u} , for the one-dimensional subspace $p(\mathbb{C}^N)$ in \mathbb{C}^N . A nice feature of the identifications $N = 2$ is that if the unit-vectors \mathbf{u} are restricted to \mathbb{R}^2 , sitting in \mathbb{C}^2 in the usual way, then the corresponding real submanifold in the Bloch sphere S^2 is the great circle: the points $(x, y, z) \in S^2$ given by $y = 0$. To Pauli, S^2 , as it sits in \mathbb{R}^3 , helps clarify the issue of quantum observables and states. Pauli works with three spin-matrices for the three coordinate directions, x , y , and z . They represent observables for a spin-1/2 particle. States are positive functionals on observables, so Pauli gets a point in \mathbb{R}^3 as the result of applying a particular state to the three matrices. The pure states give values in S^2 . Recall that pure states in quantum theory correspond to rank-one projections or to equivalence classes of unit-vectors. When this viewpoint is applied to the wavelet formulation, we get an economical way of identifying the essential wavelet numbers, i.e., the masking coefficients for the subdivisions of wavelet theory.

Most of the geometry we recall here carries over mutatis mutandis to unit-vectors in \mathbb{C}^N . In the wavelet case, the number N represents a fixed scaling. Now

the Grassmannian is $G(1, N)$, $N > 2$, but the analogue of the Bloch sphere is a little more complicated. It is worked out in geometry books, such as that of R. O. Wells Jr. [Wel80]. What is considerably more complicated mathematics is the parameterization of the variety of states corresponding to several particles. It involves the notion of *entanglement* from quantum theory. For the present purpose, we use a finite set of projections p corresponding to the special $(zp + p^\perp)$ -factors in a factorization of our polyphase matrix $G(z)$, describing a particular wavelet.

The existence of certain differentiable wavelets was discovered in the 1980s; see especially [Dau92]. I. Daubechies's, Y. Meyer's, and A. Cohen's pioneering discoveries opened a floodgate. Since then, new and powerful methods and techniques have emerged that provide constructive algorithms for optimal choices of resolutions and wavelets. These more recent methods, motivated by signal processing, include (i)–(iv) above, and they give existence generally for signals of *compact support*, represented by compactly supported functions in $L^2(\mathbb{R}^d)$. They are also at the core of the many success stories of wavelet algorithms [Coh03], [Mey00] and of harmonic analysis [HeWe96].

When the first wavelet constructions came out in the early 1980s, their significance was in fact not readily accepted or understood, in some cases not believed. Perhaps the first example of J.-O. Strömberg [Str82] was stillborn! What really opened up the subject and enriched both theory and applications was the connection to signal processing. From that came the multiresolutions, the pyramid algorithms, and the applications to data compression, to still image encoding, and more. Connections to signal processing are made in [Dau92] and especially in [Mal99]. Clearly the geometric relations (10), so transparent from Figure 1(a)–(b), invite the following generalization. It takes the form of subdivision operations with masking coefficients, popular in numerical analysis:

$$(13) \quad \varphi(x) = \sqrt{|\det A|} \sum_{k \in \mathbb{Z}^d} a_k \varphi(Ax - k),$$

$$(14) \quad \psi_i(x) = \sqrt{|\det A|} \sum_{k \in \mathbb{Z}^d} b_k^{(i)} \varphi(Ax - k),$$

$$i = 1, \dots, |\det A| - 1.$$

The numbers (a_k) in (13) are called *masking coefficients* because of the use of (13) in graphics algorithms: there the a -numbers represent the masks in the successive subdivision steps of the algorithm. Here (a_k) and $(b_k^{(i)})$ are scalar sequences indexed by the lattice \mathbb{Z}^d , and A is a fixed $d \times d$ matrix over \mathbb{Z} . We will assume that the eigenvalues λ of A satisfy $|\lambda| > 1$. This generalizes Haar's expansive scaling $x \rightarrow 2x$. It is known generally that

the wavelets of *compact support* may be described this way using specific systems (a_k) and $(b_k^{(i)})$ of *finite* sequences, i.e., sequences which are identically zero outside a finite subset, Λ say, of \mathbb{Z}^d : specifically, $a_k = 0$ for all $k \in \mathbb{Z}^d \setminus \Lambda$, and similarly for the sequence $(b_k^{(i)})_{k \in \mathbb{Z}^d}$, $i = 1, \dots, |\det A| - 1$. Iteration algorithms are used in the solution to the system (13)–(14). When solutions φ and ψ_i to this system can be found in $L^2(\mathbb{R}^d)$, then the starting point of the hierarchical wavelet algorithm is the recursive buildup of the subspaces \mathcal{V}_n and \mathcal{W}_n of (3)–(4) with the use of formulas (19)–(21) below. This can be done with finite matrix algorithms known as subdivision algorithms; see [BrJo02], [Mal99], and [Wic94]. The recursive construction may be visualized in the next example but in a context of fractals.

Example. Cloud Nine as a Reptile. (The name *reptile* refers to a tiling property which is self-reproducing in the sense that the picture repeats itself at all scales.) As an example of (13), take $d = 2$ and

$$A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}. \text{ Then } N = \det A = 5, \text{ and the five}$$

lattice points $D = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \pm 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \pm 2 \end{pmatrix} \right\}$ represent

all the five residue classes for the quotient group $\mathbb{Z}^2/A\mathbb{Z}^2$. The formula $\varphi(x) = \sum_{d \in D} \varphi(Ax - d)$ is then a special case of the scaling equation (13), and it is at the same time a natural extension of Haar's equation (10). It has a solution $\varphi = \chi_{\mathbf{T}}$, and the compact set $\mathbf{T} \subset \mathbb{R}^2$ is the unique solution to the so-called reptile equation $A\mathbf{T} = \bigcup_{d \in D} (\mathbf{T} + d)$; see also Figure 2. It is shown in [BrJo99] that \mathbf{T} tiles \mathbb{R}^2 with translations chosen from the lattice

$$\Lambda = \left\{ \begin{pmatrix} l \\ 2m \end{pmatrix} : l, m \in \mathbb{Z} \right\}. \text{ It follows from this that}$$

the measure of \mathbf{T} is 2. Then \mathbf{T} is a reptile or a generalized Haar wavelet. In general, the measure of a reptile is an integer. It is known, however, that in \mathbb{R}^4 there are 4-by-4 expansion matrices A over \mathbb{Z} for which generalized Haar wavelets do not exist. In 4-D for special matrices A , it may be that the reptiles generated by this scheme might not tile by a lattice which would turn them into Haar wavelets.

Georg Cantor's Chaos

Note that the generalization from the so-called *refinement equation* in the form (10) of Haar to the general case (13) is not just a jump in dimension, from 1 to d —that is not even the main point. The main issue is identifying the L^2 -solutions to (13); then (14) comes along for free. To understand this, consider the seemingly trivial modification of (10) into

$$(15) \quad \varphi(x) = \frac{3}{2} (\varphi(3x) + \varphi(3x - 2)).$$

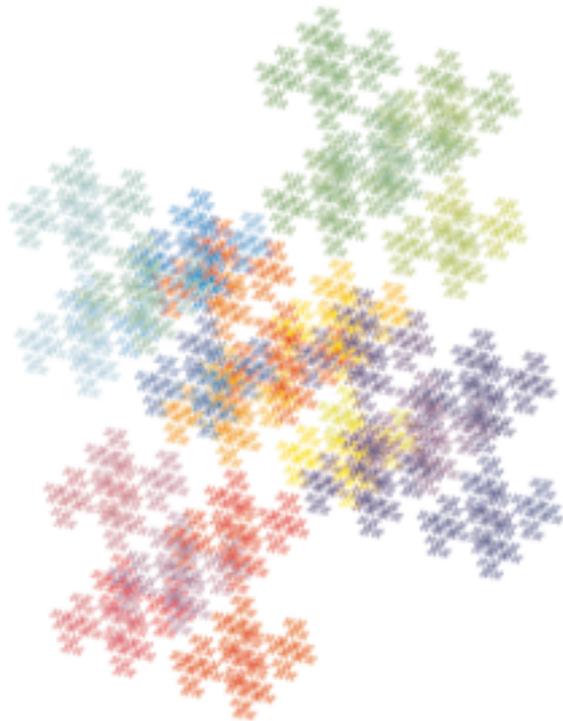


Figure 2. A Rep-Tile: The tile T from the “Cloud Nine” Example.² The picture is a result of iterating five linear maps, all having an expansive scaling matrix of determinant 5. The result is an exotic “tiling” of the plane.³ The boundary of the tile is a fractal, like the self-similar structures that occur in biological systems, in correlations of fluids at critical phase transitions, and in conditions of turbulent flow.

Forgetting functions as solutions φ , try tempered distributions. Then a Fourier transform of (15) yields

$$(16) \quad \hat{\varphi}(\xi) = e^{-i\xi/2} \prod_{n=1}^{\infty} \cos\left(\frac{\xi}{3^n}\right).$$

It can be checked that, up to normalization, this is the unique solution and that φ is the singular Cantor measure and in particular a tempered distribu-

²The name here is from [BrJo99]; the picture represents a dynamical systems orbit of order 9.

³Take a geometric figure, replicate it five times, in the original location, three units to the right, two units above, two units below, and three units to the left. Then shrink the whole collection to the same size as the original was and rotate it by a “knight’s move” angle. Repeat the process any number of times (here seven). If the initial figure is suitably chosen, the figures never overlap, and the collection tiles the plane by the same lattice at each stage of the iteration. Continuously varying colors assigned to the pieces in the order they were created show how the parts interpenetrate at every stage of iteration. See the cover of this issue for another graphical rendition of this fractal tile.

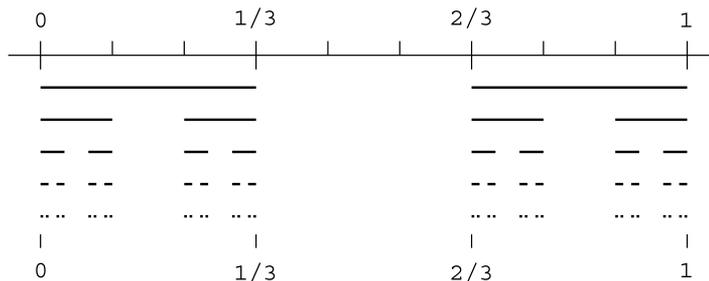


Figure 3. Iteration leading to Cantor-set support of φ in (15).

tion corresponding to the middle-third construction. Hence this φ is the unique measure satisfying

$$(17) \quad \int_{\mathbb{R}} f(x) d\varphi(x) = \frac{1}{2} \left(\int_{\mathbb{R}} f\left(\frac{x}{3}\right) d\varphi(x) + \int_{\mathbb{R}} f\left(\frac{x+2}{3}\right) d\varphi(x) \right)$$

and φ is supported in the set of Figure 3. The use of the Fourier transform in (16) makes perfectly good sense, even if φ is not a locally integrable function on \mathbb{R} . It would make sense even if we only knew a priori that φ was a compactly supported distribution: then for each $\xi \in \mathbb{R}$ we would define $\hat{\varphi}(\xi)$ as the distribution φ applied to the C^∞ -function $x \mapsto e^{-i\xi x}$. In actual fact, the iteration algorithm based on (17) which is illustrated in Figure 3 shows that φ is the Cantor measure. Hence

$$\hat{\varphi}(\xi) = \int e^{-i\xi x} d\varphi(x).$$

For the general theory of affine iteration systems, fractal limits, and their harmonic analysis, the reader is referred to [JoPe98] and [DiFr99].

Groups of Wavelets

The formulas (6) and (9) from Haar carry over to the general case as follows. Consider a $d \times d$ invertible matrix A over \mathbb{Z} . With scaling in the form

$$(18) \quad \begin{aligned} U: f &\mapsto |\det A|^{-1/2} f(A^{-1}x), \\ x &\in \mathbb{R}^d, f \in L^2(\mathbb{R}^d), \end{aligned}$$

we have

$$(19) \quad UT_y U^{-1} = T_{Ay}, \quad y \in \mathbb{Z}^d.$$

Note that $Ay \in \mathbb{Z}^d$ for $y \in \mathbb{Z}^d$ if the matrix A is integral. Moreover, generalizing (11)–(12), we get

$$(20) \quad \mathcal{V}_n = \bigvee \varphi(A^n x - k),$$

$$(21) \quad \mathcal{W}_n = \bigvee_{\substack{k \in \mathbb{Z}^d \\ i=1, \dots, |\det A|-1}} \psi_i(A^n x - k),$$

provided the coefficients in (13)–(14), called masking coefficients, satisfy the following orthogonality relations (where we have set $b_k^{(0)} := a_k$):

$$(22) \quad \sum_{k \in \mathbb{Z}^d} |b_k^{(i)}|^2 = 1$$

for all i (normalization), and

$$(23) \quad \sum_{k \in \mathbb{Z}^d} \overline{b_k^{(i)}} b_{k-Al}^{(j)} = 0$$

for all $i \neq j$, and $l \in \mathbb{Z}^d$ (orthogonality).

Introducing the d -torus as $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$, it is shown in [BrJo02] that the system (22)–(23) is equivalent to specifying a function from \mathbb{T}^d into the $N \times N$ unitary matrices, where $N = |\det A|$. Since the matrix A is fixed and integral, the matrix multiplication on \mathbb{R}^d , $x \rightarrow Ax$, passes to the quotient $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$, and the induced mapping T_A is N -to-1.

We now describe an isomorphism between the wavelet systems (22)–(23) and the *loop group elements*. (The group \mathfrak{G} of measurable functions from \mathbb{T}^d to the matrix group $U_N(\mathbb{C})$ is called the *loop group*.) Using a finite Fourier transform, we get a *function-valued inner product*, for functions on \mathbb{T}^d :

$$(24) \quad \langle p | q \rangle_A(z) = \frac{1}{N} \sum_{\substack{w \in \mathbb{T}^d \\ T_A w = z}} \overline{p(w)} q(w), \quad z \in \mathbb{T}^d,$$

where p and q vary over all scalar-valued functions on \mathbb{T}^d . In Haar's case the sum on the right-hand side in (24) is over $\pm\sqrt{z}$. The application is to the case when these functions are

$$m^{(j)}(z) := \sum_{k \in \mathbb{Z}^d} b_k^{(j)} z^k, \quad j = 1, \dots, N-1,$$

where $z^k := z_1^{k_1} z_2^{k_2} \dots z_d^{k_d}$, and $z \in \mathbb{T}^d$. Using the standard inner product on \mathbb{C}^N , it is therefore natural to extend (24) to the case of vector-valued functions $p: \mathbb{T}^d \rightarrow \mathbb{C}^N$, and revise (24) to

$$(25) \quad \langle p | q \rangle_A(z) = \frac{1}{N} \sum_{\substack{w \in \mathbb{T}^d \\ T_A w = z}} \langle p(w) | q(w) \rangle_{\mathbb{C}^N}.$$

The key step in the identification of the solutions to (22)–(23) with the group \mathfrak{G} of matrix functions $G: \mathbb{T}^d \rightarrow U_N(\mathbb{C})$ is now the following lemma.

Lemma. *The group \mathfrak{G} of all functions $G: \mathbb{T}^d \rightarrow U_N(\mathbb{C})$ from the d -torus into all the $N \times N$ unitary matrices acts transitively on vector functions m on \mathbb{T}^d as follows: $m \mapsto m^G$, where $m^G(z) := G(T_A z) m(z)$, and the pointwise product is matrix times column vector.*

Note (recalling that the inner product (25) takes values in functions on \mathbb{T}^d) that this is a *unitary* action relative to the functional inner product (25) in

the sense that the unitarity identity $\langle m^G | p^G \rangle_A = \langle m | p \rangle_A$ holds for all $G \in \mathfrak{G}$. The term

$$\|m(z)\|^2 = \langle m(z) | m(z) \rangle_{\mathbb{C}^N} = \sum_{k=0}^{N-1} |m_k(z)|^2$$

measures the total contribution to a subband filter system, the N frequency bands being indexed by the cyclic group of order N , and the summation taken over the individual bands. Hence, the lemma states that this total contribution is constant under the specified action of the group \mathfrak{G} of unitary matrix functions, the so-called polyphase matrices. For more details on the terminology of filters and matrix functions, see the list at the end of the paper.

If A is a given scaling matrix, i.e., a $d \times d$ matrix over \mathbb{Z} with eigenvalues $|\lambda| > 1$, and $N := |\det A|$, then the corresponding filters m have N subbands and are considered as functions from \mathbb{T}^d to \mathbb{C}^N . With a slight abuse of notation, we say that m is a *quadrature mirror filter* (QMF) if $\langle m_i(z) | m_j(z) \rangle_A = \delta_{i,j}$, $z \in \mathbb{T}^d$, where $\langle \cdot | \cdot \rangle_A$ is the form in (24). It follows from the lemma and an easy calculation that if m and m' are any two given QMF's corresponding to the same scaling matrix A , then the functions

$$G_{i,j}(z) = \langle m_j(z) | m'_i(z) \rangle_A, \quad z \in \mathbb{T}^d,$$

are matrix entries of a polyphase matrix G , i.e., $G \in \mathfrak{G}$, and $m' = m^G$. In other words, G transforms the first QMF m into the second m' , and G is determined uniquely from m and m' .

Examples

The significance of the approach via matrix functions is that it offers easy formulas for the numbers which encode our wavelet functions, φ , ψ (if the scale number N is 2), and φ , $\psi_1, \dots, \psi_{N-1}$ in general. Haar's two functions φ and ψ are supported in the unit interval $[0, 1]$, and Daubechies's father function φ and mother function ψ are supported in $[0, 3]$. Then there is a next generation, still for $N = 2$, where φ and ψ are both supported in $[0, 5]$. In terms of the matrix function $G(z)$, the case of $[0, 1]$ is the constant function

$$G(z) = H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

the next one is of the form $G(z) = H(zp + p^\perp)$ where p is a rank-1 projection in \mathbb{C}^2 and $p^\perp = I - p$. As p varies, we get the wavelet functions with support in $[0, 3]$. With two rank-1 projections p_1, p_2 in \mathbb{C}^2 and $G(z) = H(zp_1 + p_1^\perp)(zp_2 + p_2^\perp)$, we get the wavelet functions supported in $[0, 5]$. The simplest way of reading off the matrix function G which corresponds to a system of masking coefficients

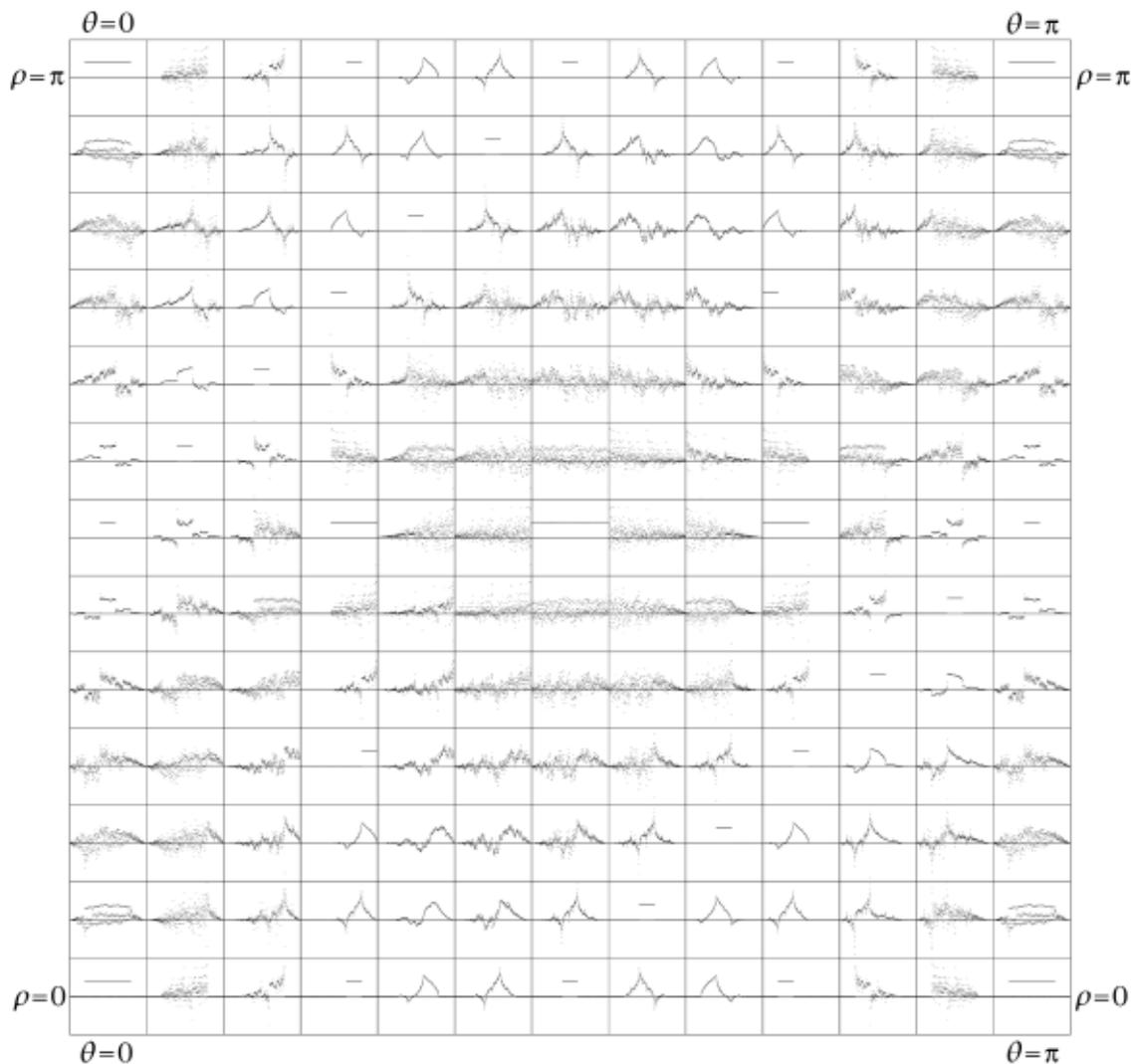


Figure 4. Two-parameter family of wavelet scaling functions.

$\begin{pmatrix} a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 \end{pmatrix}$ is to set $A = \begin{pmatrix} a_0 & a_1 \\ b_0 & b_1 \end{pmatrix}$,

$B = \begin{pmatrix} a_2 & a_3 \\ b_2 & b_3 \end{pmatrix}$, and $G(z) = A + zB$. The reader

will be able to construct the higher-order cases by generalization, or he/she may consult [BrJo02].

The World of the Spectrum

The theme of this section is that central properties of wavelets depend on the spectrum of a certain operator R , named after David Ruelle [Rue69], and also called the wavelet-transfer operator. In the present context, the operator R may be realized on sequences $(x_k)_{k \in \mathbb{Z}^d}$ as follows: $(Rx)_k = \sum_{j \in \mathbb{Z}^d} R_{Ak-j} x_j$ where $R_k = \sum_{j \in \mathbb{Z}^d} \tilde{a}_j a_{j+k}$, the numbers $(a_k)_{k \in \mathbb{Z}^d}$ are the masking coefficients in (13), and A is the expansive matrix which defines our subdivision. Let $\epsilon \in \mathbb{R}_+$. On the Hilbert space \mathcal{H}_ϵ of sequences $(x_k)_{k \in \mathbb{Z}^d}$ such that $\sum_{k \in \mathbb{Z}^d} |x_k|^2 e^{\epsilon|k|} < \infty$,

R acts as a trace class operator. Its spectrum is independent of ϵ . See (32) below for details.

Illustration: Spread out over the area of Figure 4 is a family of wavelets, each generating function (scaling function) φ determined by two parameters. Once we have the algorithm for the scaling function, an extra step automatically gets us the wavelet itself. There is a coordinate system of wavelets on the interval $[0, 5]$ with their associated scaling functions; see Figure 4. Since vectors move invariantly under unitary transformations, it is reasonable to expect that varieties of wavelets may be parameterized in a way that is analogous to how we do the classical unitary groups. This will be illustrated in the following graphics experiment.

Wavelets on the Interval $[0, 5]$ and Their Associated Scaling Functions

The scaling functions φ for these wavelets, one for each sample point represented in Figure 4, are pictured in this wall decoration, i.e., in Figure 4. The algorithm used in generating each of the functions

φ in the little rectangles of Figure 4 will be outlined in the discussion below (26). However, Figure 4 is in fact a sampling from a two-parameter variety of orthogonal wavelets. (It is only a variety, because there are isolated points where the strict orthogonality property degenerates!) More precisely, we are looking at a genus-2 surface, but with deletion of these degenerate or singular points: In the overall landscape of Figure 4, these points are visibly represented by little flat scaling functions φ with support of length 3 or 5. (Remember the “honest” Haar wavelets are supported on an interval of unit length!) What is more significant, and also clearly visible in Figure 4, is the fact that “most of” the functions φ in the variety are pretty bad! The nicest that can be said about them is that they are L^2 , and that is not all bad. (Remember the Devil’s staircase function from Figure 3 with its support on the Cantor set. As a scaling function, this “horrible” φ is not even locally integrable.)

So how *do* we find the nicer wavelets in a variety, those with more derivatives? There are two popular approaches: (i) Following [Dau92], look for vanishing moments. They can be found from factorization of certain polynomials, the Daubechies polynomials. Or: (ii) do factorization of the polyphase matrix! For that, there is a handy operator R ; it is a transfer operator (see (32) below). The spectrum of R is pictured in Figure 5; the eigenvalues of R are functions of the two parameters that refer to Figure 4. In the valleys of the mountainous landscape of level-surfaces from Figure 5, i.e., the eigenvalues as they depend on two parameters, is where we pick out the nicest of the functions in the chaos of possibilities from the wavelets in Figure 4. The operator R is nowadays called the wavelet-transfer operator, but it was studied first by David Ruelle in the 1960s in connection with a completely different problem (the phase transition question of quantum statistical mechanics [Rue69]) and later in chaos theory [Rue02].

It is convenient for each $k = 1, 2, \dots$, to look at the family of all wavelet functions φ and ψ supported in the interval $[0, 2k + 1]$ on the x -axis. Within the family of all orthogonal wavelets supported in $[0, 2k + 1]$, we look for vanishing moments of order j where $1 \leq j \leq k$. We know how to find subvarieties $S_j \supset S_{j+1}$, where S_j consists of the functions ψ in the family such that ψ has vanishing moments of order j . As j increases, S_j contains wavelets of higher and higher orders of differentiability. Testing for smoothness in each S_j can be done with the so-called joint spectral radius (JSR) test. It is a JSR computed for two noncommuting square matrices built in a simple way from the system of masking coefficients $a_0, a_1, a_2, \dots, a_{2k+1}$. But since the square matrices are noncommuting, the JSR is hard to compute. If $k = 2$, there

are six masking coefficients, and the two noncommuting square matrices are

$$\frac{1}{\sqrt{2}} \begin{pmatrix} a_0 & 0 & 0 & 0 & 0 \\ a_2 & a_1 & a_0 & 0 & 0 \\ a_4 & a_3 & a_2 & a_1 & a_0 \\ 0 & a_5 & a_4 & a_3 & a_2 \\ 0 & 0 & 0 & a_5 & a_4 \end{pmatrix}$$

and

$$\frac{1}{\sqrt{2}} \begin{pmatrix} a_1 & a_0 & 0 & 0 & 0 \\ a_3 & a_2 & a_1 & a_0 & 0 \\ a_5 & a_4 & a_3 & a_2 & a_1 \\ 0 & 0 & a_5 & a_4 & a_3 \\ 0 & 0 & 0 & 0 & a_5 \end{pmatrix}.$$

An alternative test for smoothness involves the Ruelle operator R . For $k = 2$, the interval is $[0, 5]$, and for real-valued wavelets, the full variety is given by the (θ, ρ) parameters. The eigenvalues of $R_{\theta, \rho}$ may be ordered as follows:

$$1 \geq |\lambda_1(\theta, \rho)| \geq |\lambda_2(\theta, \rho)| \geq \dots$$

There is a simple known function s from \mathbb{R}_+ to \mathbb{R}_+ such that if $|\lambda_1(\theta, \rho)| < c$, then the two functions $\varphi_{\theta, \rho}$ and $\psi_{\theta, \rho}$ are in the Sobolev space $\mathcal{H}^{s(c)}$. Recall that the Sobolev exponent $s(c)$ is a measure of differentiability. But algorithms for calculating the best s are few and not especially efficient. Nonetheless, continuity and differentiability of the wavelet functions are of critical importance in applications: The a priori estimates which give the best wavelet algorithm in JPEG 2000 are done in spaces of functions of bounded variation, and they depend on the smoothness of the wavelets that are used.

In Chapters 1–2 in [BrJo02] and in [Tre01], it is pointed out that families of compactly supported wavelets admit a group-theoretic formulation. When this idea is specialized to the case of multiresolution wavelets which have both the scaling function (father function φ) and the wavelet generator (mother function ψ) itself supported in the fixed interval from 0 to 5, then the full variety of possibilities may be described by two independently varying unit vectors in \mathbb{C}^2 , in coordinates $\mathbf{v} = (v_1, v_2) \in \mathbb{C}^2$, $\|\mathbf{v}\|^2 = |v_1|^2 + |v_2|^2 = 1$. Unit vectors in \mathbb{C}^2 define pure quantum-mechanical states. The latter may be parameterized by points on the (Bloch) sphere S^2 , i.e., points $(x, y, z) \in \mathbb{R}^3$, $x^2 + y^2 + z^2 = 1$. If σ_x , σ_y , and σ_z are the Pauli spin matrices $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, and $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ in the three coordinate directions, then $(x, y, z) = (\langle \mathbf{v} | \sigma_x \mathbf{v} \rangle, \langle \mathbf{v} | \sigma_y \mathbf{v} \rangle, \langle \mathbf{v} | \sigma_z \mathbf{v} \rangle)$ is in S^2 if and only if $\|\mathbf{v}\| = 1$. For example, $\mathbf{v} = (\cos \theta, \sin \theta)$ in \mathbb{C}^2 corresponds to the point $(\sin 2\theta, 0, \cos 2\theta)$ on S^2 . Hence, viewing S^2 as embedded in \mathbb{R}^3 , the vector moves on a great circle on S^2 in the (x, z) -plane. Thus in the example with two vectors of the form $(\cos \theta, \sin \theta)$, $(\cos \rho, \sin \rho)$ with the parameters θ

and ρ varying independently, displayed in [BrJo02, (1.2.3)-(1.2.4)], the dependence of the masking coefficients $a_k(\theta, \rho)$ on 2θ and 2ρ is as expected: the wavelet parameter is not the vector \mathbf{v} itself, but rather the subspace generated by \mathbf{v} , or the projection onto \mathbf{v} , so we get the same wavelet masking coefficients for \mathbf{v} and $-\mathbf{v}$. The coefficients a_k are real-valued for a similar reason. The two functions φ and ψ depend on variations in (θ, ρ) periodically, with period π in both parameters, so the whole variety may be represented in a square $[0, \pi) \times [0, \pi)$. This is illustrated graphically in a supplement to [BrJo02], http://www.math.uiowa.edu/~jorgen/wavelet_motions.pdf. It may be used as a flip-book of wavelets, displaying a moving picture of their variation along the path shown in the guide diagram in Figure 6. This flip-book is thus a sort of stationary variant of Wim Sweldens's Wavelet Cascade (Java) Applet <http://cm.bell-labs.com/who/wim/cascade/index.html>. With the latter, the viewer may explore changes of scaling and wavelet functions with free movements of the computer mouse, but the variation is limited to the "4-tap" wavelets, i.e., those generated by a single rank-1 projection (spin vector) and supported in the interval $[0, 3]$, while [BrJo02] and its flip-book supplement deal with a bigger "6-tap" family, i.e., with support in $[0, 5]$, and generated by two independent spin vectors.

While the functions φ and ψ are square-integrable, they are continuous on $[0, 5]$ only for (θ, ρ) in a periodic subset of $\mathbb{R} \times \mathbb{R}$. (This subset is often identified by vanishing moments or by spectral conditions such as the JSR test; see [CGV99].) If we ask whether φ and ψ , as points in a function space, show continuous dependence on (θ, ρ) , then that function-space continuity must be measured in mean square, i.e., in the metric defined by the L^2 -norm of functions on $[0, 5]$. In this metric, continuity follows from [BrJo02, Theorem 2.5.8]; see details below. Actually there are a finite number of exceptions to the L^2 -continuous dependence on (θ, ρ) . They occur when the wavelet is one of the degenerate Haar cases. These are the Haar wavelets which define not strict orthonormal bases but only tight frames. They have φ of the form $\varphi = c\chi_I$, where I is a subinterval of $[0, 5]$ of length 3 or 5, and where $c = \frac{1}{3}$ in the first case, and $c = \frac{1}{5}$ in the second. Find them on Figure 4. If the reader follows the moving wavelet pictures (on a printout, or on the screen), we hope he/she will get some intuitive ideas of fundamental wavelet relationships. The rigorous mathematics relating the various continuity properties to cascade approximation, to moments, and to spectral estimates is covered in much more detail in the references below, especially [BrJo02] and [CGV99].

Once the masking coefficients a_0, a_1, \dots, a_5 are specified as functions of the parameters θ and ρ , the computation of cascade approximants of the scaling function φ is done with a series of Mathematica operations. The algorithm is designed to start with a Haar function, and the limit of the iteration will then be a scaling function $\varphi = \varphi^{(\theta, \rho)}$ depending on the two rotation angles θ, ρ , and satisfying

$$(26) \quad \varphi^{(\theta, \rho)}(x) = \sqrt{2} \sum_{k=0}^5 a_k^{(\theta, \rho)} \varphi^{(\theta, \rho)}(2x - k).$$

The reptile features of the algorithm have the effect of producing fast cascading approximations to the limit function $\varphi^{(\theta, \rho)}$. The algorithm of the solution $\varphi^{(\theta, \rho)}$ to the scaling identity (26) then proceeds as follows (see [Jor01], [Tre01], [BrJo02, §1.2] for details). The relation (26) is interpreted as giving the values of the left-hand $\varphi^{(\theta, \rho)}$ by an operation performed on those of the $\varphi^{(\theta, \rho)}$ on the right. A binary digit inversion transforms this into

$$\mathbf{f}'_{k+1}(x) = \mathbf{A}\mathbf{f}_k(x),$$

where \mathbf{A} is the 2×3 matrix

$$\mathbf{A}_{p,q} = \sqrt{2} a_{5+p-2q}^{(\theta, \rho)} = \sqrt{2} \begin{pmatrix} a_4^{(\theta, \rho)} & a_2^{(\theta, \rho)} & a_0^{(\theta, \rho)} \\ a_5^{(\theta, \rho)} & a_3^{(\theta, \rho)} & a_1^{(\theta, \rho)} \end{pmatrix}$$

constructed from the coefficients in (26), and \mathbf{f}_j and \mathbf{f}'_j are the vector functions

$$\mathbf{f}_j(x) = \begin{pmatrix} \varphi^{(\theta, \rho)}\left(x - \frac{2}{2^j}\right) \\ \varphi^{(\theta, \rho)}\left(x - \frac{1}{2^j}\right) \\ \varphi^{(\theta, \rho)}(x) \end{pmatrix},$$

$$\mathbf{f}'_j(x) = \begin{pmatrix} \varphi^{(\theta, \rho)}(x) \\ \varphi^{(\theta, \rho)}\left(x + \frac{1}{2^j}\right) \end{pmatrix}.$$

Iterations of this operation give values of an approximation to $\varphi^{(\theta, \rho)}$ on successively finer dyadic grids in the x variable.

How To

For an implementation of this computation in Mathematica, we let `loctwont` stand for the transpose of the coefficient matrix \mathbf{A} ,⁴ and normalize the Mathematica variables `a0`, `a1`, etc., to include the factor of $\sqrt{2}$ that appears in the cascade iteration:

```
loctwont[θ_, ρ_] :=
N[Transpose[{{a4[θ, ρ], a2[θ, ρ], a0[θ, ρ]}, {a5[θ, ρ], a3[θ, ρ], a1[θ, ρ]}]]

cascadestep[phitable_, θ_, ρ_] :=
Flatten[Partition[Flatten[{0, 0, phitable, 0, 0}], 3, 1].loctwont[θ, ρ]]

wavelet[phistart_, itercount_, θ_, ρ_] :=
Transpose[Table[i (2^(-itercount)), {i, 0, 5 (2^itercount) - 5}],
Nest[cascadestep[#, θ, ρ] &, phistart, itercount]]
```

⁴So named by contraction of "local two-by-n transpose."

Each cascade step works with the list of values from the previous step (`phitable`), pads it with zeroes on left and right, works it up into a matrix by overlapping divisions (the `Partition` operation), does a matrix multiplication with a matrix of masking coefficients (`loctwont`), and reduces (with `Flatten`) the resulting matrix to a list again. At each stage the implicit grid spacing is halved, and at the specified final iteration the values of these grid points are associated with the elements of the list (using `Table` and `Transpose`), so that a plot of the scaling-function approximant can be made with other Mathematica operations such as `ListPlot`.

A direct implementation in Mathematica of the computation of the wavelet ψ from the scaling function φ by the formula [BrJo02, (2.5.25)] (with the functions and coefficients depending on the rotation angle parameters θ, ρ),

$$\psi^{(\theta, \rho)}(x) = \sqrt{2} \sum_{k=0}^5 (-1)^k a_{5-k}^{(\theta, \rho)} \varphi^{(\theta, \rho)}(2x - k),$$

is (again with $\sqrt{2}$ subsumed in `a0, a1, ...`)

```
embedwavelet[phistart_, itercount_, theta_, rho_] :=
  Flatten[{Table[0, {i, -(5 (2^itercount) - 5), -1}], Nest[cascadestep[#, theta, rho] &,
    itercount], Table[0, {i, (5 (2^itercount) - 5) + 1, 2 (5 (2^itercount) - 5)}]}]
avec[theta_, rho_] := {a0[theta, rho], a1[theta, rho], a2[theta, rho], a3[theta, rho], a4[theta, rho], a5[theta, rho]}
waveletpsi[phistart_, itercount_, theta_, rho_] :=
  Transpose[{Table[i (2^(-itercount - 1)), {i, 0, 2 (5 (2^itercount) - 5)}],
    Table[Sum[((-1)^(5 - k) avec[theta, rho][[k + 1]] embedwavelet[phistart, itercount,
      rho][[j + k (2^itercount - 1)]]], {k, 0, 5}], {j, 1, 2 (5 (2^itercount) - 5)}]}
```

But using the `Table` and `Sum` operations in this way is inefficient for two reasons. First, the sum has, after suitable rearrangement, the form of a sequence correlation that can be implemented with a fast Fourier transform. And second, the way the operations above are interpreted by Mathematica is that the same scaling function (`embedwavelet`) is computed repeatedly for each term of the sum rather than being computed once and saved. Both of these inefficiencies can be remedied by the use of the `ListCorrelate` operation, which Mathematica implements internally by Fourier-transform methods.

```
correlatewavelet[phistart_, itercount_, theta_, rho_] :=
  Flatten[Transpose[Partition[PadLeft[PadRight[Nest[cascadestep[#, theta, rho] &,
    itercount], 6 (2^itercount) - 6], 11 (2^itercount) - 11], (2^itercount) -
signavec[theta_, rho_] := {-a0[theta, rho], a1[theta, rho], -a2[theta, rho], a3[theta, rho], -a4[theta, rho], a5[theta, rho]}
correlatewaveletpsi[phistart_, itercount_, theta_, rho_] :=
  Transpose[{Table[i (2^(-itercount - 1)), {i, 0, 11 (2^itercount) - 12}],
    Flatten[Transpose[Partition[ListCorrelate[signavec[theta, rho],
      correlatewavelet[phistart, itercount, theta, rho], {1, 1}, 0], 11]}]}]
```

The necessary rearrangement amounts to grouping the list of values into a matrix and transposing the matrix. This is done, at both ends of the

computation, with the `Partition` and `Transpose` operations, followed by `Flatten` to return to an ungrouped list. Note also the use of `PadLeft` and `PadRight` to add zeroes to keep the different rows of the matrix from getting mixed in the `ListCorrelate` operation; these could have been used in the direct implementation as well, where zero-padding was needed at the ends of the sum. The Fourier-transform method used in Mathematica's implementation of the `ListCorrelate` operation gives results numerically identical to those obtained by the direct calculation with the `Table` and `Sum` operations, thanks in part to Mathematica's implementation of `ListCorrelate` that uses a real transform method on real data. For the wavelet functions computed for the flip-book, the method using `ListCorrelate` works about 500 times faster than the method using `Table` and `Sum`.

We present a guide diagram in Figure 6 corresponding to the (θ, ρ) plane of Figures 4 and 5, with dots for the point (and its periodic replicas) whose wavelet function and scaling function are displayed on the current page and line. The square outlined in black is the fundamental region $[0, \pi) \times [0, \pi)$, shown with labels in Figure 6. There is a path, part of which represents a vanishing-moment curve where the wavelet function $\psi(x)$ satisfies the moment condition $\int x \psi(x) dx = 0$, and the masking coefficients a_0, \dots, a_5 satisfy the equivalent divisibility condition $m_0(z) = 2^{-\frac{1}{2}} \sum_{k=0}^5 a_k z^k = (z + 1)^2 p(z)$ for some polynomial p . The viewer familiar with [Jor01] may find amusement in watching for the appearance in the sequence of the familiar Daubechies wavelets, as well as the single point in the sequence where the additional moment and divisibility conditions $\int x^2 \psi(x) dx = 0$ and $m_0(z) = (z + 1)^3 q(z)$ are met. In Figure 4 you see variations of these wavelets, as the parameters move around and cover a period-square.

The part of the spectrum of the Ruelle operator R in (32) which determines the shape of each wavelet is a finite set of points. As each point in this part of the spectrum is a function of the two parameters, the eigenvalues may be ordered and presented in the sketch as functions of two variables. The graphs represent surfaces starting with a top eigenvalue (a Perron-Frobenius eigenvalue equal to 1 for all values of the parameters). One of the surfaces, two steps down from the top, is pictured in Figure 5. There is a constant eigenvalue $1/2$ that gives the appearance of mountains rising from lakes or Christopher Columbus's hat or a body in bathwater, if you prefer! And then the top eigenvalue equal to 1 is the ceiling. Other eigenvalue surfaces in this series are shown in figures within the book [BrJo02], and one surface with branch cuts, used as a sort of frontispiece, looks like "Half Dome" in Yosemite Valley. This closed-form eigenvalue, when compared

to the eigenvalue surfaces sorted by absolute value, coincides with the second eigenvalue from the top in part of the parameter range and with smaller eigenvalues in other parts of the range. The parameters representing the lower parts of the surfaces are important: for example, they give algorithms that are faster. And wavelets that are more regular tend to be located in the region of the period-square which have a concentration of low-lying spectrum. Hence the flat portions of the “bathwater” surface signify an abundance of “good” wavelets.

Haar Meets Cantor

The scaling identity (13) depends on a choice of masking coefficients $\{a_k\}_{k \in \mathbb{Z}^d}$. This suggests a strategy for finding the scaling function (also called the father function) φ as a solution to an operator identity: the idea is to try to get φ as the limit of an iteration of the following operator:

$$(27) \quad S: f \mapsto \sqrt{N} \sum_{k \in \mathbb{Z}^d} a_k f(Ax - k);$$

i.e., $\varphi = \lim_{n \rightarrow \infty} S^n f$ with a suitable choice of starting point f for the iteration. As suggested by (13), an integration shows that a necessary condition for this to work is that

$$(28) \quad \sum_{k \in \mathbb{Z}^d} a_k = \sqrt{N}.$$

This normalization is indeed satisfied for the two examples (10) and (15). In the first one (10), $d = 1$, $A = N = 2$, and the solution φ is the box-function φ of Haar which is certainly in $L^2(\mathbb{R})$. In the second example (15) of Cantor, $d = 1$, $A = N = 3$, and the solution φ is now the Cantor measure. This measure, while a solution, is a singular measure, and, in fact, it is not even represented locally by an integrable function.

To understand the dichotomy between the case of Haar and that of Cantor, introduce the generating function m_0 for the masking coefficients in (13), i.e., $m_0(z) = \sum_{k \in \mathbb{Z}^d} a_k z^k$. For simplicity, specialize to $d = 1$. It will be convenient for us to consider m_0 as a function on \mathbb{T} by restriction, viewing the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ as a circle in \mathbb{C} , i.e., as $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. Then the normalization condition (28) reads

$$(29) \quad m_0(1) = \sqrt{N}.$$

The other conditions (22)–(23) on the masking coefficients, i.e.,

$$(30) \quad \sum_{k \in \mathbb{Z}} \bar{a}_k a_{k-lN} = \delta_{0,l}, \quad l \in \mathbb{Z},$$

then take the functional form

$$(31) \quad \frac{1}{N} \sum_{\substack{w \in \mathbb{T} \\ w^N = z}} |m_0(w)|^2 = 1.$$

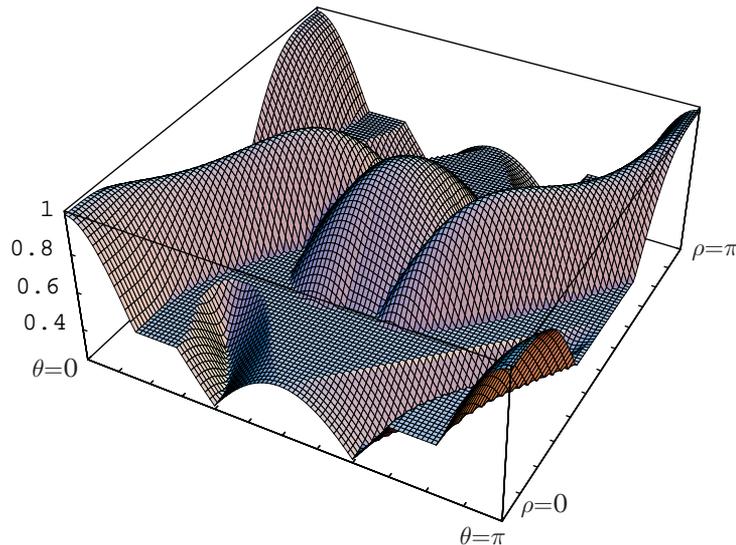


Figure 5. Eigenvalue of two-parameter family of Ruelle operators.

But this is an orthogonality condition, stated in either one of its two equivalent forms, and it is necessary for ensuring orthogonality in the ultimate construction of a wavelet basis for $L^2(\mathbb{R})$. In Haar’s example, $N = 2$ and $m_0(z) = \frac{1+z}{\sqrt{2}}$. Setting $z = e^{-i\xi}$, one finds that (31) is equivalent to the familiar trigonometric identity $\cos^2 \xi + \sin^2 \xi = 1$. But, for Cantor’s example, $N = 3$ and $m_0(z) = \frac{\sqrt{3}}{2}(1+z^2)$, and now $\frac{1}{3} \sum_{w^3=z} |m_0(w)|^2 = 3/2$.

Stated differently, if an operator R is defined as acting on functions h on \mathbb{T} by

$$(32) \quad (Rh)(z) = \frac{1}{N} \sum_{\substack{w \in \mathbb{T} \\ w^N = z}} |m_0(w)|^2 h(w),$$

then the formula (31) takes the form $R\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ denotes the function on \mathbb{T} which is constant and equal to 1. In the analysis of [Dau92] and [BrJo02], the operator R from (32) is called the wavelet transfer operator, or the Ruelle operator, and $\mathbf{1}$ is called the Perron-Frobenius eigenfunction. Hence the significant distinction between the case of Haar and that of Cantor is the numerical size of the Perron-Frobenius eigenvalue: It is 1 in the first case and 3/2 in the second. The operator R in (32) is called the *wavelet transfer operator* because of the analogy to the probabilistic view of the Perron-Frobenius matrix setting and because of the probability content of condition (31): Looking at the right-hand side in (32), for each point $z \in \mathbb{T}$ there are N distinct solutions w to $w^N = z$, and each of the corresponding numbers

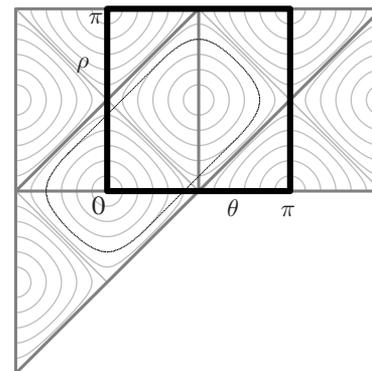


Figure 6. Guide diagram.

$\frac{1}{N} |m_0(w)|^2$ signifies conditional probabilities for transfer from some w to z , where w is one of the N possibilities. This random-walk model helps us understand and compute the spectral theory of R , as pointed out in [BrJo02].

The Ruelle transfer operator R “counts” the accumulated contribution to the cascade approximation, i.e., to the scaling function φ , when the limit exists. But unless the Perron-Frobenius eigenvalue $\lambda_{\text{PF}} = 1$, the limit “escapes” out from $L^2(\mathbb{R})$. In the middle-third-Cantor case, $\lambda_{\text{PF}} = 3/2$, which accounts for φ in this case being a singular measure, not represented by a locally integrable function on \mathbb{R} . The Hausdorff dimension d in Cantor’s case can be shown to be $d = \frac{\log 2}{\log 3}$.

Perron-Frobenius and Stability of Wavelets

If the wavelet transfer operator R in (32) is considered in the context of Lipschitz functions on \mathbb{T} , $\text{Lip}(\mathbb{T})$, then it has spectral properties which closely mirror those which are familiar classically from Perron and Frobenius for finite positive matrices. Let $R = (r_{i,j})_{i,j=1}^n$, $r_{i,j} \geq 0$. A classical theorem [Per07] states that each such matrix R has a positive eigenvalue λ_{PF} which agrees with the spectral radius of R . Specializing to $\lambda_{\text{PF}} = 1$, Perron and Frobenius showed that there are positive vectors ν and v , ν a row-vector and v a column-vector, such that $\nu R = \nu$, $Rv = v$, and $\nu \cdot v = 1$. The same turns out to hold for R in (32); only now ν is the Dirac measure at $z = 1$, called δ_1 , and v is the constant function $\mathbf{1}$ on the torus \mathbb{T} . Imitating what is known for finite matrices, we get a reduced form of the Ruelle operator R acting on the subspace $\text{Lip}_{<1} := \{h \in \text{Lip}(\mathbb{T}) : \delta_1(h) = 0\}$. We say that R has *Perron-Frobenius spectrum* if the spectral radius of R , restricted to $\text{Lip}_{<1}$, is strictly less than 1.

Introducing the Lipschitz norm

$$\begin{aligned} & \|p(e^{i\xi})\|_{\text{Lip}(\mathbb{T})} \\ &= |p(1)| + \sup_{-\pi < \xi < \eta < \pi} \frac{|p(e^{i\xi}) - p(e^{i\eta})|}{|\xi - \eta|}, \end{aligned}$$

we have:

Theorem. [BrJo02, Theorem 2.5.8] *Let m and m' be in $\text{Lip}(\mathbb{T})$ and suppose they both satisfy (29) and (31). Suppose further that each of the transfer operators R_m and $R_{m'}$ has Perron-Frobenius spectrum. Let $\lambda_{\text{red}}(m) (< 1)$ be the reduced spectral radius of R_m . Then there is a constant C independent of m and m' such that the associated scaling functions (father functions) φ_m and $\varphi_{m'}$ satisfy the following Lipschitz estimate:*

$$\|\varphi_m - \varphi_{m'}\|_{L^2(\mathbb{R})} \leq C \frac{1}{\sqrt{1 - \lambda_{\text{red}}(m)}} \|m - m'\|_{\text{Lip}(\mathbb{T})}.$$

On the other hand, the space Lip is not large enough as repository of the filter functions in the study of limits of wavelets. The wavelet ψ_S on \mathbb{R} with $\hat{\psi}_S = \chi_{(-\pi, -\frac{\pi}{2}] \cup (\frac{\pi}{2}, \pi]}$ (i.e., frequency localized), named after Shannon, has the low-pass and high-pass filter functions $\frac{1}{\sqrt{2}} m_0(e^{-i\xi}) = \chi_{[-\frac{\pi}{2}, \frac{\pi}{2})}(\xi)$ and $\frac{1}{\sqrt{2}} m_1(e^{-i\xi}) = \chi_{[\frac{\pi}{2}, \frac{3\pi}{2})}(\xi)$, and they are clearly not in $\text{Lip}(\mathbb{T})$. Nonetheless, there is a “law of large numbers for wavelets”: Let some “nice” (compactly supported) wavelet ψ be given. Two general theorems of Aldroubi et al. establish the following “limit laws”:

$$1. \underbrace{(\psi * \dots * \psi)}_{n \text{ times}} \xrightarrow{n \rightarrow \infty} \psi_S \text{ in } L^2(\mathbb{R}) \text{ with}$$

renormalization, and

$$2. \text{ Let } \psi_{D_n} \text{ denote the Daubechies wavelets given by the Daubechies polynomials } P_n \text{ [Dau92]. Then}$$

it was shown in [AbAl95] that $\psi_{D_n} \xrightarrow[n \rightarrow \infty]{L^2(\mathbb{R})} \psi_S$.

Conclusions

Wavelet analysis and wavelet applications offer a rich variety of interdisciplinary adventures. What is especially striking is the fact that crucial mathematical steps seem to get unexpectedly discovered in diverse communities of engineers, scientists, or mathematicians. If each of the groups of practitioners would make more of an effort to learn about the other, much would be gained. In wavelet theory, it has been exciting to observe the exchanges of ideas in signal-processing engineering. The mathematical developments there might have had a different aim, but nonetheless they have offered key ideas to mathematics, which have given new life to the subject of wavelets and, more generally, to harmonic analysis and approximation theory.

Terminology and Dictionary

- **multiresolution—real world:** a set of band-pass-filtered component images, assembled into a mosaic of resolution bands, each resolution tied to a finer one and a coarser one.
- mathematics:** used in wavelet analysis and fractal analysis, multiresolutions are systems of closed subspaces in a Hilbert space, such as $L^2(\mathbb{R})$, with the subspaces nested, each subspace representing a resolution, and the relative complement subspaces representing the detail which is added in getting to the next finer resolution subspace.
- **matrix function:** a function from the circle, or the one-torus, taking values in a group of N -by- N complex matrices.
- **subband filter:—engineering:** signals are viewed as functions of time and frequency, the frequency function resulting from a transform of

the time function; the frequency variable is broken up into bands, and up-sampling and down-sampling are combined with a filtering of the frequencies in making the connection from one band to the next.

—*wavelets*: scaling is used in passing from one resolution V to the next; if a scale N is used from V to the next finer resolution, then scaling by $\frac{1}{N}$ takes V to a coarser resolution V_1 , represented by a subspace of V , but there is a set of functions that serve as multipliers when relating V to V_1 , and they are called subband filters.

- **cascades**:—*real world*: a system of successive refinements which pass from a scale to a finer one, and so on; used, for example, in graphics algorithms: starting with control points, a refinement matrix and masking coefficients are used in a cascade algorithm, yielding a cascade of masking points and a cascade approximation to a picture.

—*wavelets*: in one dimension the scaling is by a number and a fixed simple function, for example, of the form $\int_0^1 \square$, is chosen as the initial step for the cascades. When the masking coefficients are chosen, the cascade approximation leads to a scaling function.

- **scaling function**: a function or a distribution φ , defined on the real line \mathbb{R} , which has the property that for some integer $N > 1$, the coarser version $\varphi\left(\frac{x}{N}\right)$ is in the closure (relative to some metric) of the linear span of the set of translated functions $\dots, \varphi(x+1), \varphi(x), \varphi(x-1), \varphi(x-2), \dots$.
- **orthogonality in the strict sense**: for $\psi \in L^2(\mathbb{R})$ to be a *dyadic wavelet*, the double-indexed family of functions $\psi_{j,k}(x) := 2^{j/2}\psi(2^jx - k)$, $j, k \in \mathbb{Z}$, must be a basis for $L^2(\mathbb{R})$; i.e., every $f \in L^2(\mathbb{R})$ must admit a unique representation

$$(33) \quad f(x) = \sum_{j,k \in \mathbb{Z}} c_{j,k} \psi_{j,k}(x).$$

If the system $\{\psi_{j,k} : j, k \in \mathbb{Z}\}$ is further assumed *orthonormal*, i.e., if

$$(34) \quad \langle \psi_{j,k} | \psi_{j',k'} \rangle = \delta_{j,j'} \delta_{k,k'},$$

then clearly

$$(35) \quad \|f\|_2^2 = \sum_{j,k \in \mathbb{Z}} |c_{j,k}|^2, \quad f \in L^2(\mathbb{R}).$$

But (33) and (35) may hold even if the orthonormal property (34) fails. For example, suppose ψ is a multiresolution wavelet in the *strict sense*, with $\{\psi_{j,k}\}$ satisfying (33) and (34); then set $\tilde{\psi}(x) := \frac{1}{3}\psi(x/3)$, and $\tilde{\psi}_{j,k}(x) := 2^{j/2}\tilde{\psi}(2^jx - k)$. We will still have

$$(36) \quad \|f\|_2^2 = \sum_{j,k \in \mathbb{Z}} |\langle \tilde{\psi}_{j,k} | f \rangle|^2, \quad f \in L^2(\mathbb{R}),$$

but now $\{\tilde{\psi}_{j,k}\}$ will *not* satisfy (34), and in fact $\|\tilde{\psi}\|_2 = 1/\sqrt{3}$. A system satisfying condition (36) is called a *tight frame*. Varying the parameters within a family of wavelets of fixed support and each satisfying the orthonormality condition, we find points of degeneracy. They correspond to wavelet bases which are only tight frames in $L^2(\mathbb{R})$, and these special wavelet generators are found precisely at points (in the parameter variety) where the eigenspace of the Perron-Frobenius eigenvalue is of dimension more than one. It is known that the degeneracy can be complicated in general, but within the varieties given by the polyphase matrices, or the spin-vector configurations, the degeneracy is known to happen on subvarieties of lower dimension. The direct correspondence between polyphase matrix and wavelet is further known to break down on the subvariety where the orthonormality degenerates into the tight frame property.

References

- [AbAl95] P. ABRV and A. ALDROUBI, Designing multiresolution analysis-type wavelets and their fast algorithms, *J. Fourier Anal. Appl.* **2** (1995), no. 2, 135-159.
- [BrJo02] O. BRATTELI and P. E. T. JORGENSEN, *Wavelets Through a Looking Glass: The World of the Spectrum*, Applied and Numerical Harmonic Analysis, Birkhäuser, Boston, 2002.
- [BrJo99] ———, Iterated function systems and permutation representations of the Cuntz algebra, *Mem. Amer. Math. Soc.* **139** (1999), no. 663.
- [Coh03] A. COHEN, *Numerical Analysis of Wavelet Methods*, Stud. Math. Appl., vol. 32, Elsevier Science Ltd., San Diego, 2003.
- [CGV99] A. COHEN, K. GRÖCHENIG, and L. F. VILLEMOS, Regularity of multivariate refinable functions, *Constr. Approx.* **15** (1999), no. 2, 241-255.
- [Dau92] I. DAUBECHIES, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Ser. Appl. Math., vol. 61, SIAM, Philadelphia, 1992.
- [DiFr99] P. DIACONIS and D. FREEDMAN, Iterated random functions, *SIAM Rev.* **41** (1999), 45-76.
- [Haa10] A. HAAR, Zur Theorie der orthogonalen Funktionensysteme, *Math. Ann.* **69** (1910), 331-371.
- [HeWe96] E. HERNÁNDEZ and G. WEISS, *A First Course on Wavelets*, Stud. Adv. Math., CRC Press, Boca Raton, FL, 1996.
- [Jor01] P. E. T. JORGENSEN, Minimality of the data in wavelet filters, *Adv. Math.* **159** (2001), 143-228.
- [JoPe98] P. E. T. JORGENSEN and S. PEDERSEN, Dense analytic subspaces in fractal L^2 -spaces, *J. Analyse Math.* **75** (1998), 185-228.
- [Mal99] S. G. MALLAT, *A Wavelet Tour of Signal Processing*, 2nd ed., Academic Press, Orlando-San Diego, 1999.
- [Mey00] Y. MEYER, Wavelets and functions with bounded variation from image processing to pure mathematics, *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur.*

Rend. Lincei (9) Mat. Appl. **11** (2000), 77–105, special issue: *Mathematics Towards the Third Millennium* (papers from the International Conference held in Rome, May 27–29, 1999).

- [PBK03] S. PARK, J. BAE, and Y. KWON, Wavelet quantum search algorithm with partial information, preprint 2003, <http://arXiv.org/abs/quant-ph/0303025>.
- [Per07] O. PERRON, Zur Theorie der Matrizen, *Math. Ann.* **64** (1907), 248–263.
- [Rue02] D. RUELLE, Dynamical zeta functions and transfer operators, *Notices Amer. Math. Soc.* **49** (2002), 887–895.
- [Rue69] D. RUELLE, *Statistical Mechanics: Rigorous Results*, W. A. Benjamin, Inc., New York-Amsterdam, 1969.
- [StNg96] G. STRANG and T. NGUYEN, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, 1996.
- [Str82] J.-O. STRÖMBERG, A modified Franklin system and higher-order spline systems on \mathbf{R}^n as unconditional bases for Hardy spaces, *Conference on Harmonic Analysis in Honor of Antoni Zygmund, Vol. II* (Chicago, 1981) (W. Beckner, A. P. Calderón, R. Fefferman, and P. W. Jones, eds.), Chapman & Hall/CRC Statistics and Mathematics, London, 1982, pp. 475–494.
- [Tre01] B. F. TREADWAY, Appendix to [Jor01].
- [Wel80] R. O. WELLS JR., *Differential Analysis on Complex Manifolds*, second ed., Grad. Texts Math., vol. 65, Springer-Verlag, New York, 1980.
- [Wic94] M. V. WICKERHAUSER, *Adapted Wavelet Analysis from Theory to Software*, A K Peters, Wellesley, MA, 1994.

About the Cover

Cloud Nine

This month's cover expands upon the “cloud nine” image in Palle Jorgensen's article. It represents the fractal tile (which Jorgensen calls a ‘reptile’) constructed by an iteration process, starting with a fundamental domain of the tiling. The central figure shows the results of several iterations applied to a single fundamental domain, together with several neighboring tiles. Along the left side are the results of the first few steps of the iteration, showing successive approximations to the final tile.

Jorgensen writes, “A reptile T arises from a limit construction, a simple algorithm which repeats the same set of affine operations in each step. They are defined from a matrix A , and a finite set of vectors D , i.e., $T = T(A, D)$ where D is a set of representatives of $\mathbb{Z}^2/A\mathbb{Z}^2$. Many interesting questions present themselves. For example, under what circumstances does T tile the plane? And if it does, can it be made to tile with lattice translations? How do you find an admissible lattice L ? The simpler planar reptiles T make a tiling by vectors in \mathbb{Z}^2 , and have area 1. They serve as higher-dimensional versions of the familiar Haar construction from one dimension. The picture on the cover has area 2, and the corresponding lattice L which makes a planar tiling is generated by the vectors $(1, 0)$ and $(0, 2)$.”

—Bill Casselman, Covers Editor
(notices-covers@ams.org)

