Scaling, wavelets, image compression, and encoding

Palle E. T. Jorgensen and Myung-Sin Song

Abstract In this paper we develop a family of multi-scale algorithms with the use of filter functions in higher dimension.

While our primary application is to images, i.e., processes in two dimensions, we prove our theorems in a more general context, allowing dimension 3 and higher.

The key tool for our algorithms is the use of tensor products of representations of certain algebras, the Cuntz algebras O_N , from the theory of algebras of operators in Hilbert space. Our main result offers a matrix algorithm for computing coefficients for images or signals in specific resolution subspaces. A special feature with our matrix operations is that they involve products and iteration of slanted matrices. Slanted matrices, while large, have many zeros, i.e., are sparse. We prove that as the operations increase the degree of sparseness of the matrices increase. As a result, only a few terms in the expansions will be needed for achieving a good approximation to the image which is being processed. Our expansions are local in a strong sense.

An additional advantage with the use of representations of the algebras O_N, and tensor products is that we get easy formulas for generating all the choices of matrices going into our algorithms.

1 Introduction

The paper is organized as follows: first motivation, history, and discussion of our applications. Since a number of our techniques use operator theory, we have a sepa-

Palle E. T. Jorgensen
Department of Mathematics, The University of Iowa, Iowa City, IA52242, USA e-mail: jorgen@math.uiowa.edu

Myung-Sin Song
Department of Mathematics and Statistics, Southern Illinois University Edwardsville, Edwardsville, IL62026, USA e-mail: msong@siue.edu

rate section with these results. We feel they are of independent interest, but we have developed them here tailored to use for image processing.

A key tool in our algorithms is the use of slanted matrices, and tensor products of representations. This material receives separate sections. The significance of the slanted property is that matrix products of slanted matrices become increasingly more sparse (i.e., the resulting matrices have wide patterns of zeros) which makes computations fast.

Motivation and Applications. We consider interconnections between three subjects which are not customarily thought to have much to do with one another: (1) the theory of stochastic processes, (2) wavelets, and (3) sub-band filters (in the sense of signal processing).

While connections between (2) and (3) have gained recent prominence, see for example [9], applications of these ideas to stochastic integration is of more recent vintage. Nonetheless, there is always an element of noise in the processing of signals with systems of filters. But this has not yet been modeled with stochastic processes, and it hasn't previously been clear which processes do the best job.

Recall however that the notion of low-pass and high-pass filters derives in part from probability theory. Here high and low refers to frequency bands, but there may well be more than two bands (not just high and low, but a finite range of bands). The idea behind this is that signals can be decomposed according to their frequencies, with each term in the decomposition corresponding to a range of a chosen frequency interval, for example high and low. Sub-band filtering amounts to an assignment of filter functions which accomplish this: each of the filters will then block signals in one band, and passes the others. This is known to allow for transmission of the signal over a medium, for example wireless. It was discovered recently (see [9]), perhaps surprisingly, that the functions which give good filters in this context serve a different purpose as well: they offer the parameters which account for families of wavelet bases, for example families of bases functions in the Hilbert space $L^2(\mathbb{R})$. Indeed the simplest quadrature-mirror filter is known to produce the Haar wavelet basis in $L^2(\mathbb{R})$.

It is further shown in [9] that both principles (2) and (3) are governed by families of representations of one of the Cuntz algebras $\mathcal{O}_N$, with the number N in the subscript equal to the number of sub-bands in the particular model. So for the Haar case, $N = 2$.

A main purpose in this paper is pointing out that fractional Brownian motion (fBm) may be understood with the data in (2) and (3), and as a result that fBm may be understood with the use of a family of representations of $\mathcal{O}_N$; albeit a quite different family of representations from those used in [9].

A second purpose we wish to accomplish is to show that the operators and representations we use in one dimension can be put together in a tensor product construction and then account for those filters which allow for processing of digital images. Here we think of both black and white, in which case we will be using a single matrix of pixel numbers. In the case of color images, the same idea applies, but then we will rather be using three matrices accounting for exposure of each of the three primary colors. If one particular family F of representations of one of the

Cuntz algebras $\mathcal{O}_N$ is used in 1D, then for 2D (images) we show that the relevant families of representations of $\mathcal{O}_N$ are obtained from F with the use of tensor product of pairs of representations, each one chosen from F.

1.1 Interdisciplinary dimensions

While there is a variety of interdisciplinary dimensions, math (harmonic, numerical, functional, ...), computer science, engineering, physics, image science, we will offer some pointers here some pointers to engineering, signal and image processing.

Engineering departments teach courses in digital images, as witnessed by such standard texts as [11]. Since 1997, there was a separate track of advances involving color images and wireless communication, the color appearance of surfaces in the world as a property that allows us to identify objects, see e.g., [17, 29], and [8, 7, 30].

From statistical inference we mention [28]. Color imaging devices such as scanners, cameras, and printers, see [26]. And on the more theoretical side, CS, [10].

2 History and Motivation

Here we discuss such tools from engineering as sub-band filters, and their manifold variations. They are ubiquitous in the processing of multi-scale data, such as arises in wavelet expansions.

A powerful tool in the processing of signals or images, in fact going back to the earliest algorithms, is that of subdividing the data into sub-bands of frequencies. In the simplest case of speech signals, this may involve a sub-division into the low frequency range, and the high. An underlying assumption for this is that the data admits such a selection of a total *finite* range for the set of all frequencies. If such a finite frequency interval can be chosen we talk about bandlimited analog signals. Now depending of the choice of analysis and synthesis method to be used, the suitability of bandlimited signals may vary. Shannon proved that once a frequency band B has been chosen, then there is a Fourier representation for the signals, as time-series, with frequencies in B which allow reconstruction from a discrete set of samples in the time variable, sampled in an arithmetic progression at a suitable rate, the Nyquist rate. The theorem is commonly called the Shannon sampling theorem, and is also known as Nyquist—Shannon—Kotelnikov, Whittaker—Shannon, WKS, etc., sampling theorem, as well as the Cardinal Theorem of Interpolation Theory.

While this motivates a variety of later analogues to digital (A/D) tools, the current techniques have gone far beyond Fourier analysis.

Here we will focus on tools based on such multi-scale which are popular in wavelet analysis. But the basic idea of dividing the total space of data into subspaces corresponding to bands, typically frequency bands, will be preserved. In passing from speech to images, we will be aiming at the processes underlying the

processing of digital images, i.e., images arising as a matrix (or checkerboard) of pixel numbers, or in case of color-images a linked system of three checkerboards, with the three representing the primary colors.

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. It has since found a variety of application to multiscale processes such as analysis on fractals. 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.

3 Operator Theory

Those key ideas multi-scale, wavelets, image processing, and the operator theory involved discussed about, and used inside our paper are covered in a number of references. Especially relevant are the following [9, 3, 2, 4, 16, 19, 20, 21], but the reader will find additional important reference lists in the books [9] and [19]. A key idea in the analysis we present here is to select a Hilbert space which will represent the total space of data; it may be $L^2(\mathbb{R}^d)$ for a suitable chosen d for dimension; or it may be anyone of a carefully selected closed subspaces. A representation of a function in d variables into an expansion relative to an orthogonal basis (or a frame basis) corresponds to a subdivision of the total space into one-dimensional subspaces. But to get started one must typically select a fixed subspace which represent a resolution of the data (or image) under consideration, then the further subdivision into closed subspaces can be accomplished with a scaling: The result is a family of closed subspaces, each representing a detail. There will then be a system of isometries which account for these subspaces, and there will be a scaling operation which makes (mathematically) precise the scale-similarity of the data in different detail-components of the total decomposition.

In wavelets, the scaling is by a number, for example 2, for dyadic wavelets. In this case, there will be two frequency bands. If the scale is instead by a positive integer $N > 2$, then there will be N natural frequency bands.

For images, or higher dimensional date, it is then natural to use an invertible $d \times d$ matrix (over the integers), say A, to model a choice of scaling. Scaling will then be

modeled with powers of A, i.e., with A^j as j ranges over the integers $\mathbb{Z}$. In this case, the number of frequency bands will be $N := |det(A)|$.

Here we review a harmonic analysis of isometries in Hilbert space. Our results are developed with view to applications to multi-scale processes. In the Hilbert space framework, this takes the form of an exhaustion of closed subspaces, a monotone family of closed subspaces where one arises from the other by an application of a scaling operator.

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 $\mathscr{H}$ is the unique sum of a shift isometry and a unitary operator, i.e., the initial Hilbert space $\mathscr{H}$ splits canonically as an orthogonal sum of two subspaces $\mathscr{H}_s$ and $\mathscr{H}_u$ in $\mathscr{H}$, one which carries the shift operator, and the other $\mathscr{H}_u$ 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 $\mathscr{H}_u$. Specifically,

$$\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_n \subset V_{n+1} \subset \cdots$$

$$\bigwedge_n V_n = \mathscr{H}_u, \text{ and } \bigvee_n V_n = \mathscr{H}.$$

However, Stéphane Mallat was motivated 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.

Subsequently, this view became popular in the wavelet community, as it offered a repository for the fundamental father and the mother functions, also called the scaling function φ, and the wavelet function ψ (see details below). 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. This 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 (GMRAs) have emerged.

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.

3.0.1 Multiresolutions

Haar's work in 1909–1910 holds implicitly the key idea which got wavelet mathematics started later with Yves Meyer, Ingrid Daubechies, Stéphane Mallat, and others (see [19] for a resent bibliograph)—namely the idea of a multiresolution. See Figures 1 and 2 for details.

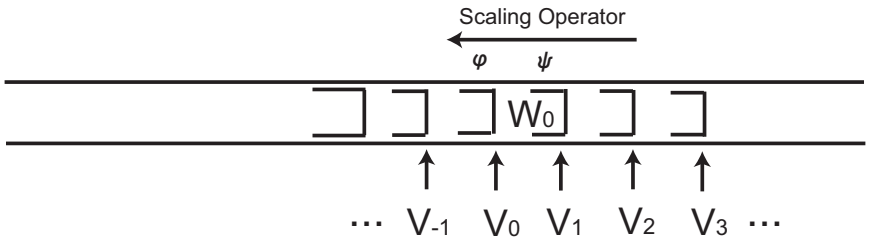


Fig. 1 Multiresolution. $L^2(\mathbb{R}^d)$-version (continuous); $\varphi \in V_0$, $\psi \in W_0$.

$f_v(x) = \sum_{k\in\mathbb{Z}} v_k \varphi(\cdot - k)$, representation by numbers $v = (v_k) \in l^2(\mathbb{Z})$. The dyadic scaling operator as a unitary operator in $L^2(\mathbb{R})$, $U = \frac{1}{\sqrt{2}} f(\frac{x}{2})$.

There are three different ways of representation, function, sequence

$$L^2(\mathbb{R}) \underset{\sim}{\supset} l^2 \simeq L^2(0,1) : f \longleftrightarrow v \longleftrightarrow V$$

$\sum_k v_k z^k = v(z)$, $l^2 \simeq L^2(\mathbb{T})$, $v = (v_k) \in l^2(\mathbb{Z})$

Representation of Fourier series:

$$v_k = \langle z^k, v\rangle_{L^2(\mathbb{T})},$$

$$v_k = \int_0^1 e^{-i2\pi kt} v(t)dt,$$

$$\sum_{k\in\mathbb{Z}} |v_k|^2 = \int_0^1 |v(t)|^2 dt.$$

Generating functions (engineering term):

$$H(z) = \sum_{k\in\mathbb{Z}} h_k z^k \quad \text{where } z = e^{i2\pi t}, \quad \text{and } z^k = e^{i2\pi kt}$$

An equivalent form of the dyadic scaling operator U in $L^2(\mathbb{R})$, $U : V^\varphi \to V^\varphi$, so U restricts to an isometry in the subspace V^φ.

$$\begin{aligned}(U\varphi)(x) &= \frac{1}{\sqrt{2}}\varphi(\frac{x}{2}) \\ &= \sqrt{2}\sum_j h_j \varphi(x-j).\end{aligned}$$

Set $\varphi_k = \varphi(\cdot - k)$; then we have

$$\begin{aligned}(U\varphi_k)(x) &= \frac{1}{\sqrt{2}}\varphi_k(\frac{x}{2}) \\ &= \frac{1}{\sqrt{2}}\varphi(\frac{x}{2}-k) \\ &= \frac{1}{\sqrt{2}}\varphi(\frac{x-2k}{2}) \\ &= \sqrt{2}\sum_j h_j \varphi(x-2k-j) \\ &= \sqrt{2}\sum_j h_j \varphi_{2k+j}.\end{aligned}$$

It follows that

$$\begin{aligned}U\sum_k v_k\varphi_k &= \sqrt{2}\sum_l\sum_k v_k h_{2k-l} \\ &= \sqrt{2}\sum_l h_{2k-l}\varphi(x-l) \quad \text{if we let } l = 2k-j.\end{aligned}$$

As a result, we get:

$$(M_H v)_l = \sqrt{2}\sum_{k\in\mathbb{Z}} h_{2k+l} v_k;$$

and

$$(M_H \delta_p)_l = \sqrt{2}\sum_{k\in\mathbb{Z}} h_{2k+l}\delta_{p,k} = h_{2p+l}.$$

Lemma 1. *Setting $W : f_v \to \sum_{k\in\mathbb{Z}} v_k\varphi(\cdot-k) \in V^\varphi$, and $(S_0 f_v)(z) := H(z) f_v(z^2)$, we then get the intertwining identity $WS_0 = UW$ on V^φ.*

Proof.

$$\begin{aligned}(WS_0 f_v)(x) &= (W f_{S_0 v})(x)\\ &= \sum_{k\in\mathbb{Z}} (S_0 v)_k \varphi(x-k)\\ &= \frac{1}{\sqrt{2}} \sum_{k\in\mathbb{Z}} v_k \varphi(\frac{x}{2}-k)\\ &= \frac{1}{\sqrt{2}} f_v(\frac{x}{2}) = (UW f_v)(x).\end{aligned}$$

Note that the isometry $S_0 = S_H$ and $S_1 = S_G$ are $\mathcal{O}_2$ Cuntz algebra.

$$\sqrt{2}\begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \\ \cdots & h_{2p-l} & h_{2p-l+2} & \cdots & \\ \cdots & h_{2p-l-1} & h_{2p-l+1} & \cdots & \\ & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$S_0^j S_1$ are isometries for $j = 0,1,2,\cdots$ that generate $\mathcal{O}_\infty$,

$$Q_j = S_0^j S_1 (S_0^j S_1)^* = S_0^j S_1 S_1^* S_0^{*j} \underset{S_1S_1^*=1-S_0S_0^*}{=} \underset{P_j}{S_0^j S_0^{*j}} - \underset{P_{j+1}}{S_0^{j+1} S_0^{*j+1}},$$

$$\sum_{j=0}^{n} Q_j = \sum_{j=0}^{n} (P_j - P_{j+1}) = I - P_{n+1} \underset{N\to\infty}{\to} I$$

since

$$P_{n+1} = S_0^{n+1} S_0^{*n+1} \to 0 \quad \text{by pure isometry lemma in [19].}$$

Recall an isometry S in a Hilbert space $\mathscr{H}$ is a shift if and only if $\lim_{N\to\infty} S^N S^{*N} = 0$. L^2 is same as resolution space $\simeq V^\varphi$. The operators are as follows: $Q_0 = S_1 S_1^*$, $Q_1 = S_0 S_1 S_1^* S_0^* = S_0 S_1 (S_0 S_1)^*$. Note that $Q_0 = I - P_0$, and $Q_j = P_j - P_{j+1}$.

$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x+k)$, $j = 0,1,2,\cdots$

$$\sum\sum c_{j,k}\psi_{j,k} = \sum\sum f_v (Q_j v)_k = \sum\sum (S_0^j S_1 S_1^* S_0^{*j} v)_k \psi_{j,k}$$

If we now take the adjoint of these matrices, these correspond to the isometries.

$$S_0^* \sim M_H^* \sim F_0,$$

$$S_1^* \sim M_G^* \sim F_1.$$

Theorem 1. *The wavelet representation is given by slanted matrices as follows:*

$\sum\sum (F_1 F_0^j v)_k \psi_{j,k} = f_v$

$$c_{j,k} = \langle \psi_{j,k}, f\rangle = \int \psi_{j,k}(x) f(x) dx.$$

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

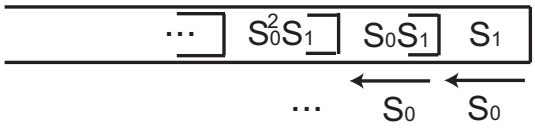


Fig. 2 Multiresolution. $l^2(\mathbb{Z})$-version (discrete); $\varphi \in V_0$, $\psi \in W_0$.

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

4 The discrete vs continuous wavelet Algorithms

4.1 The Discrete Wavelet Transform

If one stays with function spaces, it is then popular to pick the d-dimensional Lebesgue measure on $\mathbb{R}^d$, $d = 1, 2, \cdots$, 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 ψ_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

$$\psi_{i,j,k}(x) := |detA|^{j/2}\psi(A^j x + k)$$

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 $\mathscr{H}$ refers to the norm and the inner product in $\mathscr{H}$: 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 $\mathscr{H}$), 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., [6, 15] 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 $\mathscr{H}$, 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 $\mathscr{O}_N$ if n is the number of bands. For additional details, see e.g., [19].

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

$$\sum_{i=0}^{N-1} S_i S_i^* = \mathbf{1} \quad \text{(see Fig.3)} \tag{1}$$

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

$$\sum_{i=0}^{N-1} S_i S_i^* = \mathbf{1}, \text{ and } S_i^* S_j = \delta_{i,j}\mathbf{1}. \tag{2}$$

In a representation on a Hilbert space, say $\mathscr{H}$, 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 $\mathscr{H}$, i.e., the operator $I : h \to h$, for $h \in \mathscr{H}$. In operator language, the two formu-

las 1 and 2 state that each S_i is an isometry in $\mathscr{H}$, and that te respective ranges $S_i\mathscr{H}$ are mutually orthogonal, i.e., $S_i\mathscr{H} \perp S_j\mathscr{H}$ for $i \neq j$. Introducing the projections $P_i = S_iS_i^*$, we get $P_iP_j = \delta_{i,j}P_i$, and

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

Example 1. Fix $N \in \mathbb{Z}_+$. Then the easiest representation of $\mathscr{O}_N$ is the following: Let $\mathscr{H} := l^2(\mathbb{Z}_{\geq 0})$, where $\Gamma := \mathbb{Z}_{\geq 0} = \{0\} \cup \mathbb{N} = \{0,1,2,\cdots\}$.

Set $\mathbb{Z}_N = \{0,1,\cdots,N-1\} = \mathbb{Z}/N\mathbb{Z} =$ the cyclic group of order N.

We shall denote canonical ONB in $\mathscr{H} = l^2(\Gamma)$ by $|x\rangle = \delta_x$ with Dirac's formalism. For $i \in \mathbb{Z}_N$ set $S_i = \rho(s_i)$, given by

$$S_i|x\rangle = |Nx+i\rangle, \quad x \in \Gamma \tag{3}$$

then

$$S_i^*|x\rangle = \begin{cases} |\frac{x-i}{N}\rangle & \text{if } x-i \equiv 0 \mod N \\ 0 & \text{otherwise.} \end{cases}$$

The reader may easily verify the two relations in (2) by hand.

For the use of $\mathrm{Rep}(O_N, \mathscr{H})$ in signal/image processing, more complicated formulas than (3) are needed for the operators $S_i = \rho(s_i)$.

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

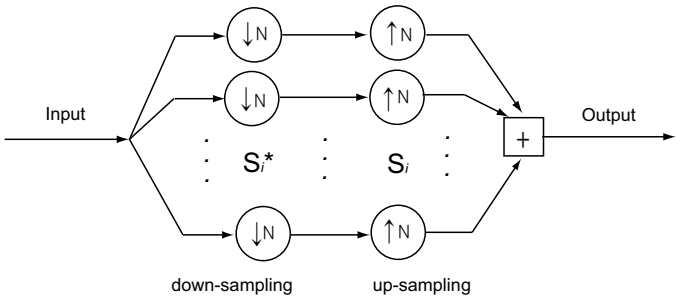


Fig. 3 Perfect reconstruction in a subband filtering as used in signal and image processing.

If the process of Figure 3 is repeated, we arrive at the discrete wavelet transform or stated in the form of images ($n = 5$)

But to get successful subband filters, we must employ a more subtle family of representations than those of (3) in Example 1. We now turn to the study of those representations.

Selecting a resolution subspace $V_0 = closure\ span\{\varphi(\cdot - k)|k \in \mathbb{Z}\}$, we arrive at a wavelet subdivision $\{\psi_{j,k}|j \geq 0, k \in \mathbb{Z}\}$, where $\psi_{j,k}(x) = 2^{j/2}\psi(2^jx - k)$, and the

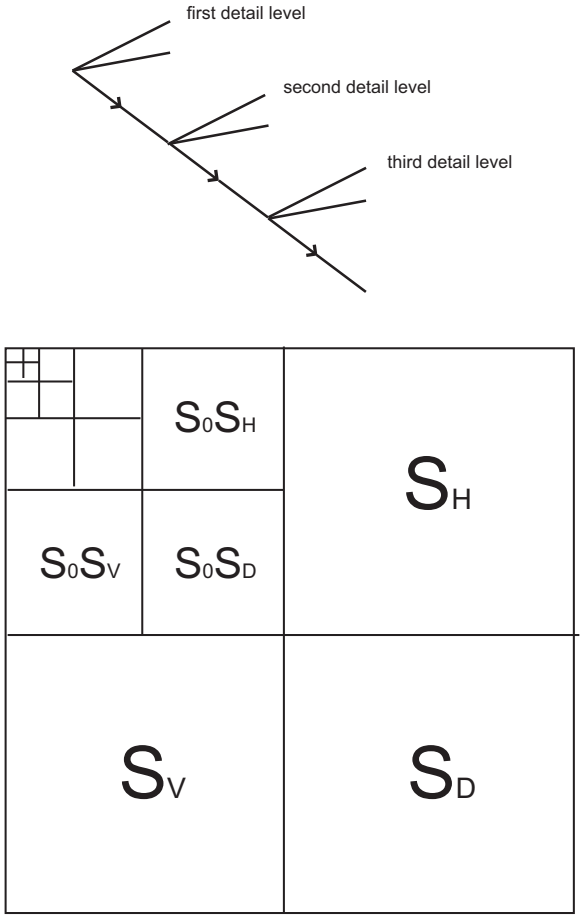


Fig. 4 The subdivided squares represent the use of the pyramid subdivision algorithm to image processing, as it is used on pixel squares. At each subdivision step the top left-hand square represents averages of nearby pixel numbers, averages taken with respect to the chosen low-pass filter; while the 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 $S_0S_0^* + S_HS_H^* + S_VS_V^* + S_DS_D^* = I$, with I denoting the identity operator in an appropriate l^2-space.

continuous expansion $f = \sum_{j,k}\langle\psi_{j,k}|f\rangle\psi_{j,k}$ or the discrete analogue derived from the isometries, $i = 1,2,\cdots,N-1$, $S_0^kS_i$ for $k = 0,1,2,\cdots$; called the discrete wavelet transform.

Fig. 5 An example of Fig. 4: $n = 2$ Jorgensen. Jorgensen's picture after level 2 wavelet decomposition. Notice that The average, horizontal, diagonal and vertical details are captured clockwise. Also, the average detail is decomposed one more time since level 2 decomposition was done.

4.1.1 Notational convention.

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

In the present contest of the Discete 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.

4.1.2 Increasing the dimension

In wavelet theory, [13] there is a tradition for reserving φ for the father function and ψ for the mother function. A 1-level wavelet transform of an $N \times M$ image can be represented as

$$\mathbf{f} \mapsto \begin{pmatrix} \mathbf{a}^1 & | & \mathbf{h}^1 \\ -- & & -- \\ \mathbf{v}^1 & | & \mathbf{d}^1 \end{pmatrix} \tag{4}$$

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

$$\begin{array}{l} \mathbf{a}^1 = V_m^1 \otimes V_n^1 : \varphi^A(x,y) = \varphi(x)\varphi(y) = \sum_i \sum_j h_i h_j \varphi(2x-i)\varphi(2y-j) \\ \mathbf{h}^1 = V_m^1 \otimes W_n^1 : \psi^H(x,y) = \psi(x)\varphi(y) = \sum_i \sum_j g_i h_j \varphi(2x-i)\varphi(2y-j) \\ \mathbf{v}^1 = W_m^1 \otimes V_n^1 : \psi^V(x,y) = \varphi(x)\psi(y) = \sum_i \sum_j h_i g_j \varphi(2x-i)\varphi(2y-j) \\ \mathbf{d}^1 = W_m^1 \otimes W_n^1 : \psi^D(x,y) = \psi(x)\psi(y) = \sum_i \sum_j g_i g_j \varphi(2x-i)\varphi(2y-j) \end{array} \tag{5}$$

where φ is the father function and ψ 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) [13].

We now introduce operators T_H and T_G in l^2 such that the expression on the RHS in (5) becomes $T_H \otimes T_H$, $T_G \otimes T_H$, $T_H \otimes T_G$ and $T_G \otimes T_G$, respectively.

We use the following representation of the two wavelet functions φ (father function), and ψ (mother function). A choice of filter coefficients (h_j) and (g_j) is made.

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:

$$\varphi(x) = 2\sum_{i\in\mathbb{Z}} h_i \varphi(2x-i);$$

$$\psi(x) = 2\sum_{i\in\mathbb{Z}} g_i \varphi(2x-i);$$

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 φ and ψ will vanish outside some finite interval on a real line.

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

$$\sum_{i\in\mathbb{Z}} \bar{h}_i h_{i+2k} = \frac{1}{2}\delta_{0,k} \tag{6}$$

is the best known.

The systems h and g are both low-pass and high-pass filter coefficients. In 5, $\mathbf{a}^1$ denotes the first averaged image, which consists of average intensity values of the original image. Note that only φ 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 φ function is used on y and ψ 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 φ function is used on x and ψ 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 ψ function, W space and g coefficients are used here. See [37, 34].

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 [33, 36]) 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 (6) 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 (6) 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.

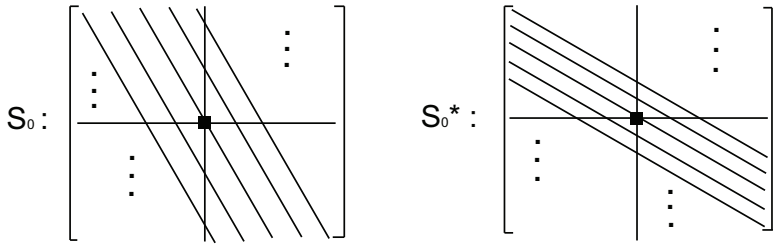
Note that the two operators S_0 and S_1 have equivalent matrix representations. Recall that by Parseval's formula we have $L^2(\mathbb{T}) \simeq l^2(\mathbb{Z})$. So representing S_0 instead as an $\infty \times \infty$ matrix acting on column vectors $x = (x_j)_{j\in\mathbb{Z}}$ we get

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

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

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

with the overbar signifying complex conjugation. This is computational significance 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.,



4.1.3 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 2 s 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.

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

$$y = F_0 x,$$

$$z = F_1 x.$$

4.2 Entropy Encoding

In this section we discuss the encoding aspect of our algorithms. While the theory here dates back to the start of information theory , see [25, 24, 27, 31, **?**] and the references cited there, its adaptation to advances in technology have been amazingly successful , see e.g., [25, 24].

An important part of digital imaging is the choice of encoding, for example the encoding of the letters in the alphabet, a, b, c, etc. As a rough principle, one selects the shortest code for the most frequently occurring letter. But to do this, both of these notions must be quantified.

It is clearly of relevance for efficiency, speed, and error detection. As it turns out, probabilities and entropy are helpful. Indeed the way Shannon quantified information a reduction in entropy by an amount A costs A units of information. We have discussed this part of the theory in more detail in [35], and here we offer just an example for illustration of the main points.

There are various entropy encoding schemes being used, and one example of it is Shannon-Fano entropy encoding. In Shannon-Fano entropy encoding, for each data on an image, i.e., pixel, a set of probabilities p_i is computed, where $\sum_{i=1}^{n} p_i = 1$.

The entropy of this set gives the measure of how much choice is involved, in the selection of the pixel value of average.

Definition 1. Shannon's entropy $E(p_1, p_2, \ldots, p_n)$ which satisfy the following:

- E is a continuous function of p_i.
- E should be steadily increasing function of n.
- If the choice is made in k successive stages, then $E =$ sum of the entropies of choices at each stage, with weights corresponding to the probabilities of the stages.

$E = -k\sum_{i=1}^{n} p_i \log p_i$. k controls the units of the entropy, which is "bits." logs are taken base 2. [5, 32]

Shannon-Fano entropy encoding is done according to the probabilities of data and the method is as follows:

- The data is listed with their probabilities in decreasing order of their probabilities.
- The list is divided into two parts that has roughly equal probability.
- Start the code for those data in the first part with a 0 bit and for those in the second part with a 1.
- Continue recursively until each subdivision contains just one data. [5, 32]

An example on a text: letters-to-codes: may be better to depict how the mechanism works. Suppose we have a text with letters a, e, f, q, r with the following probability distribution:

Letter	Probability
a	0.3
e	0.2
f	0.2
q	0.2
r	0.1

Then applying the Shannon-Fano entropy encoding scheme on the above table gives us the following assignment.

Letter	Probability	code
a	0.3	00
e	0.2	01
f	0.2	100
q	0.2	101
r	0.1	110

Note that instead of using 8-bits to represent a letter, 2 or 3-bits are being used to represent the letters in this case.

The following is an elementary example of Shannon-Fano entropy encoding

Letter	Probability	code
a	0.3	00
e	0.2	01
f	0.2	100
q	0.2	101
r	0.1	110

While this is an oversimplification, it is nonetheless a key idea used in more realistic algorithms:

- In a given text, list all letters in decreasing order of their probabilities.
- Divide the list into two parts with approximately equal probability (i.e., by the median, the total probability of each part is approximately 0.5).
- For the letters in the first part, start the code with a 0 bit, and for those in the second part with a 1.
- Recursively continue until each subdivision is left with just one letter [5].

Note that the divisions of the list are following a binary tree-rule. The initial important uses of encoding were to texts and to signals. Much more recent uses are to a host of big data sets such as data on the color images. These uses are: quantization, entropy encoding. As a result, the mathematics of encoding has seen a recent revival.

While entropy encoding is popular in engineering, [33], [36], [14], the choices made in signal processing are often more by trial and error than by theory. Reviewing the literature, we found that the mathematical foundation of the current use of entropy in encoding deserves closer attention.

5 Slanted Matrix Representations, Frame Estimates, and Computations

We will be using finite and infinite slanted matrices, and we prove two results about their tensor products, and their significance in representation theory. The significance of the slanted property is that matrix products of slanted matrices become increasingly more sparse (i.e., the resulting matrices have wide patterns of zeros) which makes computations fast. In the application this means that an image, or a more general problem from applied mathematics may be synthesized faster with the use of scale similar orthogonal bases, such as wavelet bases in $L^2(\mathbb{R}^d)$.

In this section we prove mathematical theorems supporting the applications outlined above: The operators in Fig 1 have slanted matrix representations determined by the masking sequences (h_n) and (g_n), and with the slanting changing from one operator S to the corresponding adjoint operator S^*. We then show how frame estimates are preserved under filtering with our S-systems, i.e., with the slanted matrices that realize the Cuntz relations in (a) and (b) above, The slanted matrix representation is what make computations fast. The slanting is such that repeated matrix op-

erations in the processing make for more sparse matrices, and hence for a smaller number of computational steps in digital program operations for image processing.

We begin by introducing the Cuntz operators S . The two operators come from the two masking sequences (h_n) and (g_n) , also called filter-coefficients, also called low-pass and high-pass filters.

Definition 2. If $(h_n)_{n\in\mathbb{Z}}$ is a double infinite sequence of complex numbers, i.e., $h_n \in \mathbb{C}$, for all $n \in \mathbb{Z}$; set

$$(S_0x)(m) = \sqrt{2}\sum_{n\in\mathbb{Z}} h_{m-2n}x(n) \tag{7}$$

and adjoint

$$(S_0^*x)(m) = \sqrt{2}\sum_{n\in\mathbb{Z}} \overline{h}_{n-2m}x(n);\ \text{for all } m \in \mathbb{Z}. \tag{8}$$

Then

(a) The $\infty \times \infty$ matrix representations (7) and (8) have the following slanted forms

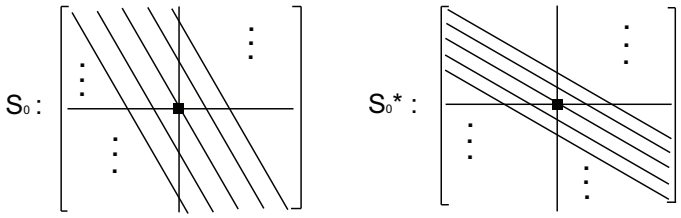


Fig. 6 S_0 and S_0^*.

(b) The set of non-zero numbers in $(h_n)_{n\in\mathbb{Z}}$ is finite if and only if the two matrices in Fig are *banded*.

(c) Relative to the inner product

$$\langle x|y\rangle_{l^2} := \sum_{n\in F} \overline{x_n}y_n \text{ in } l^2$$

(i.e., conjugate-linear in the first variable), the operator S_0 is *isometric* if and only if

$$\sum_{n\in F} \overline{h}_n h_{n+2p} = \frac{1}{2}\delta_{0,p},\ \text{for all } p \in \mathbb{Z}. \tag{9}$$

(d) If (9) holds, and

$$(S_1x)(m) = \sqrt{2}\sum_{n\in\mathbb{Z}} g_{m-2n}x(n), \tag{10}$$

then

$$S_0S_0^* + S_1S_1^* = I_{l^2} \tag{11}$$

$$S_k^*S_l = \delta_{k,l}I_{l^2} \text{ for all } k,l \in \{0,1\} \tag{12}$$

(the Cuntz relations) holds for

$$g_n := (-1)^n \bar{g}_{1-n},\ n \in \mathbb{Z}.$$

Proof. By Parseval's identity and Fourier transforms, we have the *isometric* isomorphism $l^2(\mathbb{Z}) \simeq L^2(\mathbb{T})$ where $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ is equipped with Haar measure.

Hence the assertions (a)-(d) may be checked instead in the following function representation:

$$f(z) = \sum_{n \in F} x(n) z^n, \tag{13}$$

$$m_0(z) = \sum_{n \in F} h_n z^n, \tag{14}$$

$$m_1(z) = \sum_{n \in F} g_n z^n; \tag{15}$$

setting

$$(S_j f)(z) = m_j(z) f(z^2),\ \text{for all } z \in \mathbb{T},\ \text{for all } f \in L^2(\mathbb{T}),\ j = 0, 1. \tag{16}$$

In this form, the reader may check that conditions (a)-(d) are equivalent to the following unitary principle: For almost every $z \in \mathbb{T}$ (relative to Haar measure), we have that the 2×2 matrix

$$U(z) = \begin{pmatrix} m_0(z) & m_0(-z) \\ m_1(z) & m_1(-z) \end{pmatrix} \tag{17}$$

is unitary; i.e., that $U(z)^* = U(z)^{-1}$, almost every $z \in \mathbb{T}$, where $(U^*)_{k,l} := \bar{U}_{l,k}$ denotes the adjoint matrix.

5.0.1 WARNING:

Note that the tensor product of two matrix-functions (17) does not have the same form. Nonetheless, there is a more indirect way of creating new multiresolution-wavelet filters from old ones with the use of tensor-product; see details below.

Suppose A is an index set and $(v_\alpha)_{\alpha \in A} \subset l^2(\mathbb{Z})$ is a system of vectors subject to the frame bound $(B < \infty)$

$$\sum_{\alpha \in A} |\langle v_\alpha | x \rangle_{l^2}|^2 \leq B \|x\|_{l^2}^2,\ \text{for all } x \in l^2. \tag{18}$$

Set

$$w_{j,k} := S_0^j S_1 v_\alpha,\ j \in \mathbb{N}_0 = \{0, 1, 2, \ldots\}, \alpha \in A. \tag{19}$$

If the unitarity condition (17) in the lemma is satisfied, then the induced system (19) satisfies the same frame bound (18).

Proof. Introducing Dirac's notation for rank-one operators:

$$|u\rangle\langle v||x\rangle = \langle v|x\rangle|u\rangle, \tag{20}$$

we see that (18) is equivalent to the following operator estimate

$$\sum_{\alpha\in A} |v_\alpha\rangle\langle v_\alpha| \le BI_{l^2} \tag{21}$$

where we use the standard ordering of the Hermitian operators, alias matrices.

For the system $(w_{j,\alpha})_{(j,\alpha)\in\mathbb{N}_0\times A}$ in (19), we therefore get

$$\begin{aligned}\sum_{(j,\alpha)\in\mathbb{N}_0\times A} |w_{j,\alpha}\rangle\langle w_{j,\alpha}| &= \sum_j\sum_\alpha S_0^j S_1 |v_\alpha\rangle\langle v_\alpha| S_1^* S_0^{*j} \\ &\underset{by(21)}{\le} B\sum_j S_0^j S_1 S_1^* S_0^{*j} \le B,\end{aligned}$$

Since

$$\sum_j S_0^j S_1 S_1^* S_0^{*j} = I - S_0^{n+1} S_0^{*n+1} \le I \quad \text{for all } n.$$

But the RHS in the last expression is the limit of the finite partial sums

$$\begin{aligned}\sum_{j=0}^{N} S_0^j S_1 S_1^* S_0^{*j} &= \sum_{j=0}^{N} S_0^j (I - S_0 S_0^*) S_0^{*j} && (22)\\ &= I_{l^2} - S_0^{N+1} S_0^{*N+1} && (23)\\ &\le I_{l^2} \text{ since} && (24)\end{aligned}$$

$P_{N+1} := S_0^{N+1} S_0^{*N+1}$ is a projection for all $N \in \mathbb{N}_0$. In fact

$$\cdots \le P_{N+1} \le P_N \le \cdots \le P_1 = S_0 S_0^*;$$

and P_1 denotes the projection onto $S_1 l^2$.

5.0.2 Digital Image Compression

In [23], we showed that use of Karhunen-Loève's theorem enables us to pick the best basis for encoding, thus to minimize the entropy and error, to better represent an image for optimal storage or transmission. Here, optimal means it uses least memory space to represent the data; i.e., instead of using 16 bits, use 11 bits. So the best basis found would allow us to better represent the digital image with less storage memory.

The particular orthonomal bases (ONBs) and frames which we use come from the operator theoretic context of the Karhunen-Loève theorem [1]. In approximation problems involving a stochastic component (for example noise removal in time-series or data resulting from image processing) one typically ends up with correlation kernels; in some cases as frame kernels.

Summary of the mathematics used in the various steps of the image compression flow chart in Figure 3 (on next page):

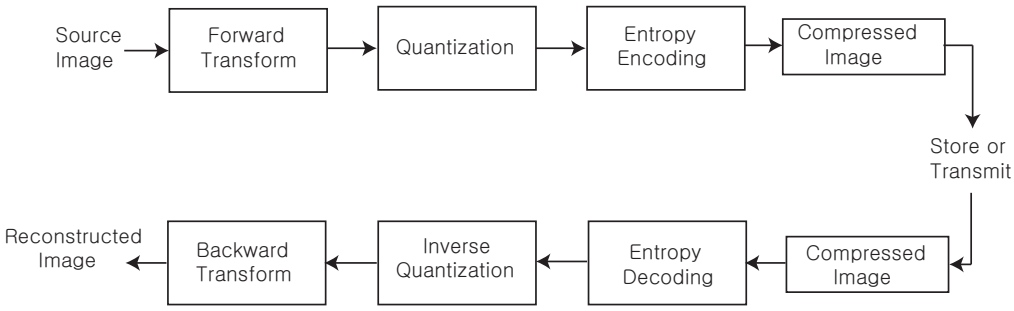


Fig. 7 Outline of the wavelet image compression process. [33]

- At the beginning of the diagram (source image) we will typically input a digital image. Of the possible forward transforms that apply to images, this proposal uses the discrete wavelet transforms.
- The quantization step refers to the following: the output from the transform will be quantized in the sense of economy of processing; for instance, with the thresholding method. After the wavelet forward transform, the image is decomposed into different details at different levels; the thresholding method will eliminate some of these details selectively resulting in lossy compression. In our recent paper [23], we initiated the use of thresholding for exactly this purpose.
- In our approach to image compression the encoding step does the following: with the quantization, the process was lossy where the step is irreversible. With the entropy encoding, if we started off with an 16 bit image we find a better representation meaning we use fewer bits to present the pixel values. This step is lossless. Entropy encoding has been used for long time in information theory, and it has found a recent revival in data compression schemes. The idea is to assign codes to symbols in order to match code lengths with the frequencies of occurrence of the symbols. By entropy we refer to a priori probability distributions of symbols. Entropy encoders compress data by replacing equal length codes with codes where the length of each code would be determined by quantitative measures of entropy. Therefore, the most frequent symbols will be assigned the shortest code. Hence the economy.

  There are number of other entropy encoding schemes, Shannon-Fano encoding, Huffman coding, arithmetic coding, etc. [5], [39], [38], [12]. These alternative encoding schemes have been successful in signal compression and in simplest static codes, but in the most recent uses of discrete algorithm on images the entropy encoding proved to have practical advantages. The lossless data compression has in common with the discrete wavelet synthesis that exact data can be reconstructed from the compressed versions. Hence, the processing is reversible. Lossless compression is used for example in zip file forward and gzip in Unix/Linux. For images, the formats png and gif also use lossless compression. Examples are executable programs and source codes.

In carrying out compression, one generates the statistical output from the input data, and the next step maps the input data into bit-streams, achieving economy by mapping data into shorter output codes.

- The result of the three steps is the compressed information which could be stored in the memory or transmitted through internet. The technology in both ECW (Enhanced Compressed Wavelet technology) and JPEG2000 allow for fast loading and display of large image files. JPEG2000 is wavelet-based image compression [19], [18]. There is a different brand new compressing scheme called Enhanced Compressed Wavelet technology (ECW). Both JPEG2000 and ECW are used in reconstruction as well.

5.1 Reconstruction

The methods described above apply to reconstruction questions from signal processing. Below we describe three important instances of this:

(i) In entropy decoding, the fewer bits assigned to represent the digital image are mapped back to the original number of bits without losing any information.
(ii) In inverse quantization, there is not much of recovery to be obtained if thresholding was used for that was a lossy compression. For other quantization methods, this may yield some recovery.
(iii) The backward transform is the wavelet inverse transform where the decomposed image is reconstructed back into an image that is of the same size as the original image in dimension but maybe of smaller size in memory.

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z} \simeq \{z \in \mathbb{C}; |z| = 1\}$. Let A be a $d \times d$ matrix over $\mathbb{Z}$ such that $\|\lambda\| > 1$ for all $\lambda \in \text{spec}(A)$. Then the order of the group $A^{-1}\mathbb{Z}^d/\mathbb{Z}^d$ is $|detA| := N$. Consider $A^{-1}\mathbb{Z}^d/\mathbb{Z}^d$ as a subgroup in $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$.

We shall need the following bijection correspondence:

Definition 3. For $z \in \mathbb{T}^d$, $z_j = e^{i2\pi\theta_j}$, $1 \le j \le d$, set

$$z^A = (e^{i2\pi\eta_j})^d, \tag{25}$$

where

$$A\theta = \eta, \tag{26}$$

i.e., $\sum_{j=1}^{d} A_{kj}\theta_j = \eta_k$, $1 \le k \le d$. Then for $z \in \mathbb{T}^d$, the set $\{w \in \mathbb{T}^d; w^A = z\}$ is in bijection correspondence with the finite group $A^{-1}\mathbb{Z}^d/\mathbb{Z}^d$.

Definition 4. Let $\mathscr{U}_N(\mathbb{C})$ be the group of all unitary $N \times N$ matrices, and let $\mathscr{U}_N^A(\mathbb{T}^d)$ be the group of all measurable function

$$\mathbb{T}^d \ni z \mapsto U(z) \in U_N(\mathbb{C}). \tag{27}$$

Let $\mathscr{M}_N^A(\mathbb{T}^d)$ be the multiresolution functions,

$$\mathbb{T}^d \ni z \mapsto M(z) = (m_j(z))_1^N \in \mathbb{C}^N; \tag{28}$$

i.e., satisfying

$$\frac{1}{N} \sum_{w\in\mathbb{T}^d, w^A=z} \overline{m_j(w)} m_k(w) = \delta_{j,k}. \tag{29}$$

Example 2. Let $\{k_j\}_{j=1}^N$ be a selection of representatives for the group $\mathbb{Z}^d/A\mathbb{Z}^d$; then

$$M_0(z) = (z^{k_j})_{j=1}^N \tag{30}$$

is in $\mathscr{M}_N^A(\mathbb{T}^d)$.

For $z \in \mathbb{T}^d$ and $k \in \mathbb{Z}^d$ we write

$$z^k = (z_1^{k_1} z_2^{k_2} \cdots z_d^{k_d}) \quad \text{multinomial.} \tag{31}$$

Lemma 2. *There is a bijection between $\mathscr{U}_N^A(\mathbb{T}^d)$ and $\mathscr{M}_N^A(\mathbb{T}^d)$ given as follows:*

(i) If $u \in \mathscr{U}_N^A(\mathbb{T}^d)$ set

$$M_U(z) = U(z^A) M_0(z) \tag{32}$$

where M_0 is the function in (30).

(ii)If $M \in \mathscr{M}_N^A(\mathbb{T}^d)$, set

$$U_M(z) = (U_{i,j}(z))_{i,j=1}^N \tag{33}$$

with

$$U_{i,j}(z) = \frac{1}{N} \sum_{w\in\mathbb{T}^d, w^A=z} \overline{M_j(w)} M_i(w). \tag{34}$$

Proof. Case (i). Given $U \in \mathscr{U}_N^A(\mathbb{T}^d)$, we compute

$$\begin{aligned}
\frac{1}{N} \sum_{w^A=z} \overline{(M_U)_i(w)} (M_U)_j(w) &\underset{\text{by (32)}}{=} \frac{1}{N} \sum_{w^A=z} \overline{(U(z)(w^k)_i} (U(z)(w^k)_j \\
&= \frac{1}{N} \sum_{w^A=z} \overline{w^{k_i}} w^{k_j} \\
&= \frac{1}{N} \sum_{w^A=z} w^{k_j - k_i} \\
&= \delta_{i,j}
\end{aligned}$$

where we have used the usual Fourier duality for the two finite groups

$$A^{-1}\mathbb{Z}^d/\mathbb{Z}^d \simeq \mathbb{Z}^d/A\mathbb{Z}^d; \tag{35}$$

i.e., a finite Fourier transform.

Case (ii). Given $M \in \mathscr{M}_N^A(\mathbb{T}^d)$, we compute $U_{i,j}(z)$ according to (34). The claim is that $U(z) = (U_{i,j}(z))$ is in $U_N(\mathbb{C})$, i.e.,

$$\sum_{l=1}^{N} \overline{U_{l,j}(z)} U_{l,j}(z) = \delta_{i,j} \quad \text{for all } z \in \mathbb{T}^d. \tag{36}$$

Proof. Proof of (36):

$$\begin{aligned}
&\sum_{l=1}^{N} \overline{U_{l,j}(z)} U_{l,j}(z) \\
&\underset{\text{by (34)}}{=} \frac{1}{N^2} \sum_{l} \sum_{w} \sum_{w'} M_i(w) \overline{M_l(w) M_j(w')} M_l(w') \\
&= \frac{1}{N} \sum_{w} \sum_{w'} \delta_{w,w'} M_i(w) \overline{M_j(w')} \\
&= \frac{1}{N} \sum_{w} M_i(w) \overline{M_j(w)} \\
&\underset{\text{by (30)}}{=} \delta_{i,j}.
\end{aligned}$$

This completes the proof of (36) and therefore the Lemma. In the summations we consider independent pairs $w, w' \in \mathbb{T}^d$ satisfying $w^A = w'^A = z$. So each ranges over a set of cardinality $N = |detA|$.

In the simplest case of $N = 2$, $d = 1$, we have just two frequency bands, are two filter functions

$$M = \begin{pmatrix} m_0 \\ m_1 \end{pmatrix}$$

see (28), (29). In that case we select the two points ± 1 in bt. In additive notation, these represents the two elements in $\frac{1}{2}\mathbb{Z}/\mathbb{Z}$ viewed as a subgroup of $\mathbb{T} \simeq \mathbb{R}/\mathbb{Z}$. The band-conditions are then $M_0(1) = 1$ and $M_1(-1) = 1$, in addition notation:

Multi-band: high-pass/low-pass. If $N > 2$, there will be more than two frequency bands. They can be written with the use of duality for the finite group $A^{-1}\mathbb{Z}^d/\mathbb{Z}^d$ in (36). Recall $|detA| = N$, and the group is cyclic of order N. The matrix for its Fourier transform matrix is denoted H_N with H Hadamard.

$$H_N = \frac{1}{\sqrt{N}} \left(e^{i\frac{2\pi jk}{N}} \right)_{j,k \in \mathbb{Z}_N}. \tag{37}$$

Lemma 3. *If $U \in \mathscr{U}_N^A(\mathbb{T}^d)$ and $M_U = U(z^A)M_0(z)$ is the multiresolution for the lemma then*

$$\mathbb{T}^d \ni z \mapsto H_N M_U(z) \tag{38}$$

satisfying the multi-band frequency pass condition.

Proof. Immediate.

Lemma 4. *Let M be a multi-band frequency pass function, see (38), and assume continuity at the zero-multi frequency, i.e., at the unit element in $\mathbb{T}^d$ (with multiplicative notation).*

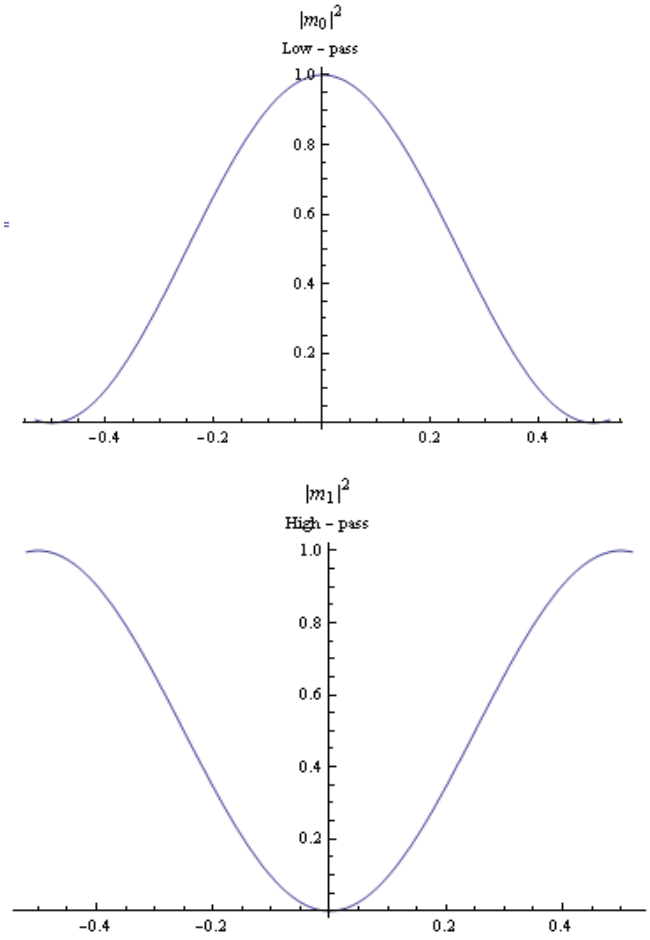


Fig. 8 The two filters of probability distributions.

(a)Then there are functions $(\psi_j)_{j=0}^{N-1}$ *in* $L^2(\mathbb{R}^d)$ *such that* $\psi_0 = \varphi$ *satisfies*

$$\widehat{\varphi}(\theta) = \prod_{k=1}^{\infty} m_0(A^{-1}\theta) \tag{39}$$

$$\widehat{\psi_j}(\theta) = m_j(A^{-1}\theta)\widehat{\varphi}(A^{-1}\theta), \quad j = 1, \cdots, N-1, \quad \textit{for all } \theta \in \mathbb{R}^d \tag{40}$$

where we have used addition notation, i.e., $\theta \in \mathbb{R}^d$ *is written as a column vector, and* $A^{-1}\theta$ *is a matrix action.*

(b)With the tripple index set $j = 1, 2, \cdots, N-1$, $k \in \mathbb{Z}$, *and* $l \in \mathbb{Z}^d$, *we get the system*

$$\psi_{j,k,l}(x) = |detA|^{k/2}\psi_j(A^k x - l). \tag{41}$$

(c) While the system in (41) in general does not form an orthonormal basis (ONB) in $L^2(\mathbb{R}^d)$, it satisfies the following Parseval-fram condition: For every $f \in L^2(\mathbb{R}^d)$, we have

$$\int_{\mathbb{R}^d} |f(x)|^2 dx = \sum_{1\le j<N} \sum_{k\in\mathbb{Z}} \sum_{l\in\mathbb{Z}^d} |\langle \psi_{j,k,l}, f\rangle_{L^2(\mathbb{R}^d)}|^2 \tag{42}$$

with the expression $\langle\cdot,\cdot\rangle_{L^2(\mathbb{R}^d)}$ on the RHS in (42) representing the usual $L^2(\mathbb{R}^d)$ inner product

$$\langle \psi, f\rangle_{L^2(\mathbb{R}^d)} = \int_{\mathbb{R}^d} \overline{\psi(x)} f(x) dx. \tag{43}$$

Proof. The essential idea is contained in [9], and the remaining details are left to the reader.

Remark 1. The condition in (42) (Parseval-frame) is weaker than ONB, referring to function $(\psi_{j,k,l})$ in (41). Rather thant asking for an ONB in $L^2(\mathbb{R}^d)$, we seek instead a Parseval-frame. But there is a variety of explicit additional conditions on the given wavelet filter from (39) and (40) which imply that $(\psi_{j,k,l})$ is in fact an ONB.

In the language of frequency bands, the low pass filter should not pass "false" frequencies. The simplest example of a non ONB Parseval wavelet is the stretched Haar wavelet

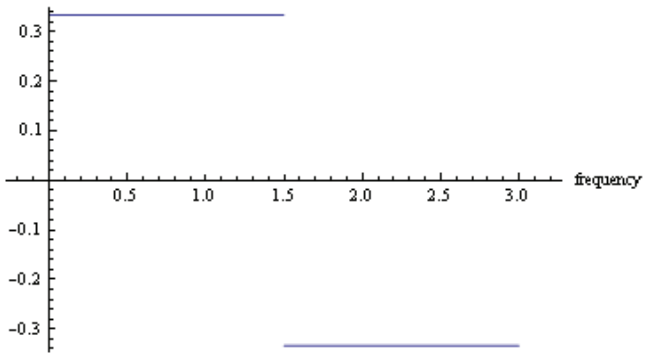


Fig. 9 Streched Haar Wavelet.

$$\psi(x) = \begin{cases} \frac{1}{3} & \text{if } 0 \le x < \frac{3}{2} \\ \frac{-1}{3} & \text{if } \frac{3}{2} \le x < 3 \\ 0 & \text{if } x \in \mathbb{R} \setminus [0,3]. \end{cases}$$

For additional details, see [9].

Lemma 5. *Let*

$$M(z) = \sum_{k\in\mathbb{Z}^d} h_k z^k, \quad \textit{with} \quad z = e^{i2\pi\theta}, \quad \theta \in \mathbb{R}^d, \tag{44}$$

$$M(\theta) = \sum_{k\in\mathbb{Z}^d} h_k e^{i2\pi k\cdot\theta}, \quad k\cdot\theta = \sum_{j=1}^{d} k_j\theta_j, \tag{45}$$

be the Fourier expansion of a filter functionl note (45) is the same as (44), but written in the additional form of Fourier analysis for $\mathbb{T}^d$. Then the function φ in (39) satisfies

$$\varphi(x) = |detA| \sum_{k\in\mathbb{Z}^d} h_k \varphi(A^t x - k) \tag{46}$$

where A^t denotes the transpose matrix to A.

Proof. The result follows from an $L^2(\mathbb{R}^d)$-Fourier transform applied to both sides in formular (46).

Lemma 6. *Let M be a low-pass filter function, and let φ be a scaling function, see (46), or equivalently (39). Then the operator*

$$T_\varphi : l^2(\mathbb{Z}^d) \to L^2(\mathbb{R}^d)$$

given by

$$T_\varphi((\xi)_k(x)) = \sum_{k\in\mathbb{Z}^d} (\xi)_k \varphi(x-k)$$

is isometric, i.e., we have

$$\sum_{k\in\mathbb{Z}^d} |\xi_k|^2 = \int_{\mathbb{R}^d} |\sum_{k\in\mathbb{Z}^d} \xi_k \varphi(x-k)|^2 dx.$$

Proof. See [9].

6 Consistency of Tensor Product for Filters and for Functions on $\mathbb{R}^d$

There are two operations we shall need to perform on infinite slanted matrices, matrix product and tensor product. The first is for the computation of the coefficients in expansions with scale similar orthogonal bases, such as wavelet bases. We prove that when these operations are performed, the resulting new slanted matrices become increasingly more sparse (i.e., the resulting matrices have wide patterns of zeros) which makes computations fast.

Since, by Lemma 5, the filter functions coming from the group $\mathscr{U}_N^A(\mathbb{T}^d)$, it follows that the corresponding filter functions $\mathscr{M}_N^A(\mathbb{T}^d)$ are also closed under tensor product.

6.1 Creating Multiresolutions and Wavelets

Using the lemmas, we may create new filters and new multiresolutions from old. In outline, the process is as follows:

- projections.
  ↓
- unitary matrices.
  ↓
- unitary matrix-functions.
  ↓
- filter functions.
  ↓
- tensor-product.
  ↓
- scaling functions.
  ↓
- multi-resolutions.
  ↓
- wavelet functions (higher dimensions with tensor-product).

Definition 5. Let $\mathscr{H}_i$, $i = 1,2$ be Hilbert spaces, and let T_i, $i = 1,2$, be linear operators in the respective spaces. By the tensor-product $T_1 \otimes T_2$ we mean the operator $T_1 \otimes T_2$ in $\mathscr{H}_1 \otimes \mathscr{H}_2$ given by

$$(T_1 \otimes T_2)(u_1 \otimes u_2) = (T_1 u_1) \otimes (T_2 u_2), \quad \text{for all } u_i \in \mathscr{H}_i, \quad i = 1,2. \tag{47}$$

An operator P in a Hilbert space $\mathscr{H}$ is called a *projection* if $P = P^* = P^2$. An operator U is called *unitary* if and only if $U^* = U^{-1}$, i.e., $UU^* = U^*U = I$ where I denotes the identity operator.

Easy Facts:

(i) The tenstor product of two projections is a projection.
(ii)The tensor product of two unitary operators is a unitary operator.
(iii)If P is a projection in a Hilbert space $\mathscr{H}$, then the function

$$U(z) := zP + (I - P) = zP + P^{\perp}, \quad (z \in \mathbb{T}) \tag{48}$$

maps $\mathbb{T}$ into the group of all unitary operators.

Proof. of (iii). The conclusion follows from representing (48) in the following operator-block matrix form

$$U(z) = \begin{bmatrix} z & 0 \\ 0 & 1 \end{bmatrix}_{P^{\perp}}^{P} \tag{49}$$

Note that the RHS in (49) is unitary if and only if $|z| = 1$.

Definition 6. (1)Let A be a $d\times d$ matrix on bz such that $\mathrm{spec}(A)\subset\{\lambda\in\mathbb{C};|\lambda|>1\}$. Let $\mathbb{Z}^d\overset{\pi_A}{\to}\mathbb{Z}^d/A\mathbb{Z}^d=:Q_A$ be the natural quotient mapping. If $\rho_A:\mathbb{Z}^d/A\mathbb{Z}^d\to bz^d$ satisfying $\pi_A\circ\rho_A=id$, we saw that

$$\mathbb{T}^d\ni z\mapsto\left(z^{\rho_A(q)}\right)_{q\in Q_A}$$

is a multiresolution filter on $\mathbb{T}^d$.

(2)If B is a $e\times e$ matric over $\mathbb{Z}$ such that $\mathrm{spec}(B)\subset\{\lambda\in\mathbb{C};|\lambda|>1\}$, we introduce $Q_B=\mathbb{Z}^e/B\mathbb{Z}^e$ and ρ_B as in (1) by analogy. We then get a $d+e$ multiresolution filter

$$\mathbb{T}^d\times\mathbb{T}^e\ni(z,w)\mapsto\left(z^{\rho_A(q)}w^{\rho_B(r)}\right)_{q\in Q_A,r\in Q_B}.$$

(3)If $N_A:=|detA|$, and $N_B:=|detB|$, then the filter in (2) takes values in $bc^{N_A+N_B}$.

Corollary 1. *The families of multiresolution filters $\mathscr{M}^A_{N_A}(\mathbb{T}^d)$ is closed under tensor product, i.e.,*

$$\mathscr{M}^A_{N_A}(\mathbb{T}^d)\otimes\mathscr{M}^B_{N_B}(\mathbb{T}^e)\subset\mathscr{M}^{A\otimes B}_{N_A\cdot N_B}(\mathbb{T}^{d+e}).$$

Proof. By the lemma, we only need to observed that the unitary matrix-functions $\mathbb{T}^d\ni z\mapsto U_A(z)\in U_{N_A}(\mathbb{C})$ are closed under tensor-product, i.e.,

$$\mathbb{T}^d\otimes\mathbb{T}^e\ni(z,w)\mapsto U_A(z)\otimes U_B(w)\in U_{N_A\cdot N_B}(\mathbb{C}).$$

6.2 Tensor-products: Applications and Examples

Lexicographical of basis vectors, for example $\mathbb{C}^2\otimes\mathbb{C}^2\to$ (11), (12), (21), (22). Scaling matrices:

$$I\otimes B\longrightarrow\begin{pmatrix}B&0\\0&B\end{pmatrix}.$$

Fourier Hadamard transform matrices:

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1&1\\1&-1\end{pmatrix}\otimes\frac{1}{\sqrt{2}}\begin{pmatrix}1&1\\1&-1\end{pmatrix}\longrightarrow\frac{1}{2}\begin{pmatrix}1&1&1&1\\1&-1&1&-1\\1&1&-1&-1\\1&-1&-1&1\end{pmatrix}$$

$$H_2\otimes H_2\neq H_4.$$

Compare with the matrix F_4 for the Fourier transform on bz_4:

$$H_4:\frac{1}{2}\begin{pmatrix}1&1&1&1\\1&i&-1&-i\\1&-1&1&-1\\1&-i&-1&i\end{pmatrix}.$$

6.3 Unitary Matrix-functionsi $U_p(z)$:

$$\mathbb{T} \ni z \longmapsto zP + P^{\perp} = \begin{pmatrix} zP & 0 \\ 0 & I-P \end{pmatrix}. \tag{50}$$

Here the function $U_P(z)$ is constructed from a fixed projection P in a Hilbert space $\mathscr{H}$, and we set $P^{\perp} = I_{\mathscr{H}} - P$ the projection onto the orthogonal complement of $P\mathscr{H}$ in $\mathscr{H}$.

Then for $(z,w) \in \mathbb{T}^2 = \mathbb{T} \times \mathbb{T}$ we have:

$$U_P(z) \times U_P(w) = \begin{pmatrix} zwP \otimes P & 0 \\ 0 & I_{\mathscr{H}\otimes\mathscr{H}} - P \otimes P \end{pmatrix}.$$

Giving (50) the matrix representation

$$\begin{pmatrix} z & 0 \\ 0 & 1 \end{pmatrix}$$

we get

$$U_P(z) \otimes U_P(w) = \begin{pmatrix} zw & 0 & 0 & 0 \\ 0 & z & 0 & 0 \\ 0 & 0 & w & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{51}$$

For the slanted matrices

$$F = \begin{pmatrix} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

we see that sparsity increases when both matrix multiplication and tensor product $\otimes$ is applied. If for example the two matrices F and F' have slanting degree 2, then the matrix product FF' has slanting degree 4. An $n-$fold product of slanted matrices of slant 2 is again slanted but of slanting degree 2^n.

The situation for tensor product is more subtle.

Example 3. Set

$$F = \begin{pmatrix} a & b & 0 \\ 0 & 0 & a \\ 0 & 0 & 0 \end{pmatrix},$$

and

$$F' = \begin{pmatrix} \alpha & \beta & 0 \\ 0 & 0 & \alpha \\ 0 & 0 & 0 \end{pmatrix}.$$

Then $F \otimes F'$ has the matrix representation

$$\begin{pmatrix}
a\alpha & a\beta & 0 & b_\alpha & b\beta & 0 & 0\,0 & 0 \\
0 & 0 & a\alpha & 0 & 0 & b\alpha & 0\,0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0\,0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0\,0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0\,0 & a\alpha \\
0 & 0 & 0 & 0 & 0 & 0 & 0\,0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0\,0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0\,0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0\,0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0\,0 & 0
\end{pmatrix}.$$

6.3.1 Slanted Matrices

Definition 7. An (infinite) matrix $F = (a_{j,k})$, $j,k \in \mathbb{Z}$, is said to be slanted of degree d if there is a function f on $\mathbb{Z}$ (depending only on F) such that

$$a_{j,k} = f_F(k - d\,j) \tag{52}$$

holds for all $j,k \in \mathbb{Z}$. We write $deg_s(F) = d$.

Lemma 7. *Let F and F' be slanted matrices, with F of slanting degree d and F' of degree e. Then the matrix product $G = FF'$ is again a slanted matrix, and for the degree we have*

$$deg_s(FF') \geq deg_s(F) \cdot deg_s(F'). \tag{53}$$

Proof. (i) By matrix-multiplicative. For the entries in $G = FF'$,

$$G = (c_{i,j}), \quad i,j \in \mathbb{Z},$$

$$c_{i,j} = \sum_{k\in\mathbb{Z}} a_{i,k}a'_{k,j} = \sum_{k\in\mathbb{Z}} f_F(k - di)\, f_{F'}(j - ek).$$

Hence the matrix entries $c_{i,j+(d\cdot e)i}$ are represented by a function g in j, i.e.,

$$c_{i,j+(d\cdot e)i} = g(j), \quad \text{for all } i,j \in \mathbb{Z}.$$

(ii)By generating functions. Because of the assumptions on the entries in the matrices $F = (a_{i,j})$ and $F' = (a'_{i,j})$ the generating functions (alias, frequency response functions) are in fact Fourier series.

As a result, a slanted matrix F represents a bounded operator T_F in $L^2(\mathbb{T}) \simeq l^2(\mathbb{Z})$. An easy computation shows that F is slanted of degree d if and only if

$$T_F M_{z^d} = M_z T, \tag{54}$$

where

$$(M_z\xi)(z) = z\xi(z), \quad \text{for all } z \in \mathbb{T};$$

and

$$(M_z d\xi)(z) = z^d \xi(z), \quad \xi \in L^2(\mathbb{T});$$

i.e., $M_{z^d} = (M_z)^d$, $d \in \mathbb{Z}_+$.

As a result, we have

$$T_F M_{z^d} = M_z T_F, \quad \text{and} \tag{55}$$

$$T_{F'} M_{z^e} = M_z T_{F'}. \quad \text{Since} \tag{56}$$

$$T_G = T_F T_{F'} \quad \text{we get} \tag{57}$$

$$\begin{aligned} M_z T_G &\underset{by(57)}{=} M_z T_F T_{F'} \\ &\underset{by(55)}{=} T_F M_{z^d} T_{F'} \\ &\underset{by(56)}{=} T_F T_{F'} M_{(z^d)^e} \\ &= T_G M_{z^{d\cdot e}}, \end{aligned}$$

proving the assertion in the lemma.

The proof of the following result is analogous to that in Lemma 7 above.

Lemma 8. *Let F and F' be slanted matrices with $\mathbb{Z}$ as index set for rows and columns, or possibly subset of $\mathbb{Z}$. If $\deg_s(F) = d$, and $\deg_s(F') = e$, then $F \otimes F'$ is slanted relative to the index set $\mathbb{Z}^2$ by vector degree (d,e). Setting*

$$(F \otimes F')_{(i,j),(k,l)} = F_{i,k} F'_{j,l} \tag{58}$$

there is a function g on $\mathbb{Z}^2$ such that

$$(F \otimes F')_{(i,j),(k,l)} = g(k - di, l - ej) \quad \textit{for all } (i,j) \in \mathbb{Z}^2 \quad \textit{and all } (k,l) \in \mathbb{Z}^2.$$

7 Rep$(\mathscr{O}_N, \mathscr{H})$

Here we show that the algorithms developed in the previous two sections may found from certain representations of an algebra in a family indexed by a positive integer N, called the Cuntz algebras $\mathscr{O}_N$.

It will be important to make a distribution between $\mathscr{O}_N$ as a C^*-algebra, and its representation by operators in some Hilbert space. As we show distinct representation of $\mathscr{O}_N$ yield distinct algorithms, distinct wavelets, and distinct matrix computations.

It is known that $\mathscr{O}_N$ is the unique (up to isomorphism) C^*-algebra on generated $(s_i)_i \in \mathbb{Z}_N$, and relations

$$s_i^* s_j = \delta_{i,j} 1, \quad \sum_{i \in \mathbb{Z}_N} s_i s_i^* = 1. \tag{59}$$

Here s_i in (59) is a symbolic expression and 1 denotes the unit-element in the C^*-algebra generated by $\{s_i; i \in \mathbb{Z}_N\}$. Hence a representation ρ of $\mathscr{O}_N$ acting on a Hilbert space $\mathscr{H}$, $\rho \in \mathrm{Rep}(\mathscr{O}_N, \mathscr{H})$, is a system of operators $S_i = S_i^{(\rho)} = \rho(s_i)$ satisfying the same relations (59), but with 1 replaced by $I_{\mathscr{H}} = \rho(1) =$ the identity operator in $\mathscr{H}$.

From (59) it follows that $\mathscr{O}_N \otimes \mathscr{O}_M = \mathscr{O}_{NM}$. Hence it follows from an analysis of tensor-product of representation that not all

$$\rho \in \mathrm{Rep}(\mathscr{O}_N, \mathscr{H})$$

is a tensor product of a pair of representations, one of $\mathscr{O}_N$ and the second of $\mathscr{O}_M$.

References

1. Ash, R. B. *Information Theory*. Dover Publications Inc., New York, (1990) Corrected reprint of the 1965 original.
2. Aldroubi, A., Cabrelli, C., Hardin, D., Molter, U.: Optimal Shift Invariant Spaces and Their Parseval Frame Generators. Appl. Comput. Harmon. Anal. **23**, 2, 273–283 (2007)
3. Albeverio, S. and Jorgensen, P. E. T. and Paolucci, A. M. Multiresolution Wavelet Analysis of Integer Scale Bessel Functions.: J. Math. Phys. **48**, 7, 073516 (2007)
4. Ball, J. A. and Vinnikov, V.: Functional Models for representations of the Cuntz algebra. Operator theory, systems theory and scattering theory: multidimensional generalizations, Oper. Theory Adv. Appl. **157**, 1–60, Birkhäuser, Basel (2005)
5. Cleary J. G. Witten I. H. Bell, T. C. Text Compression, Prentice Hall, Englewood Cliffs, (1990)
6. Baggett, L., Jorgensen, P. E. T., Merrill, K., Packer, J. A non-MRA C^r frame wavelet with rapid decay. *Acta Appl. Math.*, (2005)
7. Bose, T.,: Digital Signal and Image Processing, Wiley (2003)
8. Bose, T., Chen, M.-Q., Thamvichai, R. Stability of the 2-D Givone-Roesser model with periodic coefficients: IEEE Trans. Circuits Syst. I. Regul. Pap. **54**, 3, 566–578 (2007)
9. Bratelli, O., Jorgensen, P. E. T. *Wavelets Through a Looking Glass: The World of the Spectrum*. Birkhäuser, (2002)
10. Burger, W.: Principles of Digital Image Processing: Fundamental Techniques, Springer (2009)
11. Burdick, H. E.: Digital Imaging: Theory and Applications, Mcgraw-Hill (1997)
12. John G. Cleary, J. G., Witten, I. H. A comparison of enumerative and adaptive codes. *IEEE Trans. Inform. Theory*, **30**(2, part 2) 306–315, (1984)
13. Daubechies, I. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. (1992)
14. Donoho, D. L., Vetterli, M., DeVore, R. A., Daubechies, I. Data compression and harmonic analysis. *IEEE Trans. Inform. Theory*, **44**, 6, 2435–2476, (1998)
15. Dutkay, D. E., Roysland, K. Covariant representations for matrix-valued transfer operators. *arXiv:math/0701453*, (2007)
16. Hutchinson, J. E.: Fractals and self-similarity. Indiana Univ. Math. J. **30**, 5, 713–747 (1981)
17. Green, P., MacDonald, L.: Colour Engineering: Achieving Device Independent Colour, Wiley (2002)

18. Jaffard, S., Meyer, Y., Ryan, R. D. *Wavelets Tools for science & technology.*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, revised edition, (2001)
19. Jorgensen, P. E. T. *Analysis and probability: wavelets, signals, fractals*, volume 234 of *Graduate Texts in Mathematics*. Springer, New York, (2006)
20. Jorgensen, P. E. T., Kornelson, K., Shuman, K.: Harmonic analysis of iterated function systems with overlap. J. Math. Phys. **48**, 8, 083511 (2007)
21. Jorgensen, P. E. T., Mohari, A.: Localized bases in $L^2(0,1)$ and their use in the analysis of Brownian motion. J. Approx. Theory **151**, 1, 20–41 (2008)
22. Jorgensen, P. E. T., Song, M.-S. Comparison of discrete and continuous wavelet transforms. *Springer Encyclopedia of Complexity and Systems Science*, (2007)
23. Jorgensen, P. E. T., Song, M.-S. Entropy encoding, hilbert space, and karhunen-loève transforms. *J. Math. Phys.*, **48**, 10, 103503, (2007)
24. MacKay, D. J. C.: Information Theory, Inference and Learning Algorithms, Cambridge University Press, New York, (2003)
25. Keyl, M.: Fundamentals of Quantum Information Theory, Phys. Rep. **369**, 5, 431–548, (2002)
26. Rastislav L., Plataniotis, K. N.: Color Image Processing: Methods and Applications, CRC; 1 edition (2006)
27. Roman, S.: Introduction to Coding and Information Theory: Undergraduate Texts in Mathematics. Springer-Verlag, New York, (1997)
28. Lukac, R., Plataniotis, K. N., Venetsanopoulos, A. N.: Bayer pattern demosaicking using local-correlation approach. Computational science—ICCS 2004. Part IV, Lecture Notes in Comput. Sci. **3039**, 26–33, Berlin, Springer (2004)
29. MacDonald, L. (Editor), Luo, M. R. (Editor): Colour Image Science: Exploiting Digital Media, Wiley; 1 edition (2002)
30. Russ, J. C., *The image processing handbook*, 5th edn. (CRC Press, Boca Raton, FL, (2007)
31. Salomon, D.: Data compression. The complete reference. 4th edn. With contributions by Giovanni Motta and David Bryant. Springer-Verlag, London, (2007)
32. Shannon C.E., Weaver W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana and Chicago (1998)
33. Skodras, A., Christopoulos, C., Ebrahimi, T. Jpeg 2000 still image compression standard" ieee signal processing magazine. *IEEE Signal processing Magazine*, **18**, 36–58, (2001)
34. Song, M.-S. Wavelet image compression. In *Operator theory, operator algebras, and applications*, **414** *Contemp. Math.*, 41–73. Amer. Math. Soc., Providence, RI, (2006)
35. Song, M.-S. Entropy encoding in wavelet image compression. *Representations, Wavelets and Frames A Celebration of the Mathematical Work of Lawrence Baggett*, 293–311, (2008)
36. Usevitch, B. E. A tutorial on modern lossy wavelet image compression: Foundations of jpeg 2000. *IEEE Signal processing Magazine*, **18**, 22–35, (2001)
37. Walker, J. S. *A Primer on Wavelets and their Scientific Applications*. Chapman & Hall, CRC, (1999)
38. Witten, I. H. Adaptive text mining: inferring structure from sequences. *J. Discrete Algorithms*, **2**, 2, 137–159, (2004)
39. Witten, I. H., Neal, R. M., Cleary, J. G. Arithmetic coding for data compression. *Comm. of the ACM*, **30**, 6, 520–540, (1987)