

Tilings, Scaling Functions, and a Markov Process

Richard F. Gundy

Introduction

We discuss a class of Markov processes that occur, somewhat unexpectedly, in the construction of wavelet bases obtained from *multiresolution analyses* (MRA). The processes in question have been around for a long time. One of the first references that should be cited is a paper by Doebelin and Fortêt [10] (1937), entitled “Sur les chaînes à liaisons complètes”. In English, they are sometimes called “historical Markov processes” and have been used extensively to study the Ising model. However, their wavelet connection does not seem to have filtered into the standard texts on time-scale analysis.

The material for this article is drawn from the publications [9], [12], [13], as well as the prior contributions by various people who are cited in the appropriate places. To keep the exposition self-contained and as elementary as possible, we discuss the special case of the “quincunx matrix” in which the proofs are somewhat simpler and, in some cases, radically different from those found in the above references.

In the first section, we describe a remarkable coincidence: two discoveries, the first concerning a gambling strategy and the second concerning a wavelet basis, both leading to the same mathematics. *It is remarkable that these discoveries, each of fundamental importance, were separated by 330 years!* The wavelet discovery was a particular class of trigonometric polynomials within a class of functions called *quadrature mirror filters* (QMF functions, for short). These functions turn out to be probabilities; hence their connection to gambling.

Richard F. Gundy is professor of statistics and mathematics at Rutgers University. His email address is gundy@rci.rutgers.edu.

The names associated with the gambling strategy are Pascal and Fermat; the wavelet discovery to which we refer is due to Ingrid Daubechies.

In the subsequent sections, we are concerned with the class of 1-periodic functions $p(\xi)$, quadrature mirror filters, that generate an MRA on \mathbb{R} or \mathbb{R}^2 . Some do so, but most do not; our problem is to find out which is which. The probability methods presented here provide new information on this topic. We show that there is a one-to-one correspondence between two disparate classes of scaling functions, one defined on \mathbb{R} , the other defined on \mathbb{R}^2 . Second, the probability perspective allows us to exhibit a large class of continuous $p(\xi)$ that generate MRAs. When the function $p(\xi)$ is smooth and generates an MRA, it must satisfy known necessary conditions. However, these necessary conditions may be violated in the extreme for MRAs in which the generator $p(\xi)$ is smooth *except* at a few points. “Few” here means as few as four.

In one dimension, all of the MRAs we consider involve the dilation $\mathbb{Z} \rightarrow 2\mathbb{Z}$; in two dimensions, an interesting special case is the “quincunx” dilation, described below. In the first case, where $\xi \in \mathbb{R}^1$, the QMF function $p(\xi)$ is periodic (with period one); in the second case, when $\xi \in \mathbb{R}^1$, $p(\xi)$ is doubly periodic on the unit square. From this, one might assume that the natural fundamental domains for these functions would be $\{(0, 1), \mathbb{Z}\}$ or $\{(0, 1) \times (0, 1), \mathbb{Z}^2\}$. But the natural assumption is too naïve. In one dimension, the appropriate domain is sometimes a disconnected set called a C-tile, described below. In two dimensions, the “fundamental” fundamental domain is a fractal set called the “twin dragon”. But in the final analysis, these tiles and dragons will both be superseded

by the space $2^{\mathbb{Z}}$ of infinite binary sequences. To explain how this comes about, we will make a detour in which we examine some aspects of radix representation for \mathbb{Z} and \mathbb{Z}^2 .

Some Definitions

In the discussion that follows, we distinguish two copies of Euclidean space: $t \in \mathbb{R}^n$ is considered to be the “time” variable and $\xi \in \mathbb{R}^n$ the (Fourier transform) “frequency” variable. A QMF function $p(\xi), \xi \in \mathbb{R}^n, 0 \leq p(\xi) \leq 1$ is, first of all, a non-negative function that is periodic with period one in each variable. There is a dilation associated with $p(\xi)$, an $n \times n$ matrix $B(i, j)$ with integer entries, with the property that all the eigenvalues λ satisfy $|\lambda| > 1$. Matrices with these properties are called expanding. Since B maps $\mathbb{Z}^n \rightarrow \mathbb{Z}^n$, we may choose a full set of coset representatives $d^k, k = 0, 1, \dots, q-1, (q = |\det B|)$ from $\mathbb{Z}^n/B(\mathbb{Z}^n)$. The vectors d^k are called digits. A proof that $q = |\det(B)|$ may be found in Wojtaszczyk ([23], Proposition 5.5, page 109). For any such selection we assume that $p(\xi)$ satisfies the QMF condition

$$(1) \quad p(0) = \sum_{k=1}^m p(B^{-1}(\xi + d^k)) = 1.$$

We always assume that $p(\xi)$ is continuous in the topology of the moment. At this moment it is the usual topology. Later, it will be a totally disconnected topology. In one dimension, for the dilation 2 with $d^0 = 0, d^1 = 1$, condition (1) is simply $p(\xi/2) + p((\xi + 1)/2) = 1$, and $p(0) = 1$.

Two questions are addressed in this article: (1) How do we construct QMF functions? (2) What are necessary and sufficient conditions for a given function to generate a scaling function for an MRA in \mathbb{R}^1 and \mathbb{R}^n ? We concentrate on two special cases: (a) one dimension, with dilation 2; (b) two dimensions, with a particular matrix $B, \det B = 2$. The discussion of the general case of two-by-two matrices B with $\det B = \pm 2$ may be found in [13].

What Is a Scaling Function?

First of all, it is a real-valued function $\varphi(t), t \in \mathbb{R}$, that satisfies a *self-similarity* property with respect to a group of *affine* transformations generated by (A, \mathbb{Z}^n) , where A is an expanding matrix with integer entries. In all cases, it will turn out that the transpose of $A := B$, where B is the matrix specified in the previous section, as we shall see presently. Second, $\varphi(t)$ is a function whose translates of $\varphi(t - j), j \in \mathbb{Z}^n$, form an orthonormal family, spanning a subspace V_0 of $L^2(\mathbb{R}^n)$. We let $V_j \subset V_{j+1}, j \in \mathbb{Z}$ be the sequence of subspaces of $L^2(\mathbb{R}^n)$ with the property $f(t) \in V_0$ if and only if $f(A^j t) \in V_j$. Furthermore, assume that there is a

sequence $\{c_k; k \in \mathbb{Z}^n\} \in l^2$ such that

$$(a) \quad \varphi(t) = |\det(A)|^{1/2} \sum_{k \in \mathbb{Z}^n} c_k \varphi(At - k)$$

$$\text{with } \sum_{k \in \mathbb{Z}^n} |c_k|^2 = 1$$

$$(b) \quad \int \varphi(t) \varphi(t - k) dt = \delta(0, k), \quad k \in \mathbb{Z}^n;$$

$$(2) \quad (c) \quad \lim_{j \rightarrow \infty} \|P_j f - f\| = 0, \quad \text{for all } f(t) \in L^2(\mathbb{R}^n),$$

where P_j is the orthogonal projection onto V_j .

The triple $(\varphi, A, \{V_j, j \in \mathbb{Z}\})$ defines the MRA. The coefficients $c_k, k \in \mathbb{Z}^n$ specified in property (a) will be called *refinement coefficients*. These coefficients appear in the function $p(\xi)$ from the Introduction. See (2) below.

Some scaling functions are of the form $\varphi(t) = \chi_T(t)$ where $\chi_T(t)$ is the indicator function of a subset of \mathbb{R}^n of the form

$$T = T(A, \mathcal{D})$$

$$(3) \quad := \{t \in \mathbb{R}^n : t = \sum_{j=1}^{\infty} A^{-j} d_j, \quad d_j \in \mathcal{D}, j \geq 1\}$$

where for each $j \geq 1, d_j$ is chosen from the set $\mathcal{D} = \{d^k, k = 0, 1, \dots, q-1\}$ of coset representatives of the group $\mathbb{Z}^n/A(\mathbb{Z}^n)$. These sets are *self-affine* in the sense that $\chi_T(t) = \sum_{d_j \in \mathcal{D}} \chi_T(At - d_j)$. This is property (a) for indicator scaling functions.

Definition. A set T is a *tile* if $\sum_{j \in \mathbb{Z}^n} \chi_T(t - j) = 1$ almost everywhere. For indicators, this is property (b). The primary tile example is the Haar scaling function: here T is the unit interval $t = \sum_{j=1}^{\infty} 2^{-j} d_j, \mathcal{D} = \{0, 1\}$. In this case, T is self-affine; however, not every tile we will encounter is self-affine. (See the section “An Infinite Sequence of Games”).

The probability theory begins when we take the squared modulus of the Fourier transform $|\hat{\varphi}(\xi)|^2$ of the scaling function $\varphi(t)$ defined in (2). (For functions on \mathbb{R}^n we write the Fourier transform $\hat{\varphi}(\xi) = \int \varphi(t) \exp(-2\pi i \langle t, \xi \rangle) dt$; for sequences $c_k, k \in \mathbb{Z}^n, \hat{c}(\xi) = \sum_{k \in \mathbb{Z}^n} c_k \exp(-2\pi i \langle k, \xi \rangle)$).

The three properties in (2) may be expressed in terms of B , the transpose of A :

$$(\hat{a}) \quad |\hat{\varphi}(\xi)|^2 = p(B^{-1}\xi) |\hat{\varphi}(B^{-1}\xi)|^2.$$

$$(\hat{2}) \quad \text{Here } p(\xi) = (|\det B|)^{-1} |\hat{c}(\xi)|^2 = \left| |\det B|^{-1/2} \sum_{k \in \mathbb{Z}^n} c_k \exp(-2\pi i \langle \xi, k \rangle) \right|^2$$

where $c_k, k \in \mathbb{Z}^n$ are the refinement coefficients from display (2). The periodic function $\tilde{p}(\xi) := |\det(B)|^{-1/2} \hat{c}(\xi)$ is called a *low-pass filter*.

$$(\hat{b}) \quad \sum_{k \in \mathbb{Z}^n} |\hat{\varphi}(\xi + k)|^2 = 1 \text{ a.e.};$$

$$(\hat{c}) \quad \lim_{k \rightarrow \infty} |\hat{\varphi}(B^{-k}\xi)|^2 = 1 \text{ a.e.}$$

The conditions (\hat{a}) , (\hat{c}) imply condition (a) and (a) , (c) imply (\hat{c}) . See Madych ([18], Proposition 1, and Corollaries 2 and 3, page 266), and Hernández and Weiss ([14], Theorem 1.6, page 45, and Theorem 5.2, page 382).

Now we record two probability comments that are fundamental to what follows. First, a low-pass filter is the Fourier transform of a discrete probability distribution $\{|\det(B)|^{-1/2} c_k; k \in \mathbb{Z}^n\}$ if and only if $\varphi(t)$ is the indicator function of a tile. This follows from two observations: (a) the Lebesgue measure of a tile is necessarily one (see the section “Coding \mathbb{R}^1 and \mathbb{R}^2 into \mathbb{Z}^2 : Case 1; \mathbb{Z}^1 vs. \mathbb{Z}^2 ”); (b) a nonnegative function is orthogonal to its translates only if the supports of the translates are disjoint. If $\varphi(t)$ is a tile, it turns out that $p(\xi)$ is the characteristic function of the difference of two independent identically distributed random variables $X - X'$, where X is uniformly distributed over the digits $d^k, k = 0, 1, \dots, q - 1$. This observation will be used to good effect in the above-mentioned section. (For example, if $\varphi(t) = \chi_{(0,1)}(t)$, we may take $X - X'$ to be two independent “fair coin” Bernoulli variables; the characteristic function of the difference $X - X'$ is given by $p(\xi) = \left(\frac{1}{2}\right) |(1 + \exp(-2\pi i \xi))/\sqrt{2}|^2 = |(1 + \exp(-2\pi i \xi)/2)|^2$. The digits are $d^0 = 0, d^1 = 1, A = B = 2$, and $c_0 = c_1 = 1/\sqrt{2}$. The fact that $|\chi_{(0,1)}(\xi)|^2 = \prod_{k=1}^{\infty} p(\xi/2^k)$ is classical. In general, the squared modulus of a low-pass filter is not the characteristic function of a random variable.

On the other hand, the QMF condition does suggest a different kind of probability construction. If $\varphi(t)$ is a scaling function, then $p(\xi)$, as defined above, satisfies the QMF condition: to verify this, we write $m \in \mathbb{Z}^n$ as $m = Bm' + d^k$. From conditions (\hat{a}) , (\hat{b}) we get

$$\begin{aligned} \sum_{m \in \mathbb{Z}^n} \left| \hat{\varphi}(B^{-1}(\xi + m)) \right|^2 \\ = \sum_{m' \in \mathbb{Z}^n} \left| \hat{\varphi}(B^{-1}(\xi + d^k) + m') \right|^2 = 1 \end{aligned}$$

for each fixed digit d^k .

We take note of this and the fact that $|\hat{\varphi}(0)|^2 = \int |\varphi(t)|^2 dt = 1$, together with (\hat{a}) , to obtain

$$\begin{aligned} 1 &= \sum_{k \in \mathbb{Z}^n} |\hat{\varphi}(\xi + k)|^2 \\ &= \sum_{d_j \in D} p(B^{-1}(\xi + d^k)) = p(0), \end{aligned}$$

as shown in (1). From (\hat{a}) and the fact that $p(\xi)$ is a QMF function, we see that this implies that the function $|\hat{\varphi}(\xi)|^2$ is a (formal) infinite product:

$$(4) \quad |\hat{\varphi}(\xi)|^2 = \prod_{j=1}^{\infty} p(B^{-j}\xi).$$

This infinite product version of (\hat{a}) also suggests two interpretations for the Fourier transform

$|\hat{\varphi}(\xi)|^2$. As we have seen, when $\varphi(t) = \chi_T(t)$, $p(\xi)$ is a characteristic function of $X - X'$, so that $|\hat{\varphi}(\xi)|^2$ is the Fourier transform of the infinite convolution of probabilities corresponding to a sum of independent random variables with values in the time domain. When $\varphi(t)$ is not necessarily a tile, there is another perspective: we regard $|\hat{\varphi}(\xi)|^2$ as a probability in the frequency domain. In fact, $\prod_{j=1}^m p(B^{-j}\xi)$, $m \geq 1$, are probabilities on $R^n \times (m \text{ factors}) \times R^n$, since $p(\xi)$ satisfies the QMF condition. The foregoing sequence of partial products defines probabilities associated with a sequence of zero-sum games. The sequence of payoffs for these games forms a Markov process of the type mentioned in the Introduction. Let us see how this works in a special case.

The Pascal-Fermat Correspondence

Many historians use the date 1654 to mark the beginning of computational probability theory. It was during that year that Pascal and Fermat exchanged their thoughts on the “problem of points” in response to a question posed by the essayist and gambler Antoine Gombaud, aka Chevalier de Méré. For the purpose of this discussion, let me pose a particular form of his problem as follows: Two players, Alice and Bob, of different skills, are playing a sequence of games. Each game represents an independent trial, resulting in a score of 1 for the winner, 0 for the loser. Let us assume that the probability that Alice wins an individual game is $\alpha, 0 \leq \alpha \leq 1$. In that case, the probability that Bob wins a single game is $1 - \alpha$. They play a sequence of games until one of the players has won a fixed number of games. That number N is fixed at the outset. Can we write a formula for the probability $P(\alpha, N)$ that Alice is the overall winner? Without any hard computation, a number of things are clear:

- The number of individual trials to determine a winner cannot exceed $2N - 1$.
- Since the trials are independent, the desired probability may be computed using the binomial expansion: simply sum over all sequences of length $2N - 1$ containing at least N wins for Alice. These sequences have a total probability $P(\alpha, N)$, a polynomial in α that is strictly positive if $0 < \alpha \leq 1$.
- The degree of this polynomial is exactly $2N - 1$, since there is a positive chance that the winner will not be determined until trial $2N - 1$. The only root of this polynomial $P(\alpha, N)$ is at $\alpha = 0$, and this root has multiplicity N . This is true because we can compute $P(\alpha, N)$ in another way: restrict the sum to sequences that terminate with the N th win for Alice on the k th game, where $N \leq k \leq 2N - 1$.

For example, if $N = 2$, the first computation gives $P(\alpha, 2) = \alpha^3 + 3(1 - \alpha)\alpha^2 = \alpha^2(\alpha + (1 - \alpha)) + 2(1 - \alpha)\alpha^2$. The second computation leads to the expression $\alpha^2 + 2(1 - \alpha)\alpha^2 = \alpha^2(3 - 2\alpha)$.

- Finally, observe that $P(\alpha, N) + P(1 - \alpha, N) = 1$; that is, we have a zero-sum game.

Daubechies' QMF Function

Now let us move ahead by about three hundred and fifty years. The notion of a QMF function has emerged. As we will explain in the next paragraph, certain (but not all) QMF functions generate wavelet bases or, more precisely, scaling functions. The problem that Ingrid Daubechies [6] confronted in the 1980s was the search for a class of scaling functions $\varphi(t)$ that were (i) smooth to a specified degree and (ii) compactly supported. Since these functions are to be used for data analysis, she starts with the search for a suitable finite set of real-valued refinement coefficients $c_k, k \in \mathbb{Z}$, with the hope that these coefficients will lead to a solution to the equation (2a). (Notice that in order for $\varphi(t)$ to be compactly supported, it is necessary that $c_k \neq 0$ for only finitely many $k \in \mathbb{Z}$.) Finding a suitable sequence of coefficients is the same as finding the polynomial low-pass filter $\hat{p}(\xi) = (\sqrt{2})^{-1}\hat{c}(\xi)$. Her strategy is to proceed in two steps. First, solve what turns out to be a simpler problem: find a QMF polynomial $p(\xi)$ and form the infinite product given in (3). It is not too difficult to see that the inverse Fourier transform of $|\hat{\varphi}(\xi)|^2$ will inherit some degree of regularity from the factors to satisfy requirement (i). For example, if $p(\xi)$ has the factor $\cos^{2N}(\pi\xi)$, which is a factor of the Fourier transform of an N -order basic spline, then this spline will appear in the infinite convolution $\varphi(t)$. In addition, $p(\xi)$ should vanish only at $\xi = 1/2$, since additional zeros, depending on their nature, can be big trouble: that is, the sum in (\hat{b}) may be less than one for certain ξ due to the presence of unwanted zeros in the sequence $p(2^{-j}\xi), j \geq 1$. So, to be on the safe side, she looks for polynomials whose only zeros (at $\xi = 1/2$) are those required by the QMF condition.

Daubechies [6] produces a class of QMF trigonometric polynomials $p_N(\xi)$ by some clever combinatorics. It turns out that the class of polynomials she finds coincides, under a change of variables, with the algebraic polynomials found by Pascal and Fermat. To see why this is the case, let us suppose that we have a solution $p(\xi)$ that satisfies the above requirements. We note that any QMF polynomial is a nonnegative cosine polynomial. It is an elementary fact that such trigonometric polynomials may be written as algebraic polynomials in the variable $\alpha(\xi) = \cos^2(\pi\xi)$. For the moment, call this algebraic polynomial $P(\alpha) : P(\cos^2(\pi\xi)) = p(\xi)$. The QMF condition ($P(\alpha) + P(1 - \alpha) = 1$) allows

us to interpret $P(\alpha)$ as the probability of winning a two-player zero-sum game. The polynomial $P(\alpha)$ is required to have a single root of multiplicity N at $\alpha = 0$ in order to satisfy the smoothness requirement and to avoid the possible "big trouble" mentioned above. With the benefit of hindsight, we see that the requirements imposed on $p(\xi)$ for Daubechies' problem are met by the class of polynomials $P(\alpha, N)$ discovered by Pascal and Fermat in 1654. The case $N = 1$, where $p(\xi) = \cos^2(\pi\xi)$, gives the QMF polynomial for the Haar scaling function, the indicator function of the unit interval. Notice that $|\hat{\varphi}(\xi)|^2$ is the Fourier transform of the linear B-spline (the tent on the interval $|t| \leq 1$), which is the convolution of two indicator functions of the interval $|t| \leq 1/2$. So, just to obtain a continuous $\varphi(t)$, one must use a Pascal-Fermat polynomial with $N > 1$. Smoothness increases with N , at the expense of increasing the length of the support of φ .

In her monograph ([7], page 211) Daubechies reports that it was Yves Meyer who suggested the change of variables from ξ to α ; Meyer's endgame to derive the polynomials $P(\alpha, N)$ uses Bézout's lemma. Here, it is the Bézout lemma that is taught to first-year university students in France and to fourth-year undergraduate students in the United States.

The Pascal-Fermat connection was first described in [12].

Producing the Scaling Function

The second part of the problem is to prove that these QMF polynomials lead to the existence of compactly supported scaling functions. We may also ask whether the solution scaling functions are unique. A famous factorization fact, known as the Fejér-Riesz spectral factorization lemma, asserts that any nonnegative cosine polynomial may be expressed as the squared modulus of a cosine polynomial with *real* coefficients, restricted to the unit circle. (See Daubechies [7], page 172.) In particular, the Fejér-Riesz lemma implies that any QMF polynomial may be expressed as $p(\xi) = |(\sqrt{2})^{-1} \sum_{k=0}^N c_k \exp(-2\pi ik\xi)|^2$ for some sequence $c_k, k = 0, 1, \dots, N$. That is, we can find a (nonunique) polynomial square root $\tilde{p}(\xi)$ for the function $p(\xi)$.

All of this means that we can define the convergent infinite product $\hat{\varphi}(\xi) := \prod_{j=1}^{\infty} \tilde{p}(2^{-j}\xi)$. By Fourier inversion, we obtain a compactly supported function $\varphi(t)$. For this $\varphi(t)$, the coefficients $c_j, j = 0, 1, \dots, N$, are the ones that are specified by property (a). Next question: do the functions $\{\varphi(t - k), k \in \mathbb{Z}\}$ constitute an orthonormal family, as specified by property (b) or (\hat{b}) ? This is the most elusive of the three properties, and we postpone the full discussion it deserves. A hint: the crucial fact was emphasized above by the remark that $p(\xi) = 0$ if and only if $\xi = 1/2$. If

this is the case, then the factors $p(\xi/2^j)$, $j \geq 1$, are strictly positive, surely a necessary condition for (\hat{b}) . Property (\hat{c}) is easy to verify: it follows from the fact that the infinite product $|\hat{\varphi}(\xi)|^2$ is continuous and $p(0) = |\hat{\varphi}(0)|^2 = 1$.

Regarding the uniqueness question, a given QMF function may generate many different scaling functions. Notice that even in the simplest case, $p(\xi) = \cos^2(\pi\xi)$ and $|\hat{\varphi}(\xi)|^2 = \sin^2(\pi\xi)/(\pi\xi)^2$, the indicator function of an arbitrary unit interval, or Hilbert transform of such an indicator function, is a candidate for $\varphi(t)$. Finally, if $p(\xi)$ is any QMF function (not necessarily a polynomial) such that properties (\hat{a}) , (\hat{b}) , and (\hat{c}) hold, it is known that a square root exists (see The Wutam Consortium [24]) and so also a corresponding $\varphi(t)$ that satisfies (a) , (b) , and (c) . For this reason, the object of primary concern is the function $p(\xi)$ and the infinite product $|\hat{\varphi}(\xi)|^2$.

An Infinite Sequence of Games

What has been said about QMF polynomials applies to any periodic QMF $p(\xi)$, expressed as a Fourier series. We can re-express $p(\xi) = P(\cos^2(\pi\xi))$. Let (ξ) be the fractional part of $\xi \in \mathbb{R}$. We can define two quantities, $\alpha = \alpha(\xi) = \cos^2(\pi(\xi)/2)$ and $1 - \alpha(\xi) = \cos^2(\pi((\xi)/2 + 1/2))$ so that $P(\alpha(\xi)) + P(1 - \alpha(\xi)) = 1$. The quantity $P(\alpha(\xi))$ represents the probability that Alice wins a two-person game specified by a parameter (ξ) . Our players will play an infinite sequence of such games, and after each play, the parameter (ξ) will be changed by a certain amount, dictated by the QMF condition. In more detail, let us suppose that the two players enter the casino at a time that we arbitrarily call $t = 0$. Since the casino has been running since the beginning of time, the amusement opportunity at $t = 0$ is summarized by the pair $((\xi), p((\xi)/2))$. Here the binary digits for (ξ) give the history of wins for Alice ($\omega_k = 0$) and Bob ($\omega_k = 1$). This pair contains enough information to compute the initial probabilities α_0 and $1 - \alpha_0$; these are determined by the fractional part (ξ) and the binary coefficient ω_0 determining the parity of the integer part $[\xi]$: If $[\xi]$ is even, Alice wins the initial game ($\omega_0 = 0$) with probability $p((\xi)/2)$; if $(\omega_0 = 1)$, Bob wins with probability $p((\xi)/2 + 1/2)$. In any case, the opportunity at $t = 0$ is summarized by (ξ) , and $p((\xi)/2)$; the outcome, determined by the parity of $[\xi]$, is not known at $t = 0$. In this way, we can compute the probability that Alice wins any finite number of games. The sequence of amusement opportunities is random, depending on Alice's current fortune, and qualifies as a Markov process whose states are numbers in the unit interval. In this context, it is important to note that the sequence of wins and losses, represented by the random sequence of binary digits ω_j , $j \in \mathbb{Z}$, is *not* a two-state Markov process. If it were so,

then ω_{-1} , the first binary coefficient of (ξ) , would represent "the present". However, it is clear that $\Pr(\omega_0 = 0 | (\xi)) \neq \Pr(\omega_0 = 0 | \omega_{-1})$. We have taken the liberty to rescue the Markov property by an expedient definition of "the present" as "the entire history", specified by (ξ) . After the outcome of the game at $t = 0$ is revealed, the new state of the process is given by $(\xi)/2 + \omega_0/2$.

The interesting question concerns an infinite sequence of plays after $t = 0$: what is the probability that there is an eventual winner for this sequence of games? The eventual winner is the player who wins every game from some point forward. This can only happen if $\lim_{j \rightarrow \infty} \alpha_j = 1$ or $\lim_{j \rightarrow \infty} 1 - \alpha_j = 1$, and in this case we will say that the sequence of games is decided. The alternative is an eternal, indecisive combat. In these terms, we can state the following theorem.

Theorem 1. *A continuous QMF function $p(\xi)$, $\xi \in \mathbb{R}$ for dyadic dilation generates a scaling function if and only if, for almost every (ξ) , $0 \leq (\xi) \leq 1$, the infinite sequence of games is decided with probability one. Moreover, there should be a set of (ξ) of positive Lebesgue measure such that with positive probability, Alice is the winner, and a set of (ξ) of positive Lebesgue measure such that Bob is the winner. If (ξ) is a QMF function (that generates a scaling function) with a finite number of zeros, then the above condition can be more simply stated: for almost every (ξ) with probability one, both players have positive probability of being eventual winners.*

Commentary. If Alice finally prevails, winning every game after k trials, the coefficients in the binary expansion of k are eventually zero. So, the probability that she is the eventual winner is $\sum_{k \geq 0} |\hat{\varphi}((\xi) + k)|^2$. The Bob probability is given by the sum over $k < 0$. Of course, if the game is decided (one of the two players is the eventual winner), then $\sum_{k \in \mathbb{Z}} |\hat{\varphi}(\xi + k)|^2 = 1(\xi)$ a.e. Condition (\hat{c}) is the expression of the "Moreover" part of the theorem. This second assertion cannot be casually justified at this point. Note that this theorem does not give conditions on $p(\xi)$ that tell us whether or not the game is decided. The bare-bones conditions for a decisive winner are conditions (\hat{a}) , (\hat{b}) , and (\hat{c}) , conditions that are not easily translated to conditions on $p(\xi)$. However, if we assume that $p(\xi)$ is a polynomial, for example, then there are restrictions on its zeros that translate to necessary and sufficient conditions for the game to be decided. (See the section "Invariant Sets for QMF Functions").

The above discussion suggests that we should transfer our attention from functions of a real variable ξ to functions defined on binary sequences. This is not just a convenience. There is a more compelling reason to do this, described in this section and the following.

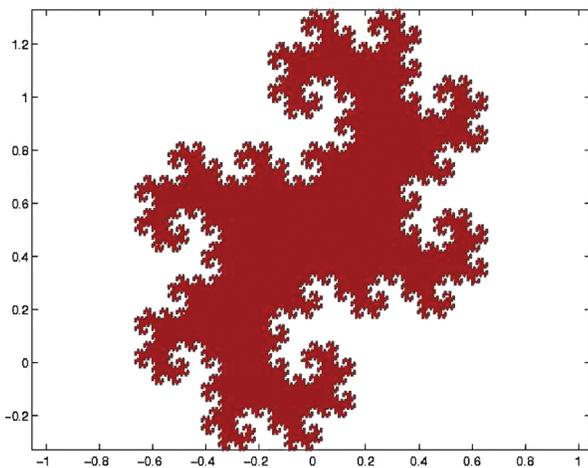


Figure 1: The Twin Dragon

$$\mathbf{T}(B, \mathcal{D}) = \sum_{j=1}^{\infty} B^{-j} d_j, \quad B = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Here $d_j = (0,0)$ or $(1,0)$.

The Same Problem for \mathbb{R}^2

The time dilations to be considered are given by matrices A with integer entries such that $\det(A) = \pm 2$. In addition, we require all eigenvalues λ to satisfy $|\lambda| > 1$. Fortunately, this class of dilations has received much attention in the wavelet literature, and, in particular, they have been completely classified by Lagarias and Wang [17]. These dilations fall into four distinct equivalence classes, where $A \sim A'$ if A is similar to A' by the action of a unimodular matrix U with integer entries. The question is the same as in the one-dimensional case in which $A = 2$: Can we give necessary and sufficient conditions for a QMF function (with respect to A) to generate a scaling function $\varphi(t), t \in \mathbb{R}^2$? The problem was posed by Resnikoff and Wells in their book [19]. Here we sketch a solution, based on what has been described above. Complete details may be found in [13]. It will suffice to tell you how this solution emerges in a particular case, one that has been discussed by several authors: the “twin dragon” dilation. (See Gröchenig and Madych [11], Lagarias and Wang [16], Lawton and Resnikoff [17], Cohen and Daubechies [2].)

The *twin dragon* T is a set in the time domain \mathbb{R}^2 that is a self-affine tile generated by the matrix (matrices) and digit set

$$\pm A = \pm \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \quad \text{with } \mathcal{D} = \{d^0 = (0,0), d^1 = (1,0)\}.$$

These matrices, together with their adjoints $\pm B$, are often referred to as the *quincunx* matrices. The set T is defined in display (2) above. Gröchenig

and Madych [11] and Lawton and Resnikoff [17] were the first to prove that T is a self-affine tile. Consequently, the indicator function of T may be viewed as a scaling function of Haar type. The strategy of their proof consists of transferring the problem to the frequency domain where they construct the appropriate QMF function $p(\xi_1, \xi_2)$ and the corresponding infinite product $|\hat{\varphi}(\xi_1, \xi_2)|^2$. The subtle part of the proof lies in showing that $|\hat{\varphi}(\xi_1, \xi_2)|^2$ satisfies $(\hat{b}), (\hat{c})$ of $(\hat{2})$. Clearly, (\hat{a}) is automatic, so the difficulty arises in verifying (\hat{b}) and (\hat{c}) . We can view T (rather, $\chi_T(t)$) as an \mathbb{R}^2 -valued random variable in the time domain that is a *sum of independent variables* of the form $A^{-j} d_j, j \geq 1, i = 1, 0$, as we pointed out in the section “Some Definitions” in which T was the unit interval in \mathbb{R}^1 . The random variables d_j take values in the digit set $\mathcal{D} = \{d^0 = (0,0), d^1 = (1,0)\}$, with probability $1/2$ each.

The characteristic function of the two-dimensional random variable $d = (d^0, d^1)$ (or Fourier transform of its probability distribution), written in the notation of the section “Some Definitions”, is $(\sqrt{2})^{-1} \hat{c}(\xi) = [1 + \exp(-2\pi i \langle d^1, \xi \rangle)]/2, \xi \in \mathbb{R}^2$, so that $p(\xi_1, \xi_2) = \cos^2(\pi \xi_1)$. In contrast to the one-dimensional case, in which $p(\xi)$ vanishes at the single point $\xi = 1/2$, the two-dimensional $p(\xi_1, \xi_2) = 0$ on the entire line $\xi_1 = 1/2$. This new element makes verifying (\hat{b}) , in particular, more difficult if we choose to copy the one-dimensional arguments. The difficulty can be overcome, but this strategy gives no clue as to how to obtain a version of Theorem 1 for general QMF functions. Here is a short historical sketch of the previous arguments, based on a theorem due to A. Cohen [1]: *In order that a polynomial QMF generate a scaling function, it is necessary and sufficient that there exists a tile T (in the sense defined in the section “Some Definitions”) consisting of a finite number of disjoint closed sets, one of which contains the origin in the interior, such that $p(B^{-j}\xi) \geq \delta > 0, j \geq 1$ for all $\xi \in T$.* We shall refer to this theorem as the *C-tile condition*. In one dimension, $p(\xi) = \cos^2(\pi \xi)$, so we can take $T = \{\xi : |\xi| \leq 1/2\}$ and $\delta = 1/4$ since $\cos^2(\pi \xi/2^{-j}) \geq 1/4, j \geq 1$. In two dimensions, the inverse matrix $B^{-1} = \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix}$ sends the vector $[1/2, -1/2]'$ to $[1/2, 0]'$. Consequently, $p(B^{-1}\xi)$ is not uniformly bounded away from 0 on the unit square. Gröchenig and Madych [11] overcome this obstacle by a clever distortion of the unit square to produce a *C-tile* in frequency domain \mathbb{R}^2 . This stratagem leads them to a proof that the twin dragon matrix A defines a self-affine tile and a corresponding Haar-like scaling function in the time domain \mathbb{R}^2 . Incidentally, this distorted square has achieved a degree of artistic recognition: it also appears in Cohen and Daubechies [2], Resnikoff and Wells [19], and Wojtaszczyk [23].

In the next section, we describe *another solution to the two-dimensional problem by direct reduction to the one-dimensional case*, thereby avoiding the difficulty just described. More to the point, we will see that Theorem 1, properly stated, makes no distinction between QMF functions $p(\xi)$ for $\xi \in \mathbb{R}^1$ or \mathbb{R}^2 and dilations 2 or A . This may be surprising since, in one dimension, the digit set contains only one linearly independent vector and this situation *persists* in two dimensions. In this case, linear independence is not the right idea. The digits are coset representatives, and it is the equality of the *indices* of the quotient groups $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}^2/A(\mathbb{Z}^2)$ that is the basis of the theorem.

Coding \mathbb{R}^1 and \mathbb{R}^2 into $2^{\mathbb{Z}}$: Case 1; \mathbb{Z}^1 vs. \mathbb{Z}^2

In the next two sections, we make a detour into the subject of radix representations for numbers, in both one and two dimensions. While this may seem to be a diversion from the main subject, it is really essential for the constructions that follow.

Here is the description of a mapping between \mathbb{R}^1 and \mathbb{R}^2 that will allow us to transport QMF functions from one space to the other. The mapping will be accomplished by sending each space into the space of binary sequences $2^{\mathbb{Z}} = \{\omega : \omega = (\dots \omega_{-1}, \omega_0, \omega_1, \dots)\}$. Having designated a zero-coordinate, we split the space into $2^{\mathbb{Z}^-} = \{\xi^- = (\dots \omega_{-2}, \omega_{-1})\}$ and $2^{\mathbb{Z}^+} = \{\xi^+ = (\omega_0, \omega_1, \dots)\}$. The integers \mathbb{Z} are to be mapped into $2^{\mathbb{Z}^+}$ by the following procedure. The nonnegative integers \mathbb{Z}^+ are given the usual binary expansion $0 \rightarrow (0, 0, \dots)$, and $0 < k \rightarrow (\omega_0, \omega_1, \dots, \omega_{n-2}, 1, 0, 0, \dots)$ where $2^{n-1} \leq k < 2^n$. The negative integers are given the “two’s complement representation”. The name says it all; take $k < 0$, $2^n < |k| \leq 2^{n+1}$. Let $(\omega_0, \omega_1, \dots, \omega_{n-1})$ be the sequence of binary coefficients for $k' = 2^{n+1} - |k|$, so $\omega_j = 0, j \geq n$. Then $k = -2^{n+1} + k'$, and we represent $k \rightarrow (\omega_0, \omega_1, \dots, \omega_{n-2}, 1, 0, 1, 1, \dots)$. Notice that $-1 \rightarrow (1, 1, \dots)$ and that when $2^n < |k| \leq 2^{n+1}$, the coefficients $\omega_{n-1} = 1, \omega_n = 0, \omega_{n+l} \equiv 1, l \geq 1$.

Now check that if we take -2 as the base, with the same digits, we obtain a finite radix expansion for all $k \in \mathbb{Z}$, that is, one in which $d_j \equiv 0$ for all $j \geq j_0$. To rephrase this observation, there is no need to create the two’s complement representation if the radix is -2 . The proof of this is just an application of the division algorithm. More surprising is the fact that *the same argument can be used to show that the matrix $-B$ also yields a finite radix expansion for every $k \in \mathbb{Z}^2$* . The fundamental fact is that in both cases, the multiplication of the dilations 2 or B by -1 converts each to a radix for “canonical representations” of all integers \mathbb{Z} (or \mathbb{Z}^2).

First, we construct a two’s complement representation for \mathbb{Z}^2 using the adjoint twin dragon matrix B as radix. Given $k \in \mathbb{Z}^2$, apply the division algorithm to obtain $k = Bk_0 + d_0$ where

d_0 is one of the digits $d_0 = (0, 0)$ or $d^1 = (1, 0)$. Can we get a B -radix expansion for such $k \in \mathbb{Z}^2$? As was the case for dilation by 2, the answer is “yes” if we admit two kinds of expansions, those that are “finite”, $d_j \equiv d^0, j \geq n$, and those that are “infinite”, that is, $d_{n-1} = d^1, d_n = d^0, d_{n+1} \equiv d_{n+2} \equiv \dots \equiv d^1$ for some $n \geq 1$. A Gaussian integer $k \in \mathbb{Z}^2$ with the latter one-sided infinite expansion is then uniquely associated with the corresponding negative integer $\tilde{k} \in \mathbb{Z}$, $2^n < |\tilde{k}| \leq 2^{n+1}, \tilde{k} \rightarrow (\omega_0, \omega_1, \dots, \omega_{n-2}, 1, 0, 1, 1, \dots)$; those $k \in \mathbb{Z}^2$ with one-sided infinite expansions with zero entries after some finite index will correspond to nonnegative integers. One certainly should know which $k \in \mathbb{Z}^2$ corresponds to -1 . It is the vector $k = (0, -1)$, the unique solution to $k = Bk + d^1$ and analogous to the equation $\tilde{k} = 2\tilde{k} + 1$, where $\tilde{k} = -1$. Let us assume that the norm $\|k\|$ is large. We apply the Euclidean division algorithm repeatedly to obtain a sequence of vectors k, k_0, k_1, \dots , such that the norms $\|k_j\|, j \geq 0$, are strictly decreasing. It is easy to verify that the norms will form a decreasing sequence provided $\|k_j\| > \sqrt{2} + 1$ since the norm $|B^{-1}| = \sqrt{2}$, and $|B^{-1}d^1| = 1$. Repeated divisions give a sequence of vectors that eventually converge to one of two fixed points. To verify this, proceed as follows:

There are twenty-one lattice points $k \in \mathbb{Z}^2$ that lie in the circle of radius $\sqrt{2} + 1$; twelve of them are attracted to $k = (0, -1)$ and the rest go to $(0, 0)$. By the way, a verification of this twenty-one count can be done without doing any division. To do it for B , draw the four-by-four square and exclude the four corners to obtain the twenty-one lattice points. In the following manner, construct two disjoint trees, one rooted at the origin, the other at $(0, -1)$, whose vertices are the twenty-one points. Beginning at the base point $(0, -1)$, perform a $-\pi/4$ rotation and dilation by $\sqrt{2} : (0, -1) \rightarrow (-1, -1) = B(0, -1)'$. From $(-1, -1)$ you have the option of performing another rotation or moving horizontally to the right one unit: the latter option leads us back to $(0, -1)$, so our advice is to rotate. Continue this procedure, rotating and moving to the right by one unit. Lateral moves by a unit must be preceded and followed by a rotation. Stop when you have reached twelve of the points inside the circle. The remaining points can be reached by starting at the origin and moving right to d^1 , followed by a rotation, etc. The radix expansions are obtained by reversing the paths.

The same counting argument can be used to show that the matrix $-B$, like the one-dimensional case -2 , yields a finite radix expansion for every $k \in \mathbb{Z}^2$. In this case, one only tracks paths from $(0, 0)$ to the boundary of the 4×4 square. This result is also a consequence of the theorem of Katai and Szabó [15], which covers a much more general class of matrices and does not involve counting.

Also take note of the fact that everything said about B and $-B$ is also true for their adjoints, A and $-A$.

Coding \mathbb{R}^1 and \mathbb{R}^2 into $2^{\mathbb{Z}}$: Case 2; Unit Interval vs. Twin Dragon

The unit interval is mapped to $2^{\mathbb{Z}^-}$ using the usual binary expansion $(\xi) \rightarrow \xi^-$. The space $2^{\mathbb{Z}^-}$ is thought of as the infinite product of two-point groups under dyadic addition, so it comes equipped with a Haar-measure $d\xi^-$. The probability space $(2^{\mathbb{Z}^-}, \mathcal{F}, d\xi^-)$, where \mathcal{F} is the σ -algebra generated by the cylinder sets, is isomorphic to the Borel unit interval $(U, \mathcal{B}, d\xi)$. Therefore, the above map is to be considered as an invertible measure preserving transformation.

Now we are going to prove that the set $T = T(B, \mathcal{D})$ is a self-affine tile in the frequency domain. Recall that $T(B, \mathcal{D})$ is a tile if the sum of the indicator functions $\sum_{k \in \mathbb{Z}^n} \chi_T(\xi + k) = 1$ almost everywhere. This implies that T has Lebesgue measure one: if χ_U is the indicator of the unit cube in \mathbb{R}^n , then

$$\begin{aligned} 1 &= \int \chi_U(\xi) d\xi = \int \chi_U(\xi) \left(\sum \chi_T(\xi + k) \right) d\xi \\ &= \int \left(\sum \chi_U(\xi - k) \right) \chi_T(\xi) d\xi = \int \chi_T(\xi) d\xi. \end{aligned}$$

The following facts are well known to the enthusiasts of this subject. (See all the authors cited above.) We will summarize these facts for the sake of completeness. (i) Any self-affine set T is compact and has positive Lebesgue measure. The compactness is proved by standard arguments. We then check that $\bigcup_{k \in \mathbb{Z}^2} T + k$ is invariant under all dilations B^n , $n \in \mathbb{Z}$, and, therefore, for any fixed ξ we can approximate $B^n \xi$, $n > 0$ by some $\xi' \in \bigcup_{k \in \mathbb{Z}^2} T + k$: we have $|\xi' - B^n \xi| \leq c$ and $|B^{-n} \xi' - \xi| \leq O(2^{-n/2})$. Thus, the above union is dense in \mathbb{R}^2 ; the Baire category theorem implies that T has positive measure. (ii) The measure $m(T \cap T + d) = 0$ since $m(T \cup T + d^1) = |\det(B)|m(T) = 2m(T)$. The same argument implies that the sets $T, T + d^1$, and $T + d^j + Ad^1$, $j = 0, 1$, are essentially disjoint. If every $k \in \mathbb{Z}^2$ had a finite radix expansion with respect to B , this argument would prove that all the sets $k + T$, $k \in \mathbb{Z}^2$, are essentially disjoint, which would imply that $T(B, \mathcal{D})$ is a self-affine tile. However, we saw that only about half of the Gaussian integers have finite B -radix representations. But suppose we had started with the dilation $-B$: in this case, we saw that every $k \in \mathbb{Z}^2$ has a finite radix representation. By the above remark, $-T = T(-B, \mathcal{D})$ is a self-affine tile. Now we must make the transition from $-T$ to T . Here again we use probability, and the first interpretation of $\varphi(t)$ from the section "Some Definitions" to the effect that the indicator function of any tile may be considered as an infinite sum of independent random variables. In particular, the indicator function of $-T$ is the

sum of the sequence of independent random variables $(-B)^{-j} d_j$, $j \geq 1$. Now let us create two sequences $(B^{-j} d_j)'$ and $((-B)^{-j} d_j)'$, $j \geq 1$, that are independent of each other and independent of the original (unprimed) sequences. Now consider the two symmetrized sequences $(B^{-j} d_j) - (B^{-j} d_j)'$ and $((-B)^{-j} d_j) - ((-B)^{-j} d_j)'$. These two sequences are identically distributed. Consequently, their characteristic functions are equal. The characteristic functions of the symmetrized variables are exactly the functions $p(B^{-j} \xi)$ and $p((-B)^{-j} \xi)$, and these two functions are identical for all $j \geq 0$. Consequently, the infinite products $|\hat{\varphi}_+(\xi)|^2 \equiv |\hat{\varphi}_-(\xi)|^2$ (see display (4) in the section "What Is a Scaling Function?") where $|\hat{\varphi}_\pm(\xi)|^2$ correspond to B and $-B$, respectively. Exactly the same arguments apply to the adjoint matrices A and $-A$. The fact that $-B$ is a self-affine tile means that (\hat{a}) and (\hat{b}) are satisfied for the matrix $-A$, and, therefore, also for A . But (\hat{a}) , (\hat{b}) for A imply that B is a self-affine tile. Consequently, $T(B, \mathcal{D})$ has Lebesgue measure one.

The important conclusion is that both the Lebesgue unit interval $(T(2, \mathcal{D}), d\xi)$ and the twin dragon pair $(T(B, \mathcal{D}), d\xi)$ are measure-isomorphic to $(2^{\mathbb{Z}^-}, d\xi^-)$, and we have the correspondence $[T(2, \mathcal{D}), \mathbb{Z}] \xrightarrow{\sim} [T(B, \mathcal{D}), \mathbb{Z}^2]$ through the intermediate space $2^{\mathbb{Z}} = (2^{\mathbb{Z}^-}; 2^{\mathbb{Z}^+})$. More important for us is the fact that any QMF $p(\xi)$ with dilation 2 or B generates a QMF on $2^{\mathbb{Z}}$ where the dilation is replaced by the left-shift θ , a two-one transformation defined as $\theta(\dots, \omega_{-2}, \omega_{-1}) = (\dots, \omega_{-3}, \omega_{-2})$. The right-shift θ^{-1} is one-to-one on the two-sided sequence space $2^{\mathbb{Z}}$. When it is followed by the projection onto $2^{\mathbb{Z}^-}$, the composition, also denoted by Θ^{-1} , has two branches, just like 2^{-1} and B^{-1} . Now it is easy to interpret Theorem 1 in these terms. The QMF function becomes a Markov transition operator with Θ^{-1} acting as the shift on the path space $2^{\mathbb{Z}}$. That is, for each initial state $\xi^- \in 2^{\mathbb{Z}^-}$, the random traveler can proceed to $(\xi^-, 0)$ with probability $p(\xi^-, 0)$ or $(\xi^-, 1)$ with probability $p(\xi^-, 1)$. (As we pointed out in the section "Producing the Scaling Function", the function $p(\xi)$ defines a transition operator for a family of processes with the two-point state space $0, 1$, but the transitions are definitely not Markovian on this state space.)

The periodicity of $p(\xi)$ translates to a condition on $p(\xi^-; \xi^+)$, expressed in these terms: $p(\xi^-; \xi^+) = p(\xi^-)$ and $p(\Theta^{-1}(\xi^-; \xi^+) = p(\xi^-, \omega_0)$. The infinite product $|\hat{\varphi}(\xi^-; \bullet)|^2$ may now be considered as a probability on the path space $2^{\mathbb{Z}}$ for each base point ξ^- . The base points have a natural measure $d\xi^-$. If we have a QMF function $p(\xi^-)$ defined on $(2^{\mathbb{Z}^-}, d\xi^-)$, the shift serves as a universal model for the various dilations with $\det B = \pm 2$. A QMF function on $2^{\mathbb{Z}^-}$ generates a scaling function on both \mathbb{R}^1 and \mathbb{R}^2 , depending on which dilation we choose, provided we can interpret Theorem 1 on the binary

sequence space $2^{\mathbb{Z}}$. Moreover, when $2^{\mathbb{Z}}$ is given the product topology, certain QMF functions that are discontinuous on \mathbb{R} or \mathbb{R}^2 become continuous. An important example is the Shannon filter $p(\xi)$, the indicator function of the set $[0, 1/4] \cup (3/4, 1]$. (See Hernández and Weiss, [14], page 62.) The restatement of Theorem 1 on the binary sequence space goes as follows:

Theorem 2. *A QMF function $p(\xi)$ generates a scaling function if and only if for $(d\xi^-)$ almost every ξ^- , the paths satisfy $\lim_{n \rightarrow \infty} [\Theta^{-n}(\xi^-, \xi^+)]^- = (\dots, 0, 0)$ or $(\dots, 1, 1)$ almost everywhere $|\hat{\varphi}(\xi^-; \bullet)|^2$. Moreover, there should exist sets of initial points ξ^- of positive $(d\xi^-)$ measure such that the $(|\hat{\varphi}(\xi^-; \bullet)|^2)$ measure is positive on sequences $\xi^+ = (\omega_0, \omega_1, \dots)$ such that $\lim_{n \rightarrow \infty} \omega_n = 0$ and such that $\lim_{n \rightarrow \infty} \omega_n = 1$. If $p(\xi^-)$ has a finite number of zeros, then $p(\xi^-)$ generates a scaling function if and only if from almost every $(d\xi^-)$ initial point (ξ^-) , there are two sets of paths, each of positive $(|\hat{\varphi}(\xi^-; \bullet)|^2)$ measure, corresponding to the two limiting sequences.*

Invariant Sets for QMF Functions

Can the probability viewpoint tell us how to recognize which QMF functions generate scaling functions? This question already arose in the discussion of Daubechies' QMF polynomials. As we pointed out above, this question was answered, in part, for certain QMF functions by A. Cohen [1]: if there exists a C -tile such that $p(\xi/2^j) \geq \delta > 0, j \geq 1$ for all $\xi \in C$, then $p(\xi)$ generates a scaling function. So, any continuous $p(\xi)$ that vanishes only at $\xi = 1/2$ must generate a scaling function since the interval $\{\xi : |\xi| \leq 1/2\}$ will serve as a C -tile. This covers Daubechies' polynomials as well as the one-dimensional Haar case. (Thus the "big trouble" alluded to earlier is avoided.) The converse is also true: if a continuous $p(\xi)$ is also smooth and generates a scaling function, then $p(\xi)$ can be used to create a C -tile. However, there are continuous QMF functions that are smooth except at a few points, for which no C -tile exists, yet that generate scaling functions. In fact, it was the investigation of some contrary claims in the literature that led us to construct a variety of bizarre counterexamples of "good" QMF functions, ones that generate scaling functions, for which no C -tile exists. Here is where the intuition from probability carries the day.

Let us ask the following question, phrased in terms of the process on state space $2^{\mathbb{Z}^-}$ generated by $p(\xi)$ and the right shift Θ^{-1} . From the standpoint of a random traveler, starting from ξ^- , directed by $p(\Theta^{-1}(\xi^-, \xi^+))$, what can happen? If the conditions of Theorem 1 do not hold, what goes wrong? Well, the paths from the initial point ξ^- must go somewhere, if they don't go to the 0-sequence $\xi^-(0)$ or the 1-sequence $\xi^-(1)$. The set of limit

points for paths from a fixed ξ^- is a closed θ -invariant set: so there is the possibility that some get kidnapped into some more exotic closed invariant subset. This is especially true if the initial point is itself a member of such a finite closed invariant subset of $2^{\mathbb{Z}^-}$ and $p(\xi) \equiv 1$ on this set. In that case, the process will remain in the closed set with probability one. On the other hand, for every point at which $p(\xi) = 1$, there is another point at which $p(\xi) = 0$. So, for example, Daubechies' QMF polynomials that only vanish at $\xi = 1/2$ provide no opportunity for the random traveler to wind up in some foreign closed invariant set other than $\xi^-(0)$ or $\xi^-(1)$.

We know that, in general, a closed invariant subset, at a positive distance from the points $\xi^-(0)$ and $\xi^-(1)$, must have product measure $(d\xi^-)$ zero since the shift is ergodic with respect to this measure. Such sets are, therefore, either finite or perfect and nowhere dense.

Here is the simplest example of a bad invariant set that will be visited by paths created by the QMF function $p(\xi) = \cos^2(3\pi\xi)$. The invariant set consists of two points, $1/3$ and $2/3$, where $p(1/3) = p(2/3) = 1$. The binary sequence representations are $\xi^-(1/3) = (\dots 1010)$ and $\xi^-(2/3) = (\dots 0101)$. The simplest pseudo-scaling function obtained from this $p(\xi)$ is the "stretched Haar scaling function", that is, the indicator function of the interval $(0 \leq t \leq 3)$. It turns out that for almost every ξ^- , some of the paths from ξ^- go to $\xi^-(0)$, some to $\xi^-(1)$, and with positive probability, some paths get kidnapped: they converge to $\{\xi^-(1/3), \xi^-(2/3)\}$. In fact, this example is typical in the sense that, if $p(\xi^-) = 1$ on a finite set $\{\theta^j(\xi^-); 0 \leq j \leq n\}$ (that is, ξ^- has a periodic binary expansion), then from every initial point ξ^- in the complement of this finite, shift-invariant set, there will be a set of paths of positive probability that converge to the set if $p(\xi)$ is smooth at every point. In general, a necessary and sufficient condition for a polynomial QMF function to generate a scaling function is the requirement that it does not take the value one on a finite shift-invariant set. This is one version of the theorem due to Albert Cohen [1]; this necessary and sufficient condition is equivalent to the C -tile condition mentioned above.

It came as some surprise to find that there was a partial failure of Cohen's theorem when the function $p(\xi)$ was not assumed to be a trigonometric polynomial. In fact, if we start with $p(\xi) = \cos^2(3\pi\xi)$ but modify the values of this function on neighborhoods of $1/3$ and $2/3$ so that the modified function approaches the value 1 along a sharp cusp, the modified function will meet the requirements of Theorem 1 to generate a scaling function. That is, the new function will still take the value 1 on the two-point invariant set, but the $\varphi(t)$ that is generated will be orthogonal to its integer

translates. The intuition is as follows: the paths from almost every ξ^- are attracted to the invariant set, as before, but they never accelerate fast enough to converge, that is, to be absorbed. In fact, the typical path will approach the invariant set, but with probability one, fall away from the set finitely often, then succumb to the ultimate attraction of either $\xi^-(0)$ or $\xi^-(1)$. These invariant sets may be called attractors for the paths from initial points: if the initial point is in the set, all paths from this point remain in the set. For initial points ξ^- outside the set, these attractors come in two varieties: accessible and inaccessible. The former are “fatal attractors”, toward which a portion of the paths from ξ^- converge. The latter attract paths but with probability one, starting from ξ^- in the complement, the path leaves any neighborhood of such an invariant set. If we assume, as we do, that $\xi^-(0)$ and $\xi^-(1)$ are the only fatal attractors, then all paths eventually converge to these points. How bad can nonfatal (inaccessible) attractors be? Given any shift-invariant set that lies at a strictly positive distance from $\xi^-(0)$ and $\xi^-(1)$, can we find a continuous QMF function for which this set is an inaccessible attractor? We don’t know the answer to this general question; however, it is possible to construct uncountable inaccessible invariant sets for a class of continuous QMF functions that generate scaling functions. The details of this construction are too involved to relate here, but they can be found in [12], together with a more rigorous account of the phenomena described in this section.

Some More History

The following remarks about the history of the ideas sketched in this article are intended as a guide to further reading.

The Problem of Points

Many authors date the beginning of computational probability to the correspondence between Pascal and Fermat concerning the problem of points. The general version goes as follows: Alice and Bob are playing an independent sequence of zero-sum games, as stated above. The winner of each game is awarded one point. The eventual winner is the player who first wins N points. Now suppose that the game is interrupted before an eventual winner is declared, and at that point, Alice needs n points to win, whereas Bob needs m points to win. The winner would have received a monetary prize of a fixed amount had the game finished. How should this prize be divided, given the incomplete results? The solution, arrived at by both Pascal and Fermat, is the computation of the probability that Alice obtains her points before Bob gets his. The

particular case discussed in this article assumes that $n = m = N$.

A reference for the mathematics of the problem of points in the form used here is Ross ([20], page 95), and for a more detailed account of the history of the Pascal-Fermat correspondence, see Devlin [8]. The concept of “an independent sequence of Bernoulli trials, with probabilities $\alpha, 1 - \alpha$ ” was certainly not in the arsenal of Pascal and Fermat, but they obtained the right answer when $\alpha = 1/2$. More important, the idea of using mathematics to make a prediction of the likelihood of a complicated future event was born in their dialogues.

Quadrature Mirror Filters

The concept of a QMF function comes from the electrical engineering literature. Given a function $f(t)$ whose Fourier transform $\hat{f}(\xi)$ is supported in a finite interval $\hat{V}_n = \{\xi: |\xi| \leq 2^n\}$, the problem is to find “filters”, $\hat{c}_0(\xi), \hat{c}_1(\xi)$, so that $c_0(\xi)\hat{f}(\xi)$ has low-frequency support $|\xi| \leq 2^{n-1}$ and $\hat{c}_1(\xi)\hat{f}(\xi)$ has high-frequency support in $2^{n-1} < |\xi| \leq 2^n$. It is enough to consider the case $n = 0$ and to produce two one-periodic functions $\hat{c}_0(\xi), \hat{c}_1(\xi)$. The most obvious example is the Shannon filters: $\hat{c}_0(\xi) = \chi(\xi; |\xi| \leq 1/4)$, and $\hat{c}_1(\xi) = 1 - \hat{c}_0(\xi)$. These functions are mirror symmetric around the quadrature phase $1/4$: $\hat{c}_0(1/4 + \xi) = \hat{c}_1(1/4 - \xi)$, so the pair $\hat{c}_0(\xi), \hat{c}_1(\xi)$ are called quadrature mirror filters. However, the coefficients in the Fourier series representing $\hat{c}_0(\xi)$ decay too slowly to allow it to be of practical use in data analysis. But the decomposition of $\hat{f}(\xi)$ into dyadic frequency bands is just one example of multiresolution analysis. Other examples give rise to pairs of filters $c_0(\xi), c_1(\xi)$ that decompose $\hat{f}(\xi) \in \hat{V}_1$ into orthogonal pieces, $\hat{f}_0(\xi) \in \hat{V}_0$ and $\hat{f}_1(\xi) \in \hat{V}_1$. To describe this decomposition, it is enough to consider the functions that generate $V_1 : V_1 = V_0 \oplus W_0$, where W_0 is the closure of all linear combinations of a single function $\psi : \psi \perp \varphi$. Since both functions belong to V_1 , we have $\hat{\varphi}(\xi) = \hat{c}_0(\xi/2)\hat{\varphi}(\xi/2)$ and $\hat{\psi}(\xi) = \hat{c}_1(\xi/2)\hat{\varphi}(\xi/2)$. By properties (\hat{a}) and (\hat{b}) (see $\hat{2}$ above) and the requirement that $\psi \perp \varphi$, we can show that $\hat{c}_1(\xi) = \exp(-2\pi i \xi) \overline{\hat{c}_0(\xi + 1/2)}$. These two filters are not, strictly speaking, quadrature mirror filters; however, $|\hat{c}_0(1/4 + \xi)|^2 = |\hat{c}_1(1/4 - \xi)|^2$. This follows from an assumption that $\hat{c}_0(\xi)$ has real coefficients, which, in turn, implies that $|\hat{c}_0(1/2 + \xi)|^2 = |\hat{c}_0(1/2 - \xi)|^2$ for $0 \leq \xi < 1/2$. Smith and Barnwell [22] introduced filters $\hat{c}_0(\xi), \hat{c}_1(\xi)$ satisfying these conditions for subband frequency filtering and called them “conjugate quadrature filters”. Daubechies ([7], page 163) reports that the CQF label didn’t stick, and these filters are called QMF in most of the literature. History seems to have compensated Smith and Barnwell in some measure:

the QMF condition (1) is often referred to as the Smith-Barnwell equation.

The Markov Process with Transition Function $p(\xi)$

As we mentioned in the Introduction, the type of historical Markov process described in this paper arises in connection with the Ising model for \mathbb{Z} -lattice systems. In this connection, the function $p(\xi)$ is viewed as the kernel of a “transfer operator”. The basic reference for this part of statistical mechanics is Ruelle ([21], Chapter 5).

Other Sources

There are a few references that have not appeared in the text that contributed to the viewpoint expressed in this article: Conze and Raugi [4] view QMF functions as Markov operators on the unit interval and interpret Cohen’s theorem in a probabilistic context. The article by Cohen and Conze [3] is another example of the application of ergodic theory to a problem in time-scale analysis. The author was also influenced by the 2005 Ph.D. thesis of E. Curry [5]. In [10], the discussion of radix representations was guided by the fundamental paper by Kataí and Szabó [15]; in particular, their theorem gives necessary and sufficient conditions for radix representations from a class of matrices that include $-B$. Each one of these references deserves study.

Acknowledgments

First of all, I want to acknowledge the contributions of my collaborators, Pavel Hitczenko and Vladimir Dobrić, to [9] and Adam Jonsson to [13]. Second, I would like to extend my warmest thanks to the editorial staff of the *Notices* and the referees of this article for their attention to detail that contributed immensely to the final version of this article. Especially, I want to recognize Mark Pinsky and Hvorje Sikić for their sharp eyes for typos and the many improvements they suggested.

References

- [1] A. COHEN, Ondelettes, Analyses multirésolution et filtres miroir en quadrature, *Inst. Henri Poincaré, Anal. Nonlinéaire* **7** (1990), 439–459.
- [2] A. COHEN and I. DAUBECHIES, Nonseparable bidimensional wavelet bases, *Revista Mat. Iberoamericana* **7** (1993), 51–137.
- [3] J. P. CONZE and A. COHEN, Régularité des bases d’ondelettes et mesures ergodiques, *Rev. Mat. Iberoamericana* **8**, no. 3 (1992), 351–365.
- [4] J. P. CONZE and A. RAUGI, Fonctions harmoniques pour un opérateur de transition et applications, *Bull. Soc. Math. France* **118** (1990), 273–310.
- [5] E. CURRY, *Characterizations of low-pass filters for multivariable wavelets and some related questions*, Ph.D. thesis (2005), Rutgers University, New Brunswick, NJ, 08854.

- [6] I. DAUBECHIES, Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **41** (1988), 909–996.
- [7] ———, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Math., SIAM, Philadelphia, PA, **61** (1992).
- [8] K. DEVLIN, *The Unfinished Game*, Basic Books, New York, NY, 2008.
- [9] V. DOBRIĆ, R. F. GUNDY, and P. HITCZENKO, Characterizations of orthonormal scaling functions: A probabilistic approach, *J. Geometric Analysis* **10**, no. 3 (2000), 413–420.
- [10] W. DOEBLIN and R. FORTÊT, Sur les chaînes à liaisons complètes, *Bull. Soc. Math. France* **65** (1937), 132–148.
- [11] K. GRÖCHENIG and W. R. MADYCH, Multiresolution analyses, Haar bases, and self-similar tilings of \mathbb{R}^n , *IEEE Trans. Inform. Theory* **38**, no. 2 (1992), 556–568.
- [12] R. F. GUNDY, Probability, ergodic theory, and low-pass filters, in Topics in Harmonic Analysis and Ergodic Theory, *Contemporary Math.*, Amer. Math. Soc. **444** (2007), 373–413.
- [13] R. F. GUNDY and A. L. JONSSON, Scaling functions on \mathbb{R}^2 for dilations of determinant ± 2 , *Applied and Computational Harmonic Analysis* **29**(1) (2010), 49–62.
- [14] E. HERNÁNDEZ and G. WEISS, *A First Course on Wavelets*, CRC Press, Boca Raton, New York, 1966.
- [15] I. KATAÍ and J. SZABÓ, Canonical number systems for complex integers, *Acta Sci. Math. (Szeged)* **37** (1975), 255–260.
- [16] J. C. LAGARIAS and Y. WANG, Haar-type orthonormal wavelet bases in \mathbb{R}^2 , *J. Fourier Analysis* **2** (1995), 1–14.
- [17] W. LAWTON and H. L. RESNIKOFF, *Multidimensional wavelet bases*, preprint of AWARE, Inc., 1991.
- [18] W. R. MADYCH, Some elementary properties of multiresolution analyses of $L^2(\mathbb{R}^n)$, in *Wavelets: A Tutorial in Theory and Applications* (C. K. Chui, ed.) Academic Press, 1993, 259–294.
- [19] H. L. RESNIKOFF and R. O. WELLS JR., *Wavelet Analysis: The Scalable Structure of Information*, Springer-Verlag, New York, Berlin, Heidelberg, 1998.
- [20] S. ROSS, *A First Course in Probability*, 7th ed., Pearson Prentice Hall, Upper Saddle River, N.J., 2006.
- [21] D. RUELLE, *Thermodynamic Formalism*, 2nd ed., Cambridge University Press, Cambridge, England, 2004.
- [22] M. J. T. SMITH and T. P. BARNWELL III, Exact reconstruction techniques for tree structured subband coders, *IEEE Trans. Acoust. Signal Speech Process.*, **34** (1986), 434–441.
- [23] P. WOJTASZCZYK, *A Mathematical Introduction to Wavelets*, London Math. Soc. Student Texts, **37**, Cambridge University Press, Cambridge, UK, New York, Melbourne, Madrid, 1997.
- [24] The Wutam Consortium, Basic properties of wavelets, *J. Fourier Anal. Appl.* **4** (1998), 575–594.