Wavelets and probability

Richard Gundy

Statistics Department, Rutgers University
Piscataway, New Jersey 08854, USA

WAVELETS AND PROBABILITY *

Richard F. Gundy
Rutgers University

Introduction.

There will be five lectures on wavelets and probability, with emphasis on mathematical problems rather than applications of wavelet techniques to applied problems. Four of the lectures will be on topics that have been the subject of my own investigations, surely the safest ground on which to make a stand.

The topics are as follows:

Wavelets and Gaussian processes.

Stopping times, adapted to wavelets, and almost everywhere convergence problems.

(a) The convergence theorems.

(b) Meyer's theorem and its extension.

Low pass filters and probability

(a) A. Cohen's theorem.

(b) The Markov process associated with a low-pass filter.

Each topic is the subject of a little essay, included in these notes. The lectures will be based on these essays and the articles quoted. The articles to which I contributed are included; otherwise the references are contained in the bibliography provided.

The first lecture, "Wavelets and Gaussian processes," represents what I have learned in a weekly seminar at Rutgers, the "anarchy seminar" as it is known to the participants. During the spring semester of 2002, Francisco Ojeda, a thesis student of Professor Vladimir Dobrič, presented the work of Benassi, Jaffard, and Roux [1], on the construction of multiresolution analyses for a class of Gaussian Markov processes. I will try to explain how this construction is made for the simplest case, that of Brownian motion.

I want to express my warmest gratitude to the collaborators whose names appear in these notes, Kazaros Kazarian, Vladimir Dobrič, and Pavel Hitczenko, and to the fellow anarchists, Vladimir Dobrič, Francisco Ojeda, Alfredo Rios, Eva Curry, and John Grothendieck, for their enthusiasm and patience with my insistent questions and constant interruptions.

[1] Benassi, A., Jaffard, S., and Roux, D. "Analyse multi-échelle des processus gaussieus markoviens d'ordre p indexés par $(0, 1)$," *Stoch. Processes and Their Appl.* **47** (1993), 275-297.

* This article contains material to be presented during the author's lecture at the workshop "Wavelets and Applications", Barcelona, July 1-6, 2002.

Wavelets and Gaussian processes.

The basic probability course, at the advanced undergraduate or beginning graduate level often includes a discussion of Brownian motion. In my opinion, the most elementary and elegant way to prove that such a process exists is to construct it using a complete orthonormal family of functions $\{\psi_i(t)\}$ for $0 \le t \le 1$. (The choice of the unit interval as parameter set is psychologically convenient: the familiar families $\{\psi_k(t)\}$ live on the compact sets.) The probability space $(\Omega, \mathcal{F}, P)$ is taken to be $\mathbf{R}^\infty = \mathbf{\Omega}$ with infinite product measure indexed by the standard Gaussian distribution on each coordinate. The coordinate variables X_k are automatically orthonormal as elements in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, and the unitary map that sends $\{\psi_k\}$ to $\{X_k\}$ allows us to send mathematical creatures living on $[0,1]$ (that can be expressed in terms of $\{\psi_k\}$) to their counterparts in $(\Omega, \mathcal{F}, P)$. In particular, if we take the indicator function of the interval $0 \le s \le t$, $t \le 1$, $\chi_{[0,t)}(\cdot)$ considered as an element of $L^2[0,1]$, and send it to $X(t,\omega)$ in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ via the unitary map, we obtain the Brownian motion process. The proof of this fact consists of verifying that $X(t)$ has the right covariance structure ($E\big(X(t)X(s)\big) = s \wedge t$) and that $X(t,\omega)$ is continuous in t for almost every ω. This is the sticky part, and can be finessed in various ways. However, to my mind, the most elegant solution is to use what is called the Ito-Nisio theorem: we can give a uniform estimate of the modulus of continuity of the approximates to $X(t,\omega)$,

$$X^n(t,\omega) = \sum_{k=0}^{n} a_k(t) X_k(\omega)$$

where $a_k(t) = \langle \chi_{[0,t)}, \psi_k \rangle$. (The most elegant approach to this estimation is via a real-variable inequality due to Garsia, Rodemich and Rumsey. See Garsia [5] for a simplified version.) This approach seems to me to establish the *unitary invariance of the continuity of* $X(t,\omega)$ in record time.

What is the shortcoming of this approach? The important aspects of $X(t)$ are revealed as t progresses (or regresses). It is the filtration $\mathcal{F}_t = \sigma(X(s), s \le t)$ that really counts, as opposed to $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$, the product σ-field generated by the coordinate variables that carries the approximates $X^n(t)$. If we wish to study $\mathcal{F}_t$, we must observe events in $\mathcal{F}_\infty = \sigma(X_0, X_1, \ldots)$; there is no way around this point. Can we reach a compromise with a judicious choice of $\{\psi_k\}$ so that the development of $X^n(t)$, as n progresses, mirrors *to some extent*, the development of $X(t)$ as t progresses. Here, the explicit choice of the basis $\{\psi_k\}$ is crucial. Non-local bases, such as the trigonometric system are definitely not appropriate.

Paul Lévy confronted this problem in another way. He constructed a polygonal approximation to the path function $X(t)$ by linear interpolation between $X\big(\frac{k}{2^n}\big)$ and $X\big(\frac{k+1}{2^n}\big)$. Thus, his approximation $X^n(t)$ equals $X(t)$ at dyadic points. Later, Ciesielski [4] observed that this polygonal approximation was, in fact, an instance of the unitary map just described when $\{\psi_k\}$ is the Haar system. (Ciesielski used the local properties of the Haar system to give a simple proof of the continuity of the path functions.) Are the Haar function "special," among all wavelet bases, for Brownian motion? Are there "special" wavelets for other processes? Let us address these questions.

First, what can we do with Haar functions that can't be done with any other system? Consider the following basic fact about the Brownian motion. The path functions $X(t,\omega)$ are not of bounded variation on any interval; in fact, they are nowhere differiable for almost every ω. However, they are of bounded *quadratic* variation on every interval, a fact of fundamental importance. By this I mean that for every finite partition $0 = t_0 < t_1 < \cdots < t_n \le 1$. $\sum_{i=1}^n \big(X_{t_i}(\omega) - X_{t_{i-1}}(\omega)\big)^2 = Q_n(\omega)$ and $E(Q_n) \le 1$. In fact, if $t_i = \frac{i}{2^n}$, $i = 0,\dots,2^n$, then $Q_{2^n}(\omega) \to 1$ for almost every ω. Here is a wavelet-Haar function proof of this fact, based on the strong law of large numbers. (If Y_i are independent identically distributed random variables, then $S_n(\omega)/n = \sum_{i=1}^n Y_i(\omega)/n \to E(Y_0)$ almost everywhere.) Consider the finite orthonormal system of indicator functions $2^{n/2}\chi_{[k/2^n,(k+1)/2^n)}$, $k = 0,\dots,2^n-1$ and their unitary images in $L^2(\Omega,\mathcal{F},P)$: $2^{n/2}X\big(\chi_{[\frac{k}{2^n},\frac{k+1}{2^n})}\big) = 2^{n/2}\big(X\big(\frac{k+1}{2^n}\big) - X\big(\frac{k}{2^n}\big)\big)$, as it turns out. Now observe that the Haar functions ψ_k, $k = 0,\dots,2^n-1$ span the same subspace of $L^2[0,1]$ as the system $2^{n/2}\chi_{[\frac{k}{2^n},\frac{k+1}{2^n})}$, $k = 0,\dots,2^n-1$. The image of ψ_k, $X(\psi_k) = X_k$, the kth coordinate function, if the Haar system is the chosen basis. Therefore, by unitary invariance, we have

$$2^n \sum_{k=0}^{2^n-1} \left(X\left(\frac{k+1}{2^n}\right) - X\left(\frac{k}{2^n}\right)\right)^2 = \sum_{k=0}^{2^n-1} \big(X_k\big)^2$$

and so

$$Q_n(\omega) = \frac{1}{2^n} \sum_{k=0}^{2^n-1} \big(X_k(\omega)\big)^2$$

which converges to 1 a.e. by the strong law of large numbers.

Since the Haar functions are well designed to study the time development of Brownian motion, one might ask whether other wavelet bases would also do the same job. Or, is there some intrinsic reason connected with Wiener measure, to prefer the Haar system? I propose to answer this question in the affirmative, based on the work of Benassi, *et al.* [1]. They show to some extent that a multiresolution analysis is, to some extent, built into any Gaussian process with a certain structure. The Brownian motion, considered as a Gaussian process, is one example with this structure.

Let me briefly review some facts about Gaussian processes. The parameter set T for the processes under consideration is either $\mathbf{R}$, $\mathbf{R}_+$ or $[0,1]$. Thus, we are given X_t, with t running over T. The state space, the range of X_t, is $\mathbf{R}^1$. The process X_t is centered, Gaussian if $E(X_t) \equiv 0$ and every finite collection $X_{t_1},\dots,X_{t_n}$ is jointly Gaussian: there exists a function $K(s,t)$ for all $(s,t) \in T \times T$, such that the quadratic form

$$Q(\mathbf{xi},\mathbf{t}) = \sum \xi_\mathbf{i}\xi_\mathbf{j}\mathbf{K}(\mathbf{t_i},\mathbf{t_j})$$

with $\mathbf{xi} = (\xi_\mathbf{1},\dots,\xi_\mathbf{n})$, $\mathbf{t} = (\mathbf{t_1},\dots,\mathbf{t_n})$, is positive definite, and

$$\hat{P}(\mathbf{xi},\mathbf{t}) = \exp\left(-\frac{\mathbf{1}}{\mathbf{2}}\mathbf{Q}(\mathbf{xi},\mathbf{t})\right)$$

is the Fourier transform of the probability measure associated with $X_{t_1},\dots,X_{t_n}$. The function $K(s,t) = E(X_sX_t)$ is called the covariance. Thus, the function $K(s,t)$ determines the process uniquely, and this function determines a Hilbert space $\mathbf{H}$ of functions obtained by completing the set of finite linear combinations of $K(s,\cdot)$ of the form

$$f(\cdot) = \sum_{i=1}^{n} a_i K(s_i, \cdot)$$

with respect to the bilinear form

$$\langle f, g\rangle = \sum a_i b_j K(s_i, s_j)$$

(see Janson [6], appendix F for a few more details). It is a consequence of this definition that the Hilbert space has a reproducing kernel $K(s,\cdot)$, that is, $K(s,\cdot) \in \mathbf{H}$ and

$$\langle K(s,\cdot), g(\cdot)\rangle = g(s).$$

Now it is not easy, in general, to recognize $\mathbf{H}$ beyond what is contained in the above description. However, sometimes $\mathbf{H}$ can be given a more explicit description. Let us consider the simplest case, where the Gaussian process is Brownian motion on $[0,1]$. In this case,

$$E\big(X(x)\cdot X(t)\big) = K(s,t) = \min(s,t)$$

and the collection of finite linear combinations

$$F(t) = \sum_{i=1}^{n} f(s_i)\min(s_i, t)$$

is precisely the collection of all piecewise linear functions on $[0,1]$ with $F(0) = 0$. A computation that requires a little patience shows that the norm

$$\sum f(s_i)f(s_j)\min(s_i, s_j) = \int_0^1 |F'(t)|^2 dt.$$

¿From this computation, we see that the reproducing kernel Hilbert space is a subspace of the Sobolev space $H^1[0,1]$, with the boundary condition given by $F(0) = 0$. This space (H_{CM}) was introduced by Cameron and Martin space [3]. Now this space H_{CM} is *local* in the following sense: it contains functions with compact support, strictly contained in $[0,1]$, and if $F \in H_{CM}$ and $F(t_0) = 0$ for some $0 < t_0 < 1$, then the restriction

$$F_{t_0}(s) = \begin{cases} F(s), & s \le t_0 \\ 0, & t_0 < s \end{cases}$$

also belongs to H_{CM}. Furthermore, it is local in that two functions with disjoint supports are orthogonal. The significance of *locality* was first pointed out by Loren Pitt [3]. Briefly, Pitt's investigation provides the following information. Given a Gaussian process $X(t)$, $t \in T$, let $H(X)$ be the subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ generated by taking the closure of finite linear combinations of $X(t_i)$ and $\mathcal{H}(T)$ be the reproducing kernel Hilbert space of functions on T given by $\phi_Y(t) = E\big(YX(t)\big)$, where $Y \in H(X)$. The reproducing kernel is $K(s,t) = E\big(X(s)X(t)\big)$. Under a mild continuity condition, Pitt's principal theorem states that a Gaussian process is Markov (in a generalized sense) if and only if $\mathcal{H}(T)$ is local. The generalized Markov property goes as follows: Recall that the usual Markov property states that the "past," and "future," represented by $\sigma\big(X(s), s < t\big)$ and $\sigma\big(X(s), s > t\big)$ respectively, are conditionally independent, given the present $\sigma\big(X(t)\big)$, $\sigma(\cdot)$ meaning "the sigma algebra generated by $(\cdot)$." However, in some cases the "present" way may be usefully enlarged by considering the intersection of

$$\tilde{\sigma} = \bigcap_{\epsilon > 0} \sigma\big(X(t+h), |h| < \epsilon\big).$$

If $X(t)$ is Brownian motion, this "germ-field" is simply the completion of $\sigma\big(X(t)\big)$. However, if $I(t)$ is the (Gaussian) process given by $I(t) = \int_0^t X(s)ds$, then the germ-field is $\sigma\big(I(t), X(t)\big)$, and the vector-valued process $\big(I(s), X(s)\big)$ is Markov in the usual sense, but $I(s)$ alone is Markov in the generalized sense. (The covariance function of $I(t)$ is

$$K(s,t) = \frac{(s \wedge t)^2 (s \vee t)}{2} - \frac{(s \wedge t)^3}{6},$$

which cannot be factored into a product $G(s \wedge t) H(s \vee t)$ as required for the strict Markov property.) See Borisov [2]. Moreover, Pitt shows that if $\mathcal{H}(T)$ is local and "rich enough," the inner product is given by a nonnegative Dirichlet form

$$\langle u, v \rangle = \sum_{\alpha,\beta} \int_T a_{\alpha,\beta} D^\alpha u D^\beta v$$

which, in nice cases, is uniformly strongly elliptic.

Now let us start over with Brownian motion, Pitt's theorems, and see how the Haar functions come out of a general procedure. The Brownian motion $X(t)$, $t \in [0,1]$ generates a Hilbert subspace $H(X)$ of $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ by taking the closure of finite linear combinations of $X(t_i)$, $i = 1, \ldots, n$. Now we transfer $H(X)$ to a Hilbert space $\mathcal{H}([0,1])$ consisting of real-valued functions on $[0,1]$ via the map

$$E\big(Y \cdot X(t)\big) = \phi_Y(t).$$

It turns out that $\mathcal{H}([0,1]) = H_{CM}$, as we saw above. This space is certainly local and the Dirichlet form is

$$\langle u, v \rangle = \int_0^1 u'(s) v'(s) ds.$$

This form is associated with $-D^2$ acting on the function in C^2 with boundary conditions $u(0) = 0$, $u'(1) = 0$ and we can check that, on these functions $-D^2$ is elliptic:

$$c\|u\|_{H^1} \leq \langle -D^2 u, u \rangle \leq \|u\|_{H^1},$$

where $\|u\|_{H1}$ is the usual Sobolev norm. Therefore we can complete the space of C^2-functions with given boundary conditions relative to this norm, and this completion is clearly the space H_{CM} that we described earlier. The gain here is that we recognize the reproducing kernel $K(s,t) = \min(s,t)$ as the Green's function for $-D^2$ with the boundary conditions stated above. Now, let us use the *local structure* of H_{CM} to construct a multiresolution analysis for H_{CM}, one that can be pulled back to $H(X)$. The first step is to notice that

$-D^2$ admits a Green's function for any appropriate boundary conditions on any subinterval of $[0,1]$. In particular, if we take boundary conditions $u(t_0) = 0$, $u(t_1) = 0$ and $u(t) \equiv 0$ outside the interval $t_0 < t < t_1$, then the Green's function for $-D^2 v = u$ is the little tent $K(s,t)$ on $t_0 < s, t < t_1$. ($K(s,t) = 0$ at t_0, t_1 and linear between (t_0, t) and (t, t_1).) If we pull back this little tent to a process in $H(X)$ (which we can do since H_{CM} is local), we obtain a miniature tied-down Brownian bridge process from t_0 to t_1. Notice that we can do this procedure on the process side: We can fix $t_0 < t_1$, start a process at t_0 and bridge it to t_1 (so that $X(t_0) = 0$, $X(t_1) = 0$), all the while remaining in $H(X)$. Now the covariance of this process, a function that lives on the interval $t_0 < t < t_1$, and belongs to H_{CM}, is precisely the little tent we constructed above.

I hope it is clear now that the Haar multiresolution analysis can be transported to H_{CM}: The Haar functions appear as rows of little tents based at the dyadic rationals. Each tent corresponds to a Brownian bridge random variable in $H(X)$. Of course the bridges are independent since the tents are orthogonal in $\mathcal{H}[0,1]$, this, in turn, being true because their derivatives are the orthogonal Haar functions. However, their orthogonality, from this point of view, is a consequence of the magic of Brownian motion, and a point of view that leads to other things.

Bibliography

1. Benassi, A., Jaffard, S., and Roux, D. "Analyse multi-échelle des processus gaussieus markoviens d'ordre p indexés par $(0,1)$," *Stoch. Processes and Their Appl.* **47** (1993), 275-297.

2. Borisov, I.S., "On a criterion for Gaussian random processes to be Markovian," *Th. Prob. Appl.* (English trans. from Russian) **27** (4) (1982), 863-865.

3. Cameron, R.H. and Martin, W.T. "Transformation of Wiener integrals under translations," *Ann. Math.* **45** (1944), 386-396.

4. Ciesielski, Z. "Hölder conditions for realizations of Gaussian processes," *Trans. Amer. Math. Soc.* **99** (1961), 403-413.

5. Garsia, A.M. "Continuity properties of Gaussian processes with multidimensional time parameter," *Proceedings of the Sixth Berkeley Symposium*, Vol. 2, 1972.

6. Janson, S. *Gaussian Hilbert Spaces.* Cambridge Tracts in Mathematics #129, Cambridge Univ. Press, 1997.

7. Pitt, L. "A Markov property for Gaussian processes with a multidimensional parameter," *Arch. Rational Mec. Anal.* **43** (1971), 367-391.

Additional remarks on pointwise convergence, decreasing rearrangements, stopping times and related topics.

I am including the manuscripts of three articles that relate to these topics. The principal result of these three papers is stated in what must be the most obscure of all possible expressions. (It appears in the concluding remarks in reference [2].) The result is as follows:

Theorem: *Let ϕ be a compactly supported (pre)scaling function for a multiresolution analysis. Suppose that $f = (\ldots f_j, f_{j+1} \ldots)$, $j \in \mathbf{Z}$ is a sequence of L^2 functions that satisfies* (a) *$f_j \in V_j$, the j-th subspace of the multiresolution analysis;* (b) *$P_j(f_{j+1}) = f_j$ where P_j is the orthogonal projection from V_{j+1} to V_j, $j \in \mathbf{Z}$. Then the following sets are equal almost everywhere:*

$$A = \big\{x : f^*(x) := \sup_j |f_j(x)| < \infty\big\},$$
$$B = \big\{x : \rho^2(f)(x) = \textstyle\sum_{j=-\infty}^{\infty}(f_j(x) - f_{j-1}(x))^2 < \infty\big\},$$
$$C = \big\{x : \liminf f_j(x) = \limsup f_j(x) < \infty\big\}.$$

Here ϕ is a compactly supported prescaling function (the translates of ϕ form a Riesz basis for V_0); the most prominent examples are the B-splines with knots at the integers, or the compactly supported orthogonal scaling functions constructed by Daubechies.

A brief history of the theorem: Kazaros Kazarian and I took up this problem in 1994, in Madrid. Our experience with martingales and harmonic functions in the half-plane suggested that such a theorem might be true for wavelet expansions. We worked on this, together and apart for about two and a half years, focusing on the B-splines. An important insight came when Kazarian noticed that the wavelet associated with ϕ had a certain "analytic" property. It was this insight, under heavy disguise, that is the "M-Z condition" in our joint paper. At the time the paper was written, I saw absolutely no way to verify this M-Z condition for Daubechies' scaling functions.

During the academic year 1996-97, Pavel Hitczenko spent his sabbatical year at Rutgers. Together with Vladimir Dobrič, the three of us discussed the problem. One day Pavel emerged from the library with a copy of Yves Meyer's paper [12] in hand. Meyer had proved a sort of unique continuation theorem for functions in V_0 generated by a (Daubechies) scaling function.

Theorem (Meyer [12]). *Let ϕ be a completely supported scaling function with support on $[0, N]$. Suppose that*

$$\sum_{k=-N+1}^{0} a_k \phi(x-k) = 0 \tag{1}$$

for all $x : 0 \le x \le 1$. Then $a_k \equiv 0$, $-N+1 \le k \le 0$.

(The theorem was subsequently given a more transparent proof by Lemarié and Malgouyres [11].) As it is stated, Meyer's theorem does not imply an "M-Z condition" for such scaling functions. What is needed is a stronger form of Meyer's theorem, where we assume only that $\sum_{-N+1}^{0} a_k\phi(x-k) = 0$ on a subset of $[0,1]$ of positive measure. The paper with Dobrič and Hitczenko [2] establishes this stronger form of Meyer's theorem, with a rather complicated argument. Subsequently, I found a simpler proof, which appears in [5].

The manuscript with Kazarian included in these notes is the same as the published article except for some corrections that were suggested by Alfredo Rios. The argument in Step 9 has been simplified and shortened. The conversations with Alfredo Rios also led to a reformulation of the principal result. To express this reformulation, we define a censoring operator by a doubly-indexed sequence of coefficients $e = (e_{j,k},\ (j,k) \in \mathbf{Z}\times\mathbf{Z},\ e_{j,k} = 0 \text{ or } 1)$. Given a sequence of L^2 functions $f = (\ldots, f_l, f_{l+1}, \ldots)$ as above, we define the e-censored sequence, denoted by $e(f)$ as follows: Let $f_j(x) - f_{j-1}(x) = \sum_k b_{j,k}\psi(2^j x - k)$, where $\psi(x)$ is the wavelet (or prewavelet) associated with the scaling (or prescaling) function ϕ. The sequence $e(f)$ is defined by

$$e(f)_j(x) - e(f)_{j-1}(x) = \sum_k e_{j,k} b_{j,k}\psi(2^j x - k).$$

Theorem. *Let f be a sequence of L^2 functions as above. Suppose that either:* (a) $\sup_j |f_j| < \infty$ *on a set A of positive measure, or* (b) $S(f) < \infty$ *on a set B of positive measure. For any $\epsilon > 0$, there exists a censoring operator e such that $e(f)(x) = f(x)$ for all $x \in A/E_\epsilon$ or all $x \in B/E_\epsilon$ where E_ϵ is a set of measure less than ϵ. Furthermore, the sequence $\epsilon(f)$ is L^2 bounded:*

$$\sup_j \|\epsilon(f)_j\|_2 < \infty.$$

As an application of this formulation of the theorem, we offer an extension of the following theorem, due to Tao [15] and Tao and Vidakovic [16]. The version below is taken from Tao and Vidakovic [16].

Definition. A function $\delta(x,\lambda)$ from $\mathbf{R}\times\mathbf{R}^+ \to \mathbf{R}$ is called a *shrinkage rule* if there exist positive constants c and ϵ such that for all $\lambda > 0$,

$$|x| - |\delta(x,\lambda)| \le c\lambda$$

and

$$|\delta(x,\lambda)| \le c|x|^{1+\epsilon}\lambda^{-\epsilon}.$$

The shrinkage operator $T_\lambda = T_\lambda^\delta$ is defined on a wavelet series $\sum c_{j,k}\psi_{j,k}(x)$ by

$$T_\lambda\big(\sum c_{j,k}\psi_{j,k}\big)(x) = \sum \delta(c_{j,k},\lambda)\psi_{j,k}(x).$$

Theorem [15]. *If $f \in L^p(\mathbf{R})$, $1 \le p < \infty$, and has wavelet expansion $f(x) \sim \sum c_{j,k}\psi_{j,k}(x)$, then*

$$\lim_{\lambda\to 0} T_\lambda(f)(x) = f(x)$$

almost everywhere.

This theorem is proved under the assumption that ψ is an orthonormal wavelet with rapid decay at infinity. However, as the reader may have guessed, if we assume that ψ is bounded and has compact support, we can drop the assumption that the series is obtained from a function in L^p, $p \geq 1$. In fact, *we see that shrinkage operators converge almost everywhere on the set where the wavelet series converges.* To verify this, we use the reformulated theorem on local convergence to obtain an L^2 function, then apply the Tao-Vidakovic argument.

Here is the proof of the Tao-Vidakovic theorem, taken from [16]. The basic step is to show that

$$\sup_{\lambda} \left| \sum \delta(c_{j,k}\lambda)\psi_{j,k}(x) \right| \leq cM(f)(x),$$

where $M(f)(x)$ is the Hardy-Littlewood maximal function of $f \in L^2(\mathbf{R})$. Let us suppose that we have this estimate. Now, as Tao remarks in [15], the operator T_λ is rather nonlinear; however, T_λ is linear if we decompose f into pieces according to the given wavelet expansion. Thus, if

$$\begin{aligned} f(x) &\sim \sum^{J_1} c_{j,k}\psi_{jk} + \sum_{J_1+1}^{\infty} c_{jk}\psi_{j,k}(x) \\ &= f_1(x) + f_2(x) \end{aligned}$$

where the first sum is over all terms up to scale J_1 and the second sum represents the remainder, then

$$T_\lambda(f) = T_\lambda(f_1) + T_\lambda(f_2).$$

Since $T_\lambda(f_1) \to f_1$ as $\lambda \to 0$, we must show that, for ϵ and $\delta(\epsilon)$ given,

$$\left| \limsup_{\lambda} T_\lambda(f_2) - \liminf_{\lambda} T_\lambda(f_2) \right| > \epsilon$$

for a set of measure less than $\delta(\epsilon)$. But this follows from the basic estimate and the L^2-norm estimate for $M(f_2)$, as usual.

Now let us verify the basic step. Since ψ has compact support, $\psi_{jk} \neq 0$ for only a finite number of indices k, so that $\psi(x)$ is supported in $[-N, N]$. Now $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$ and

$$\begin{aligned} |c_{jk}| &= \left| \int f(y)\psi_{j,k}(y)dy \right| \\ &\leq c2^{j/2}\frac{N}{2^j}\left(\frac{2^j}{2N} \int_{x-N/2^j}^{x+N/2^j} |f(y)|dy \right) \\ &\leq c2^{-j/2}Mf(x). \end{aligned}$$

The main trick, from Tao [15], is to break the full wavelet series into two parts, depending on λ and $M(f)(x)$. Let J be the scale such that

$$2^{J/2} \leq Mf(x)/\lambda \leq 2 \cdot 2^{J/2}$$

and break the series

$$\sum \delta(c_{j,k})\psi_{j,k}(x) = \sum_{j<J,k} \delta(c_{j,k}, \lambda)\psi_{jk}(x) + \sum_{j \geq J,k} \delta(c_{j,k}, \lambda)\psi_{j,k}.$$

The tail term is estimated using one of the properties of δ:

$$\begin{aligned}\sum_{j\geq J,k}\delta(c_{jk},\lambda)\psi_{jk}(x) &\leq c\sum_{j\geq J,k}\frac{|c_{j,k}|^{1+\epsilon}}{\lambda^{\epsilon}}\,2^{j/2}\\ &\leq c\sum_{j\geq J,k}(2^{-j/2}Mf(x))^{1+\epsilon}\lambda^{-\epsilon}2^{j/2}\\ &= cMf(x)\bigg(\frac{M(f)(x)}{\lambda}\bigg)^{\epsilon}\cdot\big(2^{J/2}\big)^{\epsilon}.\end{aligned}$$

(Here the sum over k is irrelavant since, at each level j, there are only a finite number $(2N)$ of indices k where $\psi_{j,k}\neq 0$.) The judicious choice of J guarantees that

$$\sum_{j\geq J_k}\delta(c_{j,k},\lambda)\psi_{j,k}(x)\leq cM(f)(x).$$

Now the first term is decomposed into

$$\sum_{j<J,k}\delta(c_{jk},\lambda)\psi_{j,k}(x)=\sum_{j<J,k}\big(\delta(c_{j,k}\lambda)-c_{j,k}\big)\psi_{jk}(x)+\sum_{j<J,k}c_{jk}\psi_{j,k}(x).$$

The first term is bounded by

$$\begin{aligned}c\sum_{j<J,k}\lambda 2^{j/2} &\leq c\lambda 2^{J/2}\\ &\leq cMf(x)\end{aligned}$$

and the second is just the orthogonal projection of f onto V_{J-1}. This too is bounded by the maximal function (see Kelly, *et al.* [7]). Just for fun, here's how that goes. This projection is obtained by expanding f using the (compactly supported) scaling function ϕ:

$$P_J(f)(x)=\sum_k c_{J,k}2^{J/2}\cdot\phi(2^Jx-k)$$

where $c_{J,k}=\int f(y)2^{J/2}\phi(2^Jy-k)dy$. As before, for fixed x, $\phi(2^Jx-k)\neq 0$ for a finite number $2N$ of indices k. This means that

$$\begin{aligned}P_J(f)(x) &\leq c2^J\int_{x-N\cdot 2^{-J}}^{x+N\cdot 2^{-J}}|f(y)|dy\\ &= cM(f)(x),\end{aligned}$$

which completes the estimations.

In conclusion, let us remark that Tom Körner [9] has shown that, in $L^2[0,2\pi]$, there exists a function f such that $\limsup_\lambda|T_\lambda(f)(x)|=\infty$ a.e. Körner uses the hard shrinkage $\delta(x,\lambda)=0$ if $|x|\leq\lambda$, and $\delta(x,\lambda)=x$ if $|x|>\lambda$. The proof builds on the work of Kolmogorov, Zahorski, and Olevski on the divergence of rearranged orthogonal series. Since "hard thresholding" would seem to be the most natural way to sum a series, Körner's construction is the most dramatic advance in this line of investigation, in my opinion.

It is interesting to contrast the *first formulation* of the theorem with what is known for real-valued trigonometrical series. Menchov proved that, given a set $E\subset[-\pi,\pi]$ of positive measure and two measurable functions f, F with $-\infty < f(x) \le F(x) = +\infty$ on E, and $f(x) = -\infty$, $F(x) = +\infty$ on the complement of E, there exists a trigonometric series

$$S_n(x) = \sum_{k=0}^{n} e_k \cos(kx + \theta_k), \qquad n \ge 0$$

such that

$$\liminf S_n(x) = f(x)$$

and

$$\limsup S_n(x) = F(x)$$

almost everywhere in $[-\pi, \pi]$. The question of convergence to $+\infty$ on a set E of positive measure was left open. Partial results were obtain by Ulyanov in the 1960's and Kahane and Katznelson in the 1970's, but it wasn't until 1988 that S. V. Konyagin finally settled the issue by showing that convergence to $+\infty$ is *not* possible. Very briefly, his proof uses two ideas. The first idea is due to W. Darsow [1], who probably worked on the problem at the University of Chicago in the 1950's. He showed that if $\liminf S_n(x) \ge 0$ on a set E of positive measure, then $S_n(x) = O(n^c)$ for some c, $1 < c < 2$, for almost every $x \in E$. From this it follows that the coefficients $c_k = O(k^2)$. This bound means that the series may be formally integrated several times to obtain a series for a continuous function.

Now Konyagin introduces a generalization of the Riemann symmetric (kth) derivative, as follows: Recall that $\Delta_h f(x) = f(x+h) - f(x-h)$. Let $\bar{h} = (h_1, \ldots, h_k)$ and $\bar{\epsilon} = (\epsilon_1, \ldots, \epsilon_k)$ where $|\epsilon_j| = 1$. Let $\Delta_{\bar{h}}(f) = \sum \epsilon_1 \cdots \epsilon_k \Delta(f)(x + \epsilon_1 h_1 + \cdots + \epsilon_k h_k)$ where the sum is taken over all $\bar{\epsilon}$ in the 2^k-cube, with $\bar{h}$ fixed. Then if $f \in C^k(\mathbf{R})$,

$$\delta_{\bar{h}}(f)(x) = \Delta_{\bar{h}}(f)(x)/2^k h_1 \cdots h_k$$

converges to $f^{(k)}(x)$ as $h \to 0$. Konyagin's idea is to introduce some variation in the $\bar{h}$: he considers the class $H_{k,\alpha}(h) = \{\bar{h} : \alpha h \le h_j \le h,\ 1 \le j \le k\}$ where $0 < \alpha < 1$. Let $\underline{D}_\alpha^{(k)} f(x)$ and $\bar{D}_\alpha^k f(x)$ be the upper and lower limits of $\delta_{\bar{h}}(f)(x)$ as $\bar{h} \in H_{k,\alpha}$, $h \to 0$.

Theorem (Konyagin). *If $f \in C(\mathbf{R})$, $k \in \mathbf{Z}_+$, $0 < \alpha < 1$, $0 < \beta < 1$, then $\underline{D}_\alpha^{(k)}(f)(x) > -\infty$ for $x \in E$, a set of positive measure, implies $\bar{D}_\beta^{(k)}(f)(x) < +\infty$ almost everywhere on E.*

The proof of this theorem relies on Ward's theorem on derivates of additive interval functions. Specifically, see the second paragraph on page 139 of Saks [13].

Konyagin applies this theorem to the function f obtained from the formally integrated series. After some modification, he is able to show that $\underline{D}_\alpha^{(k)}(f)(x) \ge 0$ on a set of positive measure, and that, essentially, $\bar{D}_\alpha^{(k)}(f)(x) \ge \limsup S_n(x)$ almost everywhere on this set. Since $\bar{D}_\alpha^{(k)}(f)(x) < +\infty$ almost everywhere on this set, the theorem is proved.

There are two features of the Menchov-Konyagin theorems that stand in contrast to the wavelet convergence theorems. First, bounded oscillation implies convergence for wavelet series under the stated assumptions, while this is not the case for trigonometric series. Second, our theorem does not address the possibility of one-sided convergence to infinity. The only theorem that I know concerning this phenomenon is due to Talalyan and Arutyunyan [14]: *there is no Haar series that converges to* $+\infty$ *on a set of positive measure.* At about the same time, independently, I found a proof of this, using martingale methods [3]; these methods were sharpened in work with Burkholder [6] where we showed, among other things, that convergence to $+\infty$ is not possible for a class of martingales that includes the Haar functions. (It is possible to prove Konyagin's theorem, quoted above, using martingale arguments.) Here is the idea from [3], restated in terms of the present discussion: Given a Haar series $\sum c_{j,k}\psi_{jk}(x)$, such that the infimum of the partial sums is bounded below at each point in a set of positive measure, we can censor the series, in the sense discussed above, so that the censored series (a) agrees with the original series except on a set of small measure; (b) the censored series is uniformly bounded below by a constant $C > 0$. Thus, by adding this constant, we have a Haar series that has nonnegative partial sums. This sequence of partial sums is a nonnegative martingale, and therefore converges almost everywhere, by Doob's theorem.

Bibliography

1. Darsow, W. "On the boundedness of trigonometric series," *J. London Math. Soc.* **35** (2) (1960), 237-238.

2. Dobrič, V., Gundy, R.F., and Hitczenko, P. "Stability properties of a compactly supported prescale function," *SIAM J. Math. Analysis* **31** (3) (2000), 574-580.

3. Gundy, R.F. "A class of martingale series," *Orthogonal Expansions and Their Continuous Analogues*, Proceedings of the Conference at Southern Illinois University, Edwardsville, IL, April 27-29 (1967), 99-102.

4. Gundy, R.F. "Cohen's characterization for low pass filters and Meyer's theorem on linear independence," in *Contemporary Math: The Harmonic and Functional Analysis of Wavelets and Frames*, AMS #247, American Mathematical Society, Providence RI, 1999, pp.249-257.

5. Gundy, R.F. "Two remarks on wavelets...," in *Cont. Math: The Harmonic and Functional Analysis of Wavelets and Frames*, AMS #247, Providence, RI, 1999, 249-257.

6. Gundy, R.F. and Burkholder, D.L. "Extrapolation and interpolation of quasi-linear operators on martingales," *Acta Math.* **124** (1970), 250-304.

7. Gundy, R.F. and Kazarian, K. "Stopping times and local convergence of spline wavelet expansions," *SIAM J. Math. Analysis* **31** (3) (2000), 561-573.

8. Kelly, S., Kon, M.A., and Raphael, L.A. "Local convergence for wavelet expansions," *J. Functional Analysis* **126** (1994), 102-138.

9. Körner, T.W. "Divergence of decreasing rearranged Fourier series," *Ann. Math.* **144** (1996), 167-180.

10. Konyagin, S. "Limits of indeterminacy," *Mat. Zametki* **446** (1988), 770-784 (Russian). (English transl., *Math. Notes* **44** (5-6), 910-920.)

11. Lemarié, P.G. and Malgouyres, G. “Support des fonctions de base dans une analyse multirésolution,” *C. R. Acad. Sci. Paris, Sér. I Math.*, **313** (1991), 377-380.

12. Meyer, Y. “Ondelettes sur l’intérvalle,” *Rev. Mat. Iberoamericana* **7** (1991), 115-133.

13. Saks, S. *Theory of the Integral*, 2nd rev. ed. (English translation by L. C. Young), Hafner Publishing Co., New York.

14. Talalyan, A.A. and Arutyunyan, F.G. “On the convergence of Haar series to $+\infty$,” *Mat. Sb.* **66** (108) (1965), 240-247.

15. Tao, T. “On the almost everywhere convergence of wavelet summation methods,” *Applied and Comp. Harmonic Analysis* **3** (1996), 384-387.

16. Tao, T. and Vidakovic, B. “Almost everywhere behavior of general shrinkage operators,” *Applied and Comp. Harmonic Analysis* **9** (2000), 72-82.

STOPPING TIMES AND LOCAL CONVERGENCE FOR SPLINE WAVELET EXPANSIONS

Richard F. Gundy* and Kazaros Kazarian**
Rutgers University and Universidad Autonoma de Madrid

*The first author would like to acknowledge the support for this research from the sabbatical program at the Universidad Autonoma de Madrid during the summer, 1994.

**The second author would like to acknowledge the support for this research from DGICYT under grants PB94-0149 and PB97-0030. Both authors would like to thank the Volkswagen Foundation for support during the month of July, 1996 at the Mathematics Research Institute, Oberwohlfach, Germany.

Key Words and Phrases: Wavelets, stopping times, martingales

AMS 1991 Subject Classification: 42C15, 60G46

Abstract.

A local convergence theorem for spline wavelet expansions is proved. This theorem relates the finiteness of the quadratic variation of the expansion with the local convergence of the expansion on sets of positive measure. A stability property of these expansions is one of the key points in the proof.

1. Introduction.

The principal purpose of this paper is to prove a local convergence theorem for spline wavelet expansions, using a combination of techniques from martingale theory and wavelet analysis. In particular, we show that the notion of a stopping time may be adapted to these wavelet expansions.

The Haar functions are the point of departure for this discussion. The Haar series may be viewed as the sequence of partial sums of orthogonal elements from the multiresolution analysis generated by the dilations and translations of $\chi_{[0,1]}(x)$, the indicator function of the unit interval. On the other hand, this sequence of partial sums forms a martingale with respect to the filtration of σ-fields generated by the (dyadic) dilations and (integer) translations of $\chi_{[0,1]}$. This coincidence is unique: no other multiresolution analysis generates martingales in a similar fashion. The following theorem concerning Haar functions is proved by martingale methods [3] (see also [5]).

THEOREM A. *Let* $f = (f_0, f_1, \ldots)$ *be the sequence of partial sums of a Haar series on* $[0,1]$. *Then the following sets are equivalent, i.e. they differ by at most a set of Lebesgue measure zero.*

$$\begin{aligned} A &= \{x : f^*(x) := \sup_n |f_n(x)| < \infty\}; \\ B &= \{x : S^2(f)(x) := \sum_0^\infty (f_n(x) - f_{n-1}(x))^2 < \infty\}; \\ C &= \{x : \limsup f_n(x) = \liminf f_n(x) \infty\}. \end{aligned}$$

This theorem is not true for other martingales, in general. Some additional stability condition must be assumed. The goal of this paper is to extend this local convergence theorem for a class of wavelet expansions satisfying a stability condition. This class includes the series arising from polynomial spline wavelets. The precise statement of the theorem, together with the stability condition, is given below.

2. Notation.

We will use a *multiresolution analysis* $V = \{V_j; -\infty < j < \infty\}$ generated by a compactly supported, continuous (pre)scale function $\phi(x)$. That is, we require that ϕ have a nonzero integral, that it satisfy a dilation equation of the form

$$\text{(a)} \qquad \phi(x) = \sum_{k=0}^{N} p_k \phi(2x-k),$$

and that

$$\text{(b)} \qquad \sum a_n^2 \cong \left\| \sum a_k \phi(x-k) \right\|_2^2$$

(the translates of ϕ are a Riesz basis). Linear combinations of translates of ϕ form a subspace of $L^2(\mathbf{R})$; the L^2-closure of this subspace is denoted by V_0. Dyadic dilations of functions in V_0 form a closed subspace of $L^2(\mathbf{R})$, denoted by V_j ($f(\cdot) \in V_j$ iff $f(2^{-j}\cdot) \in V_0$). If $V_j \subset V_{j+1}$ for all integers j it then follows that the increasing sequence of subspaces $\{V_j\}$ exhausts $L^2(\mathbf{R})$ (see [6], Chapter 2, page 48). Let P_j be the orthogonal projection operator from L^2 onto V_j.

A special case of a multiresolution analysis of this type is the one generated by $\phi(x) = \chi_{[0,1]}(x)$. In this case, the projections P_j are conditional expectations, and a sequence $P_j(f)(x)$ forms a martingale. To distinguish this special case, we use D_0 to denote the set of *all* linear combinations of translates $\chi_{[0,1]}(\cdot - k)$, $k \in \mathbf{Z}$, and $D = \{D_j; -\infty < j < \infty\}$ the set of 2^j-dilates of elements of $D_0 : D_j = \{f : f(2^{-j}x) \in D_0\}$. A *stopping time* $\tau(x)$ is a function with values in $\mathbf{Z}_+ \cup \{\infty\}$, such that the indicator function of $\{x : \tau(x) = j\}$ belongs to D_j.

3. Prewavelets with compact support.

In this paper we assume that ϕ is supported in the interval $[0, N]$. This implies the existence of a prewavelet ψ having compact support. By this, we mean a function $\psi \in V_0$ such that $\psi(2x-k)$, $k \in \mathbf{Z}$ is a Riesz basis in the orthogonal complement of V_0 in V_1, usually written as W_0. The function $\psi(2\cdot)$ is necessarily of the form

$$\psi(2x) = \sum_{-N+1}^{2N-1} c_k \phi(2x-k),$$

and since $\psi(2\cdot) \in W_0$, we have

$$\int \phi(x)\bar{\psi}(2x)dx = \sum_{-N+1}^{2N-1} \bar{c}_k \int \phi(x)\overline{\phi(2x-k)}dx = 0.$$

If we set

$$e_k = \int \phi(x)\overline{\phi(2x-k)}dx,$$

one solution to this equation is given by $c_k = (-1)^k \bar{e}_{1-k}$; the products $\bar{c}_k e_k$ and $\bar{c}_{1-k} e_{1-k}$ have opposite signs and equal magnitudes, and so cancel in pairs. The function $\psi(x)$ is compactly supported on the interval

$[-N+1, 3N-1]$ since $e_k \equiv 0$ if $k \notin [-N+1, 2N-1]$. It turns out that, with this choice, the translates $\psi(2x-k)$, $k \in \mathbf{Z}$, form a Riesz basis for W_0 (see [6]). (We pause here to note that our notation is slightly different from the usual custom, where $\psi(x)$ is taken to be a function in W_0, whose translates $\psi(x-k)$, $k \in \mathbf{Z}$, form a Riesz basis. Our notation seems to us to make the exposition run more smoothly.) We shall also use the notions of dual prescale function and dual prewavelet (see Chui [2], Chapter 5, page 170). The dual prescale function $\tilde{\phi}$ is a function that belongs to V_0, and satisfies

$$\int \phi(x-k)\overline{\tilde{\phi}(x)}dx = \delta_0(k).$$

The dual prewavelet satisfies $\tilde{\psi}(2x) \in W_0$, where

$$\int \psi(x-k)\overline{\tilde{\psi}(x)}dx = \delta_0(k).$$

We now impose an additional stability condition on the prescale function ϕ.

CONDITION (M-Z): Given any measurable subset $E \subseteq [0,1]$, of measure greater than $\delta > 0$, and any sequence $a_k;\ k = -N+1, \ldots, 0$, we have

$$\sum_{k=-N+1}^{0} |a_k| \leq c(\delta) \sup_{x \in E} \left| \sum_{k=-N+1}^{0} a_k \phi(x-k) \right|$$

where $c(\delta)$ depends on δ but is otherwise independent of E.

Remarks. Since ϕ is assumed to be bounded, the inequality may be reversed at the expense of another constant. That is,

$$c^{-1}(\delta) \sum_{k=-N+1}^{0} |a_k| \leq \sup_{x \in E} \left| \sum_{k=-N+1}^{0} a_k \phi(x-k) \right| \leq c \sum_{k=-N+1}^{0} |a_k|.$$

This "two-norm" condition is in the same spirit as a stability condition suggested by Marcinkiewicz and Zygmund in [7] in their study of convergence of series of independent random variables. Their stability condition was introduced in a conditional form in [4] to prove a generalization of Theorem A. This conditional stability condition is the basis for the results in [1].

The Haar scale function $\chi_{[0,1]}(x)$ satisfies condition (M-Z) in a vacuous way. On the other hand, if we take the N-fold convolution of $\chi_{[0,1]}$ with itself, we obtain a prescale function ϕ for the multiresolution analysis of polynomial splines of degree $N-1$. The translates $\phi(x-k);\ k = -N+1, \ldots, 0$, are linearly independent polynomials on $[0,1]$. Given $E \subseteq [0,1]$ of measure greater than $\delta > 0$, we can find $x_1, \ldots, x_N,\ x_i \in E$ such that

$$\min_{i,j} |x_i - x_j| = O_N(\delta).$$

Now write

$$\sum_{-N+1}^{0} a_k \phi(x-k) = \sum_{k=0}^{N-1} b_k x^k$$

and compute the coefficients b_k by using Cramer's rule on the Vandermonde determinant at the points $x_1, \ldots, x_N$. This gives an estimate for the b-sequence:

$$\sum_{k=0}^{N-1} |b_k| \le c(\delta) \sup_{x \in E} \left| \sum_{k=-N+1}^{0} a_k \phi(x-k) \right|,$$

and consequently, the same estimate for the a-sequence:

$$\sum_{k=-N+1}^{0} |a_k| \le c'(\delta) \sup \left| \sum_{k=-N+1}^{0} a_k \phi(x-k) \right|.$$

Results: The principal result is as follows:

THEOREM B. *Suppose that ϕ is a prescale function that satisfies condition (M-Z). Suppose that $f = (\ldots, f_{j-1}, f_j, \ldots)$ is a sequence of functions such that*

$$P_j(f_{j+1}) = f_j$$

for every $j \in \mathbf{Z}$. Then the following sets are equal almost everywhere:

$$A = \left\{ x : f^*(x) := \sup_j |f_j(x)| < \infty \right\};$$

$$B = \left\{ x : S^2(f)(x) = \sum_{j=-\infty}^{\infty} (f_j(x) - f_{j-1}(x))^2 < \infty \right\};$$

$$C = \left\{ x : \overline{\lim} f_j(x) = \underline{\lim} f_j(x) < \infty \right\}$$

(that is the Lebesgue measure of each of the sets $A \triangle B, A \triangle C, C \triangle B$ is zero).

Proof. Before giving the details of the proof, let us outline the strategy. It is enough to prove the theorem with the sets A, B, C restricted to any interval of unit length. Therefore, we assume that A, B and C are subsets of the unit interval. Given $\epsilon > 0$, we may also assume that the measure of any one of these sets is greater than $1 - \epsilon$. Both of these assumptions amount to an appropriate choice of an origin for the dilation scale. Since $f_0 \in L^2(\mathbf{R})$, we have

$$\sup_{j \le 0} |f_j(x)| < \infty;$$

$$\sum_{j \le 0} \big(f_j(x) - f_{j-1}(x)\big)^2 < \infty.$$

(See [6].) Therefore, we restrict attention to $f_j(x)$, $x \in [0,1]$ and $j \ge 0$.

The basic idea of the proof is to introduce a (dyadic) stopping time $\tau = \tau_\lambda$ with the property that $\{\tau(x) = \infty\}$ (essentially) coincides with the set A_λ (or B_λ) where $f^*(x) \le \lambda$ (or $S(f) < \lambda$). Furthermore, the stopping time should be such that the stopped sequence $f_{\tau(\cdot) \wedge j}(\cdot)$ is uniformly bounded in $L^2(\mathbf{R})$. *If it were true* that the projections P_j commute with the stopping time, i.e.

$$P_j(f_{\tau \wedge (j+1)}) = f_{\tau \wedge j},$$

then the sequence $f^\tau = (f_{\tau\wedge 1}, f_{\tau\wedge 2}, \ldots)$ would satisfy

$$S^2(f)(x) = S^2(f^\tau)(x)$$

on $\{x : \tau(x) = \infty\}$. This would imply that $A_\lambda \subseteq B$ since

$$\|S(f^\tau)\|_2 = \sup_j \|f_{\tau\wedge j}\|_2 < \infty.$$

The commutation hypothesis (that P_j commute with all dyadic stopping times) is equivalent to assuming that the P_j are dyadic conditional expectations. In this case, the sequence f_j is forced to be a martingale. Since those sequences we deal with here are not martingales, we cannot prove the theorem in this way. Nevertheless, the strategy may be rescued from obvious failure. The proof is carried out by an inductive procedure that involves a number of details. In order to clarify the description, we proceed in steps.

Step 1. We now prove that $A \subseteq B$, assuming that $A \subseteq [0,1]$ and that $\text{meas}(A) > 1-\epsilon$. Here $\epsilon > 0$ is assumed to be small, subject to constraints that will become clear as the discussion unfolds. Since $f_0(x)$ is restricted to the unit interval, we may write

$$f_0(x) = \sum_{j=-N+1}^{0} a_j \phi(x-j).$$

Choose $\lambda > 0$ and define

$$A_\lambda^c := \left\{x : \sup_{j\geq 0} |f_j(x)| > \lambda\right\}.$$

Then A_λ^c satisfies $A_\lambda^c \supset A^c$, and if λ is large enough, we will have

$$\text{meas}(A_\lambda^c) \leq 2\epsilon.$$

We construct a covering of A_λ^c by dyadic intervals, defined by means of a stopping time $\gamma(x)$ relative to the dyadic multiresolution analysis D. This stopping time will be modified by subsequent considerations before we are finished.

Step 2. We wish to project the characteristic function of the set A_λ^c onto the space D_j; let g_j be this projection. The function g_j is just the average of the characteristic function of A_λ^c over each dyadic interval of length 2^{-j}. The sequence g_j, $j \geq 0$ is a dyadic martingale. Define

$$\gamma(x) = \inf\left\{j : g_j(x) \geq 1/2\right\}, \text{ where } \gamma(x) = \infty \text{ if this set is empty.}$$

By Doob's maximal inequality, applied to the sequence g_j, $j \geq 0$, we have

$$\text{meas}\left\{x : \gamma(x) < \infty\right\} \leq 2\ \text{meas}(A_\lambda^c) \leq 4\epsilon.$$

The stopping time $\gamma(x)$ determines a family of disjoint sub-intervals of $[0,1]$ defined by $\{x : \gamma(x) = j\}$. These intervals are unions of dyadic intervals of length 2^{-j}. A connected union of maximal length is called a component of $\{x : \gamma(x) = j\}$.

Step 3. Starting with level $n_1 = \inf_x \gamma(x)$, the components of $\{x : \gamma(x) = n_1\}$ are divided into three classes: The first class consists of all components of length less than $N \cdot 2^{-n_1}$. The second class consists of all components of length greater than or equal to $N \cdot 2^{-n_1}$ but less than $10N \cdot 2^{-n_1}$. The third class consists of components of length greater than or equal to $10 \cdot N \cdot 2^{-n_1}$. We modify the intervals of the second class by extending them, to the right, by a union of dyadic intervals such that the total length of the enlarged interval (the component plus the added intervals) equals $10N \cdot 2^{-n_1}$. The entire set of intervals obtained in this way may again be divided into components: (a) short components of length less than $N \cdot 2^{-n_1}$; (b) the remaining components of length greater than or equal to $10N \cdot 2^{-n_1}$. We modify the stopping time $\gamma(x)$ to obtain another one, $\tau(x)$, in the following way: (a) $\tau(x) = n_1$ for points in any component containing the point $x = 0$ or $x = 1$ if such a component exists. (b) $\tau(x) = n_1$ for points in the long components (of length greater than or equal to $10 \cdot N2^{-n_1}$). (c) On the remaining points, $\tau(x) > n_1$ and will be defined by the inductive process. To recapitulate, we have enlarged the set $\{x : \gamma(x) = n_1\}$ to obtain a new set of components. The measure of the enlarged set is no greater than nine times the measure of the components of $\{x : \gamma(x) = n_1\}$ of intermediate length. The stopping time $\tau(x) = n_1$ on the "long" components and on the components, if they exist, containing $x = 0$ or $x = 1$. The "short" components of the enlarged set are those that do not contain $x = 0$ or $x = 1$, and whose length does not exceed $N \cdot 2^{-n_1}$.

Step 4. We now continue the induction as follows: Consider the set of "short" components, just cited, together with the set $\{x : \gamma(x) = n_1 + 1\}$. We combine the short components of the set $\{x : \gamma(x) = n_1\}$ and those components of the set $\{x : \gamma(x) = n_1 + 1\}$ that do not belong to the set $\{x : \tau(x) = n_1\}$. The union of these two sets is a collection of dyadic intervals of length $2^{-(n_1+1)}$. The components of this set are sorted into three categories as before: those of length less than $N \cdot 2^{-(n_1+1)}$, those of intermediate length, from $N \cdot 2^{-(n_1+1)}$ to $10N \cdot 2^{-(n_1+1)}$, and those of length greater than or equal to $10N \cdot 2^{-(n_1+1)}$. The components of intermediate length are enlarged so that they have length exactly $10N \cdot 2^{-(n_1+1)}$. (This enlargement is made simply by adjoining adjacent dyadic intervals, of length $2^{-(n_1+1)}$, to the right of the component in question. The new component may overlap some of the set $\{x : \tau(x) = n_1\}$, as well as other components of $\{x : \gamma(x) \leq n_1 + 1\}$, not contained in $\{x : \tau(x) = n_1\}$.) We now combine the short, enlarged, and long components of this set *together with* the components of the set $\{x : \tau(x) = n_1\}$. Since the components of $\{x : \tau(x) = n_1\}$ are all longer than $10N \cdot 2^{-n_1}$, the *components of the combined set* are either longer than or equal to $10 \cdot N \cdot 2^{-(n_1+1)}$, or shorter than $N \cdot 2^{-(n_1+1)}$.

We now define $\tau(x) = n_1 + 1$ for points x in a long component of the combined set if $\tau(x)$ has not been previously defined. We also define $\tau(x) = n_1 + 1$ if x is in a component (long or short) that contains 0 or 1 and $\tau(x)$ has not been previously defined. Thus, the only components that remain are short and isolated: that is, they are of length less than $N \cdot 2^{-(n_1+1)}$ and lie in the interior of the *complement* of the long components and contain neither 0 nor 1.

Step 5. The passage from n_1 to $n_1 + 1$ is indicative of the induction procedure. At the nth stage, the set $\{x : \tau(x) = n\}$ has been defined. The short components defined by the procedure are of length less than

$N \cdot 2^{-n}$, and lie in the interior of the unit interval, separated from the long components by a distance of at least 2^{-n}. These short components are a part of the set $\{x : \gamma(x) \le n\}$.

The short components are combined with that part of the set $\{x : \gamma(x) = n+1\}$ that remains in the set $\{x : \tau(x) > n\}$ and treated as in Step 3 to define the set $\{x : \tau(x) = n+1\}$.

Step 6. We now make a penultimate adjustment in the definition of $\tau(x)$. At each stage, if the set $\{x : \tau(x) = n\}$ produces complementary intervals belonging to the set $\{x : \tau(x) > n\}$ that are of length less than $10N \cdot 2^{-n}$, we adjoin them to the set $\{x : \tau(x) = n\}$. (That is we define $\tau(x) = n$ on these intervals also.) This addition will take place only if new points were added to the set $\{x : \tau(x) < n\}$ by additional stopping. A fixed component of $\{x : \tau(x) = n\}$ can have at most two "small" complementary contiguous intervals, and the components of $\{x : \tau(x) = n\}$ are at least of length 2^{-n}. Therefore, this addition multiplies the measure of $\{x : \tau(x) = n\}$ by a factor no greater than $16N$.

Step 7. It remains for us to estimate the measure of the set $\{x : \tau(x) < \infty\}$. First of all, let us observe that $\tau(x)$, $\tau(x) < \infty$, satisfies

$$\gamma(x) \le \tau(x) \le \gamma(x) + \log_2 N$$

since any component of $\{x : \gamma(x) = n\}$ must be "long" or "intermediate" if the scale is 2^{-m}, where $N \cdot 2^{-m} \le 2^{-n}$. Finally, the enlargement procedure has been done in such a way that

$$\text{meas}\,\{x : \tau(x) < \infty\} \le 9\,\text{meas}\,\{x : \gamma(x) < \infty\} = O(\epsilon).$$

Step 8. Now let us give a preliminary definition of the "stopped" sequence $\tilde{f}_j^\tau$. As we pointed out above, the straightforward approach, where $\tilde{f}_j^\tau(x) := f_{j\wedge\tau(x)}(x)$ is not suitable. However, this procedure may be modified as follows: The sequence $f_0, f_1, \ldots, f_{n_1-1}$ is not altered. For the index n_1, we consider the intervals that are *complementary* to $\{x : \tau(x) = n_1\}$, that is, the set $\{x : \tau(x) > n_1\}$. A typical complementary interval (in $\{x : \tau(x) > n_1\}$) consists of a union of dyadic intervals of length 2^{-n_1}. Each of these dyadic intervals may be classified by the stopping time γ: on a given interval of length 2^{-n_1}, $\gamma(\cdot) \equiv n_1$ or $\gamma(\cdot) > n_1$. By construction, the extreme left and right dyadic intervals (of the entire complementary interval) are of the latter type, where $\gamma(\cdot) > n_1$. The dyadic intervals in the interior may be of either type. However, the short components of $\{x : \gamma(x) = n_1\}$ are of length less than $N \cdot 2^{-n_1}$. The complementary interval may, of course, contain many such components, separated from each other by dyadic intervals where $\gamma(\cdot) > n_1$. Finally, the complementary interval contains neither endpoint $x = 0$ nor $x = 1$.

On a fixed complementary interval the function $f_{n_1}(x)$ may be represented as a finite sum $\tilde{f}_{n_1}^\tau$,

$$\tilde{f}_{n_1}^\tau(x) := \sum_{k=\ell}^{r} a_k \phi(2^{n_1}x - k).$$

Here we assume that the complementary interval has left endpoint $(\ell + N - 1) \cdot 2^{-n_1}$ and right endpoint $(r+1) \cdot 2^{-n_1}$, so that the above is the "shortest" representation of $f_{n_1}(x)$ on the complementary interval.

The sum does not necessarily give the value of $f_{n_1}(x)$ outside the complementary interval, nor are we assured that the difference

$$\begin{aligned}\tilde{d}^{\tau}_{n_1}(x) &= \tilde{f}^{\tau}_{n_1}(x) - \tilde{f}_{n_1-1}(x) \\ &= \tilde{f}^{\tau}_{n_1}(x) - f_{n_1-1}(x))\end{aligned}$$

belongs to the space W_{n_1-1}. However, we can assert that

$$\max_{k \leq n_1} |\tilde{f}^{\tau}_k(x)| \leq O(\lambda)$$

for all $x \in \mathbf{R}$. This is true because of the (M-Z) condition. The argument is as follows: For $f^{\tau}_k = f_k,\ k < n_1$, on *every* dyadic interval of length 2^{-k}, the (closed) subset of points $\{x : \sup_{j \geq 0} |f_j(x)| \leq \lambda\}$ contained in the dyadic interval has a proportion that is greater than 1/2. The (M-Z) condition then guarantees that

$$|f_k(x)| = O(\lambda)$$

uniformly in the dyadic interval. (See the "Remarks" after the statement of (M-Z).) The argument for $\tilde{f}^{\tau}_{n_1}$ ($= f_{n_1}$ on the complementary interval) is similar but a bit more complicated. The function $|\tilde{f}^{\tau}_{n_1}(x)| = O(\lambda)$ on any interval where $\gamma(x) > n_1$ for the reason just stated. On the stretches (short components), of length less than $N \cdot 2^{-n_1}$, where $\gamma(x) = n_1$, we cannot apply this argument directly. However, the values of $\tilde{f}^{\tau}_{n_1}(\cdot)$ in this stretch are majorized by a constant times the sum of the moduli of all coefficients a_k that enter into the representation

$$\tilde{f}^{\tau}_{n_1}(x) = \sum_{k=\ell}^{r} a_k \phi(2^{n_1}x - k)$$

on this stretch. Because the stretch is "short," the translates $\{k\}$ specific to this stretch also appear in the representation of the function $\tilde{f}^{\tau}$ on the dyadic intervals of length 2^{-n_1}, I_0 and I_1, that bound the stretch. (Some translates are associated with I_0, some associated with I_1.) On each interval I_i, $i = 0, 1$, we know that $\gamma(\cdot) > n_1$, so that $|\tilde{f}^{\tau}_{n_1}(x)| = O(\lambda)$ on these intervals. By the (M-Z) condition, the sum of the moduli of the corresponding coefficients is of the same order. On the stretch, $|\tilde{f}^{\tau}_{n_1}(x)|$ is also majorized, up to a constant, by the sum of all of these coefficients. Therefore, we may conclude that $|\tilde{f}^{\tau}_{n_1}(x)| = O(\lambda)$ on the entire complementary interval. The constants depend only on N, the bound on $|\phi(x)|$, and the constants from the (M-Z) condition.

Outside the complementary interval, we claim that $\tilde{f}^{\tau}_{n_1}(x)$ satisfies the same estimate. In fact, the support of any sum

$$\tilde{f}^{\tau}_{n_1}(x) = \sum_{k=\ell}^{r} a_k \phi(2^{n_1}x - k)$$

is contained in an interval consisting of the complementary interval, together with two intervals, to the left and right, of the complementary interval. Since $\phi(2^{n_1}x)$ has support on $[0, N \cdot 2^{-n_1}]$, the additional intervals are at most of length $(N-1)2^{-n_1}$. However, we know that the intervals where $\tau(x) = n_1$ are of length at least $10N \cdot 2^{-n_1}$, so the supports of the sums defining $\tilde{f}^{\tau}_{n_1}$ are disjoint. Since $|\tilde{f}_{n_1}(x)| = O(\lambda)$ on each complementary interval, the (M-Z) condition implies that this estimate holds on the entire support. Therefore, $|\tilde{d}^{\tau}_{n_1}| = O(\lambda)$ also.

Step 9. As noted above, the difference

$$\tilde{d}^{\tau}_{n_1}(x) = \tilde{f}^{\tau}_{n_1}(x) - \tilde{f}^{\tau}_{n_1-1}(x)$$

belongs to V_{n_1}, but not necessarily to W_{n_1-1}, the orthogonal complement of V_{n_1-1} in V_{n_1}, even though it represents d_{n_1} on each complementary interval. To remedy this, our strategy will be to obtain an expression for the difference $\tilde{d}^{\tau}_{n_1}(x)$ in terms of the prewavelets $\psi(2^{n_1}x - 2k)$, and to estimate the magnitude of the new representation. The stopping time τ will be altered again to obtain another stopping time ρ, such that the measure of $\{x : \rho(x) < \infty\}$ is larger than that of $\{x : \tau(x) < \infty\}$ by a fixed multiple. With this stopping time, the difference $f^{\rho}_{n_1} - f^{\rho}_{n_1-1}$ will be contained in W_{n_1-1}. The procedure will then be carried out for $n \geq n_1$, and we will be able to estimate the quantity $\sup_n \|f^{\rho}_n\|_2$.

On each complementary interval, $\tilde{d}^{\tau}_{n_1}(x)$ has the representation

$$\tilde{d}^{\tau}_{n_1}(x) = \sum_{k=\ell}^{r} e_k \phi(2^{n_1}x - k),$$

with ℓ, r varying from interval to interval. In fact, $\tilde{d}^{\tau}_{n_1}$ is obtained by "censoring" the coefficients in $f_{n_1} - f_{n_1-1}$, retaining only those e_k where $\phi(2^{n_1} - k)$ has some support in a complementary interval. Although $f_{n_1} - f_{n_1-1}$ belongs to W_{n_1-1}, the censored version $\tilde{d}^{\tau}_{n_1}$ does not necessarily have this property. We can, however project $\tilde{d}^{\tau}_{n_1}$ onto W_{n_1-1}, more specifically, we can project $\tilde{d}^{\tau}_{n_1}$ onto the subspace of W_{n_1-1} spanned by $\phi(2^{n_1}x - k)$, $\ell - N + 1 \leq k \leq r - 2N + 1$. This subspace is itself a subspace of the span of $\phi(2^{n_1}x - k)$, $\ell \leq k \leq r$. We can express the projection of $\tilde{d}^{\tau}_{n_1}$ onto W_{n_1-1} as follows:

$$d^{\tau}_{n_1}(x) = \sum_{k=\ell+N-1}^{r-2N+1} b_k \psi(2^{n_1}x - k).$$

The coefficients b_k are linear combinations of the coefficients e_k according the "sub-band filtering algorithm": each b_k is obtained by taking a moving average of the e_k, using the high pass filter coefficients c_k that specify $\psi(x)$, and then downsampling the resulting sequence. (This procedure is described in Chui [2], pages 19-20, for example.) We observe some features of d^{τ}_n: (a) the estimate $|e_k| = O(\lambda)$ implies that $|b_k| = O(\lambda)$ by what we have just described. This implies that $|d^{\tau}_{n_1}(x)| = O(\lambda)$; (b) the function $d^{\tau}_{n_1}$ is the restriction of $d_{n_1}(x)(= f_{n_1}(x) - f_{n_1-1}(x)) = \sum_{k=-\infty}^{\infty} b_k \psi(2^{n_1}x - k)$ to a finite series where the indices k belong to the interval $\ell + N - 1 \leq k \leq r - 2N + 1$; (c) the function

$$d^{\tau}_{n_1}(x) = f^{\tau}_{n_1}(x) - f^{\tau}_{n_1-1}(x)$$

on a *sub*interval of any complementary interval extending from $(\ell+4N-3)\cdot 2^{-n_1}$ to $(r-3N+1)2^{-n_1}$, an interval of length $\big((r-\ell) - 7N + 4\big)2^{-n_1}$. Recall that $\tilde{d}^{\tau}_{n_1} = d_{n_1}$ on a complementary interval, but not necessarily on the support of $\tilde{d}^{\tau}_{n_1}$. On intervals of length $(N-1)2^{-n_1}$ to the left and right of a complementary interval, it may happen that $\tilde{d}^{\tau}_{n_1}(x) \neq d_{n_1}(x)$ because of the restricted summation. A particular complementary interval (where $\tilde{d}^{\tau}_{n_1} = d_{n_1}$ is guaranteed) then has length $\big((r+1)-(\ell+N-1)\big)\cdot 2^{-n_1} = \big((r-\ell)-N+2\big)2^{-n_1}$. However, when we project $\tilde{d}^{\tau}_{n_1}$ onto W_{n_1-1}, we are assured that the projection $d^{\tau}_{n_1} = d_{n_1}$ on a subinterval of the complementary interval extending from $(\ell+4N-3)2^{-n_1}$ to $(r-3N+2)2^{-n_1}$. (We wish to remind the reader again that $d^{\tau}_{n_1}(x) = d_{n_1}(x)$ on an interval that begins with the index of the *last* interval in support of the *first* $\psi(2^{n_1}x - k_\ell)$ where $k_\ell = \ell + N - 1$, and extends to and includes the *first* interval in the support of the *last* $\psi(2^{n_1}x - k_r)$ where $k_r = r - 2N + 1$. Thus, the indices of the intervals are, respectively, $\ell + (N-1) + (2N-1) + (N-1) = \ell + 4N - 3$, and $r - (2N-1) - (N-1) = r - 3N + 2$. The length of the interval with these indices is $(r-\ell) - 7N + 4$, as pointed out above.)

All of this means that the sum

$$d^{\tau}_{n_1}(x) = \sum_{k=\ell}^{r} b_m \psi(2^{n_1}x - m)$$

represents $d_{n_1}(x)$ for all x in the complementary interval except possibly at the two extremes of the interval: we must exclude intervals no longer than $(3N+4)2^{-n_1}$ on the left and an interval of length $(3N)2^{-n_1}$, on the right. The measure of the exceptional points $\{x : d^{\tau}_{n_1}(x) \neq d_{n_1}(x)\}$ is the sum of contributions from each complementary interval. Each such contribution is at most $8N \cdot 2^{-n_1}$, and as such, the total measure of the exceptional set is less than at most $8N$ times the measure of $\{x : \tau(x) = n_1\}$. We use these exceptional intervals to define a modification ρ of the stopping time τ. Each component of $\{x : \tau(x) = n_1\}$ is expanded to include an interval of length $(3N+4)2^{-n_1}$ on the right (the left extreme of the complementary interval), and an interval of length $(3N) \cdot 2^{-n_1}$ on the left (the right extreme of the complementary interval). Since each complementary interval is of length at least $(10N)2^{-n_1}$, we have diminished, but not

eliminated any complementary interval. The expanded component becomes a component of $\{x : \rho(x) = n_1\}$.

Step 10. We repeat this procedure for each $n > n_1$. On each component of the set $\{x : \tau(x) > n\}$, the analysis made in Steps 7, 8 and 9 may be applied. The functions $\tilde{f}^{\tau}_n(x)$ and $d^{\tau}_n(x)$ satisfy the same estimates. That is, $|\tilde{f}^{\tau}_n(x)| = O(\lambda)$ and $|d^{\tau}_n(x)| = O(\lambda)$.

Now let us write

$$f^{\tau}_n(x) = f_0(x) + \sum_{k=1}^{n} d^{\tau}_k(x).$$

The functions $d^{\tau}_n(x) = d_n(x)$ except on a set of measure comparable to the measure of the set $\{x : \tau(x) \leq n\}$. We prove this by induction, as follows: For $n = n_1$, we have shown that $d^{\tau}_{n_1}(x) = d_{n_1}(x)$ except on the set $\{x : \rho(x) = n_1\}$. Now fix a complementary interval of the set $\{x : \tau(x) = n_1\}$ (which contains a complementary interval of $\{x : \rho(x) = n_1\}$). There are two cases to consider: (i) The set $\{x : \tau(x) = n_1 + 1\}$ does not intersect the fixed complementary interval. (ii) The set $\{x : \tau(x) = n_1 + 1\}$ does intersect, creating a smaller complementary interval (or intervals). In case (i), the difference $d^{\tau}_{n_1+1}(x) = d_{n_1+1}(x)$ except on intervals of length $(3N)2^{-(n_1+1)}$ and $(2N)2^{-(n_1+1)}$ on the left and right extremes of the complementary interval. However, these exceptional intervals are already contained in the complement of $\{x : \rho(x) = n_1\}$, so that no adjustment is needed in this case. The same is true for any $n > n_1$ as long as the set $\{x : \tau(x) = n\}$ does not intersect the complementary interval of $\{x : \tau(x) = n_1\}$. In other words, $d_n^\tau(x) = d_n(x)$ except on the set $\{x : \rho(x) = n_1\} \cap \{x : \tau(x) > n\}$.

In case (ii), the set $\{x : \tau(x) = n_1 + 1\}$ intersects the complementary interval of $\{x : \tau(x) = n_1\}$, creating a complementary interval (or intervals) to the set $\{x : \tau(x) \leq n_1 + 1\}$. In this case, we expand the components of $\{x : \tau(x) = n_1 + 1\}$ on the left and right by $(3N)2^{-(n_1+1)}$ and $(3N + 4)2^{-(n_1+1)}$, respectively, if these components disconnect the complementary interval of $\{x : \tau(x) = n_1\}$. If the component of $\{x : \tau(x) = n_1 + 1\}$ does not disconnect (e.g. it falls, for example, at the right end of the complementary interval) we expand unilaterally (on the left, in the example). This expansion process creates a new set of exceptional points $\{x : \rho(x) = n_1 + 1\}$, whose measure is comparable to the measure of $\{x : \tau(x) = n_1 + 1\}$.

We continue this process for all $n > n_1$, and so define the stopping time ρ.

Step 11. Now we "stop" the sequence f_n^τ, $n \geq 0$, using the stopping time ρ in the same way done in Step 8. At the level n, the complementary intervals of $\{x : \rho(x) \leq n\}$ are isolated and for each interval we define

$$f_n^\rho(x) = \sum_{j=0}^{n} \sum_{s=\ell}^{r} b_{j,s} \psi(2^j x - s).$$

As in Step 8 we take the shortest sum, so that $f_n^\rho(x) = f_n(x)$ on the complementary interval. The coefficients b_{2s} are, of course, the same as those in the expansion of d_n^τ. The new differences are denoted d_n^ρ.

We may now estimate the $\sup_n \|f_n^\rho\|_2$ as follows: On the set $\{x : \rho(x) = \infty\} \subset \{x : \tau(x) = \infty\}$,

$$(f^\rho)^*(x) = (\tilde{f}^\tau)^*(x) = O(\lambda),$$

as indicated in Step 10.

Therefore it remains for us to estimate $(f^\rho)^*(x)$ on the set $\{x : \rho(x) < \infty\}$. This set is a union of disjoint sets $\{x : \rho(x) = n\}$, $n = n_1, n_1 + 1, \ldots$; we estimate $(f^\rho)^*$ on each of these sets. Each set $\{x : \rho(x) = n\}$ is itself a union of components I_n^j, each of which is at least of length 2^{-n}. (Recall that a component of $\{x : \rho(x) = n\}$ contains a component of $\{x : \tau(x) = n\}$ of length less than $N \cdot 2^{-n}$ that is either contiguous with a component of $\{x : \tau(x) \leq n - 1\}$, of length greater than $10 \cdot N \cdot 2^{-(n-1)}$, or is an "endpoint interval," one that contains zero or one. Since we are restricting attention to $[0, 1]$, the endpoint components may be considered to have infinite length. Otherwise, any component of $\{x : \tau(x) = n\}$ (and $\{x : \rho(x) = n\}$) has length greater than or equal to $10 \cdot N \cdot 2^{-n}$.) By construction,

$$\max_{k<n} |f_k^\rho(x)| = O(\lambda),$$

so that it is necessary to estimate the magnitude of sums

$$\left| \sum_{k=n}^{m} d_k^\rho(x) \right|, \qquad m = n, n + 1, \ldots$$

on a fixed component of $\{x : \rho(x) = n\}$. Each of the differences

$$\left|d^{\rho}_{n+k}\right| = O(\lambda), \qquad k = 0, 1, \ldots.$$

Now consider that part of the support of d^{ρ}_{n+k} that lies in the component I^j_n of $\{x : \rho(x) = n\}$ under examination.

(a) If $k > 4\log_2 N$ (so that $N \cdot 2^{-(n+k)} < 2^{-n}$), then, that part of the support of d^{ρ}_{n+k} that intersects the component interval I^j_n, is contained in at most two disjoint intervals, each of length $N \cdot 2^{-(n+k)}$, lying at either end of the component interval I^j_n. Thus, the supports of d^{ρ}_{n+k} form a decreasing sequence of sets within each component of $\{x : \rho(x) = n\}$. Therefore, we may majorize

$$\sum_{k > \log_2 4N} \left|d^{\rho}_{n+k}(x)\right|$$

in terms of the relative distance function $\Delta^j_n(x)$, defined on each component I^j_n of $\{x : \rho(x) = n\}$. Let $|I^j_n|$ be the length of I^j_n and

$$\Delta^j_n(x) = \frac{\text{distance}\,(x,\ \text{complement of } I^j_n)}{|I^j_n|}.$$

The above considerations lead us to the estimate

$$\sum_{k > \log_2 4N} |d^{\tau}_{n+k}(x)| = O(\lambda)|\log_2 \Delta^j_n(x)|$$

for $x \in I^j_n$.

(b) If $k \le \log_2 4N$, the best estimate is simply

$$|d^{\rho}_{n+k}(x)| = O(\lambda).$$

With these estimates, we may estimate $\|(f^{\rho})^*\|_2$ as follows: On each component I^j_n,

$$\begin{aligned} |(f^{\rho})^*(x)|^2 &\le O(\lambda^2) + \Bigg(\sum_{k \ge n} |d^{\rho}_k(x)|\Bigg)^2 \\ &\le O(\lambda^2)\big[1 + \log_2^2(N) + \log_2^2 \Delta^j_n(x)\big]. \end{aligned}$$

Therefore

$$\int_{I^j_n} \left|(f^{\rho})^*(x)\right|^2 dx = O(\lambda^2)|I^j_n| + O(\lambda^2)\int_{I^j_n} \log_2^2 \Delta^j_n(x) dx.$$

The last integral is estimated by the quantity

$$\left(\int_0^1 \log_2^2 |x|^{-1} dx\right)|I^j_n|.$$

If we sum these estimates over j and n, we obtain

$$\begin{aligned} \int |(f^{\rho})^*(x)|^2 dx &= \int_{\{\rho(x) < \infty\}} + \int_{\{\rho(x) = \infty\}} |(f^{\rho})^*(x)|^2 \\ &= O(\lambda^2). \end{aligned}$$

Step 12. The "stopped" sequence $f_n^\rho(x)$, $n \geq 0$, agrees with the original sequence $f_n(x)$, $n \geq 0$ except for points in a set of measure comparable to the measure of $\{x : \rho(x) < \infty\}$. This set has small measure (less than $O(\epsilon)$). Furthermore, the estimate given in the previous step shows that $\|(f^\rho)^*\|_2 = O(\lambda)$, and f_n^ρ, $n \geq 0$ satisfies the property $P_n f_{n+1}^\rho = f_n^\rho$. Therefore

$$\begin{aligned}\|S(f^\rho)\|_2 &= \sup_n \|f_n^\rho\|_2 \\ &\leq \|(f^\rho)^*\|_2 \\ &= O(\lambda).\end{aligned}$$

Since $S(f^\tau)(x) = S(f)(x)$ on the set where $f_n^\rho(x) \equiv f_n(x)$, $n = 0, 1, \ldots$ we may conclude that $\text{meas}(A \setminus B) = O(\epsilon)$. Furthermore, it is known that L^2-bounded wavelet expansions converge a.e. in this context (see [7]), so that $\text{meas}(A \setminus C) = O(\epsilon)$.

Step 13. The entire procedure may be repeated using B as the initial set. The conclusion of this argument is $\text{meas}(B \setminus A) = O(\epsilon)$, and the proof of the Theorem is now complete.

Bibliography

1. D. L. Burkholder, R. F. Gundy, Extrapolation and interpretation of quasilinear operators on martingales, *Acta Math.* **124** (1970), 250-304.

2. C. K. Chui, *Introduction to Wavelets*, Academic Press, Inc., New York, N.Y., 1992.

3. R. F. Gundy, Martingale theory and the pointwise convergence of certain orthogonal series, *Trans. Amer. Math. Soc.* **124** (2) (1966), 228-248.

4. R. F. Gundy, The martingale version of a theorem of Marcinkiewicz and Zygmund, *Ann. Math. Statist.* **38** (1967), 725-734.

5. F. G. Arutyunyan, On the series in the Haar system, *Akad. Nauk. Armyan. SSR Dokl.* **42** (1966), 134-140.

6. E. Hernández, G. Weiss, *A First Course on Wavelets*, CRC Press, Boca Raton FL, 1996.

7. S. E. Kelly, M. A. Kon, L. A. Raphael, Pointwise convergence of wavelet expansions, *Bull. Amer. Math. Soc.* **30** (1) (1994), 87-94.

8. J. Marcinkiewicz, A. Zygmund, Sur les fonctions indépendantes, *Fund. Math.* **29** (1937), 60-90.

R. Gundy
Department of Statistics
Rutgers University
New Brunswick, NJ 08903
email: gundy@rci.rutgers.edu

K. Kazarian
Department of Mathematics
Universidad Autonoma de Madrid
Madrid, Spain
email: kazaros.kazarian@uam.es

Stability Properties for a Compactly Supported Prescale Function

V. Dobrić, R. F. Gundy, and P. Hitczenko*

Lehigh University, Rutgers University, and North Carolina State University

Key Words and Phrases: Wavelets, linear independence, local convergence

AMS 1991 Subject Classification: 42C15

*P. Hitczenko would like to acknowledge the National Science Foundation for partial support of this research, which was carried out while he was visiting Rutgers University. In particular, he would like to thank R. F. Gundy, R. Wilson, and the Departments of Mathematics, and Statistics, for their hospitality.

Abstract. We show that if ϕ is a continuous, minimally supported prescale function, then its translates are linearly independent on any set of positive measure in the unit interval. This generalizes results of Y. Meyer and P. G. Lemarié.

This result implies that a stability condition, introduced by Gundy and Kazarian for the study of local convergence of spline wavelet expansions, is satisfied for all expansions arising from multiresolution analyses generated by such prescale functions ϕ.

1. Introduction. In [7], P. G. Lemarié proved that if a multiresolution analysis contains a compactly supported function, then it contains a minimal (pre)scale function. More specifically, there exists a function ϕ of compact support such that

(1) the integer translates, $\phi(\cdot - k),\ k \in \mathbf{Z}$, are a Riesz basis for the space V_0;

(2) every function in V_0 that is compactly supported may be written as a finite linear combination of translates of ϕ.

(Throughout this paper, we will use the term "scale function" to mean the above, although some authors refer to this as "prescale function.") The most basic examples of minimal scale functions of this type are the B-splines, and the compactly supported scale functions constructed by I. Daubechies [1]. An important property of these minimal-scale functions was first proved by Y. Meyer [10] for Daubechies' functions, and subsequently stated by Lemarié [7] in the general case: *The translates of ϕ, restricted to the unit interval, form a linearly independent set.*

The purpose of this paper is to prove the following stronger version of the above for a minimal scale function ϕ that is continuous on the unit interval: the translates of ϕ are linearly independent over any subset of positive measure contained in the unit interval. The stronger version is of interest because it may be used to obtain a local convergence theorem for multiresolution analyses with continuous minimal scale functions. The first version of this type of local convergence theorem was proved by Gundy and Kazarian [4] for a class of wavelet expansions that include the spline wavelets. Their theorem assumed a stability condition (condition (M-Z) of [4]). It turns out that this stability condition is, in fact, a property of all multiresolution analyses with continuous minimal scale functions, as a consequence of the above strong linear independence of these functions.

2. Notation. We suppose that a multiresolution analysis is given. That is, we have a sequence of subspaces of $L^2(\mathbf{R})$, $V_j,\ j \in \mathbf{Z}$, such that $V_j \subset V_{j+1}$, and $f(\cdot) \in V_j$ iff $f(2^{-j}\cdot) \in V_0$. Furthermore, we are given a function $\phi \in V_0$ such that the integer translates $\phi(\cdot - k)$ form a Riesz basis for V_0: any function $f(\cdot) \in V_0$ has a representation

$$f(x) = \sum a_k \phi(x - k)$$

with

$$\sum a_k^2 \cong \|f\|_2^2 \quad \text{i.e.,} \quad c \sum a_k^2 \leq \|f\|_2^2 \leq C \sum a_k^2, \quad 0 < c < C.$$

If ϕ has a nonzero integral, then it follows that the increasing sequence of subspaces exhausts $L^2(\mathbf{R})$ (see [5, Chapter 2]). Let P_j be the orthogonal projection operator from L^2 onto V_j.

Now we impose another restriction on the multiresolution analysis. We require that the space V_0 contain a nontrivial continuous function that is compactly supported. With this additional assumption, the techniques of Lemarié [7] may be used to show that there exists a minimally supported, real-valued, continuous function $\phi \in V_0$ such that every compactly supported function in V_0 admits a representation as a finite linear combination of integer translates of ϕ. If we agree to normalize ϕ by setting its integral equal to one, then ϕ is unique, up to integer translates. (Our class of multiresolution analyses does include the spline wavelets, the compactly supported Daubechies wavelets, and those obtained from these classes by integration, as indicated in Lemarié [7].)

3. Linear Independence of Translates. In this section, we state the theorem on linear independence.

THEOREM 1. *Let ϕ be a continuous, minimal (pre)scale function supported on the interval* $[0, N]$. *Then the translates $\phi(\cdot + k),\ k = 0, \ldots, N-1$ are linearly independent over any set of positive measure of the unit interval.*

Remarks. As we noted above, this line of investigation was initiated by Y. Meyer [10] and pursued by P. G. Lemarié in [7]. These authors treated the case where the "set of positive measure" was the entire unit interval. Lemarié and Malgouyres [8] gave another simplified proof that showed that the translates were linearly independent on any subinterval of the unit interval. Finally, Lemarié [7] showed that this property characterizes minimal scale functions. Those authors made no continuity assumptions.

Proof. We give a proof by contradiction as follows: If the translates are linearly dependent over a set of positive measure, we show that they are dependent over a set of measure one in the unit interval. Since the function ϕ is continuous, this means that the translates of ϕ are dependent over the unit interval itself, thus contradicting the theorem of Meyer.

Throughout the proof, we will use matrices $\mathbf{P_0}$ and $\mathbf{P_1}$. To define these matrices, let us write the dilation equation for ϕ as

$$\phi(x/2) = \sum_{k=0}^{N} p_k \phi(x-k), \qquad \text{with} \ \ p_0, p_N \neq 0.$$

With this notation, let us define the $(N-1) \times (N-1)$ matrix $\mathbf{P}$ whose first row consists of the vector of odd numbered coefficients, p_{2k+1}, followed by the appropriate number of zeros to give the vector $N-1$ components. The second row of $\mathbf{P}$ is defined in the same way, using the even numbered coefficients, p_{2k}, followed by the appropriate number of zeros. Third and fourth rows are obtained from the first two rows by a cyclic permutation of the indices: each entry is shifted to the right, with the final entry, a zero, moving to first position. This procedure is continued until $N-1$ rows are obtained. (Thus if $N = 2k$, the second row will contain the $k+1$ entries $p_0, p_2, \ldots, p_{2k}$ followed by $k-2$ zeros. The last row will contain $k-1$ zeros followed by the k coefficients $p_1, p_3, \ldots, p_{2k-1}$. If $N = 2k+1$, then the last row of the matrix consists of

$k-1$ zeros, followed by the $k+1$ entries $p_0, p_2, \ldots, p_{2k}$.) Now define the two $N \times N$ matrices

$$\mathbf{P_0} = \begin{pmatrix} p_0 & \mathbf{p_t} \\ 0 & \mathbf{P} \end{pmatrix} \qquad \text{and} \qquad \mathbf{P_1} = \begin{pmatrix} \mathbf{P} & 0 \\ \mathbf{p_b} & p_N \end{pmatrix}$$

where $\mathbf{p_t}$ is the $N-1$ vector consisting of the even numbered p_k, starting with p_2, followed by the appropriate number of zeros; $\mathbf{p_b}$ consists of zeros followed by the coefficients p_k where k has the same parity as N, where the final entry of the vector $\mathbf{p_b}$ is the coefficient p_{N-2}.

The roles of $\mathbf{P_0}$ and $\mathbf{P_1}$ are as follows: consider a general linear combination of translates $\sum c_k\phi(x+k)$. If we take account of the fact that ϕ is supported on $[0, N]$ and restrict attention to $x \in [0, 1]$, this sum is, in fact, finite and may be expressed as $\sum_{k=0}^{N-1} c_k\phi(x+k)$. If we apply the dilation equation to express each $\phi(\cdot + k)$ in terms of a sum of translates of $\phi(2\cdot)$, the resulting double sum is a certain linear combination of translates of $\phi(2x)$ and $\phi(2x-1)$, depending on whether x is in $[0, \frac{1}{2}]$ or in $[\frac{1}{2}, 1]$. The coefficients of this linear combination are given by the matrices $\mathbf{P_0}$ or $\mathbf{P_1}$, acting on the vector $\mathbf{c} = (c_0, \ldots, c_{N-1})$. (These matrices are implicit in the reconstruction-decomposition schemes in the wavelet literature, and appear explicitly, in the 3×3 case in Daubechies [2, section 7.2].) Let $\mathbf{\Phi}(x) = \big(\phi(x), \phi(x+1), \ldots, \phi(x+N-1)\big)^t$ for $x \in [0, 1]$, and $\epsilon_k(x)$, $k = 1, 2, \ldots$ be the digits in the binary expansion of x. That is, $x = \sum \epsilon_k/2^k$, with $\epsilon_k = 0$ or 1, $k \geq 1$. Let T be the plus-one shift on the ϵ-sequence: $T : (\epsilon_1, \epsilon_2, \ldots) \to (\epsilon_2, \epsilon_3, \ldots)$. We write $Tx = \sum_{k=1}^{\infty} \epsilon_{k+1}/2^k$. We summarize the above in the following lemma.

LEMMA 1. *For* $\mathbf{c} = (c_0, c_1, \ldots, c_{N-1})$ *we have*

$$\mathbf{c} \circ \mathbf{\Phi}(x) = (\mathbf{P}_{\epsilon_1}\mathbf{c}^t) \circ \mathbf{\Phi}(Tx).$$

More generally, for any m,

$$\mathbf{c} \circ \mathbf{\Phi}(x) = (\mathbf{P}_{\epsilon_1} \cdots \mathbf{P}_{\epsilon_m}\mathbf{c}^t) \circ \mathbf{\Phi}(T^m x).$$

Proof. Recall that the support of $\phi(\cdot + m)$ is the interval $[-m, -m+N]$. For $x \in [0, 1]$,

$$\begin{aligned} \mathbf{c} \circ \mathbf{\Phi}(x) &= \sum_{k=0}^{N-1} c_k\phi(x+k) = \sum_k c_k \Big(\sum_j p_j \phi\big(2(x+k) - j\big)\Big) \\ &= \sum_m \Big(\sum_k c_k p_{2k-m} \Big) \phi(2x+m) = \sum_m \Big(\sum_k c_k p_{2k-m} \Big) \phi(Tx + \epsilon_1 + m). \end{aligned}$$

The inner sum is taken over all k with the provision that $p_{2k-m} = 0$ if $2k - m$ is not one of the integers $0, 1, \ldots, N$. The outer sum with index m changes according to whether $0 < 2x \leq 1$ or $1 < 2x \leq 2$, due to the support condition mentioned above. In the first case, when $\epsilon_1 = 0$, we have $0 \leq m \leq N-1$; in the second case, when $\epsilon_1 = 1$, $0 \leq m+1 \leq N-1$. Thus, the transformation takes two forms, with matrices $\mathbf{P_0}$ and $\mathbf{P_1}$. The rest of Lemma 1 is easily proved by induction.

Fix $\mathbf{c} = (c_0, \ldots, c_{N-1})$ and let $K_\mathbf{c} = \{x : \mathbf{c} \circ \mathbf{\Phi}(x) = 0\}$. The continuity of $\mathbf{\Phi}(x)$ implies that $K_\mathbf{c}$ is closed. ¿From now on, we assume that $\mathbf{c} \neq 0$, and that $K_\mathbf{c}$ has positive measure in $[0, 1]$; we seek to contradict Meyer-Lemarié theorem [10], [7]. There are two cases to consider.

Case 1. There exists a finite sequence $\mathbf{P}_{\epsilon_\mathbf{k}}$, $k = 1, \ldots, m$ such that $\mathbf{P}_{\epsilon_\mathbf{1}} \cdots \mathbf{P}_{\epsilon_\mathbf{m}} \mathbf{c}^t = 0$. If this is the case, we have our contradiction since $\mathbf{c} \circ \mathbf{\Phi} \equiv 0$ on the dyadic interval

$$\{x : \epsilon_1(x) = \epsilon_1, \ldots, \epsilon_m(x) = \epsilon_m\}.$$

Case 2. The vector $\mathbf{c}$ is such that $\mathbf{P}_{\epsilon_\mathbf{1}} \cdots \mathbf{P}_{\epsilon_\mathbf{m}} \mathbf{c}^t \neq 0$ for every finite sequence $\epsilon_1, \ldots, \epsilon_m$. In this case, we say that $\mathbf{c}$ is a "never zero" vector.

LEMMA 2. *Let* $\mathbf{c}$ *be a never zero vector. Then, for every* η, $0 < \eta < 1$, *there exists a never zero vector* $\mathbf{b}$ *such that* $m(K_\mathbf{b}) > 1 - \eta$.

Proof. Since $K_\mathbf{c}$ has positive measure, we can find a dyadic interval $I_j = \{x : \epsilon_1(x) = \epsilon_1, \ldots, \epsilon_j(x) = \epsilon_j\}$ such that $m(K_\mathbf{c} \cap I_j)/2^{-j} > 1 - \eta$. This is a consequence of the fact that every Lebesgue measurable set may be approximated by a finite union of dyadic intervals. Set $\mathbf{b} = \mathbf{P}_{\epsilon_\mathbf{1}} \cdots \mathbf{P}_{\epsilon_\mathbf{j}} \mathbf{c}^t$. Then $T^j(K_\mathbf{c} \cap I_j) \subseteq K_\mathbf{b}$ by Lemma 1. The set $K_\mathbf{b}$ has measure greater than $1 - \eta$ since $m\big(\{x : x = T^j y \text{ for some } y \in I_j\}\big) = 1$. This proves Lemma 2.

Now set

$$\mathcal{A}_\mathbf{c} = \left\{ \mathbf{a} \in \mathbf{R}^N : \mathbf{a} = \frac{\mathbf{b}}{\|\mathbf{b}\|_2} \text{ for some } \mathbf{b} = \mathbf{P}_{\epsilon_\mathbf{1}} \cdots \mathbf{P}_{\epsilon_\mathbf{j}} \mathbf{c}, \quad j \in \mathbf{Z} \right\}.$$

By Lemma 2, we have

$$\sup_{\mathbf{a} \in \mathcal{A}_\mathbf{c} m(K_\mathbf{a}) = 1} .$$

Now we claim that there is an $\mathbf{a} \in \mathbf{R}^N$ with $\|\mathbf{a}\|_2 = 1$ such that $m(K_\mathbf{a}) = 1$. To this end, we topologize the class $\mathcal{K}$ of all non-empty compact subsets of $[0,1]$ with the Hausdorff metric ρ:

$$\rho(A, B) = \sup_{x \in [0,1]} |d(x, A) - d(x, B)|$$

where $d(x, D) = \inf\{|x - y| : y \in D\}$. This metric is equivalent to

$$\sigma(A, B) = \inf\{\epsilon > 0, A \subset V_\epsilon(B) \text{ and } B \subset V_\epsilon(A)\}$$

where $V_\epsilon(D) = \{z \in [0,1] : d(z, D) < \epsilon\}$. It is known that $(\mathcal{K}, \rho)$ is a compact metric space. A complete discussion of these facts may be found in Kornum [6, section 6.2].

Let $\{K_{\mathbf{a}_n}\}$ be a sequence of sets in $\mathcal{K}$ such that $m(K_{\mathbf{a}_n})$ tends to one. ¿From this sequence we may extract a convergent subsequence $\{K_{\mathbf{a}_{n_k}}\}$ with limit K. Since $\|\mathbf{a}_{n_k}\|_2 = 1$, we may extract a convergent subsequence of $\mathbf{a}_{n_k}$, call it $\{\mathbf{a}_n\}$, so that finally, we obtain sequences $K_{\mathbf{a}_n} \to K$ and $\mathbf{a}_n \to \mathbf{a}$. Since $E \to m(E)$ is an uppersemicontinuous function on $(\mathcal{K}, \rho)$, that is if $E_n \to E$ then $\overline{\lim}\, m(E_n) \leq m(E)$, it follows that $m(K) = 1$. Second, we claim that $K \subset K_\mathbf{a}$. Since $K_{\mathbf{a}_n} \to K$, we have

$$\sup_{y \in [0,1]} |d(y, K) - d(y, K_{\mathbf{a}_n})| \to 0.$$

Therefore, if $x \in K$, $d(x, K_{\mathbf{a}_n}) \to 0$. Choose $x_n \in K_{\mathbf{a}_n}$ so that $|x - x_n| = d(x, K_{\mathbf{a}_n})$. Then, by the Cauchy-Schwarz inequality and the definition of $K_{\mathbf{a}}$,

$$\begin{aligned}|\mathbf{a} \circ \Phi(x)| &\le |(\mathbf{a} - \mathbf{a}_n) \circ \mathbf{\Phi}(x)| + |\mathbf{a}_n \circ (\mathbf{\Phi}(x) - \mathbf{\Phi}(x_n))| + |\mathbf{a}_n \circ \mathbf{\Phi}(x_n)| \\ &\le \|\mathbf{a} - \mathbf{a}_n\|_2 \cdot \|\mathbf{\Phi}(x)\|_2 + \|\mathbf{\Phi}(x) - \mathbf{\Phi}(x_n)\|_2.\end{aligned}$$

Since both terms on the right tend to zero, we have the inclusion $K \subset K_{\mathbf{a}}$. Therefore, $m(K_{\mathbf{a}}) = 1$.

4. Local Convergence of Wavelet Expansions. In [3], the following local convergence theorem is proved for Haar series, using martingale methods.

THEOREM A. *Let $\{f_j\}$ be a sequence of functions such that*

(a) $f_j \in V_j$ *where $\{V_j\}$ is the Haar multiresolution analysis;*

(b) $P_j(f_{j+1})(x) = f_j(x)$ *for $j \ge 0$.*

Set $f(x) = (f_0(x), f_1(x), \ldots)$ and let $S(f)(x) = \left(\sum (f_{j+1}(x) - f_j(x))^2 + f_0^2(x)\right)^{1/2}$; $f^*(x) = \sup_j |f_j(x)|$. Then, the following sets are equivalent almost everywhere:

(a) $\{x : \lim_{j\to\infty} f_j(x)$ *exists and is finite*$\}$;

(b) $\{x : S(f)(x) < +\infty\}$;

(c) $\{x : f^*(x) < \infty\}$.

Gundy and Kazarian [4] extended this local convergence theorem to the class of multiresolution analyses arising from the basic splines. In fact, the proof did not appear to use properties specific to the spline family. The basic stability condition essential to the proof is a two-norm condition, reminiscent of a condition first proposed by Marcinkiewicz and Zygmund [9] in their study of series of independent random variables. This condition, called condition (M-Z), is as follows: Let ϕ be a compactly supported scale function, supported on $[0, N]$. We suppose that, for every δ, $0 < \delta < 1$, there exist constants B_δ and C_δ such that for every measurable subset $E \subset [0, 1]$ of measure greater than δ, and any sequence a_k; $k = 0, 1, \ldots, N-1$, we have

$$\begin{aligned}C_\delta \sum_{k=0}^{N-1} |a_k| &\le \sup_{x\in E} \left| \sum_{k=0}^{N-1} a_k \phi(x + k) \right| \\ &\le B_\delta \sum_{k=0}^{N-1} |a_k|.\end{aligned}$$

The constants B_δ, C_δ depend only on ϕ and δ. The condition holds for the class of B-spline scale functions, as pointed out in [4]. However, the scope of the condition was not known, and left as an open problem in [4]. The following proposition we label as a corollary of Theorem 1.

COROLLARY 1. *Let $\{V_j\}$ be a multiresolution analysis such that V_0 contains a continuous function of compact support. Then the minimal scale function ϕ satisfies condition (M-Z).*

Before proving the corollary, we state the following theorem, in which we use the definitions in Theorem A.

THEOREM 2. (Theorem B of [4]) *Let $\{V_j\}$ be a multiresolution analysis that contains a continuous function of compact support. Then the following sets are equivalent almost everywhere:*

(a) $\{x : \lim_{j\to\infty} f_j(x) \text{ exists and is finite}\}$;

(b) $\{x : S(f)(x) < +\infty\}$;

(c) $\{x : f^*(x) < \infty\}$.

Proof of Corollary 1. Notice that B_δ may be taken to be $\|\phi\|_\infty$. Since ϕ is continuous on $[0,1]$, the issue is to show the existence of C_δ that is uniform over all sets $E \subset [0,1]$ of measure greater than δ. First, observe that, since the translates of ϕ are linearly independent over E (Theorem 1), there is a constant $C(E) > 0$, such that

$$C(E)\sum_{k=0}^{N-1}|a_k| \le \sup_{x\in E}\left|\sum_{k=0}^{N-1}a_k\phi(x+k)\right|.$$

This follows from the fact that the right-hand side defines a norm on $\mathbf{R}^N$: the linear independence of the translates of ϕ on the set E guarantees that the right-hand side is strictly positive on $\mathbf{R}^N \setminus \{0\}$. Since the left-hand side is also a norm, the existence of a constant is assured by the equivalence of norms on finite dimensional spaces. Now we must show that

$$\inf\{C(E) : m(E) \ge \delta\} > 0.$$

It is enough to show this for closed sets. To this end, we show that $C : (\mathcal{K}, \rho) \to \mathbf{R}$ defined by

$$C(E) = \inf_{\mathbf{a}\neq 0}\sup_{x\in E}\frac{|\mathbf{a}\circ\mathbf{\Phi}(x)|}{\|\mathbf{a}\|_1}$$

is a continuous function. Let $\epsilon > 0$ be given, and let $\{A_n\}$ be a sequence of sets converging to A in $\mathcal{K}$. The function $\mathbf{\Phi}$ is uniformly continuous on $[0,1]$; that is,

$$\|\mathbf{\Phi}(x) - \mathbf{\Phi}(y)\|_2 \le \epsilon$$

whenever $|x-y| \le \eta(\epsilon)$. Let n_0 be an integer such that

$$A_n \subset V_\eta(A) \qquad \text{and} \qquad A \subset V_\eta(A_n)$$

for all $n \ge n_0$. Notice that

$$C(A) \le C(\overline{V_\eta(A_n)}),$$

and that, by continuity, there exists $x = x(n,\eta,\mathbf{a}) \in \overline{V_\eta(A_n)}$ such that

$$C(\overline{V_\eta(A_n)}) = \inf_{\mathbf{a}\neq 0}\frac{|\mathbf{a}\circ\mathbf{\Phi}(x)|}{\|\mathbf{a}\|_1}.$$

Choose $y = y(n, \eta, \mathbf{a}) \in A_n$ so that $|x - y| \le \eta(\epsilon)$. Then

$$\begin{aligned}
C(\overline{V_\eta(A_n)}) &\le \inf_{\mathbf{a}\neq 0} \left(\frac{\|\mathbf{a}\|_2}{\|\mathbf{a}\|_1} \|\mathbf{\Phi}(x) - \mathbf{\Phi}(y)\|_2 + \frac{|\mathbf{a} \circ \mathbf{\Phi}(y)|}{\|\mathbf{a}\|_1} \right) \\
&\le \inf_{\mathbf{a}\neq 0} \left(\frac{\|\mathbf{a}\|_2}{\|\mathbf{a}\|_1} \epsilon + \sup_{y \in A_n} \frac{|\mathbf{a} \circ \mathbf{\Phi}(y)|}{\|\mathbf{a}\|_1} \right) \\
&\le \inf_{\mathbf{a}\neq 0} \left(\epsilon + \sup_{y \in A_n} \frac{|\mathbf{a} \circ \mathbf{\Phi}(y)|}{\|\mathbf{a}\|_1} \right) \\
&= \epsilon + C(A_n).
\end{aligned}$$

If we reverse the roles of A_n and A in the above argument, we see that $C(A_n) \le C(A) + \epsilon$. Thus, C is continuous on $\mathcal{K}$.

The collection $\{E \in \mathcal{K} : m(E) \ge \delta\}$ is a closed set in $\mathcal{K}$ since $m(\cdot)$ is upper semi-continuous on $(\mathcal{K}, \rho)$. Therefore, by continuity of C, there exists an $E_0 \in \mathcal{K}$, $m(E_0) \ge \delta$ such that $C(E_0) \le C(E)$ for all $E \in \mathcal{K}$ with $m(E) \ge \delta$. Now, by the compactness of the unit sphere in ℓ_1^N, there exists a $\mathbf{a}$ such that

$$C(E_0) = \sup_{x \in E_0} \frac{|\mathbf{a} \circ \mathbf{\Phi}(x)|}{\|\mathbf{a}\|_1}.$$

By Theorem 1, $C(E_0) > 0$.

5. Concluding Remarks. The quadratic variation functional $S(f)$ of Theorem 2 is invariant under changes of scale functions ϕ for V_0 and ψ for V_1: $S(f)$ is defined from the sequence of projections $\{P_j\}$ without specific reference to the choice of scale function. However, $S(f)$ is an "incomplete" square function in the sense that if the prewavelet family $\{\psi(2^j \cdot -k)\}$ is orthogonalized in k, to obtain a family $\{\tilde{\psi}(2^j \cdot -k)\}$ that is orthonormal in both variables $j, k \in \mathbf{Z}$, then one could consider a quadratic variation functional $\mathbf{S}(\mathbf{f})(x) = \left(\sum_{j,k} (a_{j,k}\tilde{\psi}(2^j x - k))^2\right)^{1/2}$. If we have a multiresolution analysis that admits a compactly supported, continuous orthonormal family $\{\tilde{\psi}(2^j \cdot -k)\}$, then one can show that $S(f)$ and $\mathbf{S}(\mathbf{f})$ are finite on the same set, up to a set of measure zero. The proof of this fact follows the same lines as the proof in [4]. Since the details are given there, we will not repeat them here.

Bibliography

1. I. Daubechies, Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **41** (1988), 909-996.

2. I. Daubechies, *Ten Lectures on Wavelets*, CBS-NSF Regional Conference in Applied Math., *SIAM* **61** (1992).

3. R. F. Gundy, Martingale theory and the pointwise convergence of certain orthogonal series, *Trans. Amer. Math. Soc.* **124** (2) (1966), 228-248.

4. R. F. Gundy, K. Kazarian, Stopping times and local convergence for spline wavelet expansions, Rutgers University Technical Report #97-002.

5. E. Hernández, G. Weiss, *A First Course on Wavelets*, CRC Press, Boca Raton FL, 1996.

6. P. Kornum, *Construction of Borel Measures on Metric Spaces*, Aarhus Universitet, Matematisk Institut, 1980.

7. P. G. Lemarié, Fonctions à support compact dans les analyses multi-résolutions, *Rev. Mat. Iberoamericana* **7** (2) (1991), 157-182.

8. P. G. Lemarié, G. Malgouyres, Support des fonctions de base dans une analyse multi-résolution, *C. R. Acad. Sci. Paris*, Séries I, **313** (1991), 377-380.

9. J. Marcinkiewicz, A. Zygmund, Sur les fonctions indépendentes, *Fund. Math.* **29** (1937), 60-90.

10. Y. Meyer, Ondelettes sur l'intervalle, *Rev. Mat. Iberoamericana* **7** (2) (1991), 115-133.

V. Dobrić
Lehigh University
Bethlehem, PA 18015
email vd00@lehigh.edu

R. Gundy
Rutgers University
New Brunswick, NJ 08903
email gundy@rci.rutgers.edu

P. Hitczenko
North Carolina State University
Raleigh, NC 27695
email pawel@math.ncsu.edu

Two Remarks Concerning Wavelets: Cohen's Criterion for Low-Pass Filters and Meyer's Theorem on Linear Independence

Richard F. Gundy

ABSTRACT. Part I: It is known that a continuous periodic function $m(\xi)$ may be a low-pass filter for an orthogonal scale function, yet fail to satisfy the condition given by A. Cohen for polynomials. We show that if $m(\xi)$ is a low-pass filter and if $m(\xi)$ is Hölder continuous, then Cohen's condition is satisfied. Part II: We give a simple proof of the strongest form of Meyer's theorem concerning linear independence of translates of a scaling function restricted to $[0, 1]$.

Introduction

This paper is divided into two parts. The first part is devoted to a discussion of Albert Cohen's well-known condition for a trigonometric polynomial $m(\xi)$ to be a low-pass filter. In the second part, we give an elementary proof of a strong form of Yves Meyer's theorem on linear independence of translates of a compactly supported prescaling function.

I. Cohen's Condition for $m(\xi)$

In [2], A. Cohen introduced conditions on a trigonometric polynomial $m(\xi)$ for it to be a low-pass filter corresponding to a scale function $\phi(x)$ for a multi-resolution analysis. We recall the appropriate definitions. Let $m(\xi)$ be a trigonometric polynomial that satisfies the conditions:

(a) $m(0) = 1$;

(b) $|m(\xi + 1/2)|^2 + |m(\xi)|^2 = 1$ for $0 \leq \xi \leq 1$.

Let

$$\hat{\phi}(\xi + k) := \prod_{j=1}^{\infty} m\big(\frac{\xi + k}{2^j}\big).$$

If

$$(1) \qquad \sum_{k \in \mathbf{Z}} |\hat{\phi}(\xi + k)|^2 = 1 \quad \text{a.e.},$$

AMS 1991 Subject Classification. Primary 42C15.

then

$$\phi(x) = \int_{-\infty}^{\infty} e^{2\pi i \xi x} \hat{\phi}(\xi)\, d\xi$$

is a scale function. In fact, the "almost everywhere" condition is superfluous when $m(\xi)$ is a polynomial: the exceptional set is empty since the sum on the left-hand side of (1) is a polynomial.

When $m(\xi)$ is merely continuous this is not necessarily the case. This possibility seems to have been overlooked in much of the wavelet literature, starting with A. Cohen [2], page 93 (6). In fact, in [4] it is shown that there exists a continuous function $m(\xi)$ that is C^∞ except at two points in $[0,1]$, such that the condition (1) fails to hold at two (exceptional) points. Furthermore, it is also true that any such function fails to satisfy Cohen's condition, which we recall below. (Thus, the condition introduced by Cohen is not necessary when $m(\xi)$ is only assumed to be continuous.) The purpose of this note is to discuss the case when $m(\xi)$ is a one-periodic function that is minimally smooth, in that it is assumed to satisfy a Hölder condition

$$|m(\xi) - m(\xi')| \le O(|\xi - \xi'|^\alpha), \qquad 0 < \alpha \le 1.$$

THEOREM 1. *Let $m(\xi)$ be continuous and satisfy a Hölder condition of order $\alpha > 0$. If the sum*

$$\sum_{k \in \mathbf{Z}} |\hat{\phi}(\xi + k)|^2 = 1 \quad a.e.,$$

then the exceptional set is empty.

(For $\alpha > 1/2$, this follows from Lemma 2 on page 365 of Kahane, Lemarié-Rieusset [7].) Now let us recall the condition introduced by A. Cohen in [2]. We state this in the form given by us in ([4], Theorem 1):

$$\begin{aligned} &\text{For every } \xi \in [0,1] \text{ there exists a } k = k(\xi) \in \mathbf{Z} \text{ such that} \\ &|\hat{\phi}(\xi + k)|^2 \ge \delta > 0 \text{ for some } \delta > 0 \text{ independent of } \xi. \end{aligned} \tag{2}$$

(Notice that $|\hat{\phi}(\xi + k)|^2$ is continuous in ξ when $m(\xi)$ satisfies a Hölder condition.)

This is not the usual form of Cohen's condition. It is however, equivalent to the orginal condition, as we have shown in [4].

Following the discussion in [4], we interpret $|\hat{\phi}(\xi + k)|^2$ as a probability on a binary sequence space Ω that contains $\mathbf{Z}$ as an embedded subset. Thus,

$$P_\xi(k) := |\hat{\phi}(\xi + k)|^2$$

and

$$P_\xi^N(k) := \prod_{j=1}^{N+1} |m(\frac{\xi + k}{2^j})|^2$$

for $\xi \in [0,1]$.

The probability P_ξ, on Ω, is the (Kolmogorov) extension of the family $P_\xi^N,\ N = 1, 2, \ldots$. The limiting P_ξ is concentrated on $\mathbf{Z}$ if and only if the sequence P_ξ^N is "tight" in the sense of [3]. (Each P_ξ^N may be considered as a probability on $-2^N \le k < 2^N$.) Now, as we indicated above, even when $P_\xi(k)$ is continuous in ξ, it may happen that

$$P_\xi(\mathbf{Z}) = \sum_{k \in \mathbf{Z}} |\hat{\phi}(\xi + k)|^2 = 1$$

for almost every $\xi \in [0,1]$, but $P_{\xi_0}(\mathbf{Z}) < 1$ for some ξ_0. In this case, a compactness argument shows that there exists a $\xi_0 \in [0,1]$ such that $P_{\xi_0}(\mathbf{Z}) = 0$. (See [4], Theorem 1, for this discussion.) This is equivalent to the fact that the sequence $P_{\xi_0}^N,\ N = 1, 2, \ldots$ is not tight; that is, the support of $P_{\xi_0}^N$ is a set of integers Λ, considered as binary sequences, such that $|k|,\ k \in \Lambda$, tends to infinity. The proof of the theorem stated above is the demonstration that the support of P_ξ^N contains this set Λ, and that $P_\xi^N(\Lambda) \ge 1 - \epsilon$ if $|\xi - \xi_0|$ is sufficiently small. Thus, P_ξ^N, $N=1,2,\dots$ is not tight on $\mathbf{Z}$, and therefore, $P_\xi(\mathbf{Z})<1$. This contradicts our assumption that $P_\xi(\mathbf{Z})=1$ a.e.

Now, if $P_\xi(\mathbf{Z})\equiv 1$, $\xi\in[0,1]$, then Cohen's condition (2) holds. Thus, the principal corollary of the theorem is as follows:

COROLLARY. *Let $m(\xi)$ be a one-periodic function that is Hölder continuous, and satisfies*
(a) $m(0)=1$*;*
(b) $|m(\xi)|^2+|m(\xi+1/2)|^2=1$.
Then Cohen's condition (2) is both necessary and sufficient for ϕ to be a scale function.

Proof of Theorem 1. We recall the notation of [4]. If $k\geq 0$, then we associate it with the sequence

$$k\cong(0,\omega_1(k),\omega_2(k),\dots)$$

where $k=\sum_{i=1}^{\infty}\omega_i(k)2^{i-1}$, $\omega_i(k)=0$ or 1. If $k<0$,

$$k\cong(1,\omega_1(-k-1),\omega_2(-k-1),\dots).$$

With this definition,

$$P_\xi^N(k):=\prod_{j=1}^{N+1}\left|m\big((\xi+k)/2^j\big)\right|^2$$

defines a probability on the $(N+1)$ cylinder sets, and the infinite product P_ξ is the unique extension that is countably additive on the cylinders.

Let P_{ξ_0} be the probability such that $P_{\xi_0}(\mathbf{Z})=0$, where $\mathbf{Z}$ is identified with the "finite" sequences, such that $\omega_i(k)\equiv 0$ for all i sufficiently large. The basic observation concerns the support of $P_{\xi_0}^N$. This measure cannot charge sequences ω such that $\omega_{i+1}(k)=\omega_{i+L}(k)=0$ for a fixed L, independent of N, and any i such that $i+L\leq N+1$. The size of L depends on $m(\xi)$. (We can take L to be the smallest integer such that $|m(\xi/2^{L-1})|>0$ for all $\xi\in[0,1]$ and $P_{\xi/2^{L-1}}(0)=\prod_{j=2^L}^{\infty}|m(\xi/2^j)|^2>0$. The Hölder continuity at 0 insures the existence of such an L.) Suppose that $k\in\mathbf{Z}$ is such an integer, and that $\omega_i(k)\equiv 0$ for $i=i_0+1,\dots,i_0+L$. If k belongs to the support of $P_{\xi_0}^N$ then

$$k_L=\sum_{i=1}^{i_0+L}\omega_i(k)2^{i-1}$$

also belongs to the support of $P_{\xi_0}^N$ for all N. This is because

$$P_{\xi_0}^N(k_L)=P_{\xi_0}^{i_0+L}(k_L)\cdot P_{\xi_0'}^{N-(i_0+L)-1}(0)$$

where $\xi_0'=(\xi_0+k_L)/2^{i_0+L+1}$. Now

$$P_{\xi_0}^{i_0+L}(k_L)=P_{\xi_0}^{i_0+L}(k)\geq P_{\xi_0}^N(k)>0$$

and $P_{\xi_0'}^{N-(i_0+L)-1}(0)\geq P_{\xi_0'}(0)>0$ since $|\xi_0|/2^{i_0+L+1}\leq 1/2^L$ and $|k_L|/2^{i_0+L+1}\leq 1/2^L$. Therefore $|\xi_0'|\leq 1/2^{L-1}$, which means $P_{\xi_0'}(0)>0$. All of this implies that if the sequence $\omega=\omega(k)$ has a string of L successive zeros among the first N coordinates, then $P_{\xi_0}^N(k)=0$. Let Λ^c be this set of integers, so that $P_{\xi_0}^N(\Lambda^c)=0$. We wish to show that $P_\xi^N(\Lambda^c)\leq\epsilon$ for all ξ sufficiently close to ξ_0.

The argument that follows is a standard probability calculation, but we will carry out the details of the argument in this context.

We begin by defining a stopping time on the set of sequences in Λ^c, that contain a succession of zeros of length L. Let $\omega=\omega(k)$ be the sequence corresponding to $k\in\mathbf{Z}$, and

$$\tau(\omega)=\min_j\left\{j:m\left(\frac{\xi_0+k}{2^j}\right)=0\right\}.$$

Then $\tau(\omega) < \infty$ on Λ^c and we may estimate $P_\xi^N(\Lambda^c)$, $\xi \neq \xi_0$ in the following way:

$$P_\xi^N(\Lambda^c) = \sum_{j=1}^N P_\xi^N(\tau = j).$$

We claim that $P_\xi^N(\Lambda^c) \le \epsilon$ if $|\xi - \xi_0|$ is sufficiently small. To see this, estimate

$$P_\xi^N(\tau = j_0) = \sum_{\{k:\tau=j_0\}} \prod_{j=1}^N \left| m\big((\xi + k)/2^j\big)\right|^2$$

by separating the sum into parts as follows: if $k > 0$,

$$k = k_1 + k_2,$$

where $k_1 = k_0 + \omega_{j_0}(k)2^{j_0-1}$ and $k_2 = k - k_1$. Write

$$P_\xi^N(k) = P_\xi^{j_0}(k_1)P_{\xi'}^{N-j_0-1}(k_2')$$

where $\xi' = (\xi + k_1)/2^{j_0+1}$ and $k_2' = k_2/2^{j_0+1}$. We fix (k_1, ξ') and sum over all k_2'. The sum

$$\sum_{k_2' \in \mathbf{Z}} P_{\xi'}^{N-j_0-1}(k_2') = 1,$$

so that what remains is the estimate of

$$\sum_{k_1:\tau=j_0} P_\xi^{j_0}(k_1) = \sum_{k_0} P_\xi^{j_0-1}(k_0) \cdot \left| m\left(\frac{\xi + k_0 + \omega_{j_0}(k)2^{j_0-1}}{2^{j_0+1}}\right)\right|^2.$$

Since $\left|m\big((\xi_0 + k_1)/2^{j_0+1}\big)\right|^2 = 0$ for those sequences (integers) ω such that $\tau(\omega) = j_0$, the Hölder continuity of $m(\xi)$ implies that

$$\left|m\big((\xi + k_1)/2^{j_0+1}\big)\right|^2 \le O\left\{\left(\frac{|\xi - \xi_0|}{2^{j_0+1}}\right)^\alpha\right\}$$

for every k_1.

Finally, $\sum_{k_0} P_\xi^{j_0-1}(k_0) \le 1$ and so

$$\sum_{k_1} P_\xi^{j_0}(k_1) = O\left\{\left(\frac{|\xi - \xi_0|}{2^{j_0+1}}\right)^\alpha\right\}$$

which implies

$$\sum_{j=1}^N P_\xi^N(\tau = j) = O(|\xi - \xi_0|^\alpha).$$

This proves that $P_\xi^N(\Lambda) \ge 1 - \epsilon$ if $|\xi - \xi_0|$ is sufficiently small. This, of course, contradicts $P_\xi(\mathbf{Z}) = 1$, and proves Theorem 1.

II. On a Theorem of Yves Meyer

In [10], Y. Meyer proved the following theorem:

THEOREM 2 [MEYER]. *Let $\phi(x)$ be a compactly supported orthonormal scaling function, supported on $[0, N]$. If*

$$\sum_{k=-N+1}^{0} a_k \phi(x - k) \equiv 0$$

for $x \in [0,1]$*, then* $\sum_{k=-N+1}^{0} a_k^2 = 0$.

Subsequently, Lemarié [8] noted that Meyer's theorem held in a much wider context. Specifically, if we are given a multiresolution analysis such that the space V_0 contains a non-trivial function of compact support, then Theorem 2 holds for $\phi \in V_0$, where ϕ is a non-trivial function with minimal support. The translates of such a function ϕ form a Riesz basis; the function ϕ is called a minimal prescaling function. (See Theorem 1 of [7].) Following this, Lemarié and Malgouyres [9] greatly simplified the proof of Meyer's result in the above context. In a paper with Kazarian [6], we needed a stronger form of this condition. Namely, we required the condition: if

$$\sum_{k=-N+1}^{0} a_k \phi(x-k) \equiv 0 \tag{1}$$

on a closed set $\mathcal{C} \subset [0,1]$ of positive measure, then $\sum_{k=-N+1}^{0} a_k 2 = 0$. In [6], we noted that this condition held when the function ϕ is a B-spline. Subsequently, with Dobrić and Hitczenko [5], we proved that the strong form stated above, holds for all continuous, minimally (compactly) supported prescale functions. The proof was rather complicated, and was further burdened with an unnecessary assumption. Recently, the author found an extremely simple proof of (1), assuming the validity of Theorem 2. Theorem 2, in our opinion, is such an attractive result that we were tempted to seek an even simpler proof of it than the one given by Lemarié and Malgouyres. The argument we found is (almost) free of all computations; in particular, it does not involve the use of Fourier transforms.

THEOREM 3. *Given a compactly supported prescaling function*

$$\phi(x) = \sum_{k=0}^{N} p_k \phi(2x-k)$$

with $p_0 \neq 0$*,* $p_N \neq 0$*, suppose that*

$$\sum_{k=-N+1}^{0} a_k \phi(x-k) \equiv 0$$

for $x \in \mathcal{C} \subset [0,1]$*, where* $\mathcal{C}$ *is a set of positive Lebesgue measure. Then* $\sum_{k=-N+1}^{0} a_k^2 = 0$.

Proof. Let us assume the validity of this theorem when $\mathcal{C} = [0,1]$.
This already implies that there is a constant $A > 0$ such that

$$\Big(\sum_{k=-N+1}^{0} |a_k|^2\Big)^{1/2} \leq A\Big(\int_0^1 \Big|\sum_{k=-N+1}^{0} a_k \phi(x-k)\Big| dx\Big)$$

since both sides represent a norm on $\mathbf{R}^N$. Let $B = ||\sum |\hat{\phi}(\xi+k)|^2||_\infty$

Now suppose that $\mathcal{C} \subset [0,1]$ and its complement $\mathcal{U}$ has small measure, meas$(\mathcal{U}) < \epsilon$. Then

$$\begin{aligned}\Big(\sum_{k=-N+1}^{0} |a_k|^2\Big)^{1/2} &\leq A\Big(\int_0^1 \Big|\sum_{k=-N+1}^{0} a_k \phi(x-k)\Big| dx\Big) \\ &= A\int_{\mathcal{U}} \big|\sum a_k \phi(x-k)\big| dx \\ &\leq AB^{1/2}\epsilon^{1/2}\Big(\sum_{k=-N+1}^{0} a_k 2\Big)^{1/2}.\end{aligned}$$

If $\sum_{k=-N+1}^{0} |a_k|^2 > 0$, and $\epsilon^{-1/2} > (AB^{1/2})$, we have a contradiction. However, the Meyer result (for the interval $[0,1]$) is scale invariant in the following sense.

(2) *If* $\sum_{k=-N+1}^{0} a_k\phi(x-k) = 0$ *on any subinterval* $(a,b) \subset [0,1]$, *then* $\sum_{k=-N+1}^{0} a_k 2 = 0$.

If we accept this stronger form of Meyer's result, noted by Lemarié and Malgouyres, we see that the above argument applies: We can find an interval (a,b) such that the relative measure of $\mathcal{C} \cap (a,b)$ is greater than $1-\epsilon$. The same reasoning leads to the conclusion that $\sum_{k=-N+1}^{0} a_k 2 = 0$.

Now let us give a proof of the Meyer-Lemarié-Malgouyres result. We suppose that ϕ is as in Theorem 3, and that $\sum_{k=-N+1}^{0} a_k\phi(x-k) = 0$ on $(a,b) \subset [0,1]$. (It is no restriction to assume that $a=0$ and $b = k\cdot 2^{-n}$, for some $k>0$, and some $n<m$. We can find an interval $(a',b') \subset (a,b)$ where a,b are dyadic rationals, then shift the origin accordingly.) The argument proceeds as follows: First, we contract the scale (from V_0 to V_m, $m \gg 0$) so that the interval $(0,b)$ contains the support of all translates $\phi(2^m x - k)$, $k = 0,1,\ldots,3N$. If we express $f(x)$ in the new scale, writing

$$\begin{aligned} f(x) &= \sum_{k=-N+1}^{0} a_k\phi(x-k) \\ &= \sum_k b_k\phi(2^m x - k), \end{aligned}$$

we may "decouple" the sum, so that

$$f(x) = f_L(x) + f_R(x).$$

Here, both $f_L(x)$ and $f_R(x)$ belong to V_m and $f_L(x) \equiv 0$ for $x>a$ and $f_R(x) \equiv 0$ for $x<b$. Now the function $f(x)$ belongs to $V_0 \subset V_m$, and is a finite sum of translates of $\phi(x)$. Its expression as an element of V_m is also a finite sum of translates of $\phi(2^m x)$. The decoupling procedure referred to above, expresses f as a sum of two finite sums $f_L(x)$ and $f_R(x)$, concentrated, respectively, to the left of $a=0$, and the right of $b = k\cdot 2^{-n}$. The key step is to show that $P(f_L)$ and $P(f_R)$, the projections of f_L and f_R onto V_{m-1}, are, respectively, each concentrated to the left of $a=0$ and to the right of $b=k\cdot 2^{-n}$. (By "projection," we mean the possibly nonorthogonal idempotent map that sends a function in V_m first, onto $V_{m-1} \oplus W_{m-1}$, then to V_{m-1}.) The fact that $P(f_L)$ and $P(f_R)$ are concentrated to the left and right, as indicated, means that $b_k = 0$ for a set of coefficients in the expansion of $f \in V_m$. We repeat the argument at V_{m-1} with the result that more coefficients vanish. Ultimately, we conclude that the $a_k \equiv 0$, $k = -N+1, \ldots 0$.

The first step in the decoupling is given by the following lemma. We state the lemma for V_1 and V_0 to minimize notation. It is obviously true if V_0 is replaced by V_{m-1}, V_1 by V_m.

LEMMA 1. *Suppose that* $f \in V_1$ *and that* f *is orthogonal to the set of* $3N+1$ *functions* $\phi(2x-k)$, $k = 0,1,\ldots,3N$. *Then the projection* $P(f)$ *onto* V_0 *decomposes into two disjointly supported functions* $f_L(x)$ *and* $f_R(x)$, *both belonging to* V_0, *where*

$$f_L(x) = \sum_{k=-\infty}^{-1} c_k\phi(x-k)$$

and

$$f_R(x) = \sum_{k=N+1}^{\infty} c_k\phi(x-k).$$

Proof. The projection $P(f)$ is a sum of functions of the form

$$\phi(x-j) = \sum_{k=0}^{N} p_k\phi\big(2(x-j)-k\big).$$

On the other hand, f is orthogonal to each $\phi\big(2x-(2j+k)\big)$ for $j = 0,1,\ldots,N$. Therefore, the expression for $P(f) = f_L + f_R$ is valid.

The second step is a lemma noted by Lemarié and Malgouyres [9].

LEMMA 2. *Suppose that $\phi(x)$ satisfies $\phi(x) = \sum_{k=0}^{N} p_k \phi(2x-k)$ with $p_0 \neq 0$, $p_N \neq 0$. Then $\phi(x)$ is supported on $[0, N]$ and this is the smallest interval containing its support.*

Proof. This is immediate if we use the two-scale equation for $\phi(x)$ for each scale 2^{-m}, $m \geq 1$; we verify that "inf support" is zero and "sup support" is N.

The third lemma relates the expression of $f(x)$ as a member of V_0 to its expression as an element of V_1. (Lemmas 1 and 3 could have been combined into a single statement but we chose to separate them for clarity. The reader may recall that similar estimates are made when one studies harmonic functions in $\mathbf{R}_+^{n+1}$; they lead to the construction of saw-toothed regions.)

LEMMA 3. *Let $f \in V_1$ and $P(f) \in V_0$ be its projection onto V_0, such that*

$$f(x) = \sum c_\ell \phi(2x - \ell)$$

and

$$P(f)(x) = \sum d_j \phi(x - j).$$

If $c_{2j+k} \equiv 0$, $k = 0, \ldots, N$, then $d_j = 0$.

Proof. This is also obvious since the map from $\big(\phi(x-k), \psi(x-\ell), k, \ell \in \mathbf{Z}\big)$ (where ψ is a compactly supported wavelet) to $\big(\phi(2x-k), k \in \mathbf{Z}\big)$ is nonsingular. The inverse map is determined in part by $\phi(x-j) = \sum_{k=0}^{N} p_k \phi\big(2x - (2j+k)\big)$. Thus, there exists coefficients $e_{\ell,j}$, j fixed, such that $\sum_\ell p_{\ell-2j} e_{\ell,j} = 1$ where $0 \leq \ell - 2j \leq N$, with $e_{\ell,j} \equiv 0$ outside this range. This implies that

$$\sum c_\ell e_{\ell,j} = d_j$$

and if $c_{2j+k} \equiv 0$, $k = 0, 1, \cdots, N$, then $d_j = 0$. (See section 2 of Meyer [10] for a similar argument.)

Proof of Theorem 2. Suppose that $f(x) = \sum_{k=-N+1}^{0} a_k \phi(x-k) \equiv 0$ in (a, b), where we assume $a = 0$, $b = k \cdot 2^{-n}$ for some $k > 0$. Choose $m > n$ large enough so that $\phi(2^m x - k)$, $k = 0, 1, \ldots, 3N$ are all supported in $(0, b)$. Then $f(x)$, expressed as a function in V_m, is given by

$$f(x) = \sum_{k=-\infty}^{\infty} b_k \phi(2^m x - k),$$

where $b_k = 0$ except for a finite number of indices. Since $f(x) \equiv 0$ on $(0, b)$, the function $f(x)$ is orthogonal to $\phi(2^m x - k)$, $k = 0, 1, \ldots, 3N$. Since $f \in V_0$, $P(f) = f$ and

$$f_L(x) = \sum_{k=-\infty}^{-1} c_k \phi(2^{m-1} x - k)$$

and

$$f_R(x) = \sum_{k=N+1}^{\infty} c_k \phi(2^{m-1} x - k)$$

by Lemma 1. On the other hand, since f_L and f_R have disjoint supports, it follows that

$$f_L(x) \equiv f_R(x) \equiv 0$$

on $(0, b)$. From the support considerations of Lemma 2, we obtain the fact that for all $x \,\epsilon [0, b)$(and hence for all x)

$$\sum_{k=-N+1}^{-1} c_k \phi(2^{m-1}x - k) \equiv 0$$

and

$$\sum_{k=N+1}^{2N-1} c_k \phi(2^{m-1}x - k) \equiv 0$$

so that $c_k \equiv 0$ for $k = -N+1, \ldots, 2N-1$.

Now $f = f_L + f_R$, where each component is a finite sum belonging to V_{m-1}, where

$$f_L(x) = \sum_{k=-\infty}^{-N} c_k \phi(2^{m-1}x - k)$$

and

$$f_R(x) = \sum_{k=2N}^{\infty} c_k \phi(2^{m-1}x - k).$$

Since $f \in V_0 \subset V_{m-2}$,

$$\begin{aligned} f(x) &= P(f_L)(x) + P(f_R)(x) \\ &= \sum d_j \phi(2^{m-2}x - j). \end{aligned}$$

Now we will show that $f_L - P(f_L) = f_R - P(f_R) = 0$ if $m \geq 2$. The supports of f_L and f_R are contained in the intervals $(-\infty, 0]$ and $[2^{-(m-1)}, +\infty)$, respectively. (Recall that $0 < b < 1$.) The support of f_R is contained in $[([2^{(m-1)}b] + 1) \cdot 2^{-(m-1)}, +\infty)$. On the other hand, by Lemma 3 the support of $P(f_L)$ is contained in the interval $(-\infty, N \cdot 2^{-(m-1)}]$, and the support of $P(f_R)$ is contained in $[-(N-1)2^{-(m-1)}, +\infty)$. The fact that $f_L + f_R \in V_0$ implies that $f_L - P(f_L) = -(f_R - P(f_R))$ is supported in the interval $[-(N-1)2^{-(m-1)}, N \cdot 2^{-(m-1)}]$ of length $(2N-1) \cdot 2^{-m-1}$, of length $(2N-1) \cdot 2^{-(m-1)}$. But $f_L - P(f_L) \in W_{m-2}$; it is known that any function in W_{m-2} must be supported in an interval of length at least $(2N)2^{-(m-1)}$. (See, for example, Chui [1], page 170.) Therefore $f_L - P(f_L) = f_R - P(f_R) = 0$. As we have pointed out above, this means that

$$P(f_L)(x) = \sum_{j=-\infty}^{-N} d_j \phi(2^{m-2}x - j)$$

and

$$P(f_R)(x) = \sum_{j=1}^{\infty} d_j \phi(2^{m-2} - j).$$

In other words, $d_j \equiv 0$, $-(N-1) \leq j \leq 0$. Now the above argument may be used at every scale $m \geq 2$. When $m = 2$, $P(f_R)(x) = P(f_L) \equiv 0$ by hypothesis. Therefore, $f(x) \equiv 0$ as claimed.

Acknowledgement. The author would like to thank V. Dobrić and Aline Bonami for useful conversations concerning these remarks.

References

1. Chui, C.K. *An Introduction to Wavelets*, Academic Press, San Diego, 1992.
2. Cohen, A. *Ondelettes, analyses multirésolutions, et filtres miroir en quadrature*, Ann. Inst. H. Poincaré, Anal. nonlinéaire 7 (1990), 439-459.
3. Cohen, A. *Construction de bases d'ondelettes α-Höldériennes*, Rev. Mat. Ibero- americana 6 (3-4) (1990), 91-108.
4. Dobrić, V., Gundy, R., and Hitczenko, P. *Characterizations of orthonormal scale functions: A probabilistic approach.* J. Geometric Analysis, to appear.
5. Dobrić, V., Gundy, R., and Hitczenko, P. *Stability properties for a compactly supported prescale function*, SIAM Journal on Math. Analysis, to appear.

6. Gundy, R.F. and Kazarian, K. *Stopping times and local convergence for spline wavelet expansions*, SIAM Journal on Math. Analysis, to appear.
7. Kahane, J.-P., Lemarié-Rieusset, P.-G., *Séries de Fourier et Ondelettes*, Cassini, Paris, 1998.
8. Lemarié, P.-G. *Fonctions à support compact dans les analyses multi-résolutions*, Rev. Mat. Iberoamericana 7 (2) (1991), 157-182.
9. Lemarié, P.-G. and Malgouyres, G. *Support des fonctions de base dans une analyse multi-résolution*, C. R. Acad. Sci. Paris, t.313, Série I (1991), 377-380.
10. Meyer, Y. *Ondelettes sur l'intervalle*, Rev. Mat. Iberoamericana 7 (2) (1991), 115-133.

DEPARTMENT OF STATISTICS, RUTGERS UNIVERSITY,
PISCATAWAY, NJ 08854-8019

Current address: Department of Statistics, Rutgers University, Hill Center, Busch Campus, Piscataway, NJ 08854-8019

E-mail address: gundy@rci.rutgers.edu

Additional Remarks on Low-pass Filters, Cohen's Theorem

This line of investigation began while I was reading Chapter 6, §6.3 of Ingrid Daubechies' book [4]. This particular section seems to me to have been written with a special enthusiasm: it details the proof of Albert Cohen's theorem, a necessary and sufficient condition for a trigonometric polynomial $p(\xi)$ to be a "low pass filter."

The term low-pass filter refers to the fact that $p(\xi)$ is the link between fine and coarse scales in a multiresolution analysis. The function $p(\xi)$ acts as a Fourier multiplier, sending the (Fourier transformed) scaling function $\hat{\phi}(\xi)$ in $\hat{V}_0$ to the coarser or "lower" space $\hat{V}_{-1}$ by the mapping $\hat{\phi}(2\xi) = p(\xi)\hat{\phi}(\xi)$. The signal processors call $p(\xi)$ the *transfer function* for the discrete filter $\{p_k\}$. That is, $p(\xi) = \sum p_k \exp(2\pi i k\xi)$. The important thing is that the *non-periodic* function $\hat{\phi}(\xi)$ is, formally, an infinite product

$$\hat{\phi}(\xi) = \prod_{j=1}^{\infty} p(\xi/2^j).$$

The mathematical problems come from trying to find conditions on $p(\xi)$ so that, not only does the product converge, but that $\hat{\phi}(\xi)$ satisfies

$$\sum |\hat{\phi}(\xi + k)|^2 = 1 \text{ a.e.}$$

This is the Fourier expression of the condition that $\phi(x + k)$, $k \in \mathbf{Z}$ form an orthonormal sequence. Cohen, in his thesis [2] was the first to establish necessary and sufficient conditions for a trigonometric polynomial to be a low-pass filter in the above sense.

The theorem is both deep and useful. For example, one can use it in $\mathbf{R}^2$ to prove that the fractal set known as the "twin-dragon" tiles the plane. (See Wojtaszczyk [10], page 130.) For $\mathbf{R}^1$, the necessary and sufficient condition has two expressions: one is a measure-theoretic condition, and the other is an arithmetic condition on the zeros of $p(\xi)$. The two did not seem to be related, and this was one of the features that fascinated me. Moreover, in my perusal of the literature, I came across statements, generalizations, that I could not verify. Ultimately, in collaboration with Dobrič and Hitczenko, we found the boundaries of Cohen's condition, and the surprising fact that there exist continuous $p(\xi)$ that do *not* satisfy Cohen's conditions, that are indeed, low-pass filters for orthonormal scaling functions contradicting some claims in the literature.

Subsequently, I found a way to describe all low-pass filters in terms of their action as Markov transition functions [7]. This approach had already been suggested by Lawton [9], and others [3], [8]. I simply carried the approach to its ultimate conclusion.

What can one do with this generality? Does it allow one to construct filters for real-life applications? I don't think so. However, I think the theorems do provide some insight on the structure of scaling functions, and prescaling functions.

Here is a little exercise for which the above discussion is relevant. The B-spline prescaling function

$$\phi_k := \chi_{[0,1]} * \cdots * \chi_{[0,1]}(x)$$

where the right side means the k-fold convolution of the indicator function of the unit interval with itself. The integer translates of ϕ_k form a Riesz basis in $L^2(\mathbf{R})$. That is,

$$A_k \leq \sum_j \left|\hat{\phi}_k(\xi + j)\right|^2 = \sum_j \left|\frac{\sin \pi(\xi + j)}{\pi(\xi + j)}\right|^2 \leq B_k.$$

This may be proved by induction on k using elementary estimates, starting with the fact that $\phi_1(x)$ is orthogonal to its integer translates.

Question: Given any $\phi(x)$ that is orthogonal to its integer translates, and $\phi_k(x) = \phi * \cdots * \phi(x)$, is ϕ_k a Riesz basis in $L^2(\mathbf{R}^1)$?

Bibliography

1. Cohen, A., Daubechies, I., and Feauveau, J.C. "Biorthogonal bases of compactly supported wavelets," *Comm. Pure Appl. Math.* **45** (1992), 485-560.

2. Cohen, A. and Ryan, R.D. *Signal Processing*, Chapman and Hall, 1995 (English translation of Cohen's thesis, *Ondelettes et Traitment Numérique du Signal*, Masson, Paris, 1992.

3. Conze, A. and Raugi, A. "Fonctions harmonique pour un opérateur de transition et applications," *Bull. Soc. Math. France* **118** (1990), 273-310.

4. Daubechies, I. *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Math., 1992.

5. Dobrič, V., Hitczenko, P., and Gundy, R.F. "Characterizations of orthonormal scale functions," *J. Geom. Anal.* **10** (3) (2000), 413-430.

6. Gundy, R.F. "Two remarks on wavelets...," in *Cont. Math: The Harmonic and Functional Analysis of Wavelets and Frames*, AMS #247, Providence, RI, 1999, 249-257.

7. Gundy, R.F. "Low-pass filters, martingales, and multiresolution analyses," *Appl. and Comp. Harmonic Anal.* **9** (2) (2000), 204-257.

8. Keane, M. "Strongly mixing g measures," *Inventiones Math.* **16** (1972), 309-324.

9. Lawton, W. "Necessary and sufficient conditions for constructing orthonormal wavelet bases," *J. Math. Physics* **32** (1991), 57-61.

10. Wojtaszczyk, P. *A Mathematical Introduction to Wavelets*, London Math. Soc. Student Texts #37, 1997.

CHARACTERIZATIONS OF ORTHONORMAL SCALE FUNCTIONS: A PROBABILISTIC APPROACH

by V. Dobrić, R. Gundy and P. Hitczenko

Lehigh University, Rutgers University and North Carolina State University

Key Words and Phrases: Multiresolution analysis, scale function, weak convergence of probabilities.

AMS 1991 Subject Classification: Primary 42C15

Abstract.

The construction of a multiresolution analysis starts with the specification of a scale function. The Fourier transform of this function is defined by an infinite product. The convergence of this product is usually discussed in the context of $L^2(\mathbf{R})$. Here, we treat the convergence problem by viewing the partial products as probabilities, converging weakly to a probability defined on an appropriate sequence space. We obtain a sufficient condition for this convergence, which is also necessary in the case where the scale function is continuous. These results extend and clarify those of A. Cohen, and Hernández, Wang, and Weiss. The method also applies to more general dilation schemes that commute with translations by $\mathbf{Z}^d$.

Introduction.

We will say that a function $\phi(x)$, $x \in \mathbf{R}$ is a *scaling* (or *scale*) function if $\phi(x) \in L^2(\mathbf{R})$, and

(a′) the function $\hat{\phi}(2\xi) = m(\xi)\hat{\phi}(\xi)$ with $m(\xi)$ a 2π-periodic function in $L^2(\mathbf{R})$;

(b′) $\sum_{k\in\mathbf{Z}} |\hat{\phi}(\xi + 2\pi k)|^2 = 1$ a.e., $\xi \in \mathbf{R}$;

(c′) $\lim_{j\to\infty} |\hat{\phi}(\xi/2^j)| = 1$ a.e.

This characterization of scaling functions is given in Chapter 7 of [5]. The conditions (a′) and (b′) mean that

(a) the function $\frac{1}{2}\phi\big(\frac{x}{2}\big) = \sum m_j \phi(x - j)$ with $\sum m_j 2 < \infty$;

(b) the translates of $\phi(x - k)$, $k \in \mathbf{Z}$, form an orthonormal sequence in $L^2(\mathbf{R}, dx/2\pi)$.

The condition (c′) is independent of (a′) and (b′). However, the "garden variety" scale functions are integrable, with integral one, so that $\hat{\phi}(\xi)$ is continuous, and $\hat{\phi}(0) = 1$. In these cases, (c′) is satisfied. Therefore, let us assume, for purposes of this introduction, that ϕ is integrable with integral one.

How do we construct (or recognize) such functions ϕ? Certain features are easily discernable. Since $\hat{\phi}(\xi)$ is continuous, and $\hat{\phi}(0) = 1$, the two-scale equation (a′) tells us that $\Phi(\xi) = \prod_{j=1}^{\infty} m(\xi/2^j)$, so that the properties of $\phi(x)$ (or $\Phi(\xi)$) are determined by $m(\xi)$. If we divide the sum in (b′) into two parts, according to the parity of the indices k, and use the variable 2ξ, we find

$$(*)\quad \begin{aligned} &\sum_{k\in\mathbf{Z}} |\Phi(2\xi + 2\pi k)|^2 \\ &\quad = \sum_{k\in 2\mathbf{Z}} |\Phi(2\xi + 2\pi k)|^2 + \sum_{k\in 2\mathbf{Z}+1} |\Phi(2\xi + 2\pi k)|^2 \\ &\quad = |m(\xi)|^2 \sum_{j\in\mathbf{Z}} |\Phi(\xi + 2\pi j)|^2 + |m(\xi + \pi)|^2 \sum_{j\in\mathbf{Z}} |\Phi(\xi + \pi + 2\pi j)|^2 \\ &\quad = |m(\xi)|^2 + |m(\xi + \pi)|^2 \\ &\quad = 1, \text{ a.e.} \end{aligned}$$

Therefore, in addition to 2π-periodicity, the function $m(\xi)$ must satisfy the identity expressed in the last two lines; since $|\hat{\phi}(0)|^2 = 1$, it also follows that $|m(0)|^2 = 1$. However, this identity is not sufficient to insure that (b′) is satisfied. If (b′) and (c′) together are satisfied, then $m(\xi)$ is called a *low-pass filter*. Many authors have considered the problem of finding sufficient (and necessary) conditions on $m(\xi)$ so that the infinite product

$\hat{\phi}(\xi)$ is a scale function. When $m(\xi)$ is a polynomial, two such sufficient conditions have been proposed, one by Mallat [6] and the other by Daubechies (see [3], page 182). Mallat's condition requires

$$\inf_{|\xi|\leq\pi/2} |m(\xi)| > 0$$

and Daubechies:

$$m(\xi) = \left[(1+e^{i\xi})/2\right]^N \mathcal{L}(\xi)$$

with $\sup_\xi |\mathcal{L}(\xi)| \leq 2^{N-1/2}$. The first necessary and sufficient conditions were found by Cohen [2], in the case where $m(\xi)$ is a polynomial; he later extended his considerations to the case where $m(\xi)$ is $C^1(\mathbf{R})$. The problem for more general $m(\xi)$ was considered by Hernández, Wang, and Weiss [4]. They obtained a necessary and sufficient condition when $|m(\xi)|$ takes the values 0 and 1. In the notes of Chapter 7 of the recent text by Hernández and Weiss [5], the authors propose the problem of finding necessary and sufficent conditions in the case when $m(\xi)$ is not necessarily $C^1(\mathbf{R})$. The principle purpose of this paper is to address this question. Our results are inspired by Cohen's ideas; however, we have translated his ideas into probabilistic terms. This approach seems to us to be very natural for the problem at hand, and allows us to obtain necessary and sufficient conditions in a very general context. In particular, the results of Cohen and Hernández, Wang, and Weiss are unified as special cases of a general theorem. The method applies, as well, to more general dilations schemes in $\mathbf{R}^d$. These are treated in brief, in a separate section.

To motivate the probabilistic approach, let us summarize the problem as it is usually presented. (See, for example, Daubechies ([3], Chapter 6.3) or Hernández and Weiss ([4], Chapter 7.4).) Given a candidate 2π periodic function $m(\xi)$ with $|m(\xi)|^2 + |m(\xi+\pi)|^2 = 1$, and $|m(0)|^2 = 1$, we form the sequence of partial products

$$\left|\Phi_N(\xi+2\pi k)\right|^2 = \prod_{j=1}^{N} \left|m\left(\frac{\xi+2\pi k}{2^j}\right)\right|^2.$$

Then $\lim_{N\to\infty} |\Phi_N(\xi+2\pi k)|^2 := |\Phi(\xi+2\pi k)|^2$; this limit is well defined a.e. since $|\Phi_N(\xi+2\pi k)|^2$ decreases with increasing N. The convergence in $L^2(\mathbf{R})$ is a different matter since the function $\Phi_N(\xi)$ is $(2\pi)2^N$-periodic. Therefore, except for trivial cases, $|\Phi_N(\xi)|^2$ is never integrable as a function of $\xi \in \mathbf{R}$. An obvious remedy for this defect is to restrict $\Phi_N(\xi)$ to the interval $[-2^N\pi, 2^N\pi]$. Thus, if

$$\Phi_N^*(\xi) = \begin{cases} \Phi_N(\xi) & \text{if } |\xi| \leq 2^N\pi \\ 0 & \text{otherwise,} \end{cases}$$

then $|\Phi_N^*(\xi)|^2$ also converges to $|\Phi(\xi)|^2$ pointwise a.e. To verify property (b$'$) in the definition of a scale function it turns out that it is enough to show that $\Phi_N^*(\xi)$ also converges in $L^2(\mathbf{R})$. Here matters become delicate. The L^2 convergence is complicated by the fact that there is no obvious domination. This is the point where Cohen's ideas come into play. He suggested that one should modify $\Phi_N(\xi)$ by multiplying by $\chi_{\mathbf{K}}(\xi)$, rather than $\chi_{[-\pi,\pi]}(\xi)$, where $\mathbf{K}$ is a finite union of intervals forming a compact set that is *congruent* to $[-\pi,\pi]$ in a sense described below. When such a $\mathbf{K}$ exists, the sequence $\Phi_N^{**}(\xi) = \Phi_N(\xi)\cdot\chi_{\mathbf{K}}(\xi/2^N)$ may be shown to converge in $L^2(\mathbf{R})$. With this convergence established, the convergence in $L^2(\mathbf{R})$ of the original sequence $\Phi_N^*(\xi)$ may also be proved. It was this feature of Cohen's approach that provoked our effort to find another perspective where Cohen's condition would appear in a more transparent fashion. For smooth $m(\xi)$, Cohen's condition requires that

$$\inf_{j>0} \inf_{\xi \in \mathbf{K}} |m(\xi/2^j)| > 0$$

where $\mathbf{K}$ is a compact set that is a finite union of intervals, one of which contains 0 as an interior point, such that $\mathbf{K}$ is congruent to $[-\pi, \pi]$ in the following sense:

(a) the Lebesgue measure of $\mathbf{K}$ is 2π;

(b) for every $\xi \in [-\pi, \pi]$, there is a $k \in \mathbf{Z}$ such that $\xi + 2\pi k \in \mathbf{K}$.

Notice that Cohen's condition is equivalent to a restriction on the partial products $\Phi_N(\xi)$ for $\xi \in \mathbf{K}$: Since $m(\xi)$ is smooth in a neighborhood of the origin, the partial products converge uniformly on any compact subset of $\mathbf{R}$; therefore, the condition may be stated as

$$\inf_{N \geq 1} \inf_{\xi \in \mathbf{K}} |\Phi_N(\xi)| \geq \delta > 0,$$

for some $\delta > 0$. In fact, a more succinct way to formulate the condition would be to omit the mention of $\mathbf{K}$ altogether. As we shall see, what is important is the existence of a lower bound δ for the infinite product. Furthermore, it is not the topological, but the measure-theoretic character of $\mathbf{K}$ that is important: it is enough to require the lower bound to hold almost everywhere in the following sense:

Either

(1) $|\Phi(\xi)| \geq \delta > 0$ almost everywhere in $[0, 2\pi]$

or

(2) $\sup_{k \in \mathbf{Z}} |\Phi(\xi + 2\pi k)| \geq \delta > 0$ almost everywhere in $[0, 2\pi]$.

Now let us drop the requirement that $\phi(x)$ is integrable, and focus on properties (a′) and (b′) for a function in $L^2(\mathbf{R})$. Given $m(\xi)$ as specified above, the conditions (1) and (2) are sufficient for the function $\hat{\phi}(\xi)$, satisfying (a′) to also satisfy (b′). When $\Phi(\xi)$ is continuous, then conditions (1)-(2) are necessary and sufficient for $\Phi(\xi)$ to satisfy (b′) *everywhere*, rather than almost everywhere. In fact, we show that it is possible for $\Phi(\xi)$ to be continuous, such that (b′) holds except at two points. Furthermore, *any* example where (b′) holds almost everywhere, but not everywhere, is such that (1)-(2) fails for any $\delta > 0$.

The authors would like to express their gratitude to Professors G. Weiss and H. Šikić of Washington University, for many valuable remarks on the subject of this paper. (In particular, see the section "The encounter...")

The probabilistic approach. In summary, we interpret the function $|m(\xi)|^2$ as a conditional probability defined on a space of infinite sequences. The partial products define a consistent family of probabilities on this sequence space which converge, in the usual (Kolmogorov) sense, to a probability. The existence of a scale function is equivalent to the "tightness" of this family of probabilities on "finite" sequences.

The probability space. Let $M(\xi) = |m(2\pi\xi)|^2$. Notice that $M(\xi)$ is a one-periodic function that satisfies $M(\xi) + M(\xi + 1/2) = 1$, and $M(0) = 1$. The basic probability space Ω for our discussion is the disjoint union of two spaces of infinite sequences ω with coordinates $\omega_i = 0$ or 1. We will represent elements of Ω by $\{0,1\} \times \{0,1\}^{\mathbf{N}}$; Ω^+ and Ω^- will denote sequences starting with 0 and 1, respectively. We identify integers with a subset of Ω in the following way. A positive integer k with dyadic expansion

$$k = \sum_{i=1}^{\infty} \omega_i(k) 2^{i-1}$$

is represented by the sequence

$$(0, \omega_1(k), \omega_2(k), \ldots).$$

The integer zero is identified with the sequence that is identically zero. A negative integer k is represented by coefficients of dyadic expansion of $-(k+1)$ preceded by 1 (thus, for example, the sequence $(1, 0, 0 \ldots)$ represents -1. We denote the sequences corresponding to nonnegative integers as $\mathbf{Z}^+$, and those corresponding to negative integers as $\mathbf{Z}^-$. Fix $k \in \mathbf{Z}$ and let $\mathbf{K}_N = \{\omega : \omega_i = \omega_i(k),\ 0 \le i \le N\}$ be the N dimensional Ω^+-cylinder that contains $\omega(k)$. For each $\xi \in [0,1]$ we define a probability Q_ξ^N, $0 \le \xi < 1$, on the set of all such cylinders by the following prescription. For $0 \le k \le 2^N - 1$, we set

$$Q_\xi^N(k) = \prod_{j=1}^{N} M\left(\frac{\xi + k}{2^j}\right).$$

We then have

$$\sum_{0 \le k < 2^N} \prod_{j=1}^{N} M\left(\frac{\xi + k}{2^j}\right) = 1,$$

where we used the basic fact that $M(\xi) + M(\xi + 1/2) = 1$. In the language of (conditional) probability,

$$M\left(\frac{\xi + k}{2^j}\right) = Q_\xi(\omega_j(k) \| \omega_{j-1}, \ldots \omega_1),$$

and the above sum is computed by the standard successive conditioning procedure.

With this interpretation of $M(\frac{\xi+k}{2^j})$, we see that the product defines a probability on cylinders of Ω^+, and that

$$Q_\xi^N(\mathbf{K}_N) = Q_\xi^{N+1}(\mathbf{k}_N),$$

where $\mathbf{k}_N$ is the N-dimensional cylinder corresponding to $0 \le k \le 2^N - 1$. In order to define corresponding probabilities on Ω^- let us consider a "reflected" filter

$$\widetilde{M}(\xi) = M(-\xi).$$

This filter may also be used to construct a probability on the positive integers $0 \le k < 2^N$ in the same fashion, by setting for $0 \le \eta < 1$ and $0 \le \ell < 2^N$

$$\widetilde{Q}_\eta^N(\ell) = \prod_{j=1}^{N} \widetilde{M}\left(\frac{\eta + \ell}{2^j}\right).$$

We now define measures P_ξ^N on cylinders in Ω by setting

$$P_\xi^N(k) = \begin{cases} Q_\xi^{N+1}(k), & \text{if } 0 \le k < 2^N; \\ \widetilde{Q}_{1-\xi}^{N+1}\big(-(k+1)\big), & \text{if } -2^N \le k < 0. \end{cases}$$

Notice that there is a double reflection, on the function, and on the argument, and that P_ξ^N corresponds to $N+1$ products in Q's. This specification shows that P_ξ^N, $N \ge 0$, is a consistent family ($P_\xi^N(k) = P_\xi^{N+1}(k)$ for each fixed k), since each of the families Q_ξ^N, $N \ge 1$, and $\widetilde{Q}_{1-\xi}^N$, $N \ge 1$ are consistent. To see that P_ξ^N defines a probability on the integers $-2^N \le k < 2^N$, notice that

$$\begin{aligned} \sum_{-2^N \le k < 2^N} P_\xi^N(k) &= \sum_{0 \le k < 2^N} Q^{N+1}(k) + \sum_{-2^N \le k < 0} \widetilde{Q}_{1-\xi}^{N+1}\big(-(k+1)\big) \\ &= \sum_{0 \le k < 2^N} Q_\xi^{N+1}(k) + \sum_{-2^N \le k < 0} Q_\xi^{N+1}(2^{N+1}+k) \\ &= \sum_{0 \le k < 2^{N+1}} Q_\xi^{N+1}(k) \\ &= 1. \end{aligned}$$

Therefore, P_ξ^N, $N \ge 1$ specifies a probability on the σ-field generated by the cylinders.

The encounter at Washington University. The initial version of this paper contained an error in the formulation of the definition of the family $P_\xi^N(\cdot)$, $N \ge 0$. We are extremely grateful to H. Šikić, of Washington University, for showing us this error. The discussion with Šikić occurred during a visit by one of us, to St. Louis, and resulted in a radical adjustment in the definition of P_ξ^N. It is remarkable that the conclusions of Theorem 2 survived this adjustment with minimal changes.

Now, we can restate the problem concerning the existence of a scaling function in a very succinct fashion:

THEOREM 1. *The function $m(\xi)$ is a low-pass filter for a scaling function whose Fourier transform is $\hat{\phi}(\xi)$ if and only if*

(b″) the probability P_ξ is concentrated on finite sequences for almost every ξ, $0 \le \xi < 1$. We denote this by saying $P_\xi(\mathbf{Z}) = 1$ a.e.;

(c″) there exists a set $L^+ \subset [0,1)$ of positive measure such that for $\xi \in L^+$,

$$\lim_{j\to\infty} |\hat{\phi}\big((\xi+k)/2^j\big)|^2 = 1$$

for all $k \ge 0$, and a set of positive measure $L^- \subset [0,1)$ such that for $\xi \in L^-$,

$$\lim_{j\to\infty} |\hat{\phi}\big((\xi+k)/2^j\big)|^2 = 1$$

for all $k \le -1$.

Proof of Theorem 1. If $m(\xi)$ is a low-pass filter, then $\hat{\phi}(\xi)$ satisfies (b′) which implies (b″). Conversely, the condition (b″) is another way of stating (b′).

Now we must show that (c″) and (c′) are equivalent. We use the following proposition.

PROPOSITION 1. *Let $f(\cdot)$ be a function defined on $\mathbf{R}_+$. For $0 \le \xi < 1$, and $k \ge 0$, consider the set*

$$L = \left\{ \xi : \lim_{j\to\infty} f\big((\xi + k)/2^j\big) = 1 \right\}$$

for all $k \ge 0$. This set has measure one or zero.

Proof of Proposition 1. This is a special case of the Kolmogorov zero-one law. The set L is a "tail set" in the sense that it is invariant under all transformations $\xi \to (\xi + k)/2^n$ for fixed k and n, $0 \le k < 2^n$. Such invariant sets have measure zero or one.

Now, if (c′)holds then (c″) holds. Conversely, if the (apparently) weaker condition (c″) holds, the stronger condition (c′) holds by Proposition 1. That is, the sets $L^{\pm}$ have measure one.

Remarks. The formulation of the second part of Theorem 1 was inspired by Theorem 3.16 of Papadakis, Šikić, and Weiss [7]. They propose a characterization of nonnegative periodic functions $m(\xi)$ that are low-pass filters; this characterization assumes that the infinite product $\hat{\phi}(\xi)$ satisfies (c′), and they require that the partial products, suitably truncated, converge in $L^2(\mathbf{R})$ to the limit $\hat{\phi}(\xi)$.

This requirement is equivalent to our (b″) in Theorem 1. Rather than simply assume (c′) as they did, we chose to state it in the present form for the following reason. Papadakis, Šikić, and Weiss exhibit an example attributed to M. Paluszyński, where (b′) holds (that is, the partial products converge in $L^2(\mathbf{R})$) but (c′) fails. The example is simply $\hat{\phi}(\xi) = \chi_{[0,1)}(\xi)$. Clearly, $P_\xi(\mathbf{Z}^-) = 0$ for all ξ, $0 \le \xi < 1$, and the condition (c′) does not hold: $\lim_{j\to\infty} \hat{\phi}(\xi/2^j) = 0$ if $\xi < 0$. Therefore, in search of minimal conditions, one might suggest that there exist sets of positive measure L^+ and L^- such that $P_\xi(\mathbf{Z}^+) > 0$ for $\xi \in L^+$ and $P_\xi(\mathbf{Z}^-) > 0$ for $\xi \in L^-$. These conditions are *implied* by (c″). That is, they are necessary conditions, but, in fact, fail to be sufficient. If the suggested necessary condition is strengthened to $P_\xi(\mathbf{Z}^+) > 0$ for almost every ξ, $0 \le \xi < 1$, the condition fails to be necessary. Consider the Shannon filter,

$$m(\xi) = \chi_{[0,1/4)}(\xi) + \chi_{[3/4,1)}(\xi),$$

extended periodically. Here $\hat{\phi}(\xi) = \chi_{[-1/2,1/2)}(\xi)$ and $P_\xi(\mathbf{Z}^+) = P_\xi(0) = 1$, with $P_\xi(\mathbf{Z}^-) = 0$ if $0 \le \xi < 1/2$; also $P_\xi(\mathbf{Z}^-) = P_\xi(-1) = 1$ with $P_\xi(\mathbf{Z}^+) = 0$, if $1/2 \le \xi < 1$. Therefore, the qualification "on a set of positive measure" is necessary. With this qualification, the suggested condition is not sufficient to imply (c″). We can perturb the Shannon filter so that on a set of positive measure E, such that

$$\lim |\hat{\phi}(\xi/2^j)|^2 = 0$$

for $\xi \in E$, but $P_\xi(\mathbf{Z}) = 1$ a.e. (We omit the details of this example.) The upshot of all of this is that we must have $P_\xi(\mathbf{Z}^+) > 0$ on a set of positive measure, and $P_\xi(\mathbf{Z}^-) > 0$ on a set of positive measure, as well as an almost everywhere dyadic continuity at zero. These two requirements are captured in condition (c″).

The condition (c″) has a probabilistic interpretation in terms of the underlying Markov chains associated with the functions $M(\xi)$. However, we introduced the probability notions as a tool, and our interest in the details of the probability structure are secondary. Therefore, we chose not to express condition (c$''$) in purely probabilistic terms, as we did for (b$''$).

In order to show that $P_\xi(\mathbf{Z}) = 1$, we will use Prokhorov's criterion of tightness for a sequence of probability measures.

Definition: The sequence P_ξ^N is said to be tight on $\mathbf{Z}$ in Ω, if for every $\epsilon > 0$, there is an $n(\epsilon) = n(\epsilon, \xi) > 0$, such that

$$\sum_{n(\epsilon)\le|\mathbf{k}_N|} P_\xi^N(\mathbf{k}_N) \le \epsilon \text{ for all } N \ge 0.$$

Here $|\mathbf{k}_N|$ is the index i with largest absolute value such that $\omega_i(k) = 1$.

In terms of the integers $k \in \mathbf{Z}$, we may write this tightness condition as

$$\sum_{n(\epsilon)\le k<2^N} P_\xi^N(k) + \sum_{n(\epsilon)\le k\le 2^N} P_\xi^N(-k) \le \epsilon.$$

We note that

$$\lim_{N\to\infty} P_\xi^N(\mathbf{k}_N) = P_\xi(\omega(k)) = |\hat{\phi}(\xi+k)|^2,$$

or less formally,

$$\lim_{N\to\infty} P_\xi^N(k) = P_\xi(k).$$

Finally, we write

$$P_\xi(\mathbf{Z}) = 1$$

if $\hat{\phi}(\xi)$ satisfies (b$'$).

Criterion: $P_\xi(\mathbf{Z}) = 1$ if and only if P_ξ^N are tight.

We omit the details of this argument. (See Billingsley [1].)

We are now in a position to state the principal result.

THEOREM 2. (i) *A sufficient condition for $P_\xi(\mathbf{Z}) = 1$ almost everywhere is the following (condition (C)):*

Suppose that for almost every ξ, $0 \le \xi \le 1$, there exists a $\delta > 0$ and an integer $k(\xi)$, such that $P_\xi(k(\xi)) \ge \delta$.

(ii) *Let $\xi \mapsto P_\xi(k)$ be continuous for each $k \in \mathbf{Z}$. (In other words, $|\hat{\phi}(\theta)|$ is continuous for all $\theta \in \mathbf{R}$, so that $\hat{\phi}$ satisfies the condition (c$'$) for a scale function.) If condition (C) is satisfied, then $P_\xi(\mathbf{Z}) = 1$ for every ξ, $0 \le \xi \le 1$. (That is, there is no exceptional set.) Conversely, if $P_\xi(\mathbf{Z}) = 1$ for every ξ, $0 \le \xi \le 1$, then condition (C) holds with no exceptional set.*

(iii) *There exists a function $M(\xi)$ infinitely differentiable at $\xi = 0$ such that $P_\xi(k)$ is continuous for each k, and such that $P_\xi(\mathbf{Z}) = 1$, except at two points ξ, $0 < \xi < 1$. At these exceptional points, $P_\xi(\mathbf{Z}) = 0$. In particular, condition (C) fails to hold for any $\delta > 0$.*

Remark 1. At first, the distinction between "almost everywhere" and "everywhere" in the above theorem may seem somewhat fastidious. However, these distinctions are crucial for the following reasons. If $\hat{\phi}(\xi)$

is the Fourier transform of a scale function, then the equation (b′) holds *almost everywhere*. The circumstances where (b′) holds *everywhere* are of secondary interest. In the same spirit, the natural assumption of the theorem concerns the behavior of P_ξ almost everywhere. If, however, we require $P_\xi(k)$ to be continuous in ξ for each k, then the sufficient condition (a.e.) gives the conclusion $P_\xi(\mathbf{Z}) = 1$ *everywhere*. Conversely, if $P_\xi(\mathbf{Z}) = 1$ *everywhere*, then the sufficient condition (C) holds *everywhere*. Thus, when $P_\xi(k)$ is supposed to be continuous, the sufficient condition (C) becomes necessary, but with a blemish: the natural necessary condition should read, "If $P_\xi(\mathbf{Z}) = 1$ almost everywhere, then condition (C) holds almost everywhere." However part (iii) states that this cannot hold in general, even when $P_\xi(k)$ is continuous. In particular, there are low-pass filters of class $C^0(\mathbf{R})$ generating continuous scale functions that do not satisfy Cohen's condition.

Remark 2. When $P_\xi(k)$ is continuous for each k, the condition (C) is equivalent to that given by Cohen. Since $P_0(0) = P_1(1) = 1$ and $P_\xi(0)$ ($P_\xi(1)$) is continuous, there are one-sided neighborhoods of zero and one such that $P_\xi(0) \geq \delta > 0$, $0 \leq \xi < \alpha$, and $P_\xi(1) \geq \delta > 0$, $1 - \alpha \leq \xi \leq 1$. In other words, $|\hat{\phi}(\xi)|^2 \geq \delta > 0$ for $|\xi| \leq \alpha$. Thus, the first condition for a Cohen set is satisfied. With each ξ_0 we can associate an interval, $\{\xi : |\xi - \xi_0| < \epsilon\}$ centered at ξ_0, such that $P_\xi((k(\xi_0)) \geq \delta/2$ for every ξ in the interval. Then, we find a finite subcollection ξ_i, $k(\xi_i)$, $i = 0, 1, \ldots, N$, such that the corresponding union of intervals covers the unit interval. The compact set specified by Cohen may be constructed using translations by $k_i(\xi)$, $i = 0, 1, \ldots, N$.

Now suppose that a compact set $\mathbf{K}$, with Cohen's specifications, exists. We will show that the probabilities $P_\xi^N(\cdot)$, $N \geq 1$ are tight. Choose $n(\epsilon)$ large enough so that

$$\sum_{n(\epsilon) \leq |k|} P_\xi(k) \leq \epsilon.$$

Now estimate $P_\xi^N(k)$, $n(\epsilon) \leq k < 2^N$ as follows:

$$P_\xi^N(k) = P_\xi^N(k + 2^{N+1}j)$$

where $j = j((\xi + k)/2^{N+1})$ is the integer such that $j + (\xi + k)/2^{N+1} \in \mathbf{K}$. If $j \geq 0$,

$$P_\xi^N(k) \leq \frac{1}{\delta} P_\xi^N(k + 2^{N+1}j) P_{(\xi+k)/2^{N+1}}(j)$$

and $n(\epsilon) \leq k \leq k + 2^{N+1}j$. On the other hand, if $j < 0$, and $n(\epsilon) \leq k < 2^N$, then

$$P_\xi^N(k) \leq \frac{1}{\delta} P_\xi(-(2^{N+1}|j| - k))$$

and $n(\epsilon) \leq 2^{N+1}|j| - k$ if $n(\epsilon) \leq 2^N(2|j| - 1)$. Therefore, either

$$P_\xi^N(k) \leq \frac{1}{\delta} P_\xi(k + 2^{N+1}|j|)$$

or

$$P_\xi^N(k) \leq \frac{1}{\delta} P_\xi(-(2^{N+1}|j| - k)).$$

In the first case,

$$\sum_{n(\epsilon)\le k<2^N} P_\xi^N(k) \le \frac{1}{\delta} \sum_{n(\epsilon)\le n} P_\xi(n) \le \epsilon/\delta;$$

in the second case,

$$\sum_{n(\epsilon)\le k<2^N} P_\xi^N(k) \le \frac{1}{\delta} \sum_{2^N(2|j|-1)\le|n|} P_\xi(-n).$$

Now we choose N large enough so that $2^N + 1 \ge n(\epsilon)$. A similar argument may be made for $k < 0$, with $n(\epsilon) \le |k| \le 2^N$. This shows that P_ξ^N, $N \ge N(\epsilon)$ is tight, and therefore, that the entire collection P_ξ^N is tight.

Remark 3. Hernández, Wang, and Weiss [4] treated the case where $M(\xi)$ is a $C^1(\mathbf{R})$ function, as well as the case when $M(\xi)$ is a function taking only values 0 and 1. In the latter case $P_\xi(k)$ also takes values 0 and 1, and the condition (C) of the theorem becomes

$$P_\xi(k) = 1 \text{ for some } k = k(\xi)$$

for almost every ξ, $0 \le \xi \le 1$. This condition is obviously necessary for $P_\xi(\mathbf{Z}) = 1$ a.e., in this case. Furthermore, the "almost everywhere" cannot be altered.

Proof of (i). Suppose $P_\xi(\cdot)$ satisfies the condition (C). Let us call the sequence of points $\{\xi' : \xi' = (\xi + k)/2^N \pmod 1,\ N > 0,\ k \in \mathbf{Z}\}$ the *orbit* of $\xi \in [0,1]$. We want all the probabilities $P_{\xi'}$ to satisfy condition (C), where ξ' belongs to the orbit of ξ. The set of "good" points G, where condition (C) holds has full measure, and the translates of G by dyadically rational points, G_k also have full measure. So, we take the set $\tilde{G} = \bigcap G_k$, of full measure, of points ξ that satisfy our requirement.

Now we turn to the proof of the tightness of the sequence P_ξ^N, for $\xi \in \tilde{G}$. Let $\mathbf{k}_N$ denote an N cylinder corresponding to the integer k, as specified earlier. Let $\xi' = (\xi + k)/2^{N+1} \pmod 1$ and $k(\xi')$ be an integer such that $P_{\xi'}\big(k(\xi')\big) \ge \delta$. Then the ω-sequence corresponding to $k + 2^{N+1}k(\xi')$ belongs to the N cylinder $\mathbf{k}_N$, and

$$P_\xi\big(k + 2^{N+1}k(\xi')\big) = P_\xi^N(\mathbf{k}_N) P_{\xi'}\big(k(\xi')\big).$$

Therefore,

$$P_\xi^N(\mathbf{k}_N) \le \delta^{-1} P_\xi\big(k + 2^{N+1}k(\xi')\big).$$

(This estimation is simply a transcription of Cohen's calculation.) Now, observe that the probability $P_\xi(\cdot)$ always satisfies the condition for tightness on $\mathbf{Z}$. That is, for $\epsilon > 0$, there exists an $n = n(\epsilon, \xi)$ such that

$$\sum_{n\le|k|} P_\xi(k) \le \epsilon$$

where $|k|$ is the largest (or smallest) index in the sequence $\omega(k)$ such that $\omega_i(k) = 1$. (This is always true since $P_\xi(\mathbf{Z}) \le 1$.) Therefore

$$\sum_{n \le |k|_N} P_\xi^N(\mathbf{k}_N) \le \delta^{-1} \sum_{n \le |k| \le N} P_\xi\big(k + 2^{N+1}k(\xi')\big)$$
$$\le \delta^{-1} \sum_{n \le |k|} P_\xi(k)$$
$$\le \delta^{-1} \cdot \epsilon.$$

This proves that the condition of part (i) is sufficient for tightness, and so proves that $P_\xi(\mathbf{Z}) = 1$ for $\xi \in G$.

Proof of (ii). Now we assume the condition (C) of the theorem, and that $P_\xi(k)$ is continuous on $[0,1]$ for each k. We wish to show that $P_\xi(\mathbf{Z}) = 1$ for all ξ in $[0,1]$. By part (i), the equality holds almost everywhere. If it fails at some point ξ_1, $0 < \xi_1 < 1$, then there must be a point ξ_0 where $P_{\xi_0}(\mathbf{Z}) = 0$. (Consider the possibility that $P_\xi(\mathbf{Z}) > 0$ for every ξ. Then, for every ξ, there exist $k(\xi)$ such that $P_\xi\big(k(\xi)\big) > 0$. By the continuity of $\xi \to P_\xi(k)$, the sets $\{\xi : P_\xi(k) > 0\}$ are an open cover of $[0,1]$. Therefore, there exists a $\delta > 0$ such that $P_\xi(\mathbf{Z}) \ge \delta > 0$ for every ξ.

Then the condition of part (i) holds everywhere, and the argument given in part (i) shows that $P_\xi(\mathbf{Z}) = 1$ everywhere. Therefore, there exist points ξ_0 where $P_{\xi_0}(\mathbf{Z}) = 0$.) Since $P_\xi(0)$ is continuous, and tends to one as ξ tends to zero, $P_{\xi_0}(\mathbf{Z}) = 0$ if and only if for each $k \in \mathbf{Z}$, there exists $N = N(k)$ such that $M\big((\xi_0 + k)/2^N\big) = 0$. This "sudden death syndrome" is inconsistent with the hypothesis of the theorem: we will prove that, given any $\epsilon > 0$, there exists an open set of points where

$$\max_k P_\xi(k) \le \epsilon.$$

To this end, let ξ_0 be a point in $[0,1]$ such that $P_{\xi_0}(\mathbf{Z}) = 0$. Now consider a neighborhood of ξ_0,

$$N_\eta(\xi_0) = \{\xi : |\xi - \xi_0| < \eta\},$$

where η is chosen so that

$$|M(\xi) - M(\xi')| \le \epsilon$$

for any two points (ξ, ξ') such that $|\xi - \xi'| < \eta \pmod 1$. We claim that

$$\max_k P_\xi(k) \le \epsilon$$

in the neighborhood $N_\eta(\xi_0)$. If $\xi = \xi_0 + \Delta$, where $|\Delta| < \eta$, then $M\big((\xi + k)/2^N\big) \le \epsilon$ for $N = N(k)$. This implies

$$P_\xi(k) \le \epsilon$$

for every $\xi \in N_\eta(\xi_0)$. That is, this contradicts the assumption if we choose $\epsilon < \delta$.

Now let us prove the necessity of the condition (C). Suppose that $P_\xi(\mathbf{Z}) = 1$ for every ξ in the unit interval. This implies that there exists a finite set of integers $\mathbf{Z}_\xi$ such that

$$P_\xi(\mathbf{Z}_\xi) \ge \delta(\xi) > 0.$$

By the assumption that $P_\xi(k)$ is continuous for each k, the fact that $P_\xi(\mathbf{Z}) = 1$ for every ξ, and the compactness of $[0,1]$, we can find a finite set of integers $\mathbf{Z}_0$, independent of ξ, and a fixed $\delta > 0$, such that for ξ, $0 \le \xi \le 1$,

$$P_\xi(\mathbf{Z}_0) \ge \delta.$$

This implies that

$$\max_{k \in \mathbf{Z}_0} P_\xi(k) \ge \delta / \mathrm{card}(\mathbf{Z}_0).$$

Thus, we have shown that condition (C) holds for every ξ, $0 \le \xi \le 1$.

Proof of (iii). The statement of part (ii) would be vacuous if it were not possible to construct a family P_ξ, continuous in ξ for each k, such that $P_\xi(\mathbf{Z}) = 1$ almost everywhere, but not everywhere. The following is such a construction, inspired by an example given by Cohen [2].

Let $M(\xi)$ be a continuous periodic function, with period one, such that $M(0) = 1$ and $M(\xi)$ is infinitely differentiable in neighborhoods of zero, and one half. The condition $M(\xi) + M(\xi + 1/2) = 1$ is imposed, as usual. The function $M(\xi)$ is to have only three zeros in $0 < \xi < 1$: $M(1/2) = 0$ (dictated by the usual condition), and $M(1/6) = M(5/6) = 0$. The latter two zeros mean that $M(1/3) = M(2/3) = 1$. At this point, we have the example due to Cohen, cited above. However, we insist that the function $M(\xi)$ should have cusps at the points $\xi = 1/3$ and $\xi = 2/3$, so that

$$\sum_{k=1}^{\infty} \big(1 - M(1/3 \pm \epsilon/2^k)\big) = \infty$$

and

$$\sum_{k=1}^{\infty} \big(1 - M(2/3 \pm \epsilon/2^k)\big) = \infty,$$

for any ϵ, $0 < \epsilon < 1$. (For example, we may take

$$M(\xi) \cong 1 - \big(\log(|1/3 - \xi|)\big)^{-1}$$

for ξ in a neighborhood of 1/3, with a similar specification around 2/3.)

The probability P_ξ, constructed using this $M(\xi)$, has the following properties:

(a) For any integer $k \in \mathbf{Z}$, $P_\xi(k)$ is continuous in ξ, since the infinite product converges uniformly in ξ. ($M(\xi)$ is smooth in a neighborhood of zero.)

(b) $P_\xi(\mathbf{Z}) = 0$ at $\xi = 1/3$ and at $\xi = 2/3$. In fact, at the point $\xi = 1/3$, P_ξ is concentrated on the single sequence ω such that $\omega_0 = 1$, $\omega_1 = 1$, $\omega_2 = 0, \ldots$ ($\omega_{2i} = 0$, $\omega_{2i+1} = 1$, $i \ge 0$); at the point $\xi = 2/3$, P_ξ is concentrated on $\omega = (0,0,1,0,1,\ldots)$ (i.e., $\omega_0 = 0, \omega_{2i-1} = 0, \omega_{2i} = 1$, $i \ge 1$). On the other hand, if $\xi \ne 1/3$, $\xi \ne 2/3$, the divergence of the above sums implies that $P_\xi(\omega) = 0$ for the two sequences described above. (Notice that if the cusps were placed at $\xi = 0$ and $\xi = 1$, rather than at $\xi = 1/3$ and $\xi = 2/3$, then $P_\xi(\mathbf{Z}) \equiv 0$ for all ξ, $0 < \xi < 1$.)

(c) $P_\xi(\mathbf{Z}) = 1$ for all other points in the unit interval.

To prove (c) we must show that the sequence P_ξ^N is tight.

To ease the burden of subscript notation, we will denote the cylinder $\mathbf{k}_N$ by k. With this convention, we must show that

$$\sum_{n(\epsilon)\le|k|\le 2^N} P_\xi^N(k) \le \epsilon$$

for some integer $n(\epsilon)$ and all $N \ge 0$. Now, to find the integer $n(\epsilon)$ in the definition of tightness, we make a finite number of choices, starting the process by finding $m(\epsilon)$ such that

$$\sum_{m(\epsilon)\le|k|} P_\xi(k) \le \epsilon.$$

Now choose δ small enough so that the interval $(1/3-\delta, 1/3+\delta)$ is strictly contained in $[0, 1/2]$. Let

$$A_\delta = (1/3-\delta, 1/3+\delta) \cup (-1/3-\delta, -1/3+\delta),$$

and

$$A = A(\xi,\delta) = \left\{k : |k| \le 2^N, (\xi+k)/2^{N+1} \in A_\delta\right\}.$$

Notice that if $\xi' = (\xi+k)/2^{N+1}$, $k \in \mathbf{Z}$, $|k| \le 2^N$, and $\xi' \notin A_\delta$, then the probability $P_{\xi'}(0)$, $\xi' > 0$ (or $P_{1+\xi'}(0)$, $\xi' < 0$) is uniformly bounded away from zero. In the sequel, the subscripts $-1 < \xi' < 0$ are to be interpreted as $1+\xi'$. Thus, with this notation, we have just stated that

$$\inf_{\xi'\notin A_\delta} P_{\xi'}(0) := C^{-1}(\delta) > 0.$$

Since $P_\xi(k) = P_\xi^N(k)P_{\xi'}(0)$, we may estimate as we did in part (ii) of the proof, to obtain

$$\sum_{\substack{k\notin A\\ m(\epsilon)\le|k|\le 2^N}} P_\xi^N(k) \le C(\delta) \sum_{\substack{k\notin A\\ m(\epsilon)\le|k|}} P_\xi(k).$$

We increase $m(\epsilon)$ to $p(\epsilon)$ if necessary, so that

$$\sum_{p(\epsilon)\le|k|} P_\xi(k) \le C^{-1}(\delta)\epsilon.$$

Therefore,

$$\sum_{\substack{k\notin A\\ p(\epsilon)\le|k|\le 2^N}} P_\xi^N(k) \le \epsilon.$$

The "real work" is to estimate the sum for $k \in A$, $|k| \ge p(\epsilon)$. If k satisfies these restrictions and $k > 0$, then $\omega_0(k) = 0$ and $\omega_N(k) = 1$, $\omega_{N-1}(k) = 0$, $\omega_{N-2}(k) = 1, \ldots$ with this alternating pattern continuing for a least J steps. The alternating pattern is dictated by the fact that $k/2^{N+1}$ is approximately $1/3$, which has the alternating pattern in its dyadic expansion. The fact that the approximation is δ-close ($k \in A$) means that the alternating pattern continues for at least $J = J(\delta)$ steps, with $\omega_{N-J}(k) = 1$.

If $-2^N \le -k < 0$, then our convention dictates that $\omega_0(-k) = 1$ and

$$\begin{aligned} -k &= -\Big(1 + \sum_{i=1}^{N} \omega_i(-k)2^{i-1}\Big) \\ &= -\Big(1 + \sum_{i=1}^{N} \omega_i(k-1)2^{i-1}\Big). \end{aligned}$$

We wish to compute the probability

$$\begin{aligned} P_\xi^N(-k) &= \prod_{j=1}^{N+1} M\big((\xi - k)/2^j\big) \\ &= \prod_{j=1}^{N+1} M\big((\xi + 2^{N+1} - k)/2^j\big) \end{aligned}$$

for $-k \in A$. This restriction $-k \in A$ means that $\omega_N(-k) = 1$, $\omega_{N-1}(-k) = 0, \ldots$ with the alternating ones and zeros continuing for at least $J = J(\delta)$ steps.

In any case, the restriction $|k| \ge p(\epsilon)$ means that $N \ge [\log_2 p(\epsilon)] := L$, where $[x]$ is the integer part of x. To prove tightness for the entire sequence P_ξ^N, $N \ge 1$, it suffices to prove tightness for P_ξ^N, $N \ge N(\epsilon)$. Therefore, we can restrict our attention to N such that $N - J > L$.

With this pattern in mind, we can decompose $k \in A,\ 0 < k < 2^n$, into two integers:

$$k = t_\ell + b_\ell$$

where t_ℓ is the "top" of k

$$t_\ell = \sum_{j=\ell+1}^{N} \omega_j 2^{j-1}$$

where the sequence $\omega_j,\ j = \ell, \ldots, N$, is alternately 0 and 1, as specified above. The index ℓ is determined by the following rule: We observe the sequence $\omega_{N-j},\ j = 0, 1, \ldots, \ell$ which alternates between 1 and 0, starting at $\omega_N = 1$; we stop at the index ℓ where the coefficient $\omega_\ell = 1$ and the pattern is broken for coefficients smaller than ℓ. (Thus, $(\omega_{\ell-1} = 0, \omega_{\ell-2} = 0)$ and $\omega_{\ell-1} = 1$ are the two possibilities when $\ell > 0$. If the pattern is not broken, then $\ell = 0$ and $b_\ell = 0$.) As we remarked above $0 \le \ell \le N - J$. This means that the "bottom" part of k,

$$b_\ell = \sum_{j=1}^{\ell} \omega_j 2^{j-1}$$

has arbitrary coefficients ω_j for $j < \ell$, and $\omega_\ell = 1$. Also, we note that $(\xi + b_\ell)/2^{\ell+1} \notin A_\delta$ (or $b_\ell \notin A$) for any $0 \le \ell \le N - J$, and $b_\ell \ge p(\epsilon)$ if $\ell > L$.

If $k \in A$, $2^N \le k < 0$ we may carry out a similar decomposition for the positive integer $-(k+1)$. As we have noted, $k \in A$ implies that $-(k+1)/2^{N+1}$ is approximately $1/3$. In terms of the above notation,

$$k = -(1 + b_\ell + t_\ell)$$

and

$$\begin{aligned} P_\xi^N(k) &= \tilde{Q}_{1-\xi}^{N+1}\big(-(k+1)\big) \\ &= \tilde{Q}_{1-\xi}^{N+1}(b_\ell + t_\ell). \end{aligned}$$

In this way, we see that the estimation of $P_\xi^N(k)$, for $k < 0$, may be carried out in the same way as for $k > 0$ by using the reflected filter to define probabilities on nonnegative integers.

Now suppose that $k > 0$; we may write

$$\sum_{k \in A} P_\xi^N(k) = \sum_{\ell=0}^{N-J} \sum_b P_\xi^N(b+t).$$

(Here we have omitted the subscript ℓ, so that $b = b_\ell$, $t = t_\ell$.) Write the sum on ℓ in two parts

$$\sum_{\ell=0}^{N-J} \sum_b P_\xi^N(b+t) = \sum_{\ell=0}^{L} \sum_b P_\xi^N(b+t) + \sum_{\ell=L+1}^{N-J} \sum_b P_\xi^N(b+t).$$

To estimate the first sum, we write each term

$$P_\xi^N(b+t) = P_\xi^\ell(b) P_{(\xi+b)/2^{\ell+1}}^{N-\ell-1}(t')$$

where $t' = t/2^{\ell+1}$. Notice that t' is an integer, and that the coefficients of t' satisfy $\omega_j(t') = \omega_{\ell+1+j}(t)$, $j = 0, 1, \ldots, N - \ell - 2$. This means that t' has the same pattern as t. Since the infinite sequences of alternating zeros and ones are assigned probability zero unless $\xi = 1/3$ or $2/3$ (property (b)), we have

$$P_{(\xi+b)/2^{\ell+1}}^{N-\ell-1}(t') = o(1)$$

as N tends to infinity when $\ell \le L$, uniformly in $b = b_\ell \notin A$. Therefore,

$$\begin{aligned} \sum_{\ell=0}^{L} \sum_b P_\xi^N(b+t) &= \sum_{\ell=0}^{L} \sum_b P_\xi^\ell(b) P_{(\xi+b)/2^{\ell+1}}^{N-\ell-1}(t') \\ &\le (L+1)) \cdot o(1) = o(1) \end{aligned}$$

as N tends to infinity. Recall here that neither L nor J depend on N. That is, the above sum can be made less than ϵ if $N \ge N(\epsilon)$. This imposes another restriction on the $n(\epsilon)$ we are seeking, and we incorporate this into the calculation without further mention.

Now we estimate

$$\sum_{\ell=L+1}^{N-J} \sum_b P_\xi^N(b+t) \le \sum_{\ell=L+1}^{N-J} \sum_b P_\xi^\ell(b).$$

Recall that $b \notin A$, and $p(\epsilon) \le b$ so that

$$P_\xi^\ell(b) \le C(\delta) P_\xi(b).$$

Consequently

$$\sum_{\ell=L+1}^{N-J} \sum_b P_\xi^N(b) \le C(\delta) \sum_{p(\epsilon) \le b} P_\xi(b) \le \epsilon.$$

In summary, we have shown that there exists $n(\epsilon) = \max\big(p(\epsilon), N(\epsilon)\big)$ such that

$$\sum_{n(\epsilon)\le|k|\le 2^N} P_\xi^N(k) \le 3\epsilon$$

for all N. This is sufficient and concludes the proof of part (iii) of the theorem.

The multidimensional case. The construction of scale functions corresponding to more general dilation schemes may be accomplished in much the same manner as described above for the case of dyadic dilations. Cohen's criterion may be applied without essential change. The class of dilation schemes most frequently considered are implemented by a matrix A that maps $\mathbf{Z}^d$, the integer lattice, into itself. We assume that A is *strictly expansive* in the sense that all eigenvalues λ_i are such that $|\lambda_i| > 1$. Here, a *scale function* $\phi(x)$, $x \in \mathbf{R}^d$ is a function that belongs to $L^2\big(\mathbf{R}^d/(2\pi)^d\big)$ such that

(a′) $\Phi(A^*\xi) = m(\xi)\Phi(\xi)$

for $\xi \in \mathbf{R}^d$, with $m(\xi)$ periodic on the 2^d-dimensional torus $(2\pi)^d$ and $m(0) = 1$;

(b′) $\sum_{k\in\mathbf{Z}^d} |\Phi(\xi + 2\pi k)|^2 = 1$ a.e.

These assumptions are not enough to insure that ϕ corresponds to a multiresolution analysis since $\Phi(\xi)$ is not assumed to be continuous at zero; however, this requirement is not relevant to the present discussion. (See Theorem 1.7, Chapter 2 of [5].) For a discussion of multiresolution analyses in this generality, see Wojtaszczyk ([9], Chapter 5). In particular, see Proposition 5.21 *op. cit.* for a statement of Cohen's theorem.

Given a function $\Phi(\xi)$, satisfying (b′), we have a probability $P_\xi(\cdot)$ defined from $\Phi(\xi)$, that is concentrated on the lattice $\mathbf{Z}^d$, for almost every ξ in any set that is congruent to $(2\pi)^d$. (Such sets are often called *fundamental domains* for the action of $(2\pi)\mathbf{Z}^d$ on $\mathbf{R}^d$; we shall use this term also.) The question arises: When (b′) holds, does $\Phi(\xi)$ correspond to a probability on a space containing $\mathbf{Z}^d$ in a manner similar to the case when $A = 2$, acting on $\mathbf{R}$? The "enveloping probability space" is certainly not canonical, and, the construction for the case $A = 2$ has an *ad hoc* character. This being so, can we describe a procedure for constructing this probability space that applies to any dilation? The general case presents certain technical problems associated with the fact that we do not know of a fundamental domain that is invariant under the action of $(A^{-1})^*$. As a consequence, we failed in our attempts to describe a universal sequence space Ω which is independent of ξ. However, if we restrict attention to the class of transformations that are *similarities*, we can carry out a construction that generalizes the case $A = 2$, and looks somewhat less impromptu than that described above. We hope that it illuminates what was done in that case. A *similarity* is a matrix A such that the eigenvalues λ_i have constant modulus; in our case $|\lambda_i| \equiv c > 1$. The fundamental lemma for this construction is a result due to Strichartz ([8], Lemma 5.1). We quote the lemma and include its proof for completeness.

LEMMA 1 (Strichartz). *Let B be a strictly expansive similarity defined on $\mathbf{R}^d$, such that $B(\mathbf{Z}^d) \subset \mathbf{Z}^d$. Suppose that the common value of the modulus of any eigenvalue λ satisfies $|\lambda| > 1 + d^{1/2}$. Then there exists a set of coset representatives $r_1, r_2, \ldots, r_q$ ($q = |\lambda|^d$) for the group $\mathbf{Z}^d/B(\mathbf{Z}^d)$ such that every element $k \in \mathbf{Z}^d$*

has a finite expansion

$$k = r_{i_0} + Br_{i_1} + \cdots + B^n r_{i_n}.$$

Proof. The choice of coset representatives is chosen as the set

$$\mathbf{Z}^d \cap B\big((-1/2, 1/2]^d\big).$$

This is possible since the unit cube, centered at the origin, is a fundamental domain for $\mathbf{Z}^d$ acting on $\mathbf{R}^d$. The element $k \in \mathbf{Z}^d$ has the coset representation

$$k = r_{i_0} + Br_{i_1} + \cdots + B^{n-1} r_{i_{n-1}} + B^n \tilde{r}_n$$

for some $\tilde{r}_n \in \mathbf{Z}^d$. We must show that $\tilde{r}_n \in B\big((-1/2, 1/2]^d\big)$ for some $n \geq 0$. Since B is a similarity, B maps the ball of radius $1/2$ centered at the origin, onto a ball of radius $|\lambda|/2$, centered at the origin, contained in $B\big((-1/2, 1/2]^d\big)$. We will prove that $\|\tilde{r}_n\| < |\lambda|/2$ (that is, $\tilde{r}_n$ lies in the centered ball of radius $|\lambda|/2$), and so is one of the coset representatives. Since $\|B\| = |\lambda|$ and $|r_i| \leq |\lambda| d^{1/2}/2$, we have

$$\begin{aligned}\|B^n(\tilde{r}_n)\| &\leq |k| + \left(\sum_{i=0}^{n-1} |\lambda|^i d^{1/2}\right)(|\lambda|/2) \\ &< |k| + \big[|\lambda|^n d^{1/2}/(|\lambda|-1)\big]\big(|\lambda|/2\big).\end{aligned}$$

Therefore, if we take B^{-n} on the left-hand side, we obtain

$$\|\tilde{r}_n\| < |k|/|\lambda|^n + \big[d^{1/2}/(|\lambda|-1)\big](|\lambda|/2),$$

so that $\|\tilde{r}_n\| < |\lambda|/2$ for some n, as we wished to show.

Armed with the above lemma, Strichartz proved the following theorem, using the facts about tilings of $\mathbf{R}^d$.

THEOREM 3 (Strichartz [8]). *Let B be a strictly expansive similarity transformation such that $B(\mathbf{Z}^d) \subset \mathbf{Z}^d$. Suppose that the (common) value of the modulus of any eigenvalue is greater than $1 + d^{1/2}$. Let $\{r_1, r_2, \ldots, r_q\} = \mathcal{R}$ be the set of coset representatives specified in Lemma 1. Then the set $T \subset \mathbf{R}^d$ defined by the equation*

$$B(T) = \sum_{r_i \in \mathcal{R}} (T + r_i)$$

tiles $\mathbf{R}^d$. That is, the Lebesgue measure of $(T+k) \cap (T+j)$ is zero if $k \neq j$ and $\bigcup_{k \in \mathbf{Z}^d} (T+k) = \mathbf{R}^d$.

We refer the reader to Strichartz's paper [8], and the references there, for a proof.

Now let us consider the problem of constructing a sequence space Ω, and an embedding of $\mathbf{Z}^d \mapsto \Omega$, given a strictly expansive similarity matrix A mapping $\mathbf{Z}^d$ into itself, and a candidate function $m(2\pi\xi)$, periodic with period one, for $\xi \in \mathbf{R}^d$.

A necessary (but not sufficient) condition for $M(\xi) := |m(2\pi\xi)|^2$ to be associated with a scale function (that is, a function Φ satisfying (a′) and (b′)) is that

$$\sum_{i=1}^{q} M\big(\xi + (A^*)^{-1} r_i\big) = 1 \text{ a.e.}$$

where the integers $r_i,\ i = 1, 2, \ldots, q$ are coset representatives of the group $\mathbf{Z}^d/A^*(\mathbf{Z}^d)$. This follows from properties (a′) and (b′) by an argument very similar to the one given above for the case when $A = A^* = 2$, acting on $\mathbf{Z}$. Thus, for each fixed ξ, we have a probability measure concentrated on q points in $\mathbf{Z}^d$. It is important to note that the measure is invariant under changes of coset representatives. That is, if r_i is replaced by $\tilde{r}_i = r_i + A^*(k)$, $i = 1, 2, \ldots, q$ for some $k \in \mathbf{Z}^d$, then, since $M(\xi)$ is periodic,

$$M\big(\xi + (A^*)^{-1} r_i\big) \equiv M\big(\xi + (A^*)^{-1} \tilde{r}_i\big)$$

for $i = 1, 2, \ldots, q$.

We have assumed that A is a strictly expansive similarity. Although A does not necessarily satisfy the condition of Lemma 1, that $|\lambda| > 1 + d^{1/2}$, there is a (smallest) integer p such that A^p does fulfill this condition. The subsequence of partial products

$$P_\xi^N(k) := \prod_{j=1}^{p \cdot N} M\big((A^*)^{-j}(\xi + k)\big),$$

where p is fixed and $N = 1, 2, \ldots$ defines a sequence of probabilities on $\mathbf{Z}^d$. Each of these probabilities may be considered as a probability on a sequence space Ω whose coordinates are integers that form a complete set of coset representatives for the group $\mathbf{Z}^d/(A^*)^p(\mathbf{Z}^d)$. The parameter set containing ξ is taken to be the tile T generated by $(A^*)^p$.

To be more specific, given the candidate function $M(\xi)$ we define $\widetilde{M}(\xi)$ by the product

$$\widetilde{M}(\xi) = \prod_{j=0}^{p-1} M\big((A^*)^j \xi\big).$$

Now set $B = (A^*)^p$ and consider $\widetilde{M}(\xi)$ as a candidate function with the dilation matrix B. Notice that $\widetilde{M}(\xi)$ is one-periodic and

$$\prod_{j=1}^{\infty} \widetilde{M}\big(B^{-j}\xi\big) = \prod_{j=1}^{\infty} M\big((A^*)^{-j}\xi\big).$$

We may summarize this equality be saying that $\widetilde{M}(\xi)$ is the square of the modulus of a low-pass filter for $\Phi(\xi)$ corresponding to the dilation B^*. The necessary condition given above for $\Phi(\xi)$ to be a scale function, expressed in terms of B and $\widetilde{M}$, becomes

$$\sum_{i=1}^{q^p} \widetilde{M}\big(\xi + B^{-1} r_i\big) = 1 \text{ a.e.}$$

where $r_i,\ i = 1, 2, \ldots, q^p$ is any collection of coset representatives for the group $\mathbf{Z}^d/B(\mathbf{Z}^d)$. This equality follows from its predecessor for A^*. Now we are in a position to specify Ω as a sequence space with coordinates $\omega_j(k) = r_j$ where vectors r_j are the coset representatives of $\mathbf{Z}^d/B(\mathbf{Z}^d)$ that appear in the expansion

$$k = r_0 + B r_1 + \cdots + B^n r_n$$

where $n = n(k)$ is the maximal exponent in the finite expansion provided by Lemma 1. We let ξ be the generic point in the tile T generated by B. For each such ξ, the partial products

$$P_\xi^N(k) = \prod_{j=1}^{N} \widetilde{M}\big(B^{-j}(\xi + k)\big)$$

define a sequence of consistent measures on the cylinder of Ω, as described in the one dimensional case, and the limiting measure P_ξ is defined on the σ-field generated by the cylinders. It is important to note that P_ξ^N defines a measure concentrated on finite sequences $\omega(k) \in \Omega$ with $\omega_j(k) = r_j$ and $\omega_{n+j}(k) \equiv 0$ for some n, all $j > 0$, defined by the expansion given in Lemma 1:

$$k = \sum_{j=0}^{n} B^j r_j, \quad n = n(k).$$

Furthermore, the sets

$$Z_N = \{k : n(k) = N\}$$

are nested ($Z_N \subset Z_{N+1}$) and $\mathbf{Z}^d = \lim Z_N$. The limiting measure $P_\xi(\mathbf{Z}^d) = 1$ if and only if the sequence P_ξ^N is *tight* in the sense that given $0 < \epsilon < 1$, there exists a set $Z_{N(\epsilon)}$ such that

$$P_\xi^{N(\epsilon)+j}(Z_{N(\epsilon)}) \geq 1 - \epsilon$$

for all $j > 0$.

Cohen's condition: There exists a compact set $\mathbf{K}$ containing a neighborhood of the origin, and congruent to $(1/2, 1/2]^d$ such that for $\xi \in \mathbf{K}$, $M\big((A^*)^{-n}\xi\big) > 0$ for all $n \geq 1$. The following more general condition is equivalent to Cohen's condition when $P_\xi(k)$ is continuous for each $k \in \mathbf{Z}^d$: There exists a $\delta > 0$ and $k = k(\xi) \in \mathbf{Z}^d$ such that

$$P_\xi\big(k(\xi)\big) \geq \delta > 0 \qquad \text{(Condition C)}$$

for almost every $\xi \in T$.

The proof that Condition C is sufficient for tightness is similar to the reasoning for the case $A = 2$: Given $\epsilon > 0$, find $Z_{N(\epsilon)}$ such that $P_\xi(Z^c_{N(\epsilon)}) \leq \delta\epsilon$. Then

$$\begin{aligned} P_\xi^{N(\epsilon)+j}(Z_{N(\epsilon)+j} \cap Z^c_{N(\epsilon)}) &\leq \delta^{-1} P_\xi(Z^c_{N(\epsilon)}) \\ &\leq \epsilon, \end{aligned}$$

since for $k \in Z_{N(\epsilon)+j}$ and $\ell \in \mathbf{Z}^d$

$$P_\xi\big(k + B^{N(\epsilon)+j}(\ell)\big) = P_\xi^{N(\epsilon)+j}(k) P_{B^{-(N(\epsilon)+j)}(\xi+k)}(\ell).$$

We conclude our discussion of the multidimensional case at this point.

References

1. Billingsley, P. **Convergence of Probability Measures**, Wiley, New York, NY, 1968.

2. Cohen, A. Ondelettes, analyses multirésolutions, et filtres miroir en quadrature, *Ann. Inst. H. Poincaré, Anal. nonlinéaire*, 7, 439-459, 1990.

3. Daubechies, I. **Ten Lectures on Wavelets**, (CBMS-NSF regional conference series in applied mathematics, 61) SIAM, Philadelphia, PA, 1992.

4. Hernández, E., Wang, X., and Weiss, G., Smoothing minimally supported frequency wavelets: part II, *J. Fourier Anal. Appl.*, 3 (1), 23-41, 1997.

5. Hernández, E. and Weiss, G. **A First Course on Wavelets**, CRC Press, Inc., Boca Raton, FL, 1996.

6. Mallat, S. Multiresolution approximation and wavelets, *Trans. Amer. Math. Soc.*, 315, 69-88, 1989.

7. Papadakis, M., Šikić, H., and Weiss, G., The characterization of low-pass filters and some basic properties of wavelets, scaling functions, and related concepts, *J. Fourier Anal. and Appl.*, 5 (5), 495-521, 1999.

8. Strichartz, R. Wavelets and self-affine tilings, *Constr. Approx.*, 9, 327-346, 1993.

9. Wojtaszczyk, P. **A Mathematical Introduction to Wavelets**, London Math. Soc. Student Texts 37, Cambridge Univ. Press, Cambridge, U.K., 1997.

e-mail addresses:

V. Dobric, vd00@lehigh.edu

R. Gundy, gundy@rci.rutgers.edu

P. Hitczenko, pawel@math.ncsu.edu

Low-Pass Filters, Martingales and Multiresolution Analyses

by

Richard F. Gundy

Rutgers University, Piscataway, New Jersey

Statistics Department TR #99-008

Summary

Necessary and sufficient conditions for a trigonometric polynomial to be a low-pass filter have been given by A. Cohen and W. Lawton. We give necessary and sufficient conditions for an arbitrary periodic function to be a low-pass filter, following the approach of Lawton. The technique also applies to functions that give rise to a Riesz basis.

Key Words and Phrases: Low-pass filters, martingales, multiresolution analyses

AMS 1991 Subject Classification: 42C15, 60G46

Introduction.

One way to construct a multiresolution analysis is to start with what some authors call a "prescaling function" $\phi(x),\ x \in \mathbf{R}$. That is, ϕ is assumed to belong to $L^2(\mathbf{R})$ and its translates $\phi_k(x) = \phi(x-k)$ form a Riesz basis for the subspace V_0 of $L^2(\mathbf{R})$ generated by finite linear combinations of the functions ϕ_k, $k \in \mathbf{Z}$. The Riesz basis property is often expressed by saying that if $f = \sum a_k \phi_k$ belongs to V_0, then

$$c \sum a_k^2 \le \|f\|_2^2 \le C \sum a_k^2.$$

Such a function will generate a multiresolution analysis if ϕ has the additional properties:

(1) $\phi(x) = 2\sum p_k \phi(2x-k)$ for some $\{p_k\} \in \ell^2$ (or $\hat{\phi}(\xi) = p(\xi/2)\hat{\phi}(\xi/2)$);

(2) the chain of spaces $V_j \subset V_{j+1}$, obtained from $\phi_{jk}(x) = \phi(2^j x - k)$, $j, k \in \mathbf{Z}$ by taking L^2 limits of finite linear combinations, exhausts $L^2(\mathbf{R})$.

When the sequence $\{p_k\}$ in (1) corresponds to an orthonormal Riesz basis, the periodic function

$$p(\xi) = \sum_{k \in \mathbf{Z}} p_k \exp(2\pi i k \xi)$$

is called a *low-pass filter* (or, a *quadrature mirror filter*). The corresponding function ϕ is called a *scaling function.*[1]

How does one recognize such functions ϕ? More to the point, how does one recognize those sequences $\{p_k\}$ that correspond to such a ϕ by the relation (1)? If the sequence has a finite number of nonzero terms, the question has been definitively settled in two distinct ways. A. Cohen gave the first necessary and sufficient condition in a fundamental paper [1]. He concerned himself with low-pass filters. These functions have the property that $|p(0)|^2 = 1$ and $|p(\xi)|^2 + |p(\xi + \frac{1}{2})|^2 = 1$. Cohen's condition, described below, may be viewed as a restriction on the zeros of $p(\xi)$. About the same time, W. Lawton [10] gave another condition of a different nature. He constructed a specific matrix P from the sequence $\{p_k\}$, and considered the subspace of eigenvectors of P with eigenvalue one. If this subspace is one dimensional, then $\{p_k\}$ are the coefficients of a low-pass filter. Is Lawton's condition also necessary? The question was open for a while until it was settled

[1] The term "prescaling function" seems to be widely accepted when the translates of ϕ are not necessarily orthogonal. However, if $\{p_k\}$ corresponds to such a function (a Riesz basis), the function $p(\xi)$ does not seem to have an official name. We shall say that $p(\xi)$ is the *low-pass filter for a prescaling function.* It will be understood that the space corresponding to this prescaling function is V_0 in some multiresolution analysis. The basic period of $p(\xi)$ is always one in this paper.

affirmatively in the spring of 1990 by both Cohen and Lawton, independently. (See Daubechies' excellent account of their adventure in her monograph [4], Chapter 6, §6.3.)

Suppose we ask the slightly more general question: How do we recognize those finite sequences $\{p_k\}$ that are low-pass filters for a (possibly nonorthogonal) Riesz basis? In this case, the Cohen condition is again necessary and sufficient. (See Lemarié [12], Proposition 2.) The appropriate version of the Lawton condition for Riesz bases was proved by Cohen, Daubechies, and Feauveau [2] in a paper on biorthogonal wavelets.

One concludes from this discussion that there are two distinct, but equivalent, solutions to the characterization problem when $p(\xi)$ is assumed to be a trigonometric polynomial. What happens when $p(\xi)$ is assumed to be an arbitrary periodic function? This problem is posed in the notes at the end of Chapter 7 of the text by Hernández and Weiss [7].

Two characterizations of low-pass filters for scaling functions are already known. Papadakis, Šikić, and Weiss [14] (Theorem 3.16) have formulated one of them. The other is a probabilistic version, due to Dobrić, Hitczenko, and the author [5]; for this work, we owe a debt of gratitude to H. Šikić and G. Weiss for their many critical comments. However, neither one of these papers contains results that resemble what is done in this paper.

We obtain a complete characterization of the general low-pass filter for a prescaling function, following Lawton's strategy. This approach is the one that prevails; the Cohen condition fails to be necessary in the general case, and we see no way to rescue it. This situation was already pointed out in [6].

The method of proof involves probability and some martingale theory. Roughly speaking, we regard $p(\xi)$ as a transition operator. This idea is certainly not new; its application to wavelet theory appears in Conze and Raugi [3] in 1990. They quote Keane [9] (1972) and Jamison [8] (1964), among others, who have regarded functions $p(\xi)$ as transition operators. However, for wavelet applications, the theorems in the literature are too specialized. They are burdened with assumptions that abridge the generality of their conclusions.

More background.

Suppose that $\{p_k\}$ is a sequence with a finite number of nonzero terms. For this discussion, there is no loss of generality if we assume that the nonzero terms are indexed as $p_0, p_1, \ldots, p_N$. Suppose also that a compactly supported function $\phi(x)$ exists, that it belongs to $L^2(\mathbf{R})$, and satisfies condition (1) in the above definition of a multiresolution analysis. Now consider the $(2N-1)$-dimensional vector $\{e_k\}$, $k = -N+1, \ldots, N-1$ of inner products (the autocorrelation vector) with

$$e_k = \int \phi(x)\overline{\phi(x-k)}dx.$$

Because $\{p_k\}$ relates $\phi(x)$ to $\phi(2x)$ by the convolution relation described in (1), we have a linear relation between the autocorrelation vector for $\phi(x)$ and the autocorrelation vector for $2\phi(2x)$. This relation can be expressed by a matrix, but a more convenient way to do the computation is to take Fourier transforms. (This discussion is given in detail in Daubechies [4], pp.189-190.) The sequence $\{p_k\}$ is replaced by the function $p(\xi)$, the autocorrelation vector $\{e_k\}$ by the periodic function $e(\xi)$,

$$e(\xi) := \sum_{k\in\mathbf{Z}} |\hat{\phi}(\xi + k)|^2,$$

and the matrix by a multiplication operator defined on trigonometric polynomials, of the form

$$\mathbf{P} : f \longrightarrow |p(\xi/2)|^2 f(\xi/2) + |p(\xi/2 + 1/2)|^2 f(\xi/2 + 1/2).$$

The equation (1) implies that $\mathbf{P}(e) = e$. Lawton's theorem is the following:

THEOREM 1 (W. Lawton [10]). *Suppose that $p(\xi)$ is a trigonometric polynomial with $p(0) = 1$ and that the function $f(\xi) \equiv 1$ is invariant under $\mathbf{P}$. If the space of $\mathbf{P}$-invariant polynomials is one dimensional, then the function*

$$\hat{\phi} := \prod_{k=1}^{\infty} p(\xi/2^k)$$

is the Fourier transform of a scaling function $\phi(x)$, $x \in \mathbf{R}$.

Stated in this way, the proof is almost visible. Orthonormality is a condition on the autocorrelation vector $\{e_k\} : e_k = \delta_0(k)$. On the Fourier transform side, this means that $e(\xi) = 1$ almost everywhere. Because the function $p(\xi)$ is a trigonometric polynomial, it can be shown that $e(\xi)$ is a trigonometric polynomial. Therefore, if $\mathbf{P}(1) = 1$, and 1 is the unique normalized invariant polynomial and $\mathbf{P}e = e$, we must have $e(\xi) \equiv 1$ as Lawton's theorem states.

The question of whether this condition is necessary as well as sufficient was settled in the affirmative by both Cohen and Lawton [11], independently, as we mentioned above. Subsequently, Cohen, Daubechies, and Feauveau [2] extended this necessary and sufficient condition to cover polynomials $p(\xi)$ that are low-pass filters for a prescaling function. (This is not what they stated, but is implicit in their results. See Theorem 4.3; C1 and C3 and equation 4.10 of [2].)

THEOREM 2 (Cohen, Daubechies, Feauveau). *Let $p(\xi)$ be a trigonometric polynomial with $p(0) = 1$. If $p(\xi)$ is the low-pass filter for a prescaling function, then the function $e(\xi)$ is a strictly positive $\mathbf{P}$-invariant polynomial and is the unique such polynomial with $e(0) = 1$.*

Conversely, if the operator $\mathbf{P}$ has a unique, strictly positive invariant polynomial, then $p(\xi)$ is the low-pass filter for a prescaling function.

The proof of these theorems (the necessity part) relies on the fundamental result of Cohen [1], mentioned in the introduction. His necessary and sufficient condition on $p(\xi)$ to qualify as a low-pass filter pertains to the structure of the roots of $p(\xi)$: this polynomial is not allowed to have roots that are invariant under a certain dyadic transformation. (This condition is described in detail below.) As we shall see, this approach fails in the general case, and must be replaced with something more robust. This is where martingale theory enters the picture.

Lawton's theorem: the general case.

Suppose that $p(\xi)$ is an arbitrary complex-valued periodic function. At first glance, the basic problem seems to be: find necessary and sufficient conditions on $p(\xi)$ so that there exists an L^2-function $\hat{\phi}(\xi)$ such that $\hat{\phi}(2\xi) = p(\xi)\hat{\phi}(\xi)$, and such that the translates ϕ_k of ϕ form a Riesz basis for the space V_0 in a multiresolution analysis. In fact, we really have two questions: (a) Given $p(\xi)$, when is there an L^2-function $\hat{\phi}(\xi)$ such that $\hat{\phi}(2\xi) = p(\xi)\hat{\phi}(\xi)$? (b) Given such a $\hat{\phi}(\xi)$, when are the translates ϕ_k a Riesz basis for some V_0 in a multiresolution analysis. If both (a) and (b) have positive answers, then $p(\xi)$ is itself an L^2-function, since $2^{-1}\phi(x/2) = \sum p_k\phi(x-k)$ is in V_0 and $|\hat{\phi}(\xi)|^2$ belongs to $L^\infty(\mathbf{R})$. This is the first necessary condition for $p(\xi)$ to satisfy. Let us phrase question (a) in a weaker form: Consider the operator

$$\mathbf{p} : f(\xi) \to |p(\xi/2)|^2 f(\xi/2),$$

defined on $L^1 \cap L^\infty(\mathbf{R})$. When does $\mathbf{p}$ have a nontrivial fixed point $f(\xi) \geq 0$ that belongs to $L^1 \cap L^\infty(\mathbf{R})$? If (a) has a positive answer, then $f(\xi) = |\hat{\phi}(\xi)|^2$ is such a nontrivial fixed point. This is the second necessary condition.

A slight variation of the second necessary condition produces a third. Consider the operator $\mathbf{P}$, defined on $L^\infty(\mathbf{T})$, the essentially bounded periodic functions:

$$\mathbf{P} : f(\xi) \to |p(\xi/2)|^2 f(\xi/2) + |p(\xi/2 + 1/2)|^2 f(\xi/2 + 1/2).$$

If there is a function $\hat{\phi}(\xi)$ that satisfies (a) and (b), then

$$e(\xi) := \sum_{k\in\mathbf{Z}} |\hat{\phi}(\xi + k)|^2$$

is a bounded nontrivial P invariant function. Thus, if we have positive answers to questions (a) and (b), the function $p(\xi)$ is a 1-periodic function belonging to $L^2(\mathbf{T})$, and the operators $\mathbf{p}$ and $\mathbf{P}$ have nontrival invariant functions, in $L^1 \cap L^\infty(\mathbf{R})$ and $L^\infty(\mathbf{T})$, respectively.

Suppose we have a function that satisfies these three conditions. It is known that they are not sufficient to provide answers to (a) and (b). How much (or how little) must we add to obtain a set of sufficient conditions? It turns out that there is a fourth uniqueness condition that must be added to obtain a set of sufficient conditions. Surprisingly, this fourth condition is also necessary. Consider the following:

DEFINITION. Let $g(\xi) \in L^1 \cap L^\infty(\mathbf{R})$ be a fixed reference function. An arbitrary function $f(\xi)$ is said to be a.e. dyadically g-continuous at zero if the limits

$$\lim_{j\to\infty} \frac{f(\xi/2^j)}{|g(\xi/2^j)|^2} = \lim_{j\to\infty} \frac{f(-\xi/2^j)}{|g(-\xi/2^j)|^2}$$

exist and the value (a function of ξ) is constant almost everywhere. Here, $0/0 = 1$, so that $|g(\xi)|^2$ is a.e. dyadically g-continuous. When $f(\xi)$ is g-continuous, we will denote the value of the ratio by $f(0)/|g|^2(0)$.

The importance of this idea in wavelet analysis was recognized by Papadakis, Šikić, and Weiss in their paper [14]. In particular, it was Hrvoje Šikić and Guido Weiss who, rather forcefully, focused the author's attention on this notion, in connection with a previous paper [5]. They identified this type of continuity in the case where $g(\xi) \equiv 1$, and call it "almost everywhere dyadic continuity at zero." It seems that yet another qualifier g is needed, alas.

We can now state the principal result.

THEOREM 3. *If a 1-periodic function $p(\xi)$ is a low-pass filter associated with a prescaling function, then the following are true:*

(a) *$p(\xi)$ is an L^2 function;*

(b) *the operators $\mathbf{p}$ and $\mathbf{P}$ have nontrivial fixed points, $|\hat{\phi}(\xi)|^2 \in L^1 \cap L^\infty(\mathbf{R})$ and $e(\xi) \in L^\infty(\mathbf{T})$, respectively;*

(c) *the fixed point $e(\xi)$ of the operator $\mathbf{P}$ is the unique function in the class $D_\infty(\hat{\phi})$ defined as the set of functions $h(\xi) \geq 0$ such that*

(i) *$h(\xi)$ is a.e. dyadically $\hat{\phi}$-continuous at zero, with $h(0)/|\hat{\phi}|^2(0) = 1$;*

(ii) *both $h(\xi)$ and its reciprocal $h^{-1}(\xi)$ belong to $L^\infty(\mathbf{T})$.*

Conversely, assume that $p(\xi)$ is a 1-periodic measurable function that satisfies (a), (b), and (c). Then the function $p(\xi)$ is the low-pass filter for a scaling function ϕ, defined by $\hat{\phi}(\xi) := \prod_{j=1}^{\infty} p(\xi/2^j)$.

These conditions are sharp, and examples to support this claim are given below.

The original Lawton theorem, and its generalization by Cohen, Daubechies, and Feauveau, both have a feature that Theorem 3 does not have. In those theorems, it is assumed that $p(\xi)$ is a polynomial with $p(0) = 1$, and that the operator $\mathbf{P}$ has a strictly positive invariant polynomial. (In the Lawton theorem, one assumes that $\mathbf{P}(1) = 1$.) This assumption alone implies that the operator $\mathbf{p}$ has a fixed point $|\hat{\phi}(\xi)|^2$ in $L^1 \cap L^\infty(\mathbf{R})$. This fact was noted by Mallat [13] when $\mathbf{P}(1) = 1$; the general case was proved by Cohen, Daubechies, and Feauveau [2]. A version of their theorems may be stated for $p(\xi)$ that satisfy Dini-Lipschitz conditions of the following form. Let

$$w_0(p, \delta) := |p(\delta) - p(0)|.$$

Suppose that

$$\int_0^1 w_0(p, \delta) d\delta/\delta < \infty;$$

we then say that p is *Dini-Lipschitz continuous at zero.*

Let

$$w(p, \delta) := \sup_{0 \leq \xi < 1} |p(\xi + \delta) - p(\xi)|$$

and suppose that

$$\int_0^1 w(p,\delta)d\delta/\delta < \infty;$$

we then say that p is *uniformly Dini-Lipschitz continuous.*

Let D_∞ denote the class $D_\infty(\hat{\phi})$ when $\hat{\phi}(\xi) \equiv 1$.

THEOREM 4. (a) *Let $p(\xi)$ be a 1-periodic function that is Dini-Lipschitz continuous at zero, with $p(0) = 1$. If there exists a $\mathbf{P}$-invariant function $h(\xi)$ that belongs to the class D_∞, then the operator $\mathbf{p}$ has a fixed point $|\hat{\phi}(\xi)|^2$ that belongs to $L^1 \cap L^\infty(\mathbf{R})$. The function $p(\xi)$ is a low-pass filter associated with a scaling function if and only if there is a unique $\mathbf{P}$-invariant function in D_∞.*

(b) *Let $p(\xi)$ be uniformly Dini-Lipschitz continuous with $p(0) = 1$. The function $p(\xi)$ is a low-pass filter if and only if $\mathbf{P}(1) = 1$ and $h(\xi) \equiv 1$ is the unique strictly positive $\mathbf{P}$-invariant function $h(\xi)$ with $h(0) = 1$ in the class of all uniformly Dini-Lipschitz continuous functions.*

Notice the distinction between the uniqueness class D_∞ and the corresponding class in part (b) consisting of uniformly Dini-Lipschitz continuous functions. The uniqueness in (b) is stronger than in part (a). Let us explain this as follows:

When $p(\xi)$ is a polynomial with $p(0) = 1$, the operator $\mathbf{P}$ is injective on several domains D. We emphasize the domain by writing $(\mathbf{P}, D)$ to denote the operator. Consider the three pairs $(\mathbf{P}, D_i)$, $i = 0, 1, 2, 3$ where D_0 is the set of trigonometric polynomials, D_1 the set of uniormly Dini-Lipschitz continuous periodic functions, D_2 is the set of continuous periodic functions, and D_3 the set of uniformly bounded (everywhere defined) periodic functions that are continuous at zero. The operator $(\mathbf{P}, D_i)$ satisfies $\mathbf{P}(D_i) \subset D_i$ for $i = 0, 1, 2,$ when $p(\xi)$ is a polynomial.

Suppose that $(\mathbf{P}, D_0)$ has the constant function $h(\xi) \equiv 1$ as an invariant function, and that this is the unique invariant $h(\xi) \geq 0$, $h(0) = 1$ in the space D_0. Then we can show that if $h(\xi) \geq 0$, $h(0) = 1$, is any $\mathbf{P}$-invariant function in D_3, then $h(\xi) \equiv 1$. That is, $h(\xi) \equiv 1$ is the unique such invariant function for (P, D_3) also. Since $D_0 \subset D_1 \subset D_2 \subset D_3$, the same statement can be made for (P, D_i), $i = 1, 2$. It follows trivially that this is also the case for $(\mathbf{P}, D_\infty)$.

When $p(\xi)$ is not a polynomial, but uniformly Dini-Lipschitz continuous with $p(0) = 1$, then $\mathbf{P}(D_i) \subset D_i$ for $i = 1, 2, 3$. Supppose that $(\mathbf{P}, D_1)$ has $h(\xi) \equiv 1$ as an invariant function and that this is the unique invariant $h(\xi) \geq 0$ in D_1 with $h(0) = 1$. Then we can also show the same is true for $(\mathbf{P}, D_i)$, $i = 2, 3$. This is the assertion of part (b) of Theorem 4.

The situation changes radically when we assume that $p(\xi)$, $p(0) = 1$ is continuous but only Dini-Lipschitz continuous at zero. The appropriate uniqueness class is D_∞. In [6] we have shown that there exists a $p(\xi)$ that is C^∞ at zero, with $p(0) = 1$, and continuous everywhere, such that $(\mathbf{P}, D_2)$ has $h(\xi) \equiv 1$ as an invariant function in D_2. It turns out that this is the unique invariant function in D_2. This is because $h(\xi) \equiv 1$ is no longer the unique invariant function for $(\mathbf{P}, D_3)$. There exists an invariant $h(\xi)$ in D_3 such that $h(\xi) = 1$

except at a finite number of points, where $h(\xi) = 0$ and is unique in this class. Thus, in this case, the appropriate operator is $(\mathbf{P}, D_\infty)$ and $h(\xi) \equiv 1$, considered as a representative of the class of functions $h(\xi)$: $h(\xi) = 1$ a.e. may be said to be the unique invariant function.

Cohen's theorem: The general case.

As we have pointed out above, the first necessary and sufficient conditions for a polynomial $p(\xi)$ to be a low-pass filter were stated by A. Cohen [1]. We will discuss his theorem for the case of an orthonormal scaling function, although the same condition applies to the nonorthonormal situation. A necessary condition is that $|p(0)|^2 = 1$ and $|p(\xi)|^2 + |p(\xi + \frac{1}{2})|^2 = 1$, so we assume this at the outset. The first form of the Cohen condition concerns the structure of the zero set of $p(\xi)$:

(1) The polynomial $p(\xi)$ has no zeros of the form

$$\xi_0 = k/(2^N - 1) + 1/2, \quad 1 \leq k \leq 2^N - 2, \ \pmod 1.$$

(2) The following is a more user-friendly form of the same condition: There exists a compact set K containing zero as an interior point, and a $\delta > 0$ such that

$$\inf_{\xi \in K} |p(\xi/2^j)| \geq \delta.$$

Furthermore, the set K should be 1-translation congruent to $[-\frac{1}{2}, \frac{1}{2}]$ in the sense that

$$\sum_{k \in \mathbf{Z}} \chi_K(\xi + k) = 1$$

almost everywhere.

A third, equivalent form is stated in terms of the function $\hat{\phi}(\xi) = \prod_{j=1}^{\infty} p(\xi/2^j)$.

CONDITION (C). For every $\xi \in [-\frac{1}{2}, \frac{1}{2}]$ there exists a $k = k(\xi)$ such that $|\hat{\phi}(\xi + k)|^2 \geq \delta > 0$.

In case $p(\xi)$ is a polynomial, the infinite product defining $\hat{\phi}(\xi)$ converges uniformly on compact intervals containing zero. So, if Cohen's condition (2) holds, then (C) also holds. Conversely, if $|\hat{\phi}(\xi)|^2$ is continuous (which is certainly the case if $p(\xi)$ is a polynomial) and condition (C) holds, then we may obtain K by covering $[-\frac{1}{2}, \frac{1}{2}]$ with intervals, centered at $\xi + k(\xi)$, such that $|\hat{\phi}(\xi')|^2 \geq \delta/2 > 0$ for ξ' contained in any such interval. We can extract a finite subcovering to obtain K.

Condition (C) is always sufficient for $\hat{\phi}(\xi)$ to insure that the partial products

$$\prod_{j=1}^{n} p(\xi/2^j) \chi_{[-\frac{1}{2}, \frac{1}{2}]}(\xi/2^n)$$

to converge in L^2 to a scaling function $\hat{\phi}(\xi)$, as Cohen has shown. His argument reworked with condition (C) is presented in [5]. In fact, if condition (C) holds *almost everywhere*, instead of everywhere, it is still sufficient.

The question, then, is whether condition (C), or the almost everywhere version, is necessary? That is, if the partial products converge in L^2 to a scaling function $\hat{\phi}(\xi)$, does some form of condition (C) hold? In order to avoid trivialities, let us ask the question when $\hat{\phi}(\xi)$ is *continuous*, so that compactness may come into play. If the partial products converge in L^2 to a bonafide scaling function, then, by Poisson summation,

$$e(\xi) := \sum_{k \in \mathbf{Z}} |\phi(\xi + k)|^2 = 1$$

almost everywhere. Can one find the $k(\xi)$ for almost every $\xi \in [-\frac{1}{2}, \frac{1}{2}]$ to verify condition (C)? This can certainly be done if $e(\xi) \equiv 1$. However, it can happen that there are exceptional points where $e(\xi) = 0$. An example of a bad low-pass filter producing exceptional points is constructed in [5]. The filter $p(\xi)$ is C^∞ except at four points in $[-\frac{1}{2}, \frac{1}{2}]$. The scaling function $\hat{\phi}(\xi)$ is continuous, and $e(\xi) = 0$ at the exceptional points. Furthermore, it is shown that any example of a continuous $\hat{\phi}(\xi)$ such that $e(\xi) = 0$ for a nonempty set of measure zero has the unpleasant property that condition (C) fails to hold. That is, for every $\epsilon > 0$, there exists an open interval $U(\xi; \delta)$, centered at some $\xi \in [-\frac{1}{2}, \frac{1}{2}]$ of radius $\delta = \delta(\epsilon)$ such that

$$\sup_{k \in \mathbf{Z}} |\hat{\phi}(\xi' + k)| \leq \epsilon$$

for all $\xi' \in U(\xi; \delta)$. We will discuss this example and try to motivate the construction below. For the full account, however, we refer the reader to [5].

Given the results just quoted, it is natural to ask for properties of $p(\xi)$ that imply that the exceptional set for $e(\xi)$ is empty. A sufficient condition for this to happen is to require $p(\xi)$ to be uniformly Dini-Lipschitz continuous. This result is implicit in the reference [6]. (In [6] we prove that the exceptional set for $e(\xi)$ is empty provided $p(\xi)$ satisfies a uniform Hölder condition of order α, $0 < \alpha \leq 1$. However, Maurice Hasson, of Rutgers University, pointed out that the proof is valid under the more general Dini-Lipschitz condition.)

Proofs and examples.

Proof of Theorem 3.

Necessity. As we have already pointed out, if $p(\xi)$ is the low-pass filter for a prescaling function, both operators $\mathbf{p}$ and $\mathbf{P}$ have, respectively, solutions $|\hat{\phi}(\xi)|^2$ in $L^1 \cap L^\infty(\mathbf{R})$ and $e(\xi) \geq 0$ in $L^\infty(\mathbf{T})$. Furthermore, $e^{-1}(\xi)$ belongs to $L^\infty(\mathbf{T})$. This implies that the function $\gamma(x)$, whose Fourier transform is $e^{-1/2}(\xi)\hat{\phi}(\xi)$, is a scaling function for the same multiresolution analysis. (In fact, if $\alpha(\xi)$ belongs to $L^2(\mathbf{T})$, then $\alpha(\xi)\hat{\gamma}(\xi) = \alpha(\xi)e^{-1/2}(\xi)\hat{\phi}(\xi)$ is the Fourier transform of a function in V_0, and any function in V_0 can be written in this form. Furthermore, the translates γ_k are orthogonal: The map $\hat{\phi}(\xi) \rightarrow e^{-1/2}(\xi)\hat{\phi}(\xi)$ is the standard method of orthogonalization that commutes with translations by $k \in \mathbf{Z}$.)

The function $e(\xi)$ is a.e. dyadically $\hat{\phi}$-continuous at the origin. To verify this, we quote Hernández and Weiss [7]:

Theorem 5.2 (Chapter 7, page 382 [7]). *Necessary and sufficient conditions for an $L^2(\mathbf{R})$ function $\gamma(x)$ to be a scaling function are:* (a) $\sum |\hat{\gamma}(\xi + k)|^2 = 1$ *almost everywhere;* (b) $\lim_{j \to \infty} |\hat{\gamma}(\xi/2^j)| = 1$ *almost everywhere;* (c) *there exists a periodic function $m(\xi)$ in $L^2(\mathbf{T})$ such that $\hat{\gamma}(\xi) = m(\xi/2)\hat{\gamma}(\xi/2)$.*

Since $\hat{\gamma}(\xi)$ is the Fourier transform of a scaling function, we have

$$\lim_{j\to\infty} |\hat{\gamma}(\xi/2^j)|^2 = \lim_{j\to\infty} \frac{|\hat{\phi}(\xi/2^j)|^2}{e(\xi/2^j)} = 1,$$

almost everywhere. This is another way of saying that $e(\xi)$ is a.e. dyadically $\hat{\phi}$-continuous at zero. The question of uniqueness is the only issue to be resolved; this is the heart of the matter. Let us defer this until after we have established the sufficiency of these conditions.

Sufficiency. Suppose that the operator $\mathbf{p}$ has fixed point $|\hat{\phi}(\xi)|^2$ and that $e(\xi)$ is the unique $\mathbf{P}$-invariant function in the class D_∞. Then, by Theorem 5.2 of Chapter 7 of Hernández and Weiss [7], the ratio $|\hat{\phi}(\xi)|/e^{1/2}(\xi)$ is a scaling function for a multiresolution analysis. The low-pass filter corresponding to this scaling function is

$$m(\xi) = |p(\xi)|\big(e(\xi)/e(2\xi)\big)^{1/2}$$

and $0 \le m(\xi) \le 1$. This leads us to define

$$\tilde{m}(\xi) := p(\xi)\big(e(\xi)/e(2\xi)\big)^{1/2},$$

and note that

$$\tilde{m}(\xi) = \operatorname{sgn} p(\xi) m(\xi).$$

The fact that $\tilde{m}(\xi)$ is a low-pass filter corresponding to a scaling function is a consequence of the more general fact that the class of such low-pass filters is stable under multiplication by unimodular 1-periodic measurable functions: this is part of the content of Theorem 2 of the "Brothers" Wutnam [15]. For the sake of clarity here, let us indicate the proof from [15] that $\tilde{m}(\xi)$ is indeed, a scaling function. The crucial observation is that *any unimodular 1-periodic measurable function $\mu(\xi)$ may be written in terms of another (nonunique, nonperiodic) unimodular function $t(\xi)$ as follows:*

$$\mu(\xi) = t(2\xi)t^{-1}(\xi).$$

How does one find a $t(\cdot)$? Start with the annular set $S = [-1, -1/2] \cup [1/2, 1]$ and define $t(\xi) \equiv 1$ on S. (In fact, this choice is arbitrary. As we shall see, we can take $t(\xi)$ to be any function such that $|t(\xi)| = 1$ a.e. on S.) Now extend the definition of $t(\cdot)$ to the set $2S = [-2, -1] \cup [1, 2]$:

$$t(\xi) = t(\xi/2)\mu(\xi/2), \quad \xi \in 2S.$$

In the same way, the domain of definition of $t(\cdot)$ is extended to the bands 2^jS, $j \ge 0$, and $t(\xi)$ is unimodular. Also,

$$t(2\xi) = t(\xi)\mu(\xi)$$

for $\xi \in (-\infty, -1/2] \cup [1/2, \infty)$. Now we extend the definition of $t(\xi)$ to $[-1/2, -1/4) \cup (1/4, 1/2] = 2^{-1}S$ by setting

$$t(\xi) = t(2\xi)\mu^{-1}(\xi),$$

for $\xi \in 2^{-1}S$. Again, we see that $t(\xi)$ is unimodular and satisfies the desired relation. We can continue this procedure on the bands $2^{-j}S$, and so define $t(\xi)$ for all $\xi \neq 0$.

In our case, $\mu(\xi) = \operatorname{sgn} p(\xi) = p(\xi)/|p(\xi)|$ with the convention that $0/0 = 1$. If we write $\operatorname{sgn} p(\xi) = t(\xi)t^{-1}(\xi/2)$, we may define

$$\begin{aligned}\hat{\phi}(\xi) &:= t(\xi)|\hat{\phi}(\xi)| \\ &= t(\xi)t^{-1}(\xi/2)|p(\xi/2)|t(\xi/2)|\hat{\phi}(\xi/2)| \\ &= \operatorname{sgn} p(\xi/2)|p(\xi/2)|\big(t(\xi/2)|\hat{\phi}(\xi/2)|\big) \\ &= p(\xi/2)\hat{\phi}(\xi/2).\end{aligned}$$

Since $t(\cdot)$ is unimodular, all the conditions of Theorem 5.2, cited above, remain valid, and $\hat{\phi}(\xi)$ is a bonafide scaling function for some multiresolution analysis.

Uniqueness. We assume that $p(\xi)$ is the low-pass filter for a prescaling function $\hat{\phi}(\xi)$. Then the function $e(\xi)$ belongs to the class $D_\infty(\hat{\phi})$. We assert that if $h(\xi)$ is another such function, then $h(\xi) = e(\xi)$ for almost every ξ.

Consider the ratio $|\hat{\gamma}(\xi)|^2 = |\hat{\phi}(\xi)|^2/e(\xi)$; since $\hat{\phi}(\xi)$ is a prescaling function, this function is the squared modulus of a scaling function $\hat{\gamma}(\xi)$. In particular, we have $\sum_{k\in\mathbf{Z}} |\hat{\gamma}(\xi+k)|^2 = 1$ for almost every ξ, $0 \leq \xi \leq 1$, and

$$\lim_{j\to\infty} |\hat{\gamma}(\xi/2^j)|^2 = 1$$

almost everywhere. We define the value of $|\hat{\gamma}(\xi)|^2$ at $\xi = 0$ as 1. This means that $|\hat{\gamma}(\xi + k)|^2$, $k \in \mathbf{Z}$ may be interpreted as a probability distribution on $\mathbf{Z}$, for almost every ξ. In fact, more is true. We can write $|\hat{\gamma}(\xi)|^2$ as the almost everywhere limit of a sequence of partial products $|\hat{\gamma}_N(\xi)|^2$:

$$|\hat{\gamma}(\xi + k)|^2 = \lim_{n\to\infty} |\hat{\gamma}_N(\xi + k)|^2$$

for all $k \in \mathbf{Z}$ and for almost every ξ, $0 \leq \xi \leq 1$. Now we revert to the construction in [5], which we repeat here for the sake of clarity. Let $M(\xi) = |m(2\pi\xi)|^2$, where $m(\xi)$ is defined in the sufficiency part of the proof. Notice that $M(\xi)$ is a one-periodic function that satisfies $M(\xi) + M(\xi + 1/2) = 1$, and $M(0) = 1$. The basic probability space Ω for our discussion is the disjoint union of two spaces of infinite sequences ω with coordinates $\omega_i = 0$ or 1. We will represent elements of Ω by $\{0, 1\} \times \{0, 1\}^{\mathbf{N}}$; Ω^+ and Ω^- will denote sequences starting with 0 and 1, respectively. We identify integers with a subset of Ω in the following way. A positive integer k with dyadic expansion

$$k = \sum_{i=1}^{\infty} \omega_i(k)2^{i-1}$$

is represented by the sequence

$$(0, \omega_1(k), \omega_2(k), \ldots).$$

The integer zero is identified with the sequence that is identically zero. A negative integer k is represented by coefficients of dyadic expansion of $-(k+1)$ preceded by 1 (thus, for example, the sequence $(1, 0, 0 \dots)$ represents -1. We denote the sequences corresponding to nonnegative integers as $\mathbf{Z}^+$, and those corresponding to negative integers as $\mathbf{Z}^-$. Fix $k \in \mathbf{Z}$ and let $\mathbf{k_N} = \{\omega : \omega_i = \omega_i(k),\, 0 \le i \le N\}$ be the N dimensional Ω^+-cylinder that contains $\omega(k)$. For each $\xi \in [0,1]$ we define a probability Q_ξ^N, $0 \le \xi < 1$, on the set of all such cylinders by the following prescription. For $0 \le k \le 2^N - 1$, we set

$$Q_\xi^N(k) = \prod_{j=1}^{N} M\left(\frac{\xi + k}{2^j}\right).$$

We then have

$$\sum_{0 \le k < 2^N} \prod_{j=1}^{N} M\left(\frac{\xi + k}{2^j}\right) = 1,$$

where we used the basic fact that $M(\xi) + M(\xi + 1/2) = 1$. In the language of (conditional) probability,

$$M\left(\frac{\xi + k}{2^j}\right) = Q_\xi(\omega_j(k) \| \omega_{j-1}, \dots \omega_1),$$

and the above sum is computed by the standard successive conditioning procedure.

With this interpretation of $M(\frac{\xi+k}{2^j})$, we see that the product defines a probability on cylinders of Ω^+, and that

$$Q_\xi^N(\mathbf{k_N}) = Q_\xi^{N+1}(\mathbf{k_N}),$$

where $\mathbf{k_N}$ is the N-dimensional cylinder corresponding to $0 \le k \le 2^N - 1$. In order to define corresponding probabilities on Ω^- let us consider a "reflected" filter

$$\widetilde{M}(\xi) = M(-\xi).$$

This filter may also be used to construct a probability on the positive integers $0 \le k < 2^N$ in the same fashion, by setting for $0 \le \eta < 1$ and $0 \le \ell < 2^N$

$$\widetilde{Q}_\eta^N(\ell) = \prod_{j=1}^{N} \widetilde{M}\left(\frac{\eta + \ell}{2^j}\right).$$

We now define measures P_ξ^N on cylinders in Ω by setting

$$P_\xi^N(k) = \begin{cases} Q_\xi^{N+1}(k), & \text{if } 0 \le k < 2^N; \\ \widetilde{Q}_{1-\xi}^{N+1}\big(-(k+1)\big), & \text{if } -2^N \le k < 0. \end{cases}$$

Notice that there is a double reflection, on the function, and on the argument, and that P_ξ^N corresponds to $N+1$ products in Q's. This specification shows that P_ξ^N, $N \ge 0$, is a consistent family ($P_\xi^N(k) = P_\xi^{N+1}(k)$

for each fixed k), since each of the families Q_ξ^N, $N \geq 1$, and $\widetilde{Q}_{1-\xi}^N$, $N \geq 1$ are consistent. To see that P_ξ^N defines a probability on the integers $-2^N \leq k < 2^N$, notice that

$$\begin{aligned}\sum_{-2^N \leq k < 2^N} P_\xi^N(k) &= \sum_{0 \leq k < 2^N} Q^{N+1}(k) + \sum_{-2^N \leq k < 0} \widetilde{Q}_{1-\xi}^{N+1}\big(-(k+1)\big) \\ &= \sum_{0 \leq k < 2^N} Q_\xi^{N+1}(k) + \sum_{-2^N \leq k < 0} Q_\xi^{N+1}(2^{N+1}+k) \\ &= \sum_{0 \leq k < 2^{N+1}} Q_\xi^{N+1}(k) \\ &= 1.\end{aligned}$$

Therefore, P_ξ^N, $N \geq 1$ specifies a probability on the σ-field generated by the cylinders.

Therefore, by the basic Kolmogorov theorem, the family P_ξ^N has an extension to a probability P_ξ on the Borel sets of Ω. The fact that $\hat{\gamma}(\xi)$ is a scaling function (so that $\sum |\hat{\gamma}(\xi+k)|^2 = 1$ a.e.) means that the family P_ξ^N is "tight" in the Prokorov sense, on the set of "finite" sequences, those $\omega \in \Omega$ such that $\omega(j) \equiv 0$ for $j \geq n(\omega)$ for some $n(\omega) \in \mathbf{N}$. Therefore, P_ξ is concentrated on the finite sequences; we say $P_\xi(\mathbf{Z}) = 1$ for almost every ξ, $0 \leq \xi \leq 1$. This means that if we fix ξ where $P_\xi(\mathbf{Z}) = 1$, then the coordinate process $X_j(\omega) = \omega(j)$ converges almost everywhere to zero relative to P_ξ. Now we can ask the question: What is the conditional probability distribution of $X_j(\omega)$ given $X_{j-1}(\omega), \ldots, X_0(\omega)$? For sample points ω that are finite sequences, we can compute according to the prescription just given:

$$\begin{aligned}P_\xi(X_0 = 0) &= P_\xi^N(X_0 = 0) \\ &= P_\xi^0(X_0 = 0)\end{aligned}$$

since the probabilities P_ξ^N are consistent. By the prescription,

$$P_\xi^0(X_0 = 0) = |m(\xi/2)|^2$$

and

$$P_\xi^0(X_0 = 1) = |m\big((\xi-1)/2\big)|^2$$

where $|m(\xi)|^2 = |p(\xi)|^2 e(\xi)/e(2\xi)$.

The conditional probabilities:

$$\begin{aligned}P_\xi(X_1 = 1 \| X_0 = 0) &= |m\big((\xi/2+1)/2\big)|^2; \\ P_\xi(X_1 = 0 \| X_0 = 0) &= |m(\xi/22)|^2,\end{aligned}$$

and

$$\begin{aligned}P_\xi(X_1 = 1 \| X_0 = 1) &= |m\big((\xi-1)/22 - 1/2\big)|^2; \\ P_\xi(X_1 = 0 \| X_0 = 1) &= |m\big((\xi-1)/22\big)|^2.\end{aligned}$$

We see that the coordinate process $X_j(\omega)$ generates a Markov process $\xi_{j+1}(\omega) = \big(\xi + k_j(\omega)\big)/2^{j+1}$ on $[-1, 1]$, where $k_j(\omega)$ is the sum of the coordinate variables:

$$|k_j(\omega)| := X_0(\omega) + \sum_{i=1}^{j} X_i(\omega) 2^{i-1}$$

and

$$\operatorname{sgn} k_j(\omega) = 1 - 2X_0(\omega).$$

We set $\xi_0(\omega) = \xi$. If ω corresponds to a negative integer, the process $\xi_j(\omega) < 0$ for $j > 0$, and $\xi_j(\omega)$ converges to zero from the left. If ω corresponds to a nonnegative integer, $\xi_j(\omega) \geq 0$ and converges to zero from the right. The transition probability for this process is $|m(\xi)|^2$; that is,

$$P_\xi(\xi_{j+1} \| \xi_j) = \begin{cases} |m(\xi_j/2)|^2. \\ |m\big((\xi_j \pm 1)/2\big)|^2 \end{cases}$$

Now let us discuss the uniqueness question. Let $h(\xi)$ be a $\mathbf{P}$-invariant function in the class $D_\infty(\hat{\phi})$ and consider the ratio $r(\xi) = h(\xi)/e(\xi)$. We must prove that $r(\xi) = 1$ almost everywhere. The function $r(\xi)$ is in the class D_∞; that is, $\lim_{j\to\infty} r(\xi/2^j)$ tends to the limit $r(0) = 1$ for almost every ξ, and the sequence $r(\xi/2^j) > 0$ for all $j \geq 0$, for almost every ξ. Now note that $r(\xi)$ is invariant under the operator $\mathbf{M}$ defined on $L^\infty(\mathbf{T})$ by

$$\mathbf{M} : h(\xi) \longrightarrow |m(\xi/2)|^2 h(\xi/2) + |m(\xi/2 + 1/2)|^2 h(\xi/2 + 1/2).$$

This is simply a consequence of the definitions. If we compose $r(\xi)$ with the Markov process $\xi_j(\omega)$, the composition $r\big(\xi_j(\omega)\big)$ is a martingale. (The function $r(\xi)$ is "harmonic" with respect to $\mathbf{M}$.) That is, the conditional expectation

$$\begin{aligned} E\big(r(\xi_{j+1}) \| r(\xi_j), \ldots, r(\xi_0)\big) &= E\big(r(\xi_{j+1}) \| r(\xi_j)\big) \\ &= \mathbf{M}(r)(\xi_j) \\ &= r(\xi_0). \end{aligned}$$

The martingale $r(\xi_j)$ is strictly positive, bounded, and P_ξ-almost surely convergent to the constant $r(0)$ since $\xi_j(\omega) \to 0$ almost surely (P_ξ). However, by the Lebesgue dominated convergence theorem,

$$\begin{aligned} r(0) &= E\big(\lim_{n\to\infty} r(\xi_n) \| r(\xi_j)\big) \\ &= \lim_{n\to\infty} E\big(r(\xi_n) \| r(\xi_j)\big) \\ &= r(\xi_j) \end{aligned}$$

for all $j \geq 0$. This means that $r(0) = r\big(\xi_j(\omega)\big) \equiv r(\xi_0)$ almost surely (P_ξ). That is,

$$r(0) = h(\xi)/e(\xi).$$

Since we know that $r(0) = 1$, this means $h(\xi) = e(\xi)$ almost everywhere, and the uniqueness assertion of Theorem 3 is proved.

Proof of Theorem 4. (a) We assume that $p(\xi)$ is Dini-Lipschitz continuous at zero, $p(0) = 1$ and that there exists a function $h(\xi)$ that is $\mathbf{P}$-invariant and belongs to D_∞. This implies that the operator $\mathbf{p}$ has a fixed point. When $h(\xi) = 1$, the argument appears in Mallat's paper [13]. The general case is proved along the same lines, and this argument has already been used in the proof of sufficency for Theorem 3. The difference here is that the notion of $\hat{\phi}$-continuous makes no sense until $|\hat{\phi}(\xi)|^2$ is defined. However, in this case we simply define the partial products

$$|\hat{\gamma}_N(\xi)|^2\chi_{[-\frac{1}{2},\frac{1}{2})}(\xi/2^N) = \prod_{j=1}^{N} |m(\xi/2^j)|^2\chi_{[-\frac{1}{2},\frac{1}{2})}(\xi/2^n)$$

with $m(\xi) = |p(\xi)|\big(h(\xi)/h(2\xi)\big)^{1/2}$. These partial products converge to

$$|\hat{\gamma}(\xi)|^2 = \prod^{\infty} |p(\xi/2^j)|^2/h(\xi)$$

because $h(\xi)$ is a.e. dyadically continuous at zero, with $h(0) = 1$. The sequence of partial products defined above is also uniformly bounded in $L^1 \cap L^\infty(\mathbf{R})$. (This is the result of a Fubini argument that is, by now, quite standard. (See Hernández and Weiss [7], Theorem 4.8, Chapter 7.) However, this is even more obvious if we take the point of view developed in [5] and used in the proof of uniqueness. There, we have seen that the sequence of partial products, $|\hat{\gamma}_{N+1}(\xi)|^2$ restricted to $-2^N \leq \xi < 2^N$, may be written as a sequence of probabilities P_ξ^N, concentrated on the integers $-2^N \leq k < 2^N$, with parameter ξ, $0 \leq \xi \leq 1$. Written this way,

$$\int |\hat{\gamma}_{N+1}(\xi)|^2\chi_{[-\frac{1}{2},\frac{1}{2})}(\xi/2^{N+1})d\xi = \int_0^1 \sum_k P_\xi^N(k)d\xi = 1.$$

Therefore, $|\hat{\gamma}(\xi)|^2$ belongs to $L^1 \cap L^\infty(\mathbf{R})$, and

$$\int |\hat{\gamma}(\xi)|^2 d\xi \leq 1.$$

This means that

$$\begin{aligned}|\hat{\phi}_N(\xi)|^2 &:= \prod_{j=1}^{N} |p(\xi/2^j)|^2 \\ &= |\hat{\gamma}_N(\xi)|^2\big(h(\xi/2^n)/h(\xi)\big)^{-1}\end{aligned}$$

and so

$$\begin{aligned}|\hat{\phi}(\xi)|^2 &:= \prod^{\infty} |p(\xi/2^j)|^2 \\ &= h(\xi)|\hat{\gamma}(\xi)|^2\end{aligned}$$

also belongs to $L^1 \cap L^\infty(\mathbf{R})$. The function $|\hat{\phi}(\xi)|^2$ is obviously $\mathbf{p}$-invariant, so that the proof of part (a) of Theorem 4 is complete.

(b) Here we assume that $p(\xi)$ is uniformly Dini-Lipschitz continuous with $p(0) = 1$ and that $\mathbf{P}(1) = 1$. If $p(\xi)$ is a scaling function, then we know from Theorem 3 that the unique $\mathbf{P}$-invariant function $e(\xi)$ that belongs to D_∞ satisfies $e(\xi) = 1$ almost everywhere. However, the result of [6] is that $e(\xi) \equiv 1$. The natural domain for the operator $\mathbf{P}$ is the class of functions that are uniformly Dini-Lipschitz continuous, so that the uniqueness condition is made with respect to this class. If $p(\xi)$ is a polynomial, then the natural domain for $\mathbf{P}$ is the class of polynomials; if we phrase the theorem in these terms, we obtain the original Lawton theorem.

Remarks and examples.

(i) The $\hat{\phi}$-continuity condition in Theorem 3 is necessary. If $\hat{\phi}(\xi)$ is a scaling function and $h(\xi)$ is a periodic function such that both $h(\xi)$ and $h^{-1}(\xi)$ belong to $L^\infty(\mathbf{T})$, then $\alpha(\xi) = h(\xi)\hat{\phi}(\xi)$ is a prescaling function. The function $|\alpha(\xi)|^2$ may be a.e. dyadically discontinuous at zero: $\lim_{j\to\infty} |h(\xi/2^j)|^2$ may not exist for any ξ. The function $|h(\xi)|^2$ is, of course, a.e. dyadically $\hat{\phi}$-continuous at zero.

(ii) The class $D_\infty(\hat{\phi})$ cannot be replaced by a similar class of continuous functions, even when $p(\xi)$ itself is continuous. We have made this observation in the discussion following Theorem 4.

(iii) Perhaps, Dini-Lipschitz continuity is not the last word as a condition for Theorem 4. However, some regularity is needed. In part (a), if we assume that $p(\xi)$ is continuous with $p(0) = 1$, it can happen that the operator $\mathbf{p}$ has an invariant function $|\hat{\phi}(\xi)|^2$ that belongs to $L^1 \cap L^\infty(\mathbf{R})$, and $\mathbf{P}$ has a unique invariant function in D_∞. However, $p(\xi)$ need not be a low-pass filter for a prescaling function. An example of this nature, where $p(\xi)$ is discontinuous, appears in Papadakis, Šikić, and Weiss, and is attributed to M. Paluszyński. Let $p(\xi)$ be the periodic extension of $\chi_{[0,\frac{1}{2})}(\xi)$. Then the infinite product $|\hat{\phi}(\xi)|^2 = \chi_{[0,1)}(\xi)$, and $\mathbf{P}(1) = 1$. Now this example can be modified so that $p(\xi)$ is continuous, and such that $|\hat{\phi}(\xi)|^2 \equiv 0$ for $\xi < 0$, but $|\hat{\phi}(\xi)|^2$ is continuous for $\xi > 0$ and right continuous at zero, with $|\hat{\phi}(0)|^2 = 1$. We simply modify the previous example so that $p(\xi)$ is continuous at $\xi = 1/2$ and $\xi = 1$, but so that the infinite product $|\hat{\phi}(\xi)|^2$ has the desired properties. Now we still have $|p(\xi)|^2 + |p(\xi + 1/2)|^2 = 1$ (the function $p(\xi)$ is a "generalized filter" in the terminology of Papadakis, Šikić, and Weiss [14]). This means that we can construct a probability on Ω, as before. In this case, $P_\xi(\mathbf{Z}) < 1$, and $P_\xi(\mathbf{Z}^-) = 0$ where $\mathbf{Z}^-$ denotes the negative integers. However, we can choose $p(\xi)$ so that for $0 < \xi < 1$ we have $0 < p(\xi) < 1$ except at $\xi = 1/2$, where $p(\xi) = 0$. This means that $P_\xi^N(\mathbf{Z}^-) > 0$ for all N; but $\lim_{N\to\infty} P_\xi^N(\mathbf{Z}^-) = 0$, for every ξ, $0 \le \xi \le 1$. Also, we can verify that $P_\xi(k) > 0$ for every $k \ge 0$, and every $\xi \ne 1/2$. With this choice of $p(\xi)$, the operator $\mathbf{P}$ has a very strong uniqueness property: $\mathbf{P}(1) = 1$, and this is the unique normalized periodic $\mathbf{P}$-invariant function in $L^\infty(\mathbf{T})$. This can be seen as follows: Let $h(\xi)$ be an invariant function in $L^\infty(\mathbf{T})$. The probability P_ξ is split into two parts: the mass on Ω_+ is concentrated on $\mathbf{Z}^+$, but the mass on Ω_- is diffuse. Therefore, with $P_\xi(\ \|\Omega_+)$ probability one, $\xi_0(\omega), \xi_1(\omega), \ldots$ converges to zero from the right. (Recall that $P_\xi(\mathbf{Z}) = 1$.) The sequence $h(\xi_0), h(\xi_1), \ldots$ is a bounded (therefore, convergent) martingale. Because $P_\xi(k) > 0$ for every $k \ge 0$, the martingale convergence implies that

$$\lim_{j\to\infty} h\big((\xi + k)/2^j\big)$$

exists and is finite for every $k \ge 0$, for every ξ, $0 < \xi < 1$. The limit is a "tail event," and therefore, constant for almost every ξ. (See Proposition 1 of [5].) If we suppose that $\lim_{j\to\infty} h(\xi/2^j) = 1$, we have that $h(\xi) \equiv 1$ for $0 \le \xi < 1$. Since $h(\xi)$ is assumed to be periodic, $h(\xi) \equiv 1$. Now observe that we can dispense with the conditional probability $P_\xi(\ \|\Omega_+)$: the process $h(\xi_0), h(\xi_1), \ldots, h(\xi_j) \equiv 1$ is a martingale with respect to the unconditional probability P_ξ, and $h(\xi) \equiv 1$ is the unique $\mathbf{P}$-invariant function in $L^\infty(\mathbf{T})$. In spite of this strong uniqueness property, $p(\xi)$ fails to be a low-pass filter on two counts: $P_\xi(\mathbf{Z}) < 1$, and 1 is not

$\hat{\phi}$-continuous.

The author would like to thank the referees for their careful reading of the manuscript and useful remarks for improving the exposition.

Bibliography

1. A. Cohen, "Ondelettes, analyses multirésolutions, et filtres miroir en quadrature," *Ann. Inst. H. Poincaré, Anal. nonlinéaire* **7** (1990), 439-459, 1990.

2. A. Cohen, I. Daubechies and J. C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Comm. Pure Appl. Math.* **45** (1992), 485-560.

3. A. Conze and A. Raugi, "Fonctions harmonique pour un opérateur de transition, et applications," *Bull. Soc. Math. France* **118** (1990), 273-310.

4. I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Conference series, SIAM, Philadelphia, PA (Fourth printing), 1995.

5. V. Dobrić, R. F. Gundy and P. Hitczenko, "Characterizations of orthonormal scale functions: a probabilistic approach," *J. Geom. Anal.*, to appear.

6. R. F. Gundy, "Two remarks concerning wavelets: Cohen's criterion for low-pass filters and Meyer's theorem on linear independence." In *The Functional and Harmonic Analysis of Wavelets and Frames*, pp. 249-257; ed. by L.W. Baggett and D.R. Larson. "Contemporary Mathematics," vol. 247, American Mathematical Society, Providence, RI, 1999.

7. E. Hernández and G. Weiss, *A First Course on Wavelets*, CRC Press, Inc., Boca Raton, FL, 1996.

8. B. Jamison, "Asymptotic behavior of successive iterates of continuous functions under a Markov operator," *J. Math. Analysis and Applications* **9** (1964), 203-214.

9. M. Keane, "Strongly mixing g measures," *Inventiones Math.* **16** (1972), 309-324.

10. W. Lawton, "Tight frames of compactly supported wavelets," *J. Math. Phys.* **31** (1990), 1898-1901.

11. W. Lawton, "Necessary and sufficient conditions for constructing orthonormal wavelet bases," *J. Math. Phys.* **32** (1991), 57-61.

12. P.-G. Lemarié, "Fonctions à support compact dans les analyses multi-résolutions," *Rev. Math. Iberoamericana* **7** (2) (1991), 157-182.

13. S. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbf{R})$," *Trans. Amer. Math. Soc.* **315** (1) (1989), 69-88.

14. M. Papadakis, H. Šikić and G. Weiss, "The characterization of low-pass filters and some basic properties of wavelets, scaling functions, and related concepts," *J. Fourier Anal. and Appl.* **5** (5) (1999), 495-521.

15. The Wutam Consortium, "Basic properties of wavelets," *J. Fourier Anal. and Appl.* **4** (4-5) (1998).