IEEE Transactions on Information Theory, Vol. 38, No. 2, March 1992, p. 876–881.

THE CHARACTERIZATION OF CONTINUOUS, FOUR-COEFFICIENT SCALING FUNCTIONS AND WAVELETS*

David Colella† and Christopher Heil†

Abstract. We examine four-coefficient dilation equations and give results converse to a theorem of Daubechies-Lagarias. These results complete the characterization of those four-coefficient dilation equations having a continuous solution.

A *wavelet basis* for $L^2(\mathbf{R})$ is an orthonormal basis $\{2^{n/2}\psi(2^n x - m)\}_{n,m\in\mathbf{Z}}$ generated from a single function ψ, the *wavelet*. The classical example is the *Haar system*, where $\psi = \chi_{[0,1/2)} - \chi_{[1/2,1)}$. Wavelet bases have many applications, e.g., image and speech processing. The variety of applications demands that wavelet bases having specific properties be available. It is therefore important to have means by which wavelets with desired properties can be constructed. One method is to solve a *dilation equation* or *two-scale difference equation* $f(t) = \sum_{k=-\infty}^{\infty} c_k\, f(2t-k)$. A functional solution f to a dilation equation is called a *scaling function*. The wavelet ψ can be realized as $\psi(t) = \sum (-1)^k\, c_{1-k}\, f(2t-k)$ whenever the scaling function f defines a multiresolution analysis, cf. Example 1. Recently, numerous papers have addressed the issue of constructing scaling functions, e.g., the various subdivision schemes in [1], [2], [3], [4], [5], [6], [7]. The purpose of this paper is to characterize four-coefficient dilation equations having continuous solutions. One part of this characterization (Theorem 1) was provided by Daubechies and Lagarias [4], [5]. We obtained converse results (Theorem 2) which completes this characterization and which shows that the bounds for smoothness obtained in [5] are exact [8].

In the case of four-coefficient systems, it is sufficient to consider only those $\{c_k\}$ for which at most c_0, c_1, c_2, c_3 are nonzero, i.e., dilation equations of the form

$$f(t) \;=\; c_0\, f(2t) \;+\; c_1\, f(2t-1) \;+\; c_2\, f(2t-2) \;+\; c_3\, f(2t-3). \tag{1}$$

We are interested in only those solutions f to (1) which are integrable, hence $\operatorname{supp} f \subset [0,3]$. We further require that the coefficients be real and satisfy $c_0 + c_2 = c_1 + c_3 = 1$. This condition is necessary for the existence of a multiresolution analysis. Thus, we have a two-parameter family of equations, and we select the independent parameters to be c_0 and c_3. Each scaling function f is then associated with a point in the (c_0, c_3)-plane. Since the time-reversed function $f(3-t)$ is a solution to the system determined by (c_3, c_0), any result obtained regarding this plane will be symmetric about the line $c_0 = c_3$. Finally, recall that a function f is Hölder continuous with Hölder exponent α if there exists a constant C for which $|f(x) - f(y)| \le C\, |x-y|^\alpha$ for all x, y. An extension of this definition to vector-valued functions is immediate.

---

* Manuscript received February 25, 1991; revised June 28, 1991.

† D. Colella is with The MITRE Corporation, 7525 Colshire Drive, McLean, VA 22102.

† C. Heil is with The MITRE Corporation, 7525 Colshire Drive, McLean, VA 22102 and the Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139 and was partially supported by National Science Foundation Postdoctoral Research Fellowship DMS-9007212.

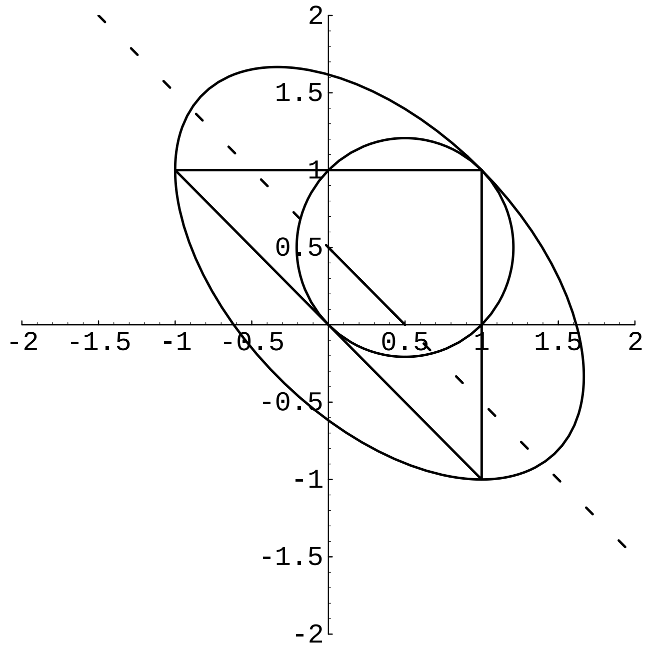

Fig. 1. *Circle of orthogonality, ellipse, dashed/solid line, and triangle.*

Example 1. To illustrate the range of phenomena that can occur even for the four-coefficient setting, we exhibit in Figure 1 several geometric objects in the (c_0, c_3)-plane: the ellipse $c_0^2 + c_3^2 - c_0 - c_3 + c_0c_3 = 1$; the circle $(c_0 - 1/2)^2 + (c_3 - 1/2)^2 = 1/2$; a triangle; and a solid/dashed line. In [9] we show that there are no integrable solutions to (1) for (c_0, c_3) on or outside the ellipse, with the exception of the point $(c_0, c_3) = (1, 1)$. There is an integrable solution to (1) for every point on and inside the circle. Neither of these results is sharp, and they are improved upon in [9]. Furthermore, Cohen [10] and Lawton [11] have characterized the N-coefficient systems $\{c_k\}_{k=0}^{N}$ that give a multiresolution analysis. In the (c_0, c_3)-plane, these are identified as all points on the circle except $(1, 1)$. We refer to the circle as the *circle of orthogonality* and associated scaling functions as *orthogonal scaling functions*. We see below that the triangle gives a first approximation to the region of points resulting in continuous scaling functions. The line identifies a necessary condition in order that a scaling function be differentiable. The intersection of this line and the circle is the well-known Daubechies scaling function D_4. Differentiable solutions to (1) do exist, and are indicated by the solid line segment in Figure 1. No four-coefficient orthogonal scaling function is differentiable (D_4 is differentiable "almost everywhere"). Finally, $(c_0, c_3) = (1, 0)$ gives rise to the scaling function $\chi_{[0,1)}$ whose wavelet basis is the Haar system. The points $(0, 1)$ and $(0, 0)$ are translates of this system, and $(1, 1)$ is a "stretched" version resulting in the non-orthogonal scaling function $\chi_{[0,3)}$.

We begin by outlining related work of [4], [5], where a sufficient condition is given for the continuity of scaling functions. We assume familiarity with the notation and main results in those papers. Given a dilation equation (1), continuity of the scaling function f can be examined in the following way. First define the matrix

$$M = \begin{pmatrix} c_1 & c_0 \\ c_3 & c_2 \end{pmatrix}.$$

The eigenvalues of M are 1 and $1 - c_0 - c_3$. In order for f to be continuous we must have $f(0) = f(3) = 0$. Using (1) with $t = 1, 2$, we obtain the values for $f(1)$ and $f(2)$ as components of the eigenvector associated with the eigenvalue 1 for M, cf. [12], therefore $f(1) = c_0$ and $f(2) = c_3$. (When $c_0 + c_3 \neq 0$, it is customary to normalize these values so that $f(1) + f(2) = 1$, cf. [5].) For f to be differentiable, M must also have the eigenvalue $1/2$, so that $c_0 + c_3 = 1/2$. This is the equation of the line plotted in Figure 1. No four-coefficient scaling function can have more than one continuous derivative.

Next define the matrices

$$T_0 = \begin{pmatrix} c_0 & 0 & 0 \\ c_2 & c_1 & c_0 \\ 0 & c_3 & c_2 \end{pmatrix} \qquad \text{and} \qquad T_1 = \begin{pmatrix} c_1 & c_0 & 0 \\ c_3 & c_2 & c_1 \\ 0 & 0 & c_3 \end{pmatrix},$$

and note that M is a submatrix of both T_0 and T_1. Assuming f exists, we let $v : [0, 1] \to \mathbf{R}^3$ be the vector-valued function

$$v(x) = \begin{pmatrix} f(x) \\ f(x+1) \\ f(x+2) \end{pmatrix}. \tag{2}$$

Then $v(0)$, $v(1)$ are right eigenvectors associated with the eigenvalue 1 for T_0, T_1, respectively. Let τ denote the left shift operator on $[0, 1]$, so that $\tau x = 2x$ for $0 \le x < 1/2$ and $\tau x = 2x - 1$ for $1/2 \le x \le 1$. Expanding x into its binary representation $x = .d_1 d_2 \ldots$, we have $v(x) = T_{d_1} v(\tau x)$ for all $x \in [0, 1]$. The definition $\tau(1/2) = 0$ can be replaced by $\tau(1/2) = 1$ without affecting this property of v since $T_0 v(1) = T_1 v(0)$. By induction, $v(x) = T_{d_1} \cdots T_{d_m} v(\tau^m x)$ for every $m > 0$. Using the matrices T_0, T_1 and the vector representation v, we can then determine the value of f at any dyadic point $\{k/2^n\}_{k,n \in \mathbf{Z}}$. For example, $(f(2^{-m}), f(1+2^{-m}), f(2+2^{-m}))^{\mathrm{t}} = v(2^{-m}) = T_0^{m-1} T_1 v(0) = T_0^m v(1)$. The scaling function f is continuous if and only if v is continuous and satisfies

$$v_1(0) = v_3(1) = 0, \tag{3}$$
$$v_{i+1}(0) = v_i(1) \quad \text{for } i = 1, 2, 3, \tag{4}$$
$$v(x) = T_{d_1} v(\tau x) \quad \text{for } x \in [0, 1]. \tag{5}$$

In the case that f is continuous, $v(2^{-m}) \to v(0) = (0, c_0, c_3)$. We compute

$$v(2^{-m}) - v(0) = \frac{1}{(1 - 2c_0 - c_3)} \begin{pmatrix} (1 - 2c_0 - c_3)\, c_0^{m+1} \\ (2c_0 - 1)\, c_0^{m+1} + c_3 (1 - c_0 - c_3)^{m+1} \\ c_3\, c_0^{m+1} + c_3 (1 - c_0 - c_3)^{m+1} \end{pmatrix}, \tag{6}$$

and therefore v continuous at zero forces $|c_0|, |1-c_0-c_3| < 1$. By symmetry, the continuity of v at 1 implies $|c_3|, |1-c_0-c_3| < 1$. Since the eigenvalues of T_0 and T_1 combined are 1, c_0, c_3, and $1-c_0-c_3$, continuity of f implies that all eigenvalues of T_0, T_1 other than 1 are less than one in absolute value and that the Hölder exponent of v is at most $-\log_2(\max\{|c_0|, |c_3|, |1-c_0-c_3|\})$. The region of points $\{(c_0, c_3) : |c_0|, |c_3|, |1-c_0-c_3| < 1\}$ is the interior of the triangle shown in Figure 1. Some orthogonal scaling functions are therefore discontinuous, and, in fact, unbounded.

Conversely, a scaling function f can sometimes be constructed from a vector function v. Given (c_0, c_3), set $v(0) = (0, c_0, c_3)^{\mathrm{t}}$ and $v(1) = (c_0, c_3, 0)^{\mathrm{t}}$, the right eigenvectors for the eigenvalue 1 for T_0, T_1. Then define v on the set of dyadic points on $[0,1]$ using (5). Conditions on the coefficients $\{c_k\}$ that give uniform continuity for v on the dyadics allows us to extend v to a continuous function on $[0,1]$. If v also satisfies (3)–(4), the continuous scaling function f is then obtained from v by (2). More specifically, we seek bounds on the Hölder exponent of v, which is identical to the Hölder exponent of f. The function f constructed in this way is the unique integrable solution to (1), up to multiplication by a constant, and we say that f, respectively v, is the *scaling function*, respectively, *vector scaling function, associated with the point* (c_0, c_3).

Obtaining the appropriate bounds for the Hölder exponent requires an examination of the joint spectral radius (defined below) for the matrix operators T_0, T_1 restricted to the two-dimensional subspace $V = \{u \in \mathbf{R}^3 : u_1 + u_2 + u_3 = 0\}$, cf. [5]. In particular, since $(1,1,1)$ is a common left eigenvalue for T_0 and T_1, the action of T_i on V is equivalent to the 2×2 matrices S_0, S_1 acting on $\mathbf{R}^2$, where

$$S_0 = \begin{pmatrix} c_0 & 0 \\ -c_3 & 1-c_0-c_3 \end{pmatrix} \qquad \text{and} \qquad S_1 = \begin{pmatrix} 1-c_0-c_3 & -c_0 \\ 0 & c_3 \end{pmatrix}.$$

Except for the eigenvalue 1, the eigenvalues of S_0, S_1 are the same as those of T_0, T_1, respectively. Before we can state the result of [5], we need some definitions.

Let $\|\cdot\|$ be a norm on $\mathbf{R}^3$, with corresponding operator norm $\|A\| = \sup\{\|Au\|/\|u\| : u \neq 0\}$ for a matrix A. Typical norms are $\|u\|_p = (|u_1|^p + |u_2|^p + |u_3|^p)^{1/p}$ for fixed $p \geq 1$, and $\|u\|_\infty = \max\{|u_1|, |u_2|, |u_3|\}$ for $p = \infty$.

Definition 1. The *joint spectral radius* $\hat{\rho}(A_0, A_1)$ of two matrices A_0, A_1 is defined by $\hat{\rho}(A_0, A_1) = \limsup \lambda_m$, where $\lambda_m = \max_{d_j=0,1} \|A_{d_1} \cdots A_{d_m}\|^{1/m}$.

This definition generalizes the usual spectral radius of a single matrix A, which is given by $\rho(A) = \limsup \|A^m\|^{1/m} = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$. An extension to larger collections of matrices is made in the obvious way. To our knowledge, the joint spectral radius was first discussed by Rota and Strang [13]. Some recent papers on the subject include [14], [15]. The joint spectral radius is independent of the choice of norm $\|\cdot\|$, and of the choice of basis, that is, $\hat{\rho}(A_0, A_1) = \hat{\rho}(B_0, B_1)$ whenever $B_i = BA_iB^{-1}$ for any fixed matrix B, cf. [5]. We are interested in $\hat{\rho}(S_0, S_1) = \hat{\rho}(T_0|_V, T_1|_V)$. The importance of the joint spectral radius in this work is not accidental, for example, cf. [6], where a probabilistic approach is considered.

The first step in characterizing the continuous scaling functions is provided by the following theorem.

THEOREM 1 [5]. *Given* (c_0, c_3), *if* $\hat{\rho}(S_0, S_1) = \hat{\rho}(T_0|_V, T_1|_V) < 1$ *then the associated scaling function is Hölder continuous with Hölder exponent* $\alpha \geq -\log_2 \hat{\rho}(S_0, S_1) - \varepsilon$ *for every* $\varepsilon > 0$.

Let $\sigma_m = \max_{d_j=0,1} \rho(S_{d_1} \cdots S_{d_m})^{1/m}$, so that σ_m is the largest absolute eigenvalue that occurs among all products of length m of the matrices S_0, S_1. Then $\sigma_m \leq \hat{\rho}(S_0, S_1) \leq \lambda_m$ for every m. Berger and Wang [14] prove that $\hat{\rho}(S_0, S_1) = \limsup \sigma_m$. It therefore follows that $\sup \sigma_m = \hat{\rho}(S_0, S_1) = \lim \lambda_m$. In particular, $\hat{\rho}(S_0, S_1) \geq \sigma_1$ and in general one expects $\hat{\rho}(S_0, S_1)$ to be strictly larger than σ_1. Since $\sigma_1 = \max\{|c_0|, |c_3|, |1 - c_0 - c_3|\}$, the set $\{(c_0, c_3) : \sigma_1 < 1\}$ is the triangular region in Figure 1.

Now consider the set $CS = \{(c_0, c_3) : \hat{\rho}(S_0, S_1) < 1\}$. Theorem 1 implies that the dilation equation (1) has a continuous solution for every point $(c_0, c_3) \in CS$. A major goal of this paper is to (numerically) determine the region where continuous solutions exist. We do this by bounding $\hat{\rho}(S_0, S_1)$ by λ_m and σ_m for various m. Since λ_m is norm dependent, let $\lambda_m(p)$ denote that value of λ_m obtained using the norm $\|\cdot\|_p$, and define the set $C_m(p) = \{(c_0, c_3) : \lambda_m(p) < 1\}$. Every $C_m(p)$ is a subset of CS, so each point in $C_m(p)$ is associated with a continuous scaling function. Since for each fixed p, $C_m(p)$ increases to the set CS, we can approximate CS with $C_m(p)$ for various m and p. The regions $C_1(1)$, $C_1(2)$, $C_1(\infty)$, determined by considering products of length one and the $\|\cdot\|_1$-norm, $\|\cdot\|_2$-norm, and $\|\cdot\|_\infty$-norm, respectively, are plotted in [16]. Figure 2 shows a numerical estimation of the set $C_{16}(1)$.

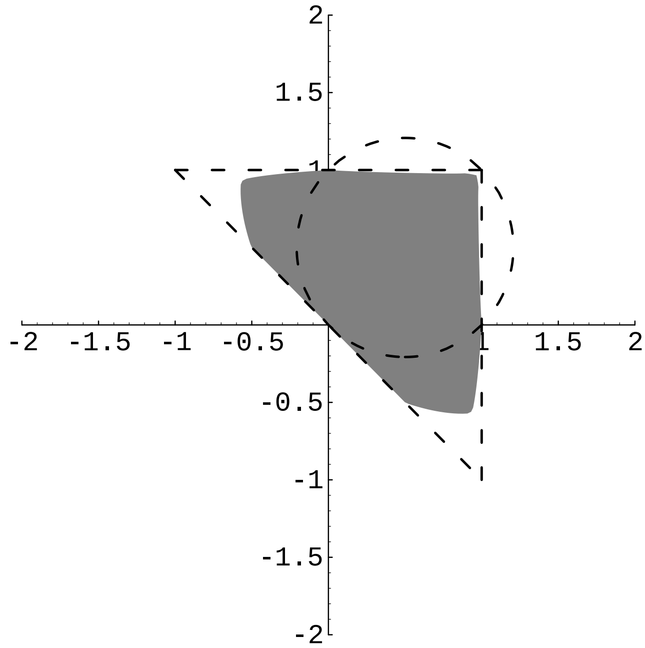

Fig. 2. *The set* $C_{16}(1) = \{(c_0, c_3) : \lambda_{16} < 1\}$ (*shaded area*).

The calculation of λ_m or σ_m can become computationally intensive even for small values of m. This makes the exact determination of the joint spectral radius nearly impossible, except in special cases. One case when it is possible is the following.

LEMMA 1. *If $\{S_0, S_1\}$ can be simultaneously symmetrized, i.e., there exists an invertible matrix B such that BS_iB^{-1} is real and symmetric for $i = 0, 1$, then $\hat{\rho}(S_0, S_1) = \sigma_1$.*

Tedious but elementary calculations allow us to determine those points of the (c_0, c_3)-plane for which simultaneous symmetrization is possible.

PROPOSITION 1. *If $\sigma_1 < 1$ and if $R = c_0\, c_3\, (1 - c_0 - c_3)\, (1 - 2c_0 - 2c_3) \geq 0$ then the scaling function associated with (c_0, c_3) is continuous with Hölder exponent exactly $-\log_2 \sigma_1$.*

Lemma 1 holds also when S_0, S_1 can be simultaneously Hermitianized. One might then hope that the hypothesis of Proposition 1 can be weakened to include that case as well. However, for the matrices S_0, S_1 simultaneous symmetrization is possible exactly when simultaneous Hermitianization is possible.

EXAMPLE 2. The region SS, defined by $\sigma_1 < 1$ and $R \geq 0$, is depicted in Figure 3.

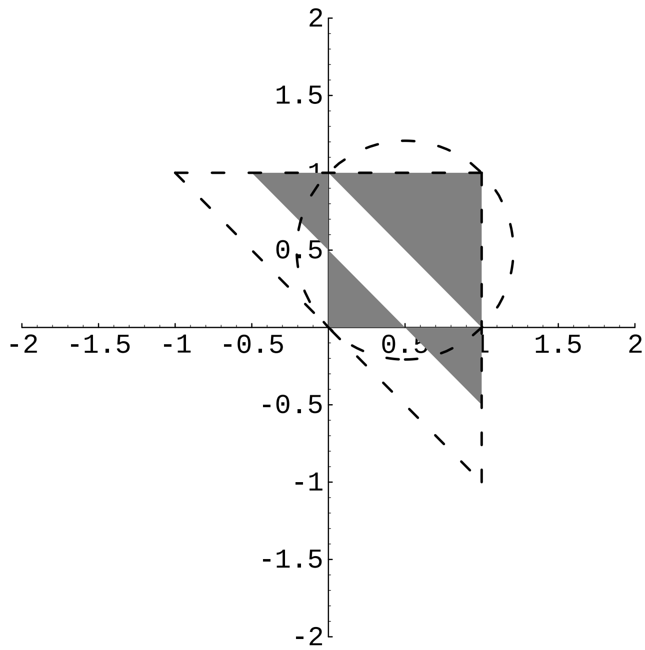

Fig. 3. *The region SS where simultaneous symmetrization is possible and leads to continuous scaling functions (shaded area).*

If simultaneous symmetrization is not possible and it is not practical to compute λ_m directly, it may still be possible to bound $\hat{\rho}(S_0, S_1)$ from above by computing a small number of matrix product norms. The following result can be found in [5].

Lemma 2. *Let $\{P_j\}$ be a set of building blocks of products of the matrices S_i, i.e.,*

(a) *each P_j is some product of m_j of the matrices S_i, and*

(b) *there is some $r \geq 0$ such that if P is any product of the matrices S_i then $P = P_{j_1} \cdots P_{j_k} Q$, where Q is some product of at most r of the matrices S_i.*
Then $\hat{\rho}(S_0, S_1) \leq \sup \|P_j\|^{1/m_j}$.

The following recursive algorithm can be used to implement Lemma 2.

Algorithm 1. Given $\hat{\rho} > \hat{\rho}(S_0, S_1)$. For each of the matrices S_0, S_1 in turn, implement the following recursion.

Given a product $P = S_{d_1} \cdots S_{d_m}$. If $\|P\|^{1/m} < \hat{\rho}$ then keep P as a building block. Otherwise, repeat this step with each of the products PS_0, PS_1.

This recursion will clearly terminate since there must exist some λ_m with $\hat{\rho}(S_0, S_1) \leq \lambda_m < \hat{\rho}$. The resulting set of building blocks will satisfy Lemma 2 with $r = l - 1$, where l is the length of the longest product in the set of building blocks. An analogous recursive algorithm for bounding $\hat{\rho}(S_0, S_1)$ from below based on the eigenvalues of products of matrices is not possible, since the eigenvalues of a product are not, in general, related to the eigenvalues of the matrices in the product.

Example 3. We consider points on the circle of orthogonality. If we could always simultaneously symmetrize S_0 and S_1 then the resulting Hölder exponent would be exactly $-\log_2 \sigma_1$ for every point. On the circle, this expression is maximized when σ_1 is minimized, at $(c_0, c_3) = (0.6, -0.2)$. Then, $\sigma_1 = 0.6$, so the Hölder exponent would be ≈ 0.737 (note the Hölder exponent of D_4 is ≈ 0.550). Unfortunately, at $(0.6, -0.2)$ we have $R = -0.0144 < 0$, so symmetrization is not possible. We estimate the Hölder exponent by direct bounding of $\hat{\rho}(S_0, S_1)$. Considering all 2^{13} products of length 13, we find that $\hat{\rho}(S_0, S_1) \leq \lambda_{13}(1) \approx 0.682455 < 1$. The associated scaling function is therefore continuous, with Hölder exponent $\alpha \geq -\log_2 \lambda_{13}(1) \approx 0.551$, i.e., f is smoother than D_4. (Note $-\log_2 \lambda_m(1) < 0.550$ for $m = 1, \ldots, 12$.) Since the only orthogonal scaling function which lies on the line $c_0 + c_3 = 1/2$, and so the only one which "could be" differentiable, is D_4, it is surprising that there exists a four-coefficient, orthogonal scaling function smoother than D_4.

Algorithm 1 can be used to improve the above estimate. Allowing a maximum product length of 73 in that algorithm results in the bound $\hat{\rho}(S_0, S_1) \leq 0.660500$, which implies $\alpha \geq 0.598370$. This search required 14156 matrix norm computations, of which 7079 matrix products were selected as building blocks. A direct computation of $\lambda_{73}(1)$ would not improve this estimate (e.g., $\|S_0^2 S_1 S_0^{14} S_1 S_0^{14} S_1 S_0^{13} S_1 S_0^{12} S_1 S_0^{12} S_1\|^{1/73} = 0.663200$), and would, in fact, be impossible to carry out.

Since the above computation involved products of significant length, it is difficult to believe that the Hölder exponent of f could be much greater than 0.598370. In order to show this, we bound α from above by bounding $\hat{\rho}(S_0, S_1)$ from below. Direct computation of $2^{31} - 1$ matrix products reveals that $\max\{\sigma_1, \ldots, \sigma_{30}\} = \sigma_{13} = \rho(S_1 S_0^{12})^{1/13} \approx 0.659679$. Therefore, by Theorem 2 below, $\alpha \leq -\log_2 \sigma_{13} \approx 0.600164$. Furthermore, since the Hölder exponent of continuity for any scaling function is at least $-\log_2 \sigma_1$, we can easily identify points on the circle of orthogonality whose scaling functions cannot have Hölder exponent greater than 0.598370, the lower bound for the Hölder exponent for f. In particular, the smoothest orthogonal scaling function must come from a point somewhere near $(0.6, -0.2)$. We do not know if there exists an orthogonal scaling function smoother than f.

Theorem 1 implies continuity for the scaling function whenever $\hat{\rho}(S_0, S_1)$ is strictly less than one. We now consider the converse problem, i.e., whether this joint spectral radius condition exactly characterizes the continuous scaling functions.

Fix the point (c_0, c_3) and assume f is an integrable solution to (1). Note first that if $f(0) \neq 0$ (so f is discontinuous), then c_0, one of the eigenvalues for T_0, is one. Thus $\hat{\rho}(T_0|_V, T_1|_V) \geq 1$. A similar argument holds for $f(3)$ and c_3. We therefore assume $f(0) = f(3) = 0$, and that c_0, $c_3 \neq 1$.

Since our results depend on the eigenvalues of matrices, and since eigenvalues are in general complex, we need to consider the complex space $\mathbf{C}^3$. Let A be a matrix with real or complex entries and having eigenvalue λ. Then $U_\lambda = \{u \in \mathbf{C}^3 : (A-\lambda)^k u = 0 \text{ for some } k > 0\}$ is an A-invariant subspace of $\mathbf{C}^3$, for if $u \in U_\lambda$ then $Au = (A - \lambda)u + \lambda u \in U_\lambda$. The standard Jordan decomposition for A ensures that every vector $v \in \mathbf{C}^3$ can be written uniquely as a sum of vectors $v_\lambda \in U_\lambda$ for the distinct eigenvalues λ of A. We say that a vector $v \in \mathbf{C}^3$ *has a component in* U_λ if $v_\lambda \neq 0$.

THEOREM 2. *Let v be the vector scaling function associated with the point (c_0, c_3). Suppose $T = T_{d_1} \cdots T_{d_m}$ is any fixed product of the matrices T_0, T_1, and λ is an eigenvalue for $T|_V$. Let x be that point whose binary decimal expansion is $x = .d_1 \ldots d_m d_1 \ldots d_m \ldots$.*

(a) *If $|\lambda| > 1$, and if there exists some $z \in [0, 1]$ such that $v(z)$ has a component in U_λ then v is unbounded.*

(b) *If v is continuous and there exists some $z \in [0, 1]$ such that $v(x) - v(z)$ has a component in U_λ then $|\lambda| < 1$ and the Hölder exponent of v is at most $-\log_2 |\lambda|^{1/m}$.*

For four-coefficient systems, the requirement in Theorem 2 that $v(z)$, respectively $v(x) - v(z)$, has a component in the eigenspace U_λ for some z is always satisfied whenever $\lambda \neq 0$. As we shall see, Proposition 2 below implies this fact for all but one special case. When $\lambda = 0$, part (a) above does not apply and part (b) remains valid with essentially no bound on the Hölder exponent.

PROPOSITION 2. *Assume that $\{v(z) : z \in [0, 1]\}$ is not contained in any line in $\mathbf{C}^3$. Let T and x be defined as in Theorem 2, and let λ be any eigenvalue of $T|_V$. Then there exists a $z \in [0, 1]$ such that $v(z)$, respectively $v(x) - v(z)$, has a component in U_λ.*

We now proceed to show $\{v(z) : z \in [0, 1]\}$ is not contained in a line in $\mathbf{C}^3$ unless $1 - c_0 - c_3 = 0$. Recalling that $v(0) = (0, c_0, c_3)^{\mathrm{t}}$ and $v(1) = (c_0, c_3, 0)^{\mathrm{t}}$, we compute $v(1/2) = T_0 v(1) = (c_0^2, c_0(1 - c_0) + c_3(1 - c_3), c_3^2)^{\mathrm{t}}$. If it were the case that $v(0)$, $v(1/2)$, and $v(1)$ did lie on a line in $\mathbf{C}^3$ then $v(0) - v(1/2)$ would be a multiple of $v(0) - v(1)$. This implies immediately that $1 - c_0 - c_3 = 0$.

When $1 - c_0 - c_3 = 0$ then $\{v(z) : z \in [0, 1]\}$ does indeed lie on a line in $\mathbf{C}^3$. For example, if $c_0 = c_3 = 1/2$ then $v(z) = (z, 1, 1 - z)^{\mathrm{t}}$. We consider this case separately. Let $1 - c_0 - c_3 = 0$, and suppose $T = T_{d_1} \cdots T_{d_m}$. Then 1 is an eigenvalue for T since $(1, 1, 1)$ is a common left eigenvector for T_0 and T_1 for this eigenvalue. Let w_1 be a right eigenvector for T for the eigenvalue 1. Since $(c_3^2, -c_0\,c_3, c_0^2)$ is a common left eigenvector for T_0 and T_1 for the eigenvalue 0, that value is also an eigenvalue for T. Let w_3 be a right eigenvector for T for the eigenvalue 0. To find the remaining eigenvalue of T, note that $w_2 = (-c_0, c_0 - c_3, c_3)^{\mathrm{t}}$ is a right eigenvector for T_0 for the eigenvalue c_0 and is a right eigenvector for T_1 for the eigenvalue c_3. Therefore if k of the d_i are 0 and the remaining $m - k$ are 1 then w_2 is a right eigenvector for T for the eigenvalue $c_0^k\,c_3^{m-k}$. The three eigenvalues of T are distinct since c_0, $c_3 \neq 0$, 1, and so w_1, w_2, w_3 is a basis for $\mathbf{C}^3$. Therefore, there exist scalars α_i, β_i, γ_i such that $v(0) = \alpha_0\,w_1 + \beta_0\,w_2 + \gamma_0\,w_3$ and $v(1) = \alpha_1\,w_1 + \beta_1\,w_2 + \gamma_1\,w_3$. However, $v(0) - v(1) = (0, c_0, c_3)^{\mathrm{t}} - (c_0, c_3, 0)^{\mathrm{t}} = w_2$, and so $\alpha_0 = \alpha_1$, $\beta_0 = \beta_1 + 1$, and $\gamma_0 = \gamma_1$. Hence, either $v(0)$ or $v(1)$ must have a component in the eigenspace of T corresponding to the eigenvalue $c_0^k\,c_3^{m-k}$. This is one eigenvalue of $T|_V$. The other eigenvalue of $T|_V$ is zero, and although neither $v(0)$ nor $v(1)$ need have a component in the zero eigenspace, this eigenspace is irrelevant for the conclusions of Theorem 2.

A summary of the results for four-coefficient systems is given in the next theorem.

THEOREM 3. *Let f be the scaling function associated with the point (c_0, c_3).*

(a) *If $\hat\rho(S_0, S_1) < 1$ then f is continuous.*

(b) *If $\hat\rho(S_0, S_1) > 1$ then f is unbounded.*

(c) *If f is continuous then $\hat\rho(S_0, S_1) \leq 1$ with $\sigma_m < 1$ for every m and f is Hölder continuous with Hölder exponent $-\log_2 \hat\rho(S_0, S_1) - \varepsilon \leq \alpha \leq -\log_2 \hat\rho(S_0, S_1)$ for every $\varepsilon > 0$.*

Theorem 3(c) is almost a complete converse of Theorem 3(a). If it was the case that a solution existed to a dilation equation with $\sup \sigma_m = \hat\rho(T_0|_V, T_1|_V) = 1$ and $\sigma_m < 1$ for every m, then Theorem 3(a) would not imply that this solution was continuous, and Theorem 3(c) would not imply that it was discontinuous. However, it is shown in [8] that a solution for this special case must be discontinuous, and therefore continuity occurs if and only if $\hat\rho(S_0, S_1) < 1$. We do not know if this special case is possible.

EXAMPLE 4. For each m, define $E_m = \{(c_0, c_3) : \sigma_m \geq 1\}$. The set E_1 is the boundary and exterior of the triangle in Figure 1. By Theorem 2, no scaling function determined by a point in E_m can be continuous. In Figure 4 we display the curve which represents the boundary of the set E_{16}. This curve serves as an outer approximation for CS, i.e., every point on or outside the curve has $\hat\rho(S_0, S_1) \geq 1$. The solid region in Figure 4 is the largest subset of CS which we have explicitly determined, namely, the union of the sets SS, $C_{16}(1)$, and $C_{16}(2)$. The boundary of this region serves as an inner approximation for CS. The actual boundary of CS therefore lies somewhere between the boundary of E_{16} and the boundary of the region $SS \cup C_{16}(1) \cup C_{16}(2)$.

Generalizations of Theorem 2 and Proposition 2 remain valid for N-coefficient systems and are proved in [8]. The work of Berger and Wang [6] indicate that the hypotheses regarding the components of $v(z)$ in Theorem 2 cannot be removed when $N > 3$, e.g., consider the remarks on "stretched dilation equations" found in that paper. It is conjectured in [8] that for each fixed N the collection of coefficients $\{c_0, c_1, \ldots, c_N\}$ where this condition fails is a set of measure zero.

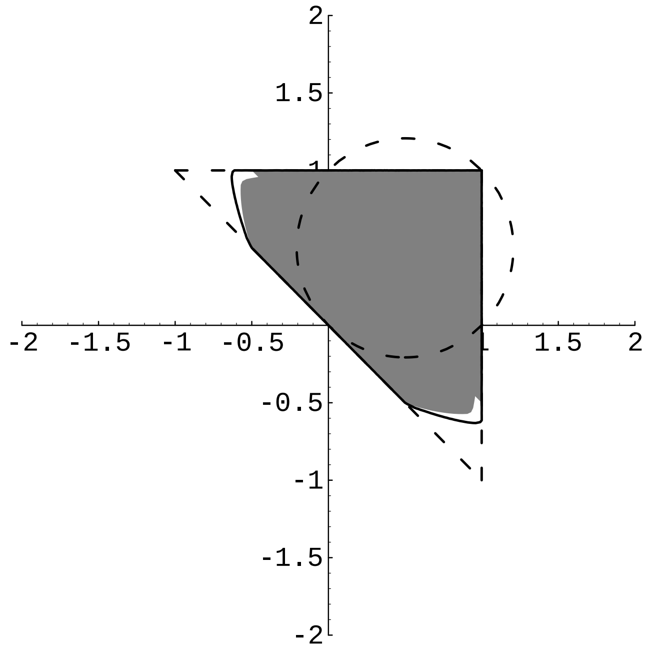

Fig. 4. *The set* $SS \cup C_{16}(1) \cup C_{16}(2)$ *(shaded area); boundary of the set* $E_{16} = \{(c_0, c_3) : \sigma_{16} \geq 1\}$ *(solid line).*

Acknowledgment. The authors thank Dr. George Benke and Mr. Robert Benedetto of The MITRE Corporation and Dr. Gil Strang of MIT for many stimulating discussions and valuable insights on this subject, and the referees for their helpful comments and suggestions.

REFERENCES

[1] A. Cavaretta, W. Dahmen, and C.A. Micchelli, *Stationary subdivision*, Amer. Math. Soc. Memoirs, to appear.

[2] G. Deslauriers and S. Dubuc, *Symmetric iterative interpolation processes*, Constr. Approx., 5 (1989), pp. 49–68.

[3] N. Dyn, J.A. Gregory, and D. Levin, *Analysis of uniform binary subdivision schemes for curve design*, Constr. Approx., 7 (1991), pp. 127–147.

[4] I. Daubechies and J. Lagarias, *Two-scale difference equations: I. Global regularity of solutions*, SIAM J. Math. Anal. (1991).

[5] I. Daubechies and J. Lagarias, *Two-scale difference equations: II. Local regularity, infinite products and fractals*, SIAM J. Math. Anal. (1991).

[6] M. A. Berger and Y. Wang, *Multi-scale dilation equations and iterated function systems*, preprint.

[7] C.A. Micchelli and H. Prautzsch, *Uniform refinement of curves*, Lin. Alg. Appl., 114/115 (1989), pp. 841–870.

[8] D. Colella and C. Heil, *Characterizations of scaling functions, I. Continuous solutions*, preprint.

[9] D. Colella and C. Heil, *Characterizations of scaling functions, II. Distributional and functional solutions*, preprint.

[10] A. Cohen, *Ondelettes, analyses multirésolutions et filtres miroirs en quadrature*, Ann. Inst. H. Poincare, 7 (1990), pp. 439–459.

[11] W. Lawton, *Necessary and sufficient conditions for constructing orthonormal wavelet bases*, J. Math. Phys., 32 (1991), pp. 57–61.

[12] G. Strang, *Wavelets and dilation equations: a brief introduction*, SIAM Rev., 31 (1989), pp. 614–627.

[13] G. C. Rota and G. Strang, *A note on the joint spectral radius*, Indag. Math., 22 (1960), pp. 379–381.

[14] M. A. Berger and Y. Wang, *Bounded semi-groups of matrices*, preprint.

[15] I. Daubechies and J. Lagarias, *Sets of matrices all infinite products of which converge*, Lin. Alg. Appl., to appear.

[16] C. Heil, *Methods of solving dilation equations*, Proc. of the 1991 NATO ASI on Prob. and Stoch. Methods in Anal. with Appl., J. Byrnes, ed., to appear.