

CHARACTERIZATIONS OF SCALING FUNCTIONS: CONTINUOUS SOLUTIONS*

DAVID COLELLA[†] AND CHRISTOPHER HEIL[‡]

Abstract. A dilation equation is a functional equation of the form $f(t) = \sum_{k=0}^N c_k f(2t - k)$, and any nonzero solution of such an equation is called a scaling function. Dilation equations play an important role in several fields, including interpolating subdivision schemes and wavelet theory. This paper obtains sharp bounds for the Hölder exponent of continuity of any continuous, compactly supported scaling function in terms of the joint spectral radius of two matrices determined by the coefficients $\{c_0, \dots, c_N\}$. The arguments lead directly to a characterization of all dilation equations that have continuous, compactly supported solutions.

Key words. dilation equation, joint spectral radius, scaling function, two-scale difference equation, wavelet

AMS subject classifications. 26A16, 39A10

Functional equations of the form

$$(1) \quad f(t) = \sum_{k=0}^N c_k f(2t - k)$$

play an important role in several fields, including wavelet theory and interpolating subdivision schemes. Such equations are referred to as *dilation equations* or *two-scale difference equations*, and any nonzero solution f is called a *scaling function*. The coefficients $\{c_0, \dots, c_N\}$ may be real or complex; if they are real then the scaling function f will be real-valued.

In this paper we obtain sharp bounds for the Hölder exponent of continuity of any continuous, compactly supported scaling function. Our arguments lead directly to a characterization of all dilation equations that have continuous, compactly supported solutions. These methods also enable us to examine how certain properties of scaling functions, such as the Hölder exponent, behave as a function of the coefficients, and we provide several examples to illustrate the basic structure present. Our work was inspired by an early preprint of [DL2], in which sufficient conditions for the existence of continuous, compactly supported scaling functions were obtained and lower bounds for the Hölder exponent of continuity were derived. In that paper the assumption was made that the coefficients satisfy

$$(2) \quad \sum_k c_{2k} = \sum_k c_{2k+1} = 1.$$

Conditions and bounds were then expressed in terms of the joint spectral radius $\hat{\rho}(T_0|_V, T_1|_V)$ of two matrices T_0, T_1 (determined by the coefficients $\{c_0, \dots, c_N\}$) restricted to a certain subspace V of \mathbf{C}^N . We have extended these results in the

* Received by the editors January 27, 1992; accepted for publication (in revised form) July 30, 1992.

[†] The MITRE Corporation, McLean, Virginia 22102 (colella@mitre.org).

[‡] Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and The MITRE Corporation, McLean, Virginia 22102. Current address: School of Mathematics, Georgia Tech, Atlanta, Georgia 30332-0160 (heil@math.gatech.edu). This author's work was partially supported by National Science Foundation grant DMS-9007212.

sense that all dilation equations possessing continuous, compactly supported solutions, without restriction on the coefficients, are now described in terms of a joint spectral radius $\hat{\rho}(T_0|_W, T_1|_W)$, with the subspace V replaced by a (possibly different) subspace W , and that sharp bounds for the Hölder exponent follow from the value of $\hat{\rho}(T_0|_W, T_1|_W)$. The replacement of V by W is not without cost: W depends explicitly on the coefficients $\{c_0, \dots, c_N\}$, while V is independent of them. This dependency nonetheless yields a number of interesting facts about the behavior of scaling functions as a function of the coefficients $\{c_0, \dots, c_N\}$. For example, we show that the maximum Hölder exponent of continuity is not continuous as a function of the coefficients. We give several methods of determining the subspace W explicitly. The results presented here implicitly characterize those dilation equations having compactly supported, n -times differentiable solutions, and are generalizable to positive integer dilation factors other than two and to higher dimensions. An application of the results of this paper to the specific case $N \leq 3$ can be found in [CH1]. The sequel [CH2] discusses more general characterizations, including discontinuous and noncompactly supported scaling functions. Finally, we wish it noted that a revised version of [DL2] brought to our attention the work [MP2], which was the first work to provide necessary and sufficient conditions for the existence of continuous, compactly supported scaling functions. However, those results are not stated in terms of a joint spectral radius and do not yield estimates for the Hölder exponent of continuity; in addition, the behavior of the properties of scaling functions as a function of the coefficients is not examined.

We say that a function h on \mathbf{R} is (globally) *Hölder continuous* if there exist constants α, K such that $|h(x) - h(y)| \leq K|x - y|^\alpha$ for all $x, y \in \mathbf{R}$. The constants α and K are referred to as a *Hölder exponent* and *Hölder constant* for h , respectively. We refer to

$$\alpha_{\max} = \sup \{ \alpha : h \text{ is Hölder continuous with exponent } \alpha \}$$

as the *maximum Hölder exponent* of h , although it should be noted that the supremum need not be attained, i.e., h need not be Hölder continuous with exponent $\alpha = \alpha_{\max}$. We make similar definitions for *local Hölder continuity* at a point x , i.e., x is fixed in the inequality and only y varies.

The application of dilation equations to subdivision schemes is discussed in the important papers [CDM], [DD], [Du], [DGL], [MP1], and [MP2]. Our own interest in dilation equations arose from their application to wavelet theory. We briefly outline this relation below; the basic results can be found in the research papers [D], [L], [M], or in the expository surveys [H], [S]. Assume that the coefficients $\{c_0, \dots, c_N\}$ satisfy (2) and

$$(3) \quad \sum_k c_k \bar{c}_{k+2j} = \begin{cases} 2, & \text{if } j = 0, \\ 0, & \text{if } j \neq 0, \end{cases}$$

where we take $c_k = 0$ if $k < 0$ or $k > N$. It can then be shown that there exists an integrable and square-integrable scaling function f . Define the *wavelet*

$$(4) \quad g(t) = \sum_k (-1)^k c_{N-k} f(2t - k),$$

and construct $g_{n,k}(t) = 2^{n/2} g(2^n t - k)$ by dilation and translation of g . If the scaling function f is orthogonal to its integer translates (i.e., $\int f(t) \overline{f(t-k)} dt = 0$ for $k \neq 0$), then $\{g_{n,k}\}_{n,k \in \mathbf{Z}}$ will form an orthonormal basis for $L^2(\mathbf{R})$ after a suitable normalization of g . It can be shown that for nearly all choices of coefficients $\{c_0, \dots, c_N\}$

satisfying (2) and (3) the associated scaling function f is orthogonal to its integer translates and therefore determines a wavelet orthonormal basis.

Example 1. Consider the dilation equation where $N = 1$ and $c_0 = c_1 = 1$. Both (2) and (3) are satisfied, and the associated scaling function $f = \chi_{[0,1]}$ is orthogonal to its integer translates (where χ_E denotes the characteristic function of the set E). The wavelet g given by (4) is then $g = \chi_{[0,1/2]} - \chi_{[1/2,1]}$. The orthonormal basis $\{g_{n,k}\}$ determined by this g is known as the *Haar system*. \square

The Haar system has the desirable property that the wavelet g is compactly supported. However, the fact that this wavelet is discontinuous severely limits its usefulness. An important problem for wavelet theory is therefore the construction of smooth, compactly supported wavelets. By (4), it suffices to construct smooth, compactly supported scaling functions. The construction of such scaling functions is also important in subdivision theory; however, in that framework (3) is irrelevant and (2) may or may not be assumed. Several authors have proved conditions for the existence of smooth, compactly supported scaling functions, with varying interpretations of “smooth”, and with varying restrictions on the coefficients. As suggested by the “self-similar” nature of the dilation equation, a continuous scaling function f is often “fractal” in nature, in the sense that if it is n -times differentiable then its n th derivative is Hölder continuous with Hölder exponent strictly less than one. Daubechies and Lagarias have proved that compactly supported, infinitely differentiable scaling functions are impossible [DL1].

The methods used to prove the conditions referred to above generally fall into the following three categories or combinations thereof: Fourier transform methods (e.g., [D], [DL1], [E], [M], [V]), iterated function system methods (e.g., [W], [D], [DL1], [DD], [DGL]), and dyadic interpolation methods (e.g., [W], [DL2], [MP2]). In this paper we use the dyadic interpolation method to characterize all dilation equations that have continuous, compactly supported solutions, without any restrictions on the coefficients. However, since a number of facts that will be important to us are more easily proved using the Fourier transform technique, we briefly review that method. This technique is based on the equivalent form of the dilation equation on the Fourier transform side, namely,

$$(5) \quad \hat{f}(2\gamma) = m_0(\gamma) \hat{f}(\gamma),$$

where $m_0(\gamma) = (1/2) \sum c_k e^{ik\gamma}$ and $\hat{f}(\gamma) = \int f(t) e^{i\gamma t} dt$. If $m_0(0) = (1/2) \sum c_k = 1$ then it follows that

$$(6) \quad \hat{\mu}(\gamma) = \prod_{j=1}^{\infty} m_0(2^{-j}\gamma)$$

converges uniformly on compact sets to a continuous function and is a solution to (5). The inverse Fourier transform μ of this function is therefore a solution to (1), at least in the sense of distributions. One can show that μ has support contained in $[0, N]$, and is the only distributional solution to (1) with a continuous Fourier transform (up to multiplication by a constant). In particular, there can be at most one integrable solution to a dilation equation satisfying $\sum c_k = 2$ (up to multiplication by a constant and with uniqueness interpreted as usual as equality almost everywhere), and if one exists it will have the form (6) and have compact support. More generally, Daubechies and Lagarias have proved the following theorem.

THEOREM 1 [DL1]. *Let coefficients $\{c_0, \dots, c_N\}$ be given. If there exists an integrable, compactly supported solution f to the dilation equation (1), then $\text{supp}(f) \subset$*

$[0, N]$ and there exists an integer $n \geq 0$ such that the following statements hold.

- (a) $\sum c_k = 2^{n+1}$.
- (b) f is unique up to multiplication by a constant and has Fourier transform $\hat{f}(\gamma) = \gamma^n \prod_{j=1}^{\infty} 2^{-n} m_0(2^{-j}\gamma)$.
- (c) There is an integrable, compactly supported solution F to the dilation equation determined by the coefficients $\{2^{-n}c_0, \dots, 2^{-n}c_N\}$ and (with the proper choice of scale) f is the n th distributional derivative of F .

In particular, continuous, compactly supported scaling functions can exist only when $\sum c_k = 2^{n+1}$, and for $n > 0$ are the (usual) n th derivatives of scaling functions satisfying $\sum c_k = 2$. Dilation equations satisfying $\sum c_k = 2$ are therefore in some sense fundamental. Additional assumptions on the coefficients $\{c_0, \dots, c_N\}$ can impose enough regularity on the infinite product in (6) so that the scaling function can be proved to be continuous or n -times differentiable, and to bound from below the Hölder exponent of continuity of the scaling function or its n th derivative. Eirola [E] has nicely demonstrated that Fourier transform methods are well suited to estimating Sobolev, rather than Hölder, exponents of continuity of scaling functions.

We turn now to the dyadic interpolation method and our own results. Since we are concerned mainly with questions of continuity, we assume that scaling functions, even if discontinuous, are defined for all points in \mathbf{R} and satisfy the dilation equation at all points, not just almost everywhere, as is the case in the Fourier transform method. This has the seemingly paradoxical effect that a given dilation equation may have more than one distinct integrable, compactly supported solution. For example, if f is one such scaling function and if $c \in \mathbf{C}$ and $S \subset \mathbf{R}$ are given so that both S and its complement are measurable and invariant under the mapping $t \mapsto 2t - k$ for all $k \in \mathbf{Z}$, then

$$\tilde{f}(t) = \begin{cases} f(t), & t \in S, \\ cf(t), & t \notin S, \end{cases}$$

is also an integrable, compactly supported scaling function. However, by Theorem 1, either S or its complement must have measure zero (more fundamentally, this also follows from ergodicity considerations). This leads us to pose the following problem, which affects the interpretation of Theorem 4 at the end of this paper: is it possible that such differing representatives of a compactly supported scaling function may have differing properties, e.g., if one representative is unbounded, must all representatives be unbounded?

The dyadic interpolation method is based on this key observation: if the values of the scaling function at the integers are known then the dilation equation determines the values of the scaling function at the half-integers, and by recursion at every *dyadic point* $x = k/2^n$ where $k, n \in \mathbf{Z}$. If f is continuous then this determines its values at all points (thereby giving an easily programmable method for graphing a continuous scaling function). Daubechies and Lagarias [DL2] and Micchelli and Prautzsch [MP2] independently implemented this recursion via products of two $N \times N$ matrices, and used this implementation to obtain conditions for the existence of continuous scaling functions. We explain this now, using the notation of [DL2].

Given coefficients $\{c_0, \dots, c_N\}$, define the $N \times N$ matrices T_0 and T_1 by $(T_0)_{ij} =$

c_{2i-j-1} and $(T_1)_{ij} = c_{2i-j}$, i.e.,

$$T_0 = \begin{pmatrix} c_0 & 0 & 0 & \cdots & 0 & 0 \\ c_2 & c_1 & c_0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_N & c_{N-1} \end{pmatrix}$$

and

$$T_1 = \begin{pmatrix} c_1 & c_0 & 0 & \cdots & 0 & 0 \\ c_3 & c_2 & c_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & c_N \end{pmatrix}.$$

Given $x \in [0, 1]$ with $x \neq \frac{1}{2}$, define

$$\tau x = 2x \bmod 1 = \begin{cases} 2x, & 0 \leq x < \frac{1}{2}, \\ 2x - 1, & \frac{1}{2} < x \leq 1, \end{cases}$$

i.e., if $x \neq \frac{1}{2}$ and $x = .d_1d_2\dots$ is a binary expansion of x then $\tau x = .d_2d_3\dots$. The value of $\tau(\frac{1}{2})$ is undefined. Note that a dyadic point x has two binary expansions, one ending in infinitely many zeros, and one ending in infinitely many ones. The former expansion will be termed the *upper* or *finite* binary expansion, and the latter the *lower* binary expansion. All nondyadic points have unique binary expansions.

The following result allows questions about scaling functions to be translated into questions about certain vector-valued (specifically, \mathbf{C}^N -valued) functions on $[0, 1]$. The definition of *Hölder continuity* for vector-valued functions is analogous to the definition for ordinary functions, with the absolute value replaced by any norm on \mathbf{C}^N . The Hölder exponent of continuity for a vector-valued function is independent of the choice of norm.

PROPOSITION 1. *Let coefficients $\{c_0, \dots, c_N\}$ be given.*

(a) *Assume f is a scaling function with $\text{supp}(f) \subset [0, N]$. Define the vector-valued function $v : [0, 1] \rightarrow \mathbf{C}^N$ by*

$$(7) \quad v(x) = \begin{pmatrix} f(x) \\ f(x+1) \\ \vdots \\ f(x+N-1) \end{pmatrix}.$$

Then v satisfies

$$(8) \quad v_{i+1}(0) = v_i(1), \quad i = 1, \dots, N-1,$$

$$(9) \quad v(x) = T_0 v(\tau x), \quad 0 < x < \frac{1}{2},$$

$$(10) \quad v(x) = T_1 v(\tau x), \quad \frac{1}{2} < x < 1,$$

where $v_i(x)$ is the i th component of $v(x)$. If $f(0) = f(N) = 0$ (e.g., if f is continuous)

then v also satisfies

$$\begin{aligned}
 (11) \quad & v_1(0) = v_N(1) = 0, \\
 (12) \quad & v(0) = T_0v(0), \\
 (13) \quad & v(1) = T_1v(1), \\
 (14) \quad & v\left(\frac{1}{2}\right) = T_0v(1) = T_1v(0).
 \end{aligned}$$

If f is continuous then so is v . If f is Hölder continuous with Hölder exponent α , then the same is true of v .

(b) Assume $v : [0, 1] \rightarrow \mathbf{C}^N$ is a vector-valued function satisfying (8)–(14). Define the function f by

$$(15) \quad f(x) = \begin{cases} 0, & x \leq 0 \text{ or } x \geq N, \\ v_i(x), & i - 1 \leq x \leq i, \ i = 1, \dots, N. \end{cases}$$

Then f is a scaling function with $\text{supp}(f) \subset [0, N]$. If v is continuous then so is f . If v is Hölder continuous with Hölder exponent α then the same is true of f .

A vector-valued function $v : [0, 1] \rightarrow \mathbf{C}^N$ satisfying (8)–(14) will be called a *scaling vector*. If v is a scaling vector then (9), (10), and (12)–(14) can be summarily written

$$(16) \quad v(x) = T_{d_1}v(\tau x), \quad 0 \leq x \leq 1,$$

where $x = .d_1d_2\dots$ is any binary expansion of x , since, by (14), the ambiguity at $x = \frac{1}{2}$ is nonproblematic.

Note that if f is a scaling function supported in $[0, N]$ and v is defined by (7) then (12)–(14) follow immediately from (11). Also, note from (1) that $c_0, c_N \neq 1$ implies $f(0) = f(N) = 0$. Thus, for $c_0, c_N \neq 1$ there is an exact equivalence between scaling functions and scaling vectors. If $c_0 = 1$ then it is possible that $f(0) \neq 0$, and in this case it is easy to see that (12)–(14) may fail (a specific example is the Haar system, i.e., $N = 1, c_0 = c_1 = 1$, and $f = \chi_{[0,1]}$). Similar remarks hold if $c_N = 1$. However, it follows directly from the dilation equation that if $|c_0| \geq 1$ then f must be discontinuous at 0, and if $|c_N| \geq 1$ then f must be discontinuous at N .

If v is a scaling vector then, by (11), $v(0) = (0, a_1, \dots, a_{N-1})^t$ for some $a = (a_1, \dots, a_{N-1})^t$. It follows then from (12) that $Ma = a$, where M is the $(N - 1) \times (N - 1)$ submatrix of T_0 and T_1 defined by $M_{ij} = c_{2i-j}$, i.e.,

$$(17) \quad M = \begin{pmatrix} c_1 & c_0 & 0 & \cdots & 0 & 0 \\ c_3 & c_2 & c_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_N & c_{N-1} \end{pmatrix}.$$

Thus scaling vectors can only exist when M has 1 as an eigenvalue. This is true, for example, when (2) is satisfied, for then $(1, \dots, 1)$ is a left eigenvector for M for the eigenvalue 1.

CONSTRUCTION 1. We demonstrate that if 1 is an eigenvalue of M , then it is possible to construct a vector-valued function v , defined for dyadic $x \in [0, 1]$ only, which satisfies (8)–(14). This construction need not be unique if the eigenvalue 1 has multiplicity greater than 1. In general, it may not be possible to extend this function to a scaling vector defined at all points in $[0, 1]$ so that the associated scaling function defined by (15) is integrable. Equivalently, every dilation equation such that 1 is an

eigenvalue for M can be solved on the restricted domain of the dyadic points in \mathbf{R} , but not necessarily for all points in \mathbf{R} .

To construct v , let $a = (a_1, \dots, a_{N-1})^t$ be any right eigenvector for M for the eigenvalue 1. Define $v(0) = (0, a_1, \dots, a_{N-1})^t$ and $v(1) = (a_1, \dots, a_{N-1}, 0)^t$; then (8) and (11) are satisfied. Given a dyadic $x \in (0, 1)$, let $x = .d_1 \dots d_m$ be its upper binary expansion, and define

$$(18) \quad v(x) = T_{d_1} \cdots T_{d_m} v(0).$$

Then (9), (10), and (12)–(14) follow immediately for dyadic x . \square

As pointed out above, the matrix M must have 1 as an eigenvalue in order that a scaling vector exist; in particular, this must be the case if a continuous, compactly supported scaling function exists. Since the derivative of a differentiable scaling function is itself a scaling function for the dilation equation determined by the coefficients $\{2c_0, \dots, 2c_N\}$, it follows immediately that if a compactly supported scaling function is n -times differentiable, then $1, 2^{-1}, \dots, 2^{-n}$ must all be eigenvalues for M . In particular, $n < N - 1$, and no compactly supported scaling function can be infinitely differentiable.

Example 2. Consider the case $N = 3$. If $\{c_0, c_1, c_2, c_3\}$ satisfy (2) then $c_1 = 1 - c_3$ and $c_2 = 1 - c_0$. Restricting our attention to real-valued coefficients (as we will do in all specific examples in this paper), the collection of four-coefficient dilation equations satisfying (2) can therefore be identified with the (c_0, c_3) -plane, so that each point in the plane determines a four-coefficient dilation equation and conversely. We make this identification throughout when we discuss the case $N = 3$. Despite the fact that the dilation equation can be solved at dyadic points for any coefficient choice (c_0, c_3) , it is proved in [CH2] that it is impossible to construct integrable, compactly supported scaling functions for any point (c_0, c_3) on or outside the ellipse shown in Fig. 1, with the single exception of the point $(1, 1)$. Conversely, integrable, compactly supported scaling functions do exist for all points in the shaded region of Fig. 1. These scaling functions are continuous for those points in the shaded region that are also inside the triangle of Fig. 1, and are differentiable for those points lying on the solid portion of the dashed line [CH1]. A scaling function for the point $(1, 1)$ is $\chi_{[0,3]}$. However, the function v defined in Construction 1 for this point is highly oscillatory, e.g., $v(x)$ takes each of the values $(0, 1, 1)^t$, $(1, 0, 1)^t$, and $(1, 1, 0)^t$ on a dense set of dyadic x . \square

Let us now consider the question of continuity of scaling vectors. Given a scaling vector v and a dyadic point $x = .d_1 \dots d_m$, consider points $y = .d_1 \dots d_m d_{m+1} \dots d_n$ close to x . If v is continuous then $T_{d_1} \cdots T_{d_m}(v(0) - v(\tau^m y)) = v(x) - v(y) \rightarrow 0$ as $y \rightarrow x$. This suggests that a characterization of continuity for scaling functions requires consideration of all possible products $T_{d_1} \cdots T_{d_m}$ of T_0 and T_1 operating on all possible differences $v(x) - v(y)$. The correct tool for this turns out to be the *joint spectral radius* of T_0, T_1 restricted to a certain subspace of \mathbf{C}^N . We therefore digress to define the joint spectral radius of general matrices and to give some of its properties relevant to the results in this paper. Let $\|\cdot\|$ be any norm on \mathbf{C}^N , with corresponding operator norm $\|A\| = \sup_{u \neq 0} \|Au\|/\|u\|$ defined for $N \times N$ matrices A .

DEFINITION 1. *The joint spectral radius $\hat{\rho}(A_0, A_1)$ of two matrices A_0, A_1 is*

$$\hat{\rho}(A_0, A_1) = \limsup_{m \rightarrow \infty} \hat{\rho}_m,$$

where

$$\hat{\rho}_m = \hat{\rho}_m(A_0, A_1) = \max_{d_j=0,1} \|A_{d_1} \cdots A_{d_m}\|^{1/m}.$$

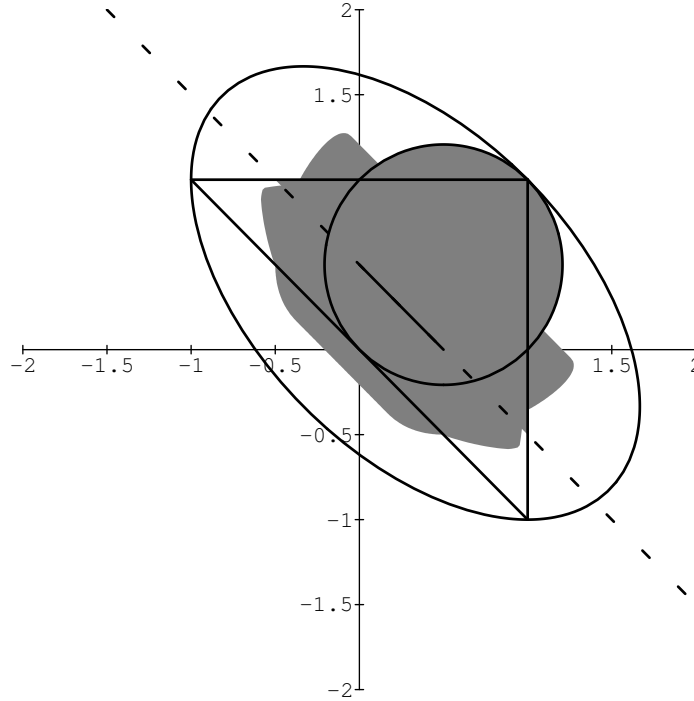


FIG. 1. The (c_0, c_3) -plane, identified with real-valued, four-coefficient dilation equations satisfying (2).

This definition generalizes the usual spectral radius $\rho(A)$ of a single matrix A , which is given by the formula

$$\rho(A) = \limsup_{m \rightarrow \infty} \|A^m\|^{1/m} = \max \{ |\lambda| : \lambda \text{ is an eigenvalue of } A \}.$$

An extension of the joint spectral radius to larger collections of matrices, and to matrices restricted to subspaces, is made in the obvious way. The joint spectral radius was introduced by Rota and Strang [RS]; some recent papers include [BW] and [DL3].

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 invertible matrix B . Set

$$\hat{\sigma}_m = \hat{\sigma}_m(A_0, A_1) = \max_{d_j=0,1} \rho(A_{d_1} \cdots A_{d_m})^{1/m},$$

so that $\hat{\sigma}_m$ is the m th root of the largest absolute eigenvalue that occurs among all products of length m of the matrices A_0, A_1 . Then $\hat{\sigma}_m \leq \hat{\rho}(A_0, A_1) \leq \hat{\rho}_m$ for every m . In particular, $\hat{\rho}(A_0, A_1) \geq \hat{\sigma}_1$ and generally one expects $\hat{\rho}(A_0, A_1)$ to be strictly larger than $\hat{\sigma}_1$. Berger and Wang [BW] prove the nontrivial result that $\hat{\rho}(A_0, A_1) = \limsup \hat{\sigma}_m$. It therefore follows that $\sup \hat{\sigma}_m = \hat{\rho}(A_0, A_1) = \lim \hat{\rho}_m = \inf \hat{\rho}_m$. As each $\hat{\sigma}_m$ is a continuous function of the entries of A_0 and A_1 , $\sup \hat{\sigma}_m$ is lower semicontinuous as a function of those entries. Similarly, $\inf \hat{\rho}_m$ is upper semicontinuous, and therefore $\hat{\rho}(A_0, A_1)$ is continuous as a function of the entries of A_0 and A_1 [HS].

The exact evaluation of the joint spectral radius is difficult, except in special cases such as the following.

LEMMA 1. *If A_0 and A_1 can be simultaneously symmetrized, i.e., there exists an invertible matrix B such that BA_iB^{-1} is real and symmetric for $i = 0, 1$, then $\hat{\rho}(A_0, A_1) = \hat{\sigma}_1 = \max\{\rho(A_0), \rho(A_1)\}$.*

The conclusion of Lemma 1 holds more generally when A_0, A_1 can be simultaneously Hermitianized.

It is easy to see that given any $\theta > \hat{\rho}(A_0, A_1)$, there exists a constant $C > 0$ such that $\hat{\rho}_m \leq C^{1/m} \theta$ for every m . This need not be true with $\theta = \hat{\rho}(A_0, A_1)$. One case where it is true with $\theta = \hat{\rho}(A_0, A_1)$ is if A_0, A_1 are simultaneously symmetrizable, for then $\hat{\sigma}_1 = \hat{\rho}_1$ (using the Euclidean space norm), and therefore $\hat{\rho}_m \leq \hat{\rho}_1 = \hat{\sigma}_1 \leq \hat{\rho}(A_0, A_1)$ for every m .

Another case in which some simplification of the joint spectral radius occurs is the following.

LEMMA 2. *If A_0, A_1 can be simultaneously block upper-triangularized, i.e., there exists an invertible matrix B such that BA_iB^{-1} has the block form*

$$BA_iB^{-1} = \begin{pmatrix} C_i^1 & & * \\ & \ddots & \\ 0 & & C_i^k \end{pmatrix}, \quad i = 0, 1,$$

for some square submatrices C_i^1, \dots, C_i^k , then $\hat{\rho}(A_0, A_1) = \max_{j=1, \dots, k} \{\hat{\rho}(C_0^j, C_1^j)\}$.

In particular, if each block C_i^j is a single number—meaning that A_0, A_1 can be simultaneously upper-triangularized—then $\hat{\rho}(A_0, A_1) = \hat{\sigma}_1 = \max\{\rho(A_0), \rho(A_1)\}$. The proof of Lemma 2 follows easily from block matrix multiplication and consideration of the eigenvalues of the products $BA_{d_1} \cdots A_{d_m} B^{-1}$.

We return now to the matrices T_0, T_1 and the scaling vector v . Assume that (2) holds; then $(1, \dots, 1)$ is a common left eigenvector for both T_0 and T_1 . Hence the $(N - 1)$ -dimensional subspace

$$V = \{u \in \mathbf{C}^N : u_1 + \cdots + u_N = 0\}$$

is invariant under both of those operators since it is the orthogonal complement of $(1, \dots, 1)$ in \mathbf{C}^N . By (18), V contains every difference $v(x) - v(y)$ for dyadic $x, y \in [0, 1]$. This observation suggests consideration of the joint spectral radius $\hat{\rho}(T_0|_V, T_1|_V)$ of T_0, T_1 restricted to the subspace V . Note that since V is invariant under both T_0 and T_1 , there exists a change-of-basis matrix B such that

$$(19) \quad BT_iB^{-1} = \begin{pmatrix} S_i & * \\ 0 & 1 \end{pmatrix}, \quad i = 0, 1,$$

where S_0, S_1 are $(N - 1) \times (N - 1)$ matrices, and $\hat{\rho}(T_0|_V, T_1|_V) = \hat{\rho}(S_0, S_1)$. A sufficient condition for continuity and a lower bound for the maximum Hölder exponent of continuity is provided by the following theorem of Daubechies and Lagarias.

THEOREM 2 [DL2]. *Assume that coefficients $\{c_0, \dots, c_N\}$ satisfying (2) are given. If $\hat{\rho}(T_0|_V, T_1|_V) < 1$ then a continuous scaling vector v exists, and is Hölder continuous with $\alpha_{\max} \geq -\log_2 \hat{\rho}(T_0|_V, T_1|_V)$. The exponent $\alpha = -\log_2 \hat{\rho}(T_0|_V, T_1|_V)$ is allowed if there exists a constant $C > 0$ such that $\hat{\rho}_m \leq C^{1/m} \hat{\rho}(T_0|_V, T_1|_V)$ for every m .*

We emphasize that the value $-\log_2 \hat{\rho}(T_0|_V, T_1|_V)$ in Theorem 2 is only a lower bound for the maximum Hölder exponent, i.e., the theorem does not imply that v

cannot be Hölder continuous for exponents $\alpha > -\log_2 \hat{\rho}(T_0|_V, T_1|_V)$.

Note that if (2) does not hold, then V need not be invariant under both T_0 and T_1 , and therefore $\hat{\rho}(T_0|_V, T_1|_V)$ need not be well defined. The necessary and sufficient conditions for the existence of continuous, compactly supported scaling functions presented below do not depend on (2).

The problem of characterizing all choices of coefficients $\{c_0, \dots, c_N\}$ for which $\hat{\rho}(T_0|_V, T_1|_V) < 1$ is difficult. Of course, $\hat{\rho}(T_0|_V, T_1|_V) \leq \hat{\rho}_m$ for each m and $\hat{\rho}_m$ depends on only finitely many matrices; however, the number of matrices involved grows exponentially with m and therefore $\hat{\rho}_m$ can be computed in practice only for small m . A recursive algorithm, based on the *building blocks* idea of [DL2], for bounding a joint spectral radius from above with minimal computation is given in [CH1].

Theorem 2 implies that continuous scaling vectors exist for every coefficient choice in the set

$$C_V = \{ \{c_0, \dots, c_N\} \text{ satisfying (2) : } \hat{\rho}(T_0|_V, T_1|_V) < 1 \}.$$

This is an open set since $\hat{\rho}(T_0|_V, T_1|_V)$ is a continuous function of the coefficients $\{c_0, \dots, c_N\}$. The lower bound $-\log_2 \hat{\rho}(T_0|_V, T_1|_V)$ for the maximum Hölder exponent given in Theorem 2 is also a continuous function of the coefficients. However, we show in Example 7 that α_{\max} itself is *not* continuous as a function of the coefficients, even for coefficients in C_V . Despite this, we can prove that if (2) holds then the corresponding scaling functions change in a continuous manner (with respect to the sup-norm) as the coefficients change within C_V . This result takes advantage of the fact that V is independent of the coefficients.

PROPOSITION 2. *Fix $\{c_0, \dots, c_N\} \in C_V$ and let f be the corresponding scaling function. For each $\varepsilon > 0$ there exists a $\delta > 0$ such that if $\{\tilde{c}_0, \dots, \tilde{c}_N\}$ satisfies (2) and $|c_i - \tilde{c}_i| < \delta$ for $i = 0, \dots, N$ then there is a continuous, compactly supported scaling function \tilde{f} corresponding to $\{\tilde{c}_0, \dots, \tilde{c}_N\}$ with $\sup |f(t) - \tilde{f}(t)| < \varepsilon$.*

Proof. For each $\eta > 0$ set

$$U_\eta = \{ \{\tilde{c}_0, \dots, \tilde{c}_N\} \text{ satisfying (2) : } |c_i - \tilde{c}_i| \leq \eta, i = 0, \dots, N \}.$$

Since $\{c_0, \dots, c_N\} \in C_V$, there must exist an integer m such that $\hat{\rho}(T_0|_V, T_1|_V) \leq \hat{\rho}_m < 1$. As $\hat{\rho}_m$ depends on only finitely many matrix products, there must be $\theta, \eta > 0$ such that if $\{\tilde{c}_0, \dots, \tilde{c}_N\} \in U_\eta$ then $\hat{\rho}(\tilde{T}_0|_V, \tilde{T}_1|_V) \leq \tilde{\rho}_m < \theta < 1$, where \tilde{T}_0, \tilde{T}_1 are the matrices corresponding to $\{\tilde{c}_0, \dots, \tilde{c}_N\}$ and $\tilde{\rho}_m = \max_{d_j=0,1} \|(\tilde{T}_{d_1} \cdots \tilde{T}_{d_m})|_V\|^{1/m}$. Therefore, for each $\{\tilde{c}_0, \dots, \tilde{c}_N\} \in U_\eta$ there exists a corresponding continuous scaling vector \tilde{v} that is Hölder continuous with exponent $\tilde{\alpha}$ satisfying

$$\tilde{\alpha} \geq \alpha = -\log_2 \theta.$$

Careful examination of the proof of Theorem 2 given in [DL2] reveals that corresponding Hölder constants \tilde{K} for \tilde{v} satisfy

$$\tilde{K} \leq \frac{2\tilde{C}\tilde{R}}{\tilde{\rho}_m},$$

where

$$\tilde{C} = \max_{r=0, \dots, m-1} \left\{ \left(\frac{\tilde{\rho}_r}{\tilde{\rho}_m} \right)^r \right\}$$

and

$$\tilde{R} = \sup_{t \in [0,1]} \|\tilde{v}(t)\| \leq \left(\frac{\tilde{C}(\max\{\|\tilde{T}_0\|, \|\tilde{T}_1\|\} + 1)}{1 - \tilde{\rho}_m} + 1 \right) \|\tilde{v}(0)\|.$$

The quantities \tilde{C} and \tilde{R} are continuous as functions of the coefficients, so $K = \sup_{\{\tilde{c}_0, \dots, \tilde{c}_N\} \in U_\eta} \tilde{K} < \infty$.

Now let ε be given and choose n large enough that $2^{-n\alpha}K < \frac{\varepsilon}{4}$. Let S_n be the set of all dyadic points in $[0, 1]$ with finite binary expansions of length n or less. The functions \tilde{v} are completely determined on S_n by $\tilde{v}(0)$ and the 2^n matrices of the form $\tilde{T}_{d_1} \cdots \tilde{T}_{d_n}$. Hence, there exists a $\delta < \eta$ such that $\sup_{x \in S_n} \|v(x) - \tilde{v}(x)\| < \frac{\varepsilon}{2}$ for all $\{\tilde{c}_0, \dots, \tilde{c}_N\} \in U_\delta$. Given $y \in [0, 1]$ arbitrary, there is an $x \in S_n$ with $|y - x| < 2^{-n}$, so

$$\begin{aligned} \|v(y) - \tilde{v}(y)\| &\leq \|v(y) - v(x)\| + \|v(x) - \tilde{v}(x)\| + \|\tilde{v}(x) - \tilde{v}(y)\| \\ &\leq K|y - x|^\alpha + \frac{\varepsilon}{2} + K|x - y|^\alpha \\ &< \varepsilon. \end{aligned}$$

The result now follows once we recall that all norms on \mathbf{C}^N are equivalent, and note that the sup-norm for scaling vectors is equivalent to the sup-norm for the associated scaling functions. \square

Theorem 2 states that the condition $\hat{\rho}(T_0|_V, T_1|_V) < 1$ implies the existence of a continuous, compactly supported scaling function. The following example, inspired by [W], shows that $\hat{\rho}(T_0|_V, T_1|_V) < 1$ is not, in general, necessary for the existence of a continuous, compactly supported scaling function.

Example 3. Fix N , choose coefficients $\{c_0, \dots, c_N\}$ satisfying (2), and let f be the associated scaling function. Let $d > 1$ be any odd integer, set $\tilde{N} = Nd$, and define coefficients $\{\tilde{c}_0, \dots, \tilde{c}_{\tilde{N}}\}$ by

$$\tilde{c}_k = \begin{cases} c_j, & \text{if } k = jd, \\ 0, & \text{if } k \neq jd. \end{cases}$$

Let \tilde{T}_0, \tilde{T}_1 be the matrices corresponding to $\{\tilde{c}_0, \dots, \tilde{c}_{\tilde{N}}\}$, and let \tilde{V} be the corresponding subspace of $\mathbf{C}^{\tilde{N}}$. By construction, the coefficients $\{\tilde{c}_0, \dots, \tilde{c}_{\tilde{N}}\}$ satisfy (2), and the associated scaling function is $\tilde{f}(t) = f(\frac{t}{d})$. It must be the case that $\hat{\rho}(\tilde{T}_0|_{\tilde{V}}, \tilde{T}_1|_{\tilde{V}}) \geq \rho(\tilde{T}_0|_{\tilde{V}}) \geq 1$ since both $(1, 1, \dots, 1)$ and $(1, 0, 0, 1, 0, 0, \dots, 1, 0, 0)$ are left eigenvectors for \tilde{T}_0 and \tilde{T}_1 for the eigenvalue 1 and $\dim(\tilde{V}) = \tilde{N} - 1$. However, if $\hat{\rho}(T_0|_V, T_1|_V) < 1$ then f , and therefore \tilde{f} , will be continuous. The dilation equation defined by the coefficients $\{\tilde{c}_0, \dots, \tilde{c}_{\tilde{N}}\}$ is referred to as a *stretched dilation equation*. Wang examines stretched dilation equations using iterated function system techniques. \square

For a stretched dilation equation, the differences $v(x) - v(y)$ are contained in a lower-dimensional subspace of V , and therefore the value of $\hat{\rho}(T_0|_V, T_1|_V)$ may be determined by vectors not directly related to the scaling vector v . This suggests that better results might be obtained by replacing V with the subspace

$$\begin{aligned} W &= \text{span}\{v(x) - v(y) : \text{dyadic } x, y \in [0, 1]\} \\ &= \text{span}\{v(x) - v(0) : \text{dyadic } x \in [0, 1]\}. \end{aligned}$$

Note that W is invariant under T_0, T_1 by construction, without the need to assume (2). If (2) does hold then $W \subset V$. Note also that W depends explicitly on the scaling vector v , while V does not. We show in Theorem 3 below that $\hat{\rho}(T_0|_W, T_1|_W) < 1$ is sufficient to ensure the existence of a continuous scaling function, and we give several examples illustrating the distinction between V and W and its consequences. Moreover, by considering W instead of V , we obtain necessary conditions for the

existence of continuous scaling functions. These necessary conditions can be illustrated by the following example.

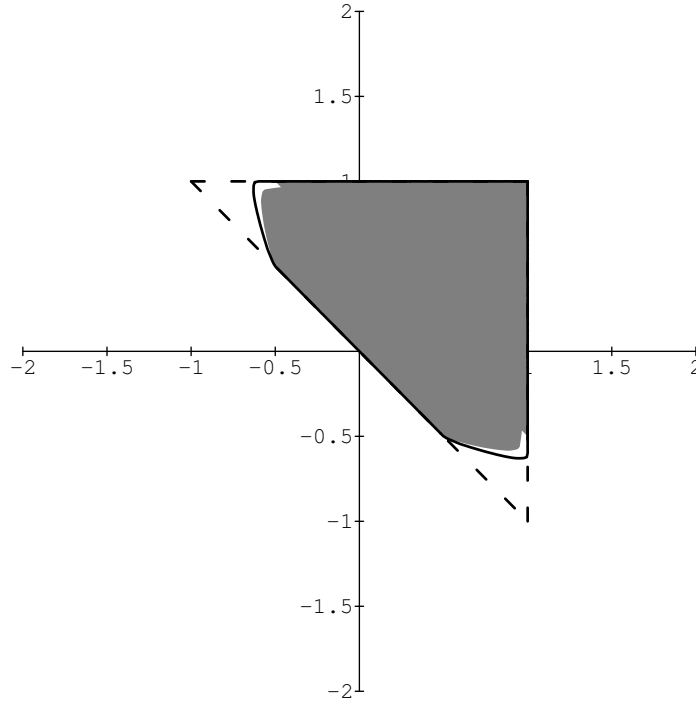


FIG. 2. Subsets of the (c_0, c_3) -plane: region where either $\hat{\rho}_{20} < 1$ or $\hat{\rho}(T_0|_V, T_1|_V) < 1$ due to simultaneous symmetrization (shaded area); boundary of the region where $\hat{\sigma}_{20} < 1$ (solid curve).

Example 4. Consider $N = 3$ as in Example 2. The shaded region in Fig. 2 is the union of the set of points (c_0, c_3) where $\hat{\rho}_{20}(T_0|_V, T_1|_V) < 1$ (obtained by numerical computation using the Euclidean norm) and the set of points (c_0, c_3) such that S_0, S_1 defined by (19) are simultaneously symmetrizable with $\hat{\rho}(T_0|_V, T_1|_V) = \hat{\rho}(S_0, S_1) < 1$. Theorem 2 therefore implies that continuous, compactly supported scaling functions exist for each point in the shaded region.

To derive a necessary condition, consider the behavior of the scaling vector v near the point 0. If v is continuous, then necessarily $v(2^{-m}) \rightarrow v(0)$ as $m \rightarrow \infty$. It is straightforward to check by evaluating $v(2^{-m}) = T_0^m v(1)$ directly that this occurs if and only if $|c_0|, |1 - c_0 - c_3| < 1$. Similarly, the continuity of v at 1 implies $|c_3|, |1 - c_0 - c_3| < 1$. The eigenvalues of $T_0|_V, T_1|_V$ combined are c_0, c_3 , and $1 - c_0 - c_3$, so if v is continuous then $\hat{\sigma}_1(T_0|_V, T_1|_V) < 1$. Moreover, this analysis shows that $\|v(2^{-m}) - v(0)\|$ is on the order of $\hat{\sigma}_1^m$, so the maximum Hölder exponent of continuity of v is at most $-\log_2 \hat{\sigma}_1$. The region of points (c_0, c_3) such that $\hat{\sigma}_1 < 1$ is the interior of the triangle shown in Fig. 2, i.e., continuous scaling vectors are restricted to points in the interior of that triangle. \square

The case $N = 3$ is simple enough that the result of Example 4 can be obtained directly by evaluating $v(2^{-m})$ as a function of the coefficients $\{c_0, \dots, c_N\}$. Although this exact evaluation cannot be done at arbitrary points, the spirit of the approach of Example 4 is used in the proof of Theorem 3(b) below to obtain that continuity

implies $\hat{\sigma}_m(T_0|_W, T_1|_W) < 1$ for every m . Specifically, the matrices T_0, T_1 are replaced by products $T = T_{d_1} \cdots T_{d_m}$, the eigenvalues $c_0, c_3, 1 - c_0 - c_3$ of $T_0|_V, T_1|_V$ by eigenvalues of $T|_W$, and the points $2^{-m}, 0$ by points X_m, Y_m depending on the product T . We must replace V by W to ensure that $v(X_m) - v(Y_m)$ will have a component in the required eigenspace of T . This was unnecessary for the case $N = 3$ since the dilation equations for that case are very restrictive, cf. Example 5. For $N = 3$, the set of points (c_0, c_3) such that $\hat{\sigma}_{20} < 1$ is the interior of the solid curve shown in Fig. 2, i.e., continuous scaling vectors are restricted not only to the interior of the triangle, but to the interior of the region bounded by the solid curve.

Since $\hat{\rho}(T_0|_W, T_1|_W) = \sup \hat{\sigma}_m$, the preceding remarks suggest that the value of $\hat{\rho}(T_0|_W, T_1|_W)$ is essentially the single determining factor for the existence of a continuous scaling function. This is made precise in the following theorem. Recall that scaling vectors can exist only when the matrix M defined by (17) has 1 as an eigenvalue.

THEOREM 3. *Assume the coefficients $\{c_0, \dots, c_N\}$ are such that 1 is an eigenvalue for M .*

(a) *Let v be the vector-valued function defined for dyadic x constructed in Construction 1. If $\hat{\rho}(T_0|_W, T_1|_W) < 1$ then v extends to a continuous scaling vector.*

(b) *If v is any continuous scaling vector then $\hat{\rho}(T_0|_W, T_1|_W) < 1$. In this case v is Hölder continuous with $\alpha_{\max} = -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$, and the exponent $\alpha = \alpha_{\max}$ is achieved if and only if there exists a constant $C > 0$ such that $\hat{\rho}_m \leq C^{1/m} \hat{\rho}(T_0|_W, T_1|_W)$ for every m .*

The proof of Theorem 3 is given below; we first make some remarks on its statement, consider how to determine W , and examine the relationship between V and W for those dilation equations satisfying (2).

First, note that the subspace W depends explicitly on v . Thus, in the hypothesis of Theorem 3(a), the vector-valued function of Construction 1 must be determined before W and $\hat{\rho}(T_0|_W, T_1|_W)$ can be evaluated.

Second, in the hypotheses of Theorem 3(b), note that if v is a continuous scaling vector, then it agrees at dyadic points with the vector-valued function given in Construction 1. Thus

$$C_W = \{ \{c_0, \dots, c_N\} \text{ satisfying (2)} : \hat{\rho}(T_0|_W, T_1|_W) < 1 \}$$

is the precise set of all dilation equations satisfying (2) that have continuous, compactly supported solutions, cf. the definition of C_V following Theorem 2. Example 7 below presents a specific case where $C_W \neq C_V$ for reasons other than trivial cases such as stretched dilation equations.

Finally, note that since the derivative of a differentiable scaling function is itself a solution of another dilation equation (with coefficients multiplied by two), Theorem 3 implicitly characterizes those dilation equations having compactly supported, n -times differentiable solutions. This higher-order characterization, and the fact that computation of the joint spectral radius can often be simplified when the coefficients satisfy $\sum c_k = 2$ and the *sum rules* $\sum (-1)^k k^j c_k = 0$ for $j = 0, \dots, n$, is elaborated on in [HC].

We now discuss several methods of determining the subspace W explicitly. Note that W is one subspace that is invariant under both T_0 and T_1 . The following proposition provides a means of recognizing which invariant subspace is W by examination of the single vector $v(1) - v(0)$. This vector is easily computable since $v(0) = (0, a_1, \dots, a_{N-1})^t$ and $v(1) = (a_1, \dots, a_{N-1}, 0)^t$ for some eigenvector $a =$

$(a_1, \dots, a_{N-1})^t$ of the matrix M for the eigenvalue 1. If the multiplicity of this eigenvalue is one, then $v(0)$ and $v(1)$ are uniquely determined up to multiplication by a constant.

PROPOSITION 3. *W is the smallest subspace of \mathbf{C}^N invariant under both T_0 and T_1 which contains the vector $v(1) - v(0)$.*

Proof. Let U be the smallest subspace of \mathbf{C}^N invariant under both T_0 and T_1 which contains the vector $v(1) - v(0)$. Then $U \subset W$, and $v(1) = v(0) + u$ for some $u \in U$. Since $T_0v(1) = T_1v(0) = v(\frac{1}{2})$ we have $v(\frac{1}{2}) - v(0) = T_0v(1) - v(0) = T_0u \in U$. Continuing in this manner we obtain that if $x = .d_1 \dots d_m \in [0, 1]$ is dyadic, then $v(x) - v(0) = T_{d_1} \dots T_{d_m}v(0) - v(0) \in U$, whence $U \supset W$. \square

The following corollary gives sufficient conditions for W to be all of V , assuming that (2) is satisfied. A *hyperplane* in \mathbf{C}^N is any translate of a subspace of \mathbf{C}^N .

COROLLARY 1. *Assume (2) holds. If $\{v(x) : \text{dyadic } x \in [0, 1]\}$ is not contained in any hyperplane in \mathbf{C}^N of dimension $N - 2$ or less, then $W = V$. In particular, if $\text{span}\{v(x) : \text{dyadic } x \in [0, 1]\} = \mathbf{C}^N$, then $W = V$.*

Proof. It follows from the hypotheses that $\{v(x) - v(0) : \text{dyadic } x \in [0, 1]\}$ is not contained in any hyperplane of dimension $N - 2$ or less. Since this set is contained in the $(N - 1)$ -dimensional subspace V , it follows that V is its span, and therefore is W . \square

If the components of $v(0)$ add to zero then $v(x) \in V$ for every dyadic x and therefore $\text{span}\{v(x) : \text{dyadic } x \in [0, 1]\} \subset V \neq \mathbf{C}^N$, yet it may still be the case that $\{v(x) : \text{dyadic } x \in [0, 1]\}$ is not contained in any hyperplane of dimension $N - 2$ or less. If the components of $v(0)$ do not add to zero then $\text{span}\{v(x) : \text{dyadic } x \in [0, 1]\} = \mathbf{C}^N$ if and only if $\{v(x) : \text{dyadic } x \in [0, 1]\}$ is not contained in any hyperplane of dimension $N - 2$ or less. Lawton [L] proved that if (2) holds then an integrable, compactly supported scaling function f satisfies $\sum f(t + k) = \int f = \hat{f}(0)$ almost everywhere. Since f must also satisfy $\hat{f}(\gamma) = \hat{f}(0) \prod_1^\infty m_0(\gamma/2^j)$, cf. (6), it must be the case that $\hat{f}(0) \neq 0$, and thus $\sum_{i=1}^N v_i(x) = \hat{f}(0) \neq 0$ almost everywhere. In particular, if (2) holds and the components of $v(0)$ add to zero then v cannot be continuous.

Corollary 1 is easily implemented only for small values of N . In general, the hypotheses of the following proposition are easier to check.

PROPOSITION 4. *Assume (2) holds, and let $T = T_{d_1} \dots T_{d_m}$ be any product of T_0, T_1 such that*

- (a) *T has distinct eigenvalues, and*
- (b) *there is some dyadic $z \in [0, 1]$ such that $v(z)$ has a component in each of the eigenspaces of T .*

Let $z = .z_1z_2 \dots$ be any binary expansion of z , and define x_0, \dots, x_{N-1} by $x_0 = z$, $x_1 = .d_1 \dots d_m z_1 z_2 \dots$, $x_2 = .d_1 \dots d_m d_1 \dots d_m z_1 z_2 \dots$, etc. Then $\{v(x_0), \dots, v(x_{N-1})\}$ forms a basis for \mathbf{C}^N , and therefore $W = V$.

Proof. Note that $v(x_i) = T^i v(z)$ for $i = 0, \dots, N - 1$. Let $\lambda_1, \dots, \lambda_N$ be the distinct eigenvalues of T ; then, by hypothesis, $v(z) = u_1 + \dots + u_N$ where each u_i is a nonzero eigenvector for T corresponding to the eigenvalue λ_i . Assume that $\{v(z), \dots, T^{N-1}v(z)\}$ was not a basis for \mathbf{C}^N ; then there exist scalars $\alpha_1, \dots, \alpha_N$, not all zero, such that

$$0 = \alpha_1 v(z) + \alpha_2 T v(z) + \dots + \alpha_N T^{N-1} v(z)$$

$$\begin{aligned}
&= \alpha_1 u_1 + \cdots + \alpha_1 u_N \\
&\quad + \alpha_2 \lambda_1 u_1 + \cdots + \alpha_2 \lambda_N u_N \\
&\quad \vdots \\
&\quad + \alpha_N \lambda_1^{N-1} u_1 + \cdots + \alpha_N \lambda_N^{N-1} u_N.
\end{aligned}$$

As $\{u_1, \dots, u_N\}$ forms a basis for \mathbf{C}^N we must therefore have $\alpha_1 + \alpha_2 \lambda_i + \cdots + \alpha_N \lambda_i^{N-1} = 0$ for $i = 1, \dots, N$, i.e., $\lambda_1, \dots, \lambda_N$ are roots of the polynomial $p(\lambda) = \alpha_1 + \alpha_2 \lambda + \cdots + \alpha_N \lambda^{N-1}$. As p has degree at most $N-1$, this is a contradiction. \square

Typically, the product T and the value of z in the hypotheses of Proposition 4 are chosen so as to simplify computations, as in the following example.

Example 5. Fix $N = 3$ as in Example 2 and consider Proposition 4 with $T = T_0$ and $z = 1$. The eigenvalues of T_0 are 1, c_0 , and $1 - c_0 - c_3$; assume these are distinct. Corresponding eigenvectors of T_0 are $u_1 = v(0) = (0, c_0, c_3)^t$, $u_2 = (1 - 2c_0 - c_3, 2c_0 - 1, c_3)^t$, and $u_3 = (0, 1, -1)^t$. Up to multiplication by a constant, $v(1) = (c_0, c_3, 0)^t$. Since $1 - 2c_0 - c_3 \neq 0$,

$$v(1) = u_1 + \frac{c_0}{1 - 2c_0 - c_3} u_2 + \frac{c_3(1 - c_0 - c_3)}{1 - 2c_0 - c_3} u_3.$$

Thus $v(1)$ has a component in each of the eigenspaces of T_0 , and therefore $W = V$, provided that $c_0, c_3, 1 - c_0 - c_3 \neq 0$.

Compare the above computation to the result proved in [CH1], that $\{v(x) : \text{dyadic } x \in [0, 1]\}$ is not contained in any line in \mathbf{C}^3 if $c_0, c_3, 1 - c_0 - c_3 \neq 0$; hence W is all of V in that case. When $1 - c_0 - c_3 = 0$ then W has dimension 1 and hence is a proper subset of V . This case is not the result of a stretched dilation equation, but rather arises from the presence of zero eigenvalues. In particular, the eigenvalues of $T_0|_V$ are c_0 and $1 - c_0 - c_3$ and the eigenvalues of $T_1|_V$ are c_3 and $1 - c_0 - c_3$. If $1 - c_0 - c_3 = 0$ then zero is a common eigenvalue of $T_0|_V$ and $T_1|_V$; moreover, the corresponding eigenspaces are identical. Note, however, that $\hat{\rho}(T_0|_V, T_1|_V) = \hat{\rho}(T_0|_W, T_1|_W)$ despite the fact that $V \neq W$, since the zero eigenvalue does not impact the value of the joint spectral radius. \square

The next example is another simple illustration of how W may differ from V without the joint spectral radius being affected.

Example 6. Given coefficients $\{c_0, \dots, c_N\}$ satisfying (2), define $\tilde{c}_i = c_i$ for $i = 0, \dots, N$ and $\tilde{c}_{N+1} = 0$. Let f, v, W , etc., be the usual items associated with $\{c_0, \dots, c_N\}$, and let $\tilde{f}, \tilde{v}, \tilde{W}$, etc., be the corresponding items associated with $\{\tilde{c}_0, \dots, \tilde{c}_{N+1}\}$. Clearly $\tilde{f} = f$, yet $\tilde{W} \neq \tilde{V}$, even if $W = V$, since $\tilde{W} \subset \{u \in \mathbf{C}^{N+1} : u_1 + \cdots + u_N = 0, u_{N+1} = 0\} \neq \tilde{V}$. Wang [W] demonstrated that $\hat{\rho}(T_0|_V, T_1|_V) = \hat{\rho}(\tilde{T}_0|_{\tilde{V}}, \tilde{T}_1|_{\tilde{V}})$, and it follows similarly that $\hat{\rho}(T_0|_W, T_1|_W) = \hat{\rho}(\tilde{T}_0|_{\tilde{W}}, \tilde{T}_1|_{\tilde{W}})$. If $W = V$ then these four numbers are equal, despite the fact that $\tilde{W} \neq \tilde{V}$. \square

The calculations in the preceding example are carried out by placing T_0 and T_1 in block upper-triangular form, as follows. Suppose we are given any dilation equation satisfying (2) such that W is a proper subspace of V , say $J = \dim(W) < N - 1$. Since both W and V are invariant under T_0 and T_1 , there will be a change-of-basis matrix B such that BT_0B^{-1} and BT_1B^{-1} have the block upper-triangular forms

$$BT_iB^{-1} = \begin{pmatrix} P_i & * & * \\ 0 & Q_i & * \\ 0 & 0 & 1 \end{pmatrix}, \quad i = 0, 1,$$

where the matrices P_i are $J \times J$ and the matrices Q_i are $(N - J - 1) \times (N - J - 1)$. Hence,

$$\hat{\rho}(T_0|_W, T_1|_W) = \hat{\rho}(P_0, P_1)$$

and

$$\hat{\rho}(T_0|_V, T_1|_V) = \max \{ \hat{\rho}(P_0, P_1), \hat{\rho}(Q_0, Q_1) \}.$$

One such basis choice is the eigenvector basis for T_0 (if it exists), in which case T_0 is diagonalized and T_1 can be block upper-triangularized.

Our next example presents a class of dilation equations for which the distinction between V and W plays an important role in determining the continuity of the associated scaling vectors. In particular, this example illustrates the fact that the maximum Hölder exponent is not continuous as a function of the coefficients $\{c_0, \dots, c_N\}$.

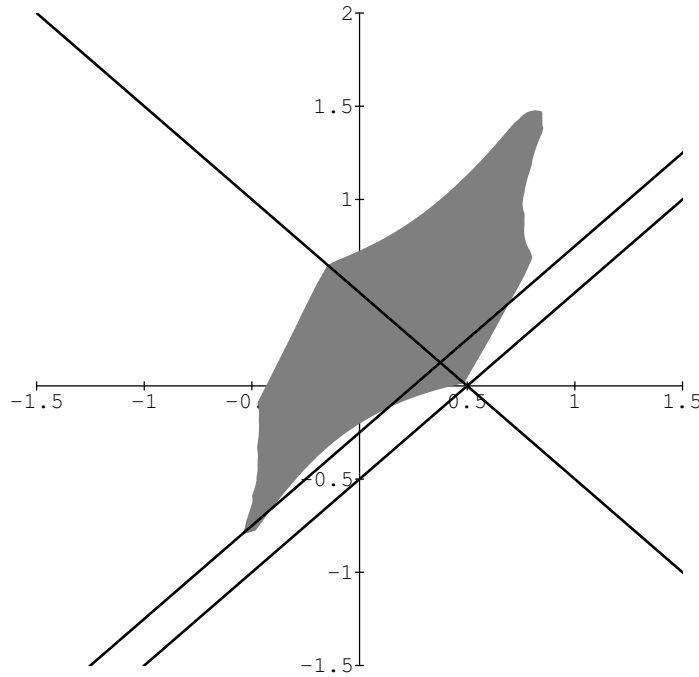


FIG. 3. The (c_0, c_1) -plane, identified with symmetric, real-valued, seven-coefficient dilation equations satisfying (2).

Example 7. Set $N = 6$ and consider the two-parameter family of all symmetric sets $\{c_0, c_1, \frac{1}{2} - c_0, 1 - 2c_1, \frac{1}{2} - c_0, c_1, c_0\}$ of real coefficients which satisfy (2). We identify these dilation equations with the (c_0, c_1) -plane. The shaded region in Fig. 3 is a numerical approximation of the region $\{(c_0, c_1) : \hat{\rho}_{16}(T_0|_V, T_1|_V) < 1\}$ (using the norm $\|u\| = |u_1| + \dots + |u_6|$) and is therefore a subset of C_V , i.e., all points in this region give rise to continuous, compactly supported scaling functions. We discuss

several subfamilies of the (c_0, c_1) -plane below; in particular, the three lines shown in Fig. 3 are discussed in parts (b), (c), and (d).

(a) The point $(\frac{1}{2}, 0)$ corresponds to a stretched dilation equation, and the associated scaling vector is $v(x) = (x, x + 1, x + 2, 3 - x, 2 - x, 1 - x)^t$. Therefore W is a line in \mathbf{C}^6 and hence is a proper subset of V . Although $\hat{\rho}(T_0|_V, T_1|_V) = 1$ we have $\hat{\rho}(T_0|_W, T_1|_W) = \frac{1}{2}$, and v is Hölder continuous with Hölder exponent 1. In particular, C_V is a proper subset of C_W .

(b) Consider now those dilation equations in the (c_0, c_1) -plane lying on the line $(c_0, c_1) = (\frac{1}{2} + \delta, -\delta)$, $\delta \in \mathbf{R}$. When $\delta = 0$ this is the stretched dilation equation discussed in part (a). When $\delta \neq 0$, Proposition 3 can be used to show that $\dim(W) = 3$. Furthermore, P_0, P_1 can be simultaneously block upper-triangularized to the form

$$\begin{pmatrix} R_i & * \\ 0 & \frac{1}{2} \end{pmatrix},$$

where R_0, R_1 are 2×2 matrices. If $\delta < -\frac{3}{8}$ or $\delta > -\frac{1}{4}$ then R_0, R_1 can be simultaneously symmetrized. The 2×2 matrices Q_0, Q_1 can be simultaneously symmetrized for all δ , whence

$$\hat{\rho}(T_0|_W, T_1|_W) = \max \left\{ \frac{1}{2}, \left| \frac{1}{2} + \delta \right|, \left| \frac{1}{2} + 2\delta \right| \right\}$$

and

$$\hat{\rho}(T_0|_V, T_1|_V) = \max \left\{ \frac{1}{2}, \left| \frac{1}{2} + \delta \right|, \left| \frac{1}{2} + 2\delta \right|, |1 + 3\delta| \right\},$$

at least for $\delta \notin [-\frac{3}{8}, -\frac{1}{4}]$. In particular,

$$\begin{cases} \hat{\rho}(T_0|_W, T_1|_W) = \hat{\rho}(T_0|_V, T_1|_V) = \frac{1}{2}, & -\frac{1}{4} < \delta < -\frac{1}{6}, \\ \hat{\rho}(T_0|_W, T_1|_W) = \frac{1}{2} < \hat{\rho}(T_0|_V, T_1|_V) < 1, & -\frac{1}{6} < \delta < 0, \\ \hat{\rho}(T_0|_W, T_1|_W) < 1 < \hat{\rho}(T_0|_V, T_1|_V), & 0 < \delta < \frac{1}{4}. \end{cases}$$

The stretched dilation equation $(\frac{1}{2}, 0)$ is therefore a boundary point of C_V , yet belongs to C_W , and there exist nonstretched dilation equations in C_W that are not in C_V .

(c) Consider next the dilation equations lying on the line $(c_0, c_1) = (\frac{1}{2} + \delta, \delta)$. Again, $\delta = 0$ corresponds to the stretched dilation equation $(\frac{1}{2}, 0)$. Applying Proposition 4 we find that $W = V$ for all $\delta \neq 0$ (we can ignore those finitely many δ for which the eigenvalues of T_0 are not distinct). Therefore $\hat{\rho}(T_0|_W, T_1|_W) \geq \rho(T_0|_W) > 1$ for $\delta \notin [(-3 - \sqrt{57})/8, -\frac{5}{4}]$, i.e., there do not exist any continuous scaling functions corresponding to $\delta \notin [(-3 - \sqrt{57})/8, -\frac{5}{4}]$. Thus, although $\hat{\rho}(T_0|_W, T_1|_W) < 1$ at the point $(\frac{1}{2}, 0)$, an arbitrarily small change in the coefficients can result in a dilation equation with $\hat{\rho}(T_0|_W, T_1|_W) > 1$, i.e., $\hat{\rho}(T_0|_W, T_1|_W)$ is *not* a continuous function of the coefficients since the subspace W can change abruptly. Proposition 2 with W in place of V is therefore false.

(d) Finally, consider those dilation equations lying on the line $(c_0, c_1) = (\frac{3}{8} + \delta, \frac{1}{8} + \delta)$. The point $(\frac{3}{8}, \frac{1}{8})$, corresponding to $\delta = 0$, also lies on the line considered in part (a), and satisfies $\hat{\rho}(T_0|_W, T_1|_W) = \frac{1}{2}$, $\hat{\rho}(T_0|_V, T_1|_V) = \frac{5}{8}$, and its associated scaling vector is continuous with Hölder exponent 1. For $\delta \neq 0$, Proposition 4 can be used to show that $W = V$. Since $\hat{\rho}(T_0|_V, T_1|_V)$ is a continuous function of the coefficients (c_0, c_1) , given ε small we can find a $\delta_0 > 0$ such that $|\hat{\rho}(T_0|_V, T_1|_V) - \frac{5}{8}| < \varepsilon$ when $0 < |\delta| < \delta_0$. Thus the scaling vectors for $0 < |\delta| < \delta_0$ are continuous, with maximum Hölder exponent $\alpha \leq -\log_2(\frac{5}{8} - \varepsilon) < 1$. However, the maximum Hölder exponent corresponding to $\delta = 0$ is exactly 1. As a function of the coefficients

(c_0, c_1) , the maximum Hölder exponent is therefore *not* continuous. By comparison, $\hat{\rho}(T_0|_V, T_1|_V) < 1$ for $0 \leq |\delta| < \delta_0$, so by Proposition 2 the associated scaling vectors deform continuously with respect to the sup-norm as δ varies. \square

In the examples we have considered, the sets W and V are equal except under restricted conditions. We therefore conjecture that for each fixed N , the set of coefficients $\{c_0, \dots, c_N\}$ satisfying (2) such that $W \neq V$ is a set of measure zero in the set of all coefficients satisfying (2).

We turn now to the proof of Theorem 3, for which we require a definition and lemma. In the statement of Proposition 4, the matrix T was assumed to have distinct eigenvalues, and hence T had a full set of linearly independent eigenvectors. Therefore, the definition of “component” used in hypothesis (b) of Proposition 4 was obvious. Now let A be an arbitrary $N \times N$ matrix with complex entries. If $\lambda \in \mathbf{C}$ is an eigenvalue of A , then $U_\lambda = \{u \in \mathbf{C}^N : (A - \lambda)^k u = 0 \text{ for some } k > 0\}$ is an A -invariant subspace of \mathbf{C}^N , for if $u \in U_\lambda$, then $Au = (A - \lambda)u + \lambda u \in U_\lambda$. By standard Jordan decomposition techniques we can write $\mathbf{C}^N = U_\lambda \oplus Z$, where Z is a unique A -invariant subspace of \mathbf{C}^N [Her]. Any vector $u \in \mathbf{C}^N$ can therefore be uniquely written $u = u_\lambda + z$ where $u_\lambda \in U_\lambda$ and $z \in Z$. We say that u has a component in U_λ if $u_\lambda \neq 0$; note this implies $u \neq 0$.

One of the basic tools of our analysis is given in the next lemma.

LEMMA 3. *Let A be an $N \times N$ matrix and let λ be any eigenvalue of A . If $u \in \mathbf{C}^N$ has a component in U_λ , then there is a constant $C > 0$ such that $\|A^n u\| \geq C |\lambda|^n$ for all $n > 0$.*

Proof. As all norms on \mathbf{C}^N are equivalent, it suffices to prove the result for the Euclidean space norm $\|u\| = (|u_1|^2 + \dots + |u_N|^2)^{1/2}$.

Let $\mathbf{C}^N = U_\lambda \oplus Z$ be the standard Jordan decomposition induced by A . The subspace U_λ can be written $U_\lambda = U_1 \oplus \dots \oplus U_p$, where each U_i is a nontrivial, A -invariant subspace of U_λ that cannot be further decomposed into A -invariant subspaces. As u has a component in U_λ it must therefore have a component in some U_i , say U_1 . Let $Z_1 = U_2 \oplus \dots \oplus U_p \oplus Z$, so $\mathbf{C}^N = U_1 \oplus Z_1$ and U_1, Z_1 are A -invariant.

Since U_1 is finite-dimensional there must exist a smallest positive integer m such that $(A - \lambda)^m w = 0$ for all $w \in U_1$. Let $u_1 \in U_1$ be such that $(A - \lambda)^{m-1} u_1 \neq 0$ and set $u_k = (A - \lambda)^{k-1} u_1$ for $k = 2, \dots, m$. Since U_1 is indecomposable, $\{u_1, \dots, u_m\}$ comprises a basis for U_1 . Moreover, the $\{u_k\}$ satisfy the relationships

$$(20) \quad \begin{aligned} Au_k &= u_{k+1} + \lambda u_k \quad \text{for } k = 1, \dots, m-1, \\ Au_m &= \lambda u_m. \end{aligned}$$

Consider now the vector u ; we have $u = v_1 + \dots + v_m + z$ where each v_k is a scalar multiple of u_k and $z \in Z_1$. Since u has a component in U_1 , at least one of the v_k is nonzero. Let k^* be that positive integer such that $v_{k^*} \neq 0$ while $v_k = 0$ for $k < k^*$. Then (20) implies that

$$(21) \quad A^n u = \lambda^n v_{k^*} + \sum_{k > k^*} P_{k,n} v_k + A^n z,$$

where each $P_{k,n}$ is a polynomial in λ of degree at most n . Now let Z_2 denote the proper subspace of \mathbf{C}^N generated by Z_1 and $\{u_k : k \neq k^*\}$. Since $\dim(Z_2) = N - 1$, there exist unique vectors $z^* \in Z_2$ and $u^* \in Z_2^\perp$ such that $v_{k^*} = u^* + z^*$. Moreover, $u^* \neq 0$ since $v_{k^*} \notin Z_2$. It follows then from (21) that for each $n > 0$,

$$A^n u = \lambda^n u^* + \lambda^n z^* + \sum_{k > k^*} P_{k,n} v_k + A^n z = \lambda^n u^* + z_n,$$

where $z_n \in Z_2$. Since u^* is orthogonal to z_n , we therefore have

$$\|A^n u\|^2 = \|\lambda^n u^*\|^2 + \|z_n\|^2 \geq |\lambda|^{2n} \|u^*\|^2. \quad \square$$

The proof of Lemma 3 can be simplified if A is diagonalizable.

Proof of Theorem 3. (a) The fact that the hypothesis $\hat{\rho}(T_0|_V, T_1|_V) < 1$ in Theorem 2 can be replaced by $\hat{\rho}(T_0|_W, T_1|_W) < 1$ and that the assumption of (2) is then no longer necessary is a simple observation following from the proof of Theorem 2 given in [DL2]. Moreover, the same proof implies that v is Hölder continuous for each exponent $0 \leq \alpha < -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$, or $0 \leq \alpha \leq -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$ if there exists a constant $C > 0$ such that $\hat{\rho}_m \leq C^{1/m} \hat{\rho}(T_0|_W, T_1|_W)$ for every m . In particular, $\alpha_{\max} \geq -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$.

(b) Assume that v is a continuous scaling vector; then v is Hölder continuous for some exponent α in the range $0 \leq \alpha \leq 1$. Let K be any corresponding Hölder constant.

Choose any $m > 0$, and let $T = T_{d_1} \cdots T_{d_m}$ be any fixed product of the matrices T_0, T_1 . Let λ be any eigenvalue of $T|_W$. By definition of W , there must be dyadic points $x, y \in [0, 1]$ such that $v(x) - v(y)$ has a component in $U_\lambda = \{u \in W : (T - \lambda)^k u = 0 \text{ for some } k > 0\}$. Let $x = .x_1 x_2 \dots$ and $y = .y_1 y_2 \dots$ be binary expansions of x and y , respectively, and define the dyadic points X_k, Y_k for $k > 0$ by

$$\begin{aligned} X_1 &= .d_1 \dots d_m x_1 x_2 \dots, & Y_1 &= .d_1 \dots d_m y_1 y_2 \dots, \\ X_2 &= .d_1 \dots d_m d_1 \dots d_m x_1 x_2 \dots, & Y_2 &= .d_1 \dots d_m d_1 \dots d_m y_1 y_2 \dots, \end{aligned}$$

etc., so $v(X_k) = T^k v(x)$ and $v(Y_k) = T^k v(y)$ for all $k > 0$. We then have by Lemma 3 that

$$\|v(X_k) - v(Y_k)\| = \|T^k(v(x) - v(y))\| \geq C |\lambda|^k$$

for some constant $C > 0$ independent of k . Since $|X_k - Y_k| \rightarrow 0$ as $k \rightarrow \infty$ and v is continuous, we must therefore have $|\lambda| < 1$. In fact, $|X_k - Y_k| = 2^{-mk} |x - y|$, so

$$\|v(X_k) - v(Y_k)\| \geq K_\lambda |X_k - Y_k|^{\alpha_\lambda}$$

for all k , where $\alpha_\lambda = -\log_2 |\lambda|^{1/m}$ and $K_\lambda = C |x - y|^{-\alpha_\lambda}$. However, by the Hölder continuity of v ,

$$\|v(X_k) - v(Y_k)\| \leq K |X_k - Y_k|^\alpha$$

for all k , so $|X_k - Y_k|^{\alpha - \alpha_\lambda} \geq K_\lambda / K$ for all k . Since $|X_k - Y_k| \rightarrow 0$ as $k \rightarrow \infty$, this implies $\alpha \leq \alpha_\lambda$.

Taking the supremum over all eigenvalues of $T = T_{d_1} \cdots T_{d_m}$ for all choices of $d_j = 0, 1$, we obtain $\hat{\sigma}_m < 1$ and $\alpha \leq -\log_2 \hat{\sigma}_m$. Thus $\hat{\rho}(T_0|_W, T_1|_W) = \sup \hat{\sigma}_m \leq 1$ and $\alpha \leq -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$, so $\alpha_{\max} \leq -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$. Combining this with the proof of part (a), we obtain $\alpha_{\max} = -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$.

Suppose that $\alpha = \alpha_{\max}$. Then given any $x = .x_1 x_2 \dots$ and $y = .y_1 y_2 \dots$, we have

$$\|T_{d_1} \cdots T_{d_m}(v(x) - v(y))\| = \|v(X) - v(Y)\| \leq K |X - Y|^\alpha = K (2^{-m} |x - y|)^\alpha,$$

where $X = .d_1 \dots d_m x_1 x_2 \dots$ and $Y = .d_1 \dots d_m y_1 y_2 \dots$. By considering a basis $\{v(x_j) - v(y_j)\}_{j=1}^J$ for W , we obtain that there is a constant $C > 0$ such that $\|T_{d_1} \cdots T_{d_m} w\| \leq C 2^{-m\alpha} \|w\|$ for every $w \in W$, and therefore $\hat{\rho}_m \leq C^{1/m} 2^{-\alpha} = C^{1/m} \hat{\rho}(T_0|_W, T_1|_W)$.

The proof is now complete except for the case of a dilation equation satisfying

$$(22) \quad \sup_m \hat{\sigma}_m = \hat{\rho}(T_0|_W, T_1|_W) = 1 \quad \text{and} \quad \hat{\sigma}_m < 1 \quad \text{for every } m.$$

Assume therefore that a continuous scaling vector exists for a dilation equation satisfying (22). Then each $\hat{\rho}_m \geq 1$, and so for every $m > 0$ there is some product of $T_0|_W, T_1|_W$ of length m having norm greater than or equal to 1. Using a diagonalization process, we can therefore construct a sequence d_1, d_2, \dots so that $S_m = T_{d_1} \cdots T_{d_m}$ satisfies $\|S_m|_W\| \geq 1$ for all m . Let $\{v(x_j) - v(y_j)\}_{j=1}^J$ be a basis for W . Then there exist vectors $u_m = \sum_{j=1}^J a_{m,j} (v(x_j) - v(y_j)) \in W$ such that $\|u_m\| = 1$ and $\|S_m u_m\| \geq 1$ for every m . Since W is finite dimensional, we can (by taking subsequences if necessary) suppose that for each j the limit $a_j = \lim_{m \rightarrow \infty} a_{m,j}$ exists. Then $S_m u_m \rightarrow u$ with $\|u\| = \lim_{m \rightarrow \infty} \|S_m u_m\| \geq 1$. Now, by the continuity of v , we have that $S_m v(x_j), S_m v(y_j) \rightarrow v(x)$ as $m \rightarrow \infty$, where $x = .d_1 d_2 \dots$, so

$$u = \lim_{m \rightarrow \infty} S_m u_m = \lim_{m \rightarrow \infty} \sum_{j=1}^J a_{m,j} S_m (v(x_j) - v(y_j)) = \sum_{j=1}^J a_j (v(x) - v(y)) = 0,$$

a contradiction. \square

The proof of Theorem 3(b) raises an interesting question: do there exist dilation equations such that (22) holds? A scaling function for such a dilation equation must be discontinuous by Theorem 3(b). More generally, the same existence question can be asked with the value 1 in (22) replaced by any positive number; we do not know of any specific examples of such dilation equations, i.e., dilation equations for which the supremum $\sup \hat{\sigma}_m$ is not achieved for any m . Note that $\sup \hat{\sigma}_m = \hat{\sigma}_1$ when simultaneous symmetrization occurs; however, the supremum may be achieved even though simultaneous symmetrization does not occur. For example, consider the dilation equation defined by $N = 3$ and $(c_0, c_3) = (\frac{1}{3}, \frac{1}{3})$. In this case $V = W$ and $\sup \hat{\sigma}_m = \hat{\sigma}_2 > \hat{\sigma}_1$. Lagarias and Wang have conjectured that the supremum must always be achieved; in [LW] they prove several results related to this conjecture.

Theorem 3 discusses only global Hölder exponents. It is possible for a local Hölder exponent at a particular point to be strictly greater than the maximum global Hölder exponent. For example, assume $\hat{\rho}(T_0|_W, T_1|_W) < 1$; then v is continuous by Theorem 3, and its maximum global Hölder exponent of continuity is $\alpha_{\max} = -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$. Fix any product $T = T_{d_1} \cdots T_{d_m}$, and let z be any rational point whose binary expansion is of the form $z = .z_1 \dots z_n d_1 \dots d_m d_1 \dots d_m \dots$. If $y > z$ with $2^{-n-(k+1)m} \leq y - z < 2^{-n-km}$, then the first $n + km$ digits of y are the same as those of z (using the upper binary expansion of z if z is dyadic). Therefore

$$\begin{aligned} \|v(y) - v(z)\| &= \|T_{z_1} \cdots T_{z_n} T^k (v(\tau^{n+km} y) - v(\tau^{n+km} z))\| \\ &\leq C \|(T_{z_1} \cdots T_{z_n})|_W\| \|T^k|_W\| \\ &\leq C' \theta^k \\ &= C' \theta^{-(n+m)/m} (2^{-n-(k+1)m})^{-\log_2 \theta^{1/m}} \\ &\leq C' \theta^{-(n+m)/m} |y - z|^{-\log_2 \theta^{1/m}}, \end{aligned}$$

where $\rho(T|_W) < \theta < 1$ and C and C' are constants independent of k . Using a similar argument from the left (with the lower binary expansion for z if z is dyadic), we conclude that v is locally Hölder continuous at z at least for each exponent α_z in the range

$$(23) \quad 0 \leq \alpha_z < \begin{cases} -\log_2 \rho(T|_W)^{1/m}, & \text{if } m > 1, \\ -\log_2 \max\{\rho(T_0|_W), \rho(T_1|_W)\}, & \text{if } m = 1. \end{cases}$$

Thus the maximum local Hölder exponent at z is bounded below by the right-hand side of (23). This quantity depends only on the single product T (or on T_0 and T_1 if z is dyadic), and therefore can be strictly greater than the maximum global exponent, which depends on all possible products of T_0 and T_1 . The proof of Theorem 3(b) demonstrates that the right-hand side of (23) is the maximum local Hölder exponent at the point $x = .d_1 \dots d_m d_1 \dots d_m \dots$. If $T_0|_W$ and $T_1|_W$ are invertible, then

$$\|v(\tau^m y) - v(x)\| \leq \|(T_{z_1} \dots T_{z_n}|_W)^{-1}\| \|v(y) - v(z)\|,$$

and therefore in this case the right-hand side of (23) is also the maximum local Hölder exponent at any z of the form $z = .z_1 \dots z_n d_1 \dots d_m d_1 \dots d_m \dots$.

Example 8. Consider $N = 3$ as in Example 2. Given coefficients (c_0, c_3) , the maximum local Hölder exponent of continuity at any dyadic point is $\alpha_{\text{dyadic}} = -\log_2 \max\{|c_0|, |c_3|, |1 - c_0 - c_3|\}$. Consider the following specific examples.

(a) In [CH1] we determined that the scaling function determined by $(c_0, c_3) = (0.6, -0.2)$ is Hölder continuous with maximum global Hölder exponent somewhere in the range $0.598 \leq \alpha_{\text{max}} \leq 0.600$. However, $\alpha_{\text{dyadic}} = -\log_2 0.6 \approx 0.737$. This is the largest maximum local Hölder exponent at dyadic points for (c_0, c_3) satisfying both (2) and (3).

(b) If $\max\{|c_0|, |c_3|, |1 - c_0 - c_3|\} < \frac{1}{2}$ then α_{dyadic} strictly exceeds 1. It follows that v is differentiable at all dyadic z with $v'(z) = 0$. If v was differentiable everywhere, then this would imply that v was identically constant, which is not the case. For example, for $(c_0, c_3) = (1/3, 1/3)$ the maximum global Hölder exponent is only $\alpha_{\text{max}} \approx 0.891$. \square

Finally, we use the techniques applied in the proof of Theorem 3 to make some observations about discontinuous scaling functions. In particular, we obtain the following theorem.

THEOREM 4. *Assume v is a scaling vector.*

(a) *If $\hat{\rho}(T_0|_W, T_1|_W) \geq 1$ with $\hat{\sigma}_m \geq 1$ for some m , then v is discontinuous. If $T_0|_W, T_1|_W$ are invertible, then discontinuities occur on a dense set of rational points in $[0, 1]$.*

(b) *If $\hat{\rho}(T_0|_W, T_1|_W) > 1$, then v is unbounded. If $T_0|_W, T_1|_W$ are invertible, then singularities occur on a dense set of rational points in $[0, 1]$.*

Proof. (a) If $\hat{\sigma}_m \geq 1$ for some m then there is some product $T = T_{d_1} \dots T_{d_m}$ and some eigenvalue λ of $T|_W$ such that $|\lambda| \geq 1$. Let X_k, Y_k be as in the proof of Theorem 3(b); it follows then that $\|v(X_k) - v(Y_k)\| \geq C|\lambda|^k$ for some constant $C > 0$ independent of k . As $X_k, Y_k \rightarrow X = .d_1 \dots d_m d_1 \dots d_m \dots$, it follows that v cannot be continuous at X . An argument similar to the one following equation (23) shows that if $T_0|_W, T_1|_W$ are invertible then discontinuities also occur at every point z of the form $z = .z_1 \dots z_n d_1 \dots d_m d_1 \dots d_m \dots$.

(b) If $\hat{\rho}(T_0|_W, T_1|_W) > 1$ then $\hat{\sigma}_m > 1$ for some m , and therefore there is some product $T = T_{d_1} \dots T_{d_m}$ and some eigenvalue λ of $T|_W$ such that $|\lambda| > 1$. The remainder of the proof follows just as in part (a). \square

As a final comment, suppose we are given a scaling vector v with $\hat{\rho}(T_0|_W, T_1|_W) > 1$. We have then by Theorem 4(b) that v is unbounded. The singularities constructed in the proof of this result occur only at rational points. Consider then the function \tilde{v} obtained by setting $\tilde{v}(x) = v(x)$ for nonrational x and $\tilde{v}(x) = 0$ for rational x . This is also a scaling vector, and $\hat{\rho}(\tilde{T}_0|_{\tilde{W}}, \tilde{T}_1|_{\tilde{W}}) = 0$. However, \tilde{v} is clearly discontinuous; note that Theorem 3(a) does not imply continuity since \tilde{v} is not given at dyadic points by the method of Construction 1. Since $\hat{\rho}(\tilde{T}_0|_{\tilde{W}}, \tilde{T}_1|_{\tilde{W}}) < 1$, Theorem 4(b) does not

imply that \tilde{v} is unbounded. We therefore ask whether it is possible that \tilde{v} be bounded. As pointed out by the referee, the subspaces W or \tilde{W} are probably not the best for dealing with this question. A more relevant subspace might be

$$\bigcap_S \text{span}\{v(x) - v(y) : x, y \in S\},$$

where the intersection is taken over all subsets $S \subset [0, 1]$ with measure one that are invariant under τ .

Acknowledgments. We thank George Benke of The MITRE Corporation and Gil Strang of MIT for many stimulating discussions and valuable insights on this subject.

REFERENCES

- [BW] M. A. BERGER AND Y. WANG, *Bounded semi-groups of matrices*, Linear Algebra Appl., 166 (1992), pp. 21–27.
- [CDM] A. CAVARETTA, W. DAHMEN, AND C. A. MICCHELLI, *Stationary Subdivision*, Mem. Amer. Math. Soc., 93 (1991), pp. 1–186.
- [CH1] D. COLELLA AND C. HEIL, *The characterization of continuous, four-coefficient scaling functions and wavelets*, IEEE Trans. Inf. Th., Special Issue on Wavelet Transforms and Multiresolution Signal Analysis, 38 (1992), pp. 876–881.
- [CH2] ———, *Characterizations of scaling functions, II. Distributional and functional solutions*, manuscript.
- [D] I. DAUBECHIES, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 41 (1988), pp. 909–996.
- [DL1] I. DAUBECHIES AND J. LAGARIAS, *Two-scale difference equations: I. Existence and global regularity of solutions*, SIAM J. Math. Anal., 22 (1991), pp. 1388–1410.
- [DL2] ———, *Two-scale difference equations: II. Local regularity, infinite products of matrices and fractals*, SIAM J. Math. Anal., 23 (1992), pp. 1031–1079.
- [DL3] ———, *Sets of matrices all infinite products of which converge*, Linear Algebra Appl., 161 (1992), pp. 227–263.
- [DD] G. DESLAURIERS AND S. DUBUC, *Symmetric iterative interpolation processes*, Constr. Approx., 5 (1989), pp. 49–68.
- [Du] S. DUBUC, *Interpolation through an iterative scheme*, J. Math. Anal. Appl., 114 (1986), pp. 185–205.
- [DGL] N. DYN, J. A. GREGORY, AND D. LEVIN, *Analysis of uniform binary subdivision schemes for curve design*, Constr. Approx., 7 (1991), pp. 127–147.
- [E] T. EIROLA, *Sobolev characterization of solutions of dilation equations*, SIAM J. Math. Anal., 23 (1992), pp. 1015–1030.
- [H] C. HEIL, *Methods of solving dilation equations*, in Prob. and Stoch. Methods in Anal. with Appl., J. S. Byrnes, J. L. Byrnes, K. A. Hargreaves, and K. Berry, eds., NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci. 372, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1992, pp. 15–45.
- [HC] C. HEIL AND D. COLELLA, *Dilation equations and the smoothness of compactly supported wavelets*, in Wavelets: Mathematics and Applications, J. Benedetto and M. Frazier, eds., CRC Press, Boca Raton, FL, 1993, pp. 161–200.
- [HS] C. HEIL AND G. STRANG, *Continuity of the joint spectral radius: Application to wavelets*, in Linear Algebra for Signal Processing, A. Bojanczyk and G. Cybenko, eds., IMA Vol. Math. Appl., Springer-Verlag, New York, to appear.
- [Her] I. N. HERSTEIN, *Topics in Algebra, 2nd ed.*, Second Edition, John Wiley and Sons, New York, 1975.
- [LW] J. C. LAGARIAS AND Y. WANG, *The finiteness conjecture for the joint spectral radius*, Linear Algebra Appl., to appear.
- [L] W. LAWTON, *Tight frames of compactly supported affine wavelets*, J. Math. Phys., 31 (1990), pp. 1898–1901.

- [M] S. G. MALLAT, *Multiresolution approximations and wavelet orthonormal bases for $L^2(\mathbf{R})$* , Trans. Amer. Math. Soc., 315 (1989), pp. 69–87.
- [MP1] C. A. MICCHELLI AND H. PRAUTZSCH, *Refinement and subdivision for spaces of integer translates of compactly supported functions*, in Numerical Analysis, D. F. Griffith and G. A. Watson, eds., Academic Press, New York, 1987, pp. 192–222.
- [MP2] ———, *Uniform refinement of curves*, Linear Algebra Appl., 114/115 (1989), pp. 841–870.
- [RS] G. C. ROTA AND G. STRANG, *A note on the joint spectral radius*, Kon. Nederl. Akad. Wet. Proc. A, 63 (1960), pp. 379–381.
- [S] G. STRANG, *Wavelets and dilation equations: a brief introduction*, SIAM Rev., 31 (1989), pp. 614–627.
- [V] L. VILLEMOES, *Energy moments in time and frequency for two-scale difference equations*, SIAM J. Math. Anal., 23 (1992), pp. 1519–1543.
- [W] Y. WANG, *Two-scale dilation equations and the cascade algorithm*, Random and Computational Dynamics, to appear.