In: Wavelets: Mathematics and Applications, J. Benedetto and M. Frazier, eds, CRC Press, Boca Raton, FL, 1994, pp. 161-200.

4

Dilation equations and the smoothness of compactly supported wavelets

Christopher Heil* and David Colella†

ABSTRACT The construction of compactly supported wavelets with specified amounts of smoothness is an important problem in wavelet theory. This problem reduces to the construction of scaling functions, i.e., solutions f of dilation equations $f(t) = \sum_{k=0}^{N} c_k\, f(2t-k)$, with specified smoothness. This article characterizes all smooth, compactly supported scaling functions in terms of a joint spectral radius of two $N \times N$ matrices T_0, T_1 constructed from the coefficients $\{c_0, \dots, c_N\}$ of the dilation equation, restricted to an appropriate subspace of $\mathbf{C}^N$. The number of continuous derivatives of the scaling function and the range of Hölder exponents of continuity of the last continuous derivative are determined by the value of this joint spectral radius. Numerous examples are provided to illustrate the results.

CONTENTS

*The first author is also Pure Mathematics Instructor at The Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and acknowledges partial support by National Science Foundation Grant DMS-9007212.

†The MITRE Corporation, McLean, Virginia 22102

4.1 Introduction.

Dilation equations play a crucial role in the construction and resulting properties of multiresolution analyses and wavelet orthonormal bases for $L^2(\mathbf{R})$, the space of square-integrable functions on the real line [BW2], [ColHe1], [ColHe3], [ColHe4], [DL1], [DL2], [E], [R1], [R2], [S], [V1], [V2], [W]. This connection is one reason for the recent burst of activity in the study of such equations and their solutions. There are also important applications of dilation equations to other areas, most notably interpolating subdivision schemes, although those are outside the scope of this chapter [CDaMi], [DeDu], [Du], [DyGLe], [DyLe], [MiP1], [MiP2].

Multiresolution analyses are discussed in detail elsewhere in this volume and therefore will not be defined here (see, for example, Chapter 1 by Strichartz, or references [D1], [D2], [He] [M], [Me]). Each multiresolution analysis determines a *scaling function* f, i.e., a solution of a *dilation equation*

$$f(t) \;=\; \sum_{k=-\infty}^{\infty} c_k\, f(2t-k), \tag{4.1}$$

which is square-integrable. The coefficients $\{c_k\}$ are square-summable real or complex numbers. This scaling function then determines a *wavelet*

$$g(t) \;=\; \sum_{k=-\infty}^{\infty} (-1)^k\, c_{1-k}\, f(2t-k), \tag{4.2}$$

such that the collection $\{g_{n,k}\}_{n,k\in\mathbf{Z}}$ with $g_{n,k}(t) = 2^{n/2}\, g(2^n t - k)$ forms an orthonormal basis for $L^2(\mathbf{R})$ after suitable normalization of g. This *wavelet orthonormal basis* is thus obtained by dilating and translating a single $L^2(\mathbf{R})$ function, and therefore the properties of the basis elements are completely determined by the corresponding properties of the wavelet.

Wavelet orthonormal bases are appealing for a variety of reasons, one being that it is possible to find wavelets g which have good localization in both time and frequency. For example, in this chapter we characterize all wavelets arising from finite-coefficient dilation equations which are smooth and compactly supported—such wavelets are necessarily well-localized in the time domain and have good decay in their Fourier transforms. Moreover, we characterize the exact degree of smoothness of these wavelets, by which we mean that we determine the total number of continuous derivatives and the possible Hölder exponents of continuity of the last derivative. As proved by Daubechies and Lagarias [DL1] (and discussed in Section 4.7.1), compact support is incompatible with infinite differentiability. A function h defined on the real line $\mathbf{R}$ is *Hölder continuous* if there exist constants α, K such that $|h(x) - h(y)| \le K\,|x-y|^{\alpha}$ 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. If h is differentiable then we can take $\alpha = 1$, but not conversely.

In order to ensure compact support for the wavelet g we assume that the number of nonzero coefficients c_k in Eq. (4.1) is finite. By translating the scaling function f if necessary, we therefore assume for the remainder of the chapter that the dilation equation has the form

$$f(t) \;=\; \sum_{k=0}^{N} c_k \, f(2t-k), \tag{4.3}$$

i.e., we assume $c_k = 0$ for $k < 0$ or $k > N$. We seek square-integrable, compactly supported solutions of Eq. (4.3). Such scaling functions are necessarily integrable, and we show below that compactly supported, integrable solutions to Eq. (4.3) are unique up to multiplication by a constant (if they exist), and are supported in the finite interval $[0, N]$. If f is such a compactly supported scaling function then the wavelet g is obtained from f by a finite sum. Thus the smoothness of g is entirely determined by the corresponding smoothness of the scaling function f. We therefore concentrate for the remainder of the chapter on the properties of compactly supported scaling functions, rather than wavelets.

The class of dilation equations induced by multiresolution analyses is quite restrictive; in particular, the following two conditions must be satisfied:

$$\sum_k c_{2k} \;=\; \sum_k c_{2k+1} \;=\; 1 \tag{4.4}$$

and

$$\sum_k c_k \, \bar{c}_{k+2j} \;=\; \begin{cases} 2, & \text{if } j = 0, \\ 0, & \text{if } j \neq 0. \end{cases} \tag{4.5}$$

In fact, not only are these two conditions necessary for the existence of a multiresolution analysis, they are also almost sufficient. By this we mean the following. Assume that coefficients $\{c_0, \dots, c_N\}$ are given which satisfy both Eqs. (4.4) and (4.5). Then it can be shown that the dilation equation will have an integrable and square-integrable solution f. If f is orthogonal to its integer translates, i.e., if $\int f(t)\, \overline{f(t-k)}\, dt = 0$ when $k \neq 0$, then f can be used to construct a multiresolution analysis [M]. Lawton [La] has shown that for each N, f will be orthogonal to its integer translates except for a set of coefficients $\{c_0, \dots, c_N\}$ of measure zero in the set of all coefficients $\{c_0, \dots, c_N\}$ satisfying Eqs. (4.4) and (4.5).

Let us illustrate the preceding remarks with a simple example. If we set $N = 1$ and define $c_0 = c_1 = 1$ then the dilation equation has the form

$$f(t) \;=\; f(2t) + f(2t-1).$$

By inspection, one compactly supported solution is $f = \chi_{[0,1)}$, the characteristic function of the interval $[0, 1)$. Note that the coefficients satisfy both of Eqs. (4.4) and (4.5), and that f is (trivially) orthogonal to its integer translates. We therefore expect that the wavelet

$$g(t) \;=\; f(2t) - f(2t-1) \;=\; \chi_{[0,1/2)}(t) - \chi_{[1/2,1)}(t)$$

will generate an orthonormal basis for $L^2(\mathbf{R})$, and in fact this is the case. This wavelet basis $\{g_{n,k}\}_{n,k\in\mathbf{Z}}$ was first constructed by Haar [Ha], and is today known as the *Haar system*. Although compactly supported, the Haar wavelet g is certainly not smooth! Smooth, compactly supported wavelets were first constructed by Daubechies [D1].

Eqs. (4.4) and (4.5) are necessary for the existence of multiresolution analyses. However, these equations do not play any significant role in many of the applications of dilation equations to other areas. Therefore, we will not impose any restrictions on the coefficients $\{c_0, \dots, c_N\}$ in our characterization of smooth, compactly supported scaling functions, i.e., we will obtain all possible smooth, compactly supported solutions to dilation equations, without regard to their applicability to multiresolution analysis. There are several different methods for constructing such solutions, most of which can be classified as combinations of one or more of three techniques, which we shall refer to as Cascade Algorithm techniques, Fourier transform techniques, and dyadic interpolation techniques. We briefly outline these techniques in the remainder of this section. We then expand somewhat on the use of the Fourier transform in Section 4.2, and use dyadic interpolation thoughout the remainder of the chapter (Sections 4.3 through 4.8). We make minor use of the Cascade Algorithm in Section 4.8.2.

The Cascade Algorithm is based on the fact that a scaling function is a fixed point for the linear operator F defined by

$$Fh(t) \;=\; \sum_{k=0}^{N} c_k\, h(2t-k).$$

As usual, the fixed point is located by iterating, or "cascading," some reasonable starting function h_0 (typically, $h_0 = \chi_{[0,1]}$) through the operator F. That is, one hopes that the functions h_n defined by $h_n = Fh_{n-1}$ will converge to a scaling function f. However, convergence of the algorithm is not guaranteed, or may occur only in a weak sense. This type of algorithm is typically used to study fractals; it is therefore not surprising that scaling functions or their last derivatives are often Hölder continuous with exponent strictly less than one. Some articles in which the Cascade Algorithm is discussed include [BW2], [DL1], [DeDu], [DyGLe], [DyLe], [D1], [D2].

The use of the Fourier transform, on the other hand, is suggested by the convolution nature of the dilation equation. We therefore take the Fourier transform of the dilation equation and obtain its equivalent Fourier transform version:

$$\hat{f}(\gamma) \;=\; m_0(\gamma/2)\, \hat{f}(\gamma/2), \tag{4.6}$$

where $m_0(\gamma) = \frac{1}{2}\sum c_k\, e^{ik\gamma}$ is the *symbol* of the coefficients $\{c_0, \dots, c_N\}$. Iterating Eq. (4.6) we obtain

$$\hat{f}(\gamma) \;=\; m_0(\gamma/2)\, \hat{f}(\gamma/2) \;=\; m_0(\gamma/2)\, m_0(\gamma/4)\, \hat{f}(\gamma/4) \;=\; \cdots,$$

which suggests that properties of $\hat{f}$ will be determined by the convergence of an infinite product involving m_0. This technique can be used, for example, to prove uniqueness results or to study the smoothness of f via consideration of the decay of $\hat{f}$, but is usually not optimal for estimating time domain properties of f such as the Hölder exponent of continuity. We discuss this technique in more detail in Section 4.2; some articles in which Fourier transform techniques are used include [ColHe4], [DL1], [E], [V1], [V2], [D1], [D2], [M].

Finally, the dyadic interpolation technique is the method that we will use to construct and characterize all smooth, compactly supported scaling functions. It is based on the observation that if the values of a scaling function f are known at, say, the integer points, then the dilation equation immediately determines their values at the half-integer points. From this we obtain next the values at the quarter-integers, and so by recursion eventually finding the values at all *dyadic points*, i.e., all points of the form $k/2^n$. If f is uniformly continuous on this restricted set of points then it can be extended to all points in a continuous way. The implementation and execution of this method occupies us for the majority of the chapter, specifically Sections 4.3–4.8. We outline below the content of each of these sections.

In Section 4.3 we show how to implement dyadic interpolation in a matrix formulation. This requires an equivalent matrix form of the dilation equation, which we obtain as follows. First, the scaling function f, which is supported in the interval $[0, N]$, is converted into a vector-valued function v defined on the interval $[0, 1]$. Given $x \in [0, 1]$, the first component $v_1(x)$ of $v(x)$ is simply the value of $f(x)$; the second component $v_2(x)$ is then $f(x + 1)$, and so forth. We refer to this vector-valued function v as a *scaling vector*. Next, given $x \in [0, 1]$, we show that the collection of N equations obtained by evaluating the dilation equation at each of the points $x, x + 1, \dots, x + N - 1$ can be expressed as a matrix equation involving the scaling vector v and an appropriate product of two matrices T_0, T_1 which are determined from the coefficients $\{c_0, \dots, c_N\}$. The product has the form $T_{d_1} \cdots T_{d_m}$, where $d_1, \dots, d_m$ are the first m digits of the binary (base two) expansion of x. We write this expansion as $x = .d_1 d_2 \cdots$; it is unique for non-dyadic points. Dyadic points, on the other hand, have two binary expansions: a "terminating" expansion ending in infinitely many zeros, which we write for convenience as $x = .d_1 \cdots d_j$, and another expansion ending in infinitely many ones. Because the matrices T_0, T_1 satisfy a certain "consistency" condition, either expansion may be used to form the product $T_{d_1} \cdots T_{d_m}$ when x is dyadic.

Section 4.4 uses this matrix form of the dilation equation to find sufficient conditions under which a continuous scaling vector (and hence a continuous, compactly supported scaling function) can be constructed. This essentially amounts to finding conditions under which f, once defined at dyadic points by dyadic interpolation, will be uniformly continuous on this restricted set. The sufficient conditions are given in terms of a growth condition on all possible products of T_0 and T_1, restricted to an appropriate subspace W of $\mathbf{C}^N$. Moreover, a lower bound for the Hölder exponent of continuity is also obtained.

In Section 4.5 we show how these growth conditions on products of T_0 and T_1 can be viewed as a *joint spectral radius* of the two matrices. We therefore digress temporarily and discuss the basic properties of the joint spectral radius in Section 4.5.1, and methods of computing it in Section 4.5.2.

We return to considering the continuity of scaling vectors and associated scaling functions in Section 4.6. In particular, we prove that the results of Section 4.4 are sharp, i.e., the joint spectral radius can be used to precisely characterize all dilation equations which have continuous, compactly supported solutions, and to give precise bounds on the possible Hölder exponents of continuity.

The results of Sections 4.4 and 4.6 completely determine all possible continuous, compactly supported scaling functions. However, if f is a differentiable scaling function for the dilation equation with coefficients $\{c_0, \dots, c_N\}$ then f' is a scaling function for the dilation equation with coefficients $\{2c_0, \dots, 2c_N\}$. Since we know when this latter dilation equation will have a continuous solution, we have an implicit characterization of all dilation equations having continuously differentiable solutions, and, by iteration, of all n-times continuously differentiable solutions. In Section 4.7.1 we make this implicit characterization explicit, also in terms of an appropriate joint spectral radius. In Section 4.7.2 we show how the assumption of extra *sum rules* on the coefficients affects that joint spectral radius.

Finally, in Section 4.8 we give several examples illustrating the results of previous sections for specific dilation equations. In Section 4.8.1 we find all dilation equations with $N = 3$ which have smooth, compactly supported solutions. In Section 4.8.2 we consider the analogous problem for $N = 6$ with the additional assumption that the coefficients $\{c_0, \dots, c_6\}$ satisfy Eq. (4.4) and are symmetric, i.e., $c_0 = c_6$, $c_1 = c_5$, and $c_2 = c_4$.

Dyadic interpolation methods have been used by many authors, and we do not take credit for inventing them. In particular, Michelli and Prautzsch [MiP2] (motivated by interpolating subdivision problems) were the first to implement dyadic interpolation via products of two matrices. However, they did not make use of the joint spectral radius, nor did they obtain estimates for the Hölder exponent. Daubechies and Lagarias [DL1], [DL2] independently discovered the same implementation, and gave sufficient conditions for the existence of smooth, compactly supported scaling functions and lower bounds for the corresonding Hölder exponents of the last derivative in terms of a joint spectral radius. We improved on their sufficient conditions and proved that those improved conditions are also necessary [ColHe1], [ColHe3]. In particular, for the case of continuous, compactly supported scaling functions, the Daubechies and Lagarias theorem is essentially a restricted version of our Theorem 4.3 in Section 4.4. Our converse is Theorem 4.11 in Section 4.6. Equivalent results have also been obtained by Berger and Wang [BW2], [W]. We note that Rioul [R1], [R2] has a different implementation of dyadic interpolation, and obtains estimates for the Hölder exponent not expressed in terms of a joint spectral radius. Also, although we discuss smoothness only in global terms, the matrix methods we describe can be used to obtain detailed local information, e.g., smoothness at a point or at a class of points [DL2]. Finally, the techniques we discuss are also applicable to dilation equations with integer scale factors other than two, and to certain higher-dimensional dilation equations. Daubechies and Lagarias [DL2]. give some indications of how this can be done.

4.2 The Fourier transform.

Let us begin our study of dilation equations by assuming that f is an integrable scaling function—this includes, of course, the class of smooth, compactly supported scaling functions which is our primary interest. The Fourier transform $\hat{f}(\gamma) = \int f(t)\, e^{i\gamma t}\, dt$ of f is then a well-defined continuous function, so Eq. (4.6) gives an equivalent form for the dilation equation on the Fourier transform side. In particular, when $\gamma = 0$ we have $\hat{f}(0) = \Delta\, \hat{f}(0)$, where

$$\Delta \;=\; m_0(0) \;=\; \tfrac{1}{2} \sum c_k,$$

and therefore $\int f(t)\, dt = \hat{f}(0) = 0$ if $\Delta \neq 1$.

By iterating Eq. (4.6), we obtain that

$$\hat{f}(\gamma) \;=\; \hat{f}(\gamma/2^n) \prod_{j=1}^{n} m_0(\gamma/2^j) \tag{4.7}$$

for each $n > 0$. We are now tempted to let $n \to \infty$, but we may have convergence problems in the infinite product if $\Delta \neq 1$. Therefore, let us temporarily impose the restriction that $\Delta = 1$. Since m_0 is a trigonometric polynomial, it is not difficult to show that the infinite product

$$P(\gamma) \;=\; \prod_{j=1}^{\infty} m_0(\gamma/2^j) \tag{4.8}$$

then converges (uniformly on compact sets) to a continuous function, and therefore

$$\hat{f}(\gamma) \;=\; \hat{f}(0)\, P(\gamma).$$

In particular, note that $\hat{f}(0) \neq 0$ if f is nontrivial. Moreover, since P is independent of f, this shows that integrable solutions to dilation equations are unique up to scale when $\Delta = 1$, and we show next that such solutions are necessarily compactly supported. We will also show below that compactly supported, integrable scaling functions are likewise unique up to scale when $\Delta \neq 1$, although in that case there may also exist non-compactly supported, integrable scaling functions.

To calculate the support of f, we return to the time side. Let $\nu = \frac{1}{2} \sum c_k\, \delta_k$, where δ_k is the point mass at the point k. Then $\hat{\nu} = m_0$ and $\mathrm{supp}(\nu) = \{0, \dots, N\} \subset [0, N]$. Let ν_n be a compressed version of ν, i.e., $\nu_n = \frac{1}{2} \sum c_k\, \delta_{2^{-n}k}$, and let μ_n be the measure $\mu_n = \nu_1 * \cdots * \nu_n$, where the asterisk denotes convolution.

Then $\hat{\mu}_n(\gamma) = \prod_{j=1}^n m_0(\gamma/2^j)$ and $\mathrm{supp}(\mu_n) \subset [0, N/2] + \cdots + [0, N/2^n] \subset [0, N]$. Since $\hat{f}(0)\,\hat{\mu}_n(\gamma) \to \hat{f}(\gamma)$ as $n \to \infty$, we therefore have $\hat{f}(0)\,\mu_n \to f$ in the weak* topology, and hence $\mathrm{supp}(f) \subset [0, N]$.

Let us consider now what happens when $\Delta \neq 1$. In this case we adjust the terms of the infinite product in Eq. (4.8) in order to obtain convergence. Specifically, we rewrite Eq. (4.7) in the following form:

$$\hat{f}(\gamma) \;=\; \hat{f}(\gamma/2^n)\,\Delta^n \prod_{j=1}^{n} \frac{m_0(\gamma/2^j)}{\Delta}, \tag{4.9}$$

and note that

$$P_\Delta(\gamma) \;=\; \prod_{j=1}^{\infty} \frac{m_0(\gamma/2^j)}{\Delta}$$

converges to a continuous function. Since $\hat{f}(0) = 0$, it follows by letting $n \to \infty$ in Eq. (4.9) that $\hat{f} \equiv 0$ if $|\Delta| \leq 1$ with $\Delta \neq 1$. Thus there are no nontrivial integrable solutions of dilation equations when $|\Delta| \leq 1$ with $\Delta \neq 1$.

Assume therefore that $|\Delta| > 1$. Since $\int f(t)\,dt = \hat{f}(0) = 0$, it follows that if f is compactly supported then its primitive

$$f_1(t) \;=\; \int_{-\infty}^{t} f(s)\,ds$$

is a nontrivial integrable function with compact support. Moreover, f_1 is a solution of the dilation equation with coefficients $\{c_0/2, \ldots, c_N/2\}$, so we must either have $|\Delta/2| > 1$ or $\Delta/2 = 1$. If $|\Delta/2| > 1$ then we can repeat the process, forming the primitive f_2 of f_1, and conclude that either $|\Delta/4| > 1$ or $\Delta/4 = 1$. This process cannot continue indefinitely, i.e., we must have $\Delta = 2^n$ for some $n > 0$. Then f_n is an integrable solution to the dilation equation defined by the coefficients $\{2^{-n}c_0, \ldots, 2^{-n}c_N\}$ and hence (by our analysis for the case $\Delta = 1$) is the unique integrable solution of that dilation equation (up to scale) and has support contained in $[0, N]$. Thus f is the n^{th} derivative of f_n, so $\mathrm{supp}(f) \subset [0, N]$ as well, and we conclude that f is the unique compactly supported, integrable solution of its dilation equation (up to scale). On the other hand, if f is not compactly supported then its primitive f_1 need not be integrable, and the preceding analysis breaks down, i.e., we cannot prove uniqueness for integrable but non-compactly supported scaling functions when $|\Delta| > 1$. In fact, there exist dilation equations with $|\Delta| > 1$ which have infinitely many integrable solutions, and others which have none [DL1]. The above analysis does show that at most one integrable solution can be compactly supported.

The above characterization of compactly supported, integrable scaling functions was first proved by Daubechies and Lagarias [DL1] (some of our discussion was also inspired by Berger's excellent set of notes [B]). To summarize:

THEOREM 4.1.
If there exists an integrable, compactly supported solution f to the dilation equation determined by the coefficients $\{c_0, \dots, c_N\}$ then there exists an integer $n \geq 0$ such that the following statements hold (with uniqueness interpreted as holding up to scale).

- **(a)** $\sum c_k = 2^{n+1}$.
- **(b)** *The dilation equation determined by the coefficients $\{2^{-n}c_0, \dots, 2^{-n}c_N\}$ has a unique integrable solution f_n, and* $\mathrm{supp}(f_n) \subset [0, N]$.
- **(c)** *f is the unique compactly supported, integrable solution to its dilation equation, and, with the proper choice of scale, f is the n^{th} derivative of f_n.*

In particular, dilation equations satisfying $\sum c_k = 2$ (i.e., $\Delta = 1$) are in some sense "fundamental" (note that Eq. (4.4) is a special case). However, this condition is *not* sufficient to ensure the existence of an integrable solution! For example, the simplest dilation equation $N = 0$ with $c_0 = 2$ is easily seen to have no integrable solutions. We have given some necessary or sufficient conditions for the existence of integrable solutions [ColHe4], and Wang [W] has conjectured a necessary and sufficient condition in terms of a type of joint spectral radius (similar to the joint spectral radius discussed in later sections of this chapter). On the other hand, distributional solutions for dilation equations satisfying $\sum c_k = 2$ can always be constructed—in fact, infinitely many distinct distributional solutions exist, but there is exactly one compactly supported distributional solution (up to scale) [ColHe4], [DL1]. Some of these non-compactly supported solutions may be realizable as functions; in fact, it is possible for a dilation equation to have both smooth, compactly supported and smooth, non-compactly supported solutions.

We have used the Fourier transform in this section only to obtain uniqueness results, but many other results are possible. For example, Daubechies [D1] estimates the decay of $\hat{f}$ by considering the infinite product in Eq. (4.8); see also the expostion in Chapter 1 by Strichartz. Estimates for the smoothness of f can be obtained from this, including even Hölder exponents of continuity, although the results are usually not optimal since the Hölder exponent is essentially a time domain rather than a frequency domain property. On the other hand, Eirola [E] and Villemoes [V1], [V2] obtain good estimates for the Sobolev exponent of continuity by using Fourier transforms, since the Sobolev exponent is essentially a frequency domain property.

4.3 Equivalence of scaling functions and scaling vectors.

In the remainder of this chapter we use the dyadic interpolation method to characterize all smooth, compactly supported scaling functions. Our first step in this direction is to convert the dilation equation into an equivalent matrix formulation.

Assume, therefore, that coefficients $\{c_0, \dots, c_N\}$ are given, and that a continuous, compactly supported solution f to the dilation equation exists. Then we know that $\operatorname{supp}(f) \subset [0, N]$, so the vector-valued function

$$v(x) \;=\; \begin{pmatrix} f(x) \\ f(x+1) \\ \vdots \\ f(x+N-1) \end{pmatrix}, \qquad x \in [0,1], \tag{4.10}$$

captures all information about f. Note that $v_1(0) = v_N(1) = 0$ since f is continuous, and $v_{i+1}(0) = v_i(1)$ for $i = 1, \dots, N-1$ by construction, where $v_i(x)$ denotes the i^{th} component of $v(x)$.

Now, if $0 \le x \le 1/2$ then it follows from the dilation equation and the fact that $\operatorname{supp}(f) \subset [0, N]$ with $f(0) = f(N) = 0$ that the value of $v(x) = (f(x), \dots, f(x+N-1))^{\text{t}}$ is completely determined from the value of $v(2x) = (f(2x), \dots, f(2x+N-1))^{\text{t}}$ in a linear manner. Therefore, there exists an $N \times N$ matrix T_0 such that $v(x) = T_0 v(2x)$ when $0 \le x \le 1/2$. Similarly, if $1/2 \le x \le 1$ then $v(x)$ is determined from $v(2x-1)$ by a linear transformation, so $v(x) = T_1 v(2x-1)$ when $1/2 \le x \le 1$ for some matrix T_1. In fact, $(T_0)_{ij} = c_{2i-j-1}$ 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

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

Note that we have *consistency* at $x = 1/2$, i.e.,

$$v(1/2) \;=\; T_0 v(1) \;=\; T_1 v(0). \tag{4.11}$$

Thus, a continuous, compactly supported scaling function f determines a continuous, vector-valued function $v : [0,1] \to \mathbf{C}^N$ which satisfies

$$v_1(0) \;=\; v_N(1) \;=\; 0, \tag{4.12}$$

$$v_{i+1}(0) \;=\; v_i(1), \quad i = 1, \dots, N-1, \tag{4.13}$$

$$v(x) \;=\; T_{d_1} v(\tau x), \tag{4.14}$$

where $x = .d_1 d_2 \cdots$ is a binary expansion of x and

$$\tau x \;=\; 2x \bmod 1 \;=\; \begin{cases} 2x, & 0 \le x < 1/2, \\ 2x-1, & 1/2 < x \le 1, \end{cases}$$

(so $\tau x = .d_2 d_3 \cdots$), and where Eq. (4.14) is interpreted for $x = 1/2$ to mean Eq. (4.11). If f is Hölder continuous with Hölder exponent α then it follows easily that v is also Hölder continuous with the same exponent, where Hölder continuity for vector-valued functions is defined by the inequality $\|v(x)-v(y)\| \le K\,|x-y|^{\alpha}$, where $\|\cdot\|$ is an arbitrary norm on $\mathbf{C}^N$.

If f is now an arbitrary compactly supported scaling function (i.e., not necessarily continuous) and v is defined as above by Eq. (4.10), then Eqs. (4.11)–(4.14) will still hold, *provided that* $f(0) = f(N) = 0$. Note that this must be the case if c_0, $c_N \ne 1$. If $|c_0| \ge 1$ or $|c_N| \ge 1$ then it is not difficult to show directly from the dilation equation that f must be discontinuous at 0 or N, respectively.

In light of the above discussion, we will call any vector-valued function $v : [0,1] \to \mathbf{C}^N$ satisfying Eqs. (4.11)–(4.14) for all $x \in [0,1]$ a *scaling vector*.

The procedure described above is reversible. That is, assume that a scaling vector v is given, and define

$$f(x) \;=\; \begin{cases} 0, & x \le 0 \text{ or } x \ge N, \\ v_i(x), & i-1 \le x \le i,\ i = 1, \dots, N. \end{cases} \tag{4.15}$$

It follows immediately that f is a compactly supported solution to the dilation equation and is continuous if v is continuous, with the same Hölder exponent of continuity as v.

In summary, (Hölder) continuous, compactly supported scaling functions and (Hölder) continuous scaling vectors are completely equivalent, in the sense that one exists if and only if the other exists. Furthermore, as long as c_0, $c_N \ne 1$, a compactly supported scaling function exists if and only if a scaling vector exists. We therefore concentrate for the next several sections on the existence and properties of scaling vectors, rather than scaling functions.

4.4 Sufficient conditions for the existence of continuous scaling vectors.

In this section we find sufficient conditions on the coefficients $\{c_0, \dots, c_N\}$ which allow construction of a continuous scaling vector v. Note first from Eq. (4.12) that if a scaling vector v exists then $v_1(0) = 0$. Since $T_0 v(0) = v(0)$ and only the first entry of the top row of T_0 is nonzero we therefore have that $(v_2(0), \dots, v_N(0))^{\mathrm{t}}$ is a right eigenvector for M for the eigenvalue 1, 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.,

$$M = \begin{pmatrix} c_1 & c_0 & 0 & \cdots & 0 & 0 \\ c_3 & c_2 & c_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_N & c_{N-1} \end{pmatrix}. \tag{4.16}$$

Thus, a scaling vector can only exist if 1 is an element of the spectrum of M. This is the case, for example, when Eq. (4.4) is satisfied, for then $(1, \ldots, 1)$ must be a left eigenvector for M for the eigenvalue 1.

Assume, therefore, that 1 is an eigenvalue for M, and let us attempt to construct a scaling vector v. Let $a = (a_1, \ldots, a_{N-1})^{\mathrm{t}}$ be any right eigenvector for M for the eigenvalue 1, and define $v(0) = (0, a_1, \ldots, a_{N-1})^{\mathrm{t}}$ and $v(1) = (a_1, \ldots, a_{N-1}, 0)^{\mathrm{t}}$. In terms of the scaling function, this gives the values of f at the integers. Equations (4.11)–(4.13) are then satisfied. Next, for each dyadic point $x = .d_1 \cdots d_m$, define

$$v(x) = T_{d_1} \cdots T_{d_m} v(0). \tag{4.17}$$

In terms of the scaling function, this amounts to plugging the values of f at the integers into the dilation equation to obtain the values of f at the half-integers, then plugging these into the dilation equation again to get the values at the quarter-integers, and so on recursively to all dyadic points.

Equation (4.17) immediately implies that Eq. (4.14) is satisfied *for dyadic* $x \in [0,1]$, and we wish to extend the domain of v to *all* $x \in [0,1]$ so that Eq. (4.14) holds for all x in this interval. For this it suffices to prove that v is uniformly continuous on its current domain of definition (the set of dyadic points in $[0,1]$), for then there would exist a *unique* extension of v to all of $[0,1]$ so that v is continuous, and Eq. (4.14) would follow for arbitrary $x \in [0,1]$ upon taking limits. Moreover, if v satisfies a Hölder condition with exponent α for dyadic x, $y \in [0,1]$ then it will also satisfy a Hölder condition with the same exponent for arbitrary x, $y \in [0,1]$.

Consider, therefore, two dyadic points $x < y$ in $[0,1]$. If x and y are close enough then they will share the first few digits in their binary expansions, e.g., if $x = .d_1 \cdots d_k$ and $2^{-m-1} \le y - x < 2^{-m}$ with $m > k$ then $x = .d_1 \cdots d_m$ and $y = .d_1 \cdots d_m d_{m+1} \cdots d_{m+j}$ for some j (with $d_i = 0$ for $i = k+1, \ldots, m$). Therefore,

$$v(y) - v(x) = T_{d_1} \cdots T_{d_m} \big(v(\tau^m y) - v(0) \big). \tag{4.18}$$

Thus, to prove continuity from the right on the set of dyadic points in $[0,1]$ we must consider the behavior of all possible products $T_{d_1} \cdots T_{d_m}$ acting on all possible differences $v(z) - v(0)$ for dyadic $z \in [0,1]$. Define, therefore,

$$W = W(a) = \operatorname{span}\{v(z) - v(0) : \text{dyadic } z \in [0,1]\}, \tag{4.19}$$

and note that W is a subspace of $\mathbf{C}^N$ which is invariant under both T_0 and T_1. In general, W depends on the choice of a; however, for most choices of coefficients the eigenvalue 1 for M will be simple, in which case W is independent of a. By a slight abuse of terminology we therefore usually refer to W as if it was unique, and allow the dependence on a to be implicit.

If Eq. (4.4) holds then W is a subspace of

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

which is also invariant in this case under both T_0 and T_1. In general, it is not difficult to determine W explicitly. For example, we have the following [ColHe3].

PROPOSITION 4.2.
W is the smallest subspace of $\mathbf{C}^N$ which is invariant under both T_0 and T_1 and contains the vector $v(1) - v(0)$.

We have conjectured [ColHe3] that, for dilation equations satisfying Eq. (4.4), $W = V$ except for a set of coefficients $\{c_0, \dots, c_N\}$ of measure zero. However, the distinction between W and V can be critical for determining whether a dilation equation has a continuous solution or not; some examples are given in Section 4.8.2.

Return now to consideration of Eq. (4.18); we want to determine the "size" of $v(y) - v(x)$. Fix, therefore, any norm $\|\cdot\|$ on $\mathbf{C}^N$, with corresponding operator norm

$$\|A\| \;=\; \sup_{u \neq 0} \frac{\|Au\|}{\|u\|}$$

defined for $N \times N$ matrices A. Then from Eq. (4.18) we obtain

$$\|v(y) - v(x)\| \;\le\; C_1 \, \|(T_{d_1} \cdots T_{d_m})|_W\|, \tag{4.20}$$

where

$$C_1 \;=\; 2 \sup\{\|v(z)\| : \text{dyadic } z \in [0,1]\}$$

(of course, we do not know *a priori* that $C_1 < \infty$, but let us assume this for the moment). The right hand side of Eq. (4.20) behaves geometrically with respect to m, so let C, $\theta > 0$ be such that

$$\max_{d_j = 0,1} \|(T_{d_1} \cdots T_{d_m})|_W\| \;\le\; C\,\theta^m \qquad \text{for all } m > 0. \tag{4.21}$$

For example, $\theta = \max\{\|T_0|_W\|, \|T_1|_W\|\}$ is one possibility, but it is often possible to choose θ much smaller. Combining Eqs. (4.20) and (4.21) we have

$$\|v(y) - v(x)\| \;\le\; C_1\, C\, \theta^{-1}\, \theta^{m+1} \;=\; C_1\, C\, \theta^{-1}\, (2^{-m-1})^{\alpha} \;\le\; K\, |x-y|^{\alpha},$$

where $K = C_1\, C\, \theta^{-1}$ and $\alpha = -\log_2 \theta$. A symmetric argument establishes the same inequality for dyadic $y < x$ with $2^{-m-1} \le x - y < 2^{-m}$. As K and α are independent of m, we have therefore nearly proved the following result.

THEOREM 4.3.
Let C, $\theta > 0$ be such that Eq. (4.21) holds. If $\theta < 1$ then v determines a continuous, compactly supported scaling function f which is Hölder continuous with Hölder exponent $\alpha = -\log_2 \theta$.

A special case of Theorem 4.3, requiring the assumption of Eq. (4.4), was first proved by Daubechies and Lagarias [DL2].

The proof of Theorem 4.3 will be complete once we demonstrate that $C_1 < \infty$ whenever Eq. (4.21) holds with $\theta < 1$. This follows immediately from the calculation

$$\begin{aligned}
&\|T_{d_1} \cdots T_{d_m} v(0) - v(0)\| \\
&\quad \le \|T_{d_1} \cdots T_{d_m} v(0) - T_{d_1} \cdots T_{d_{m-1}} v(0)\| + \cdots + \|T_{d_1} v(0) - v(0)\| \\
&\quad \le \|(T_{d_1} \cdots T_{d_{m-1}})|_W\| \, \|T_{d_m} v(0) - v(0)\| + \cdots + \|T_{d_1} v(0) - v(0)\| \\
&\quad \le (C\,\theta^{m-1} + \cdots + C)\, C_2 \\
&\quad \le \frac{C\, C_2}{1-\theta},
\end{aligned}$$

where $C_2 = \max_{j=0,1} \|T_{d_j} v(0) - v(0)\|$.

4.5 The joint spectral radius.

The hypothesis in Theorem 4.3 that Eq. (4.21) holds with $\theta < 1$ is essentially a spectral constraint on the two operators $T_0|_W$, $T_1|_W$. We devote this section to showing how this constraint can be formulated in terms of a "joint spectral radius" of $T_0|_W$ and $T_1|_W$.

4.5.1 Definition and properties.

Recall that the usual spectral radius of a single matrix A is defined by

$$\rho(A) = \limsup_{m\to\infty} \|A^m\|^{1/m}. \tag{4.22}$$

This is independent of the choice of norm $\|\cdot\|$. The spectral radius can also be computed in an eigenvalue form, i.e.,

$$\rho(A) = \limsup_{m\to\infty} \sigma(A^m)^{1/m}, \tag{4.23}$$

where

$$\sigma(B) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } B\}.$$

Of course, Eq. (4.23) trivially implies

$$\rho(A) = \sigma(A) \tag{4.24}$$

since $\sigma(A^m) = \sigma(A)^m$, and in fact Eq. (4.24) is often used as the definition of the spectral radius.

The generalization of the usual spectral radius of a single matrix to the case of a set of elements of a normed algebra was first made by Rota and Strang [RoS]. For the specific case of two matrices, the definition is as follows.

DEFINITION 4.4. *The joint spectral radius $\hat{\rho}(A_0, A_1)$ of two $N \times N$ matrices A_0, A_1 is*

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

where

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

The definition of the joint spectral radius for larger collections of matrices, or for matrices restricted to subspaces (e.g., $\hat{\rho}(T_0|_W, T_1|_W)$), is made in the obvious way.

The value of $\hat{\rho}(A_0, A_1)$ is independent of the choice of norm. Berger and Wang [BW1] proved the nontrivial analogy of Eq. (4.23), i.e.,

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

where

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

However, the analogue of Eq. (4.24) fails in general, i.e., $\hat{\rho}(A_0, A_1)$ may not equal the maximum of the absolute values of the eigenvalues of A_0 and A_1 (which is given by $\hat{\sigma}_1$). For example, consider $A_0 = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ and $A_1 = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$; since $A_1 = A_0^{\mathrm{t}}$ we have $\hat{\rho}(A_0, A_1) = \hat{\sigma}_2 = \sqrt{(3+\sqrt{5})/2} > \hat{\sigma}_1 = 1$.

Given any $\theta > \hat{\rho}(A_0, A_1)$, it is easy to see that there must exist a constant $C > 0$ such that $\hat{\rho}_m^m \le C\,\theta^m$ for every m. However, this need not be true with $\theta = \hat{\rho}(A_0, A_1)$ (if it is then we say that $A_0/\hat{\rho}(A_0, A_1)$, $A_1/\hat{\rho}(A_0, A_1)$ are *product bounded*, since all products of these two normalized matrices are then bounded in norm). In terms of the joint spectral radius, Theorem 4.3 may therefore be restated as follows.

THEOREM 4.5.
If $\hat{\rho}(T_0|_W, T_1|_W) < 1$ then there exists a continuous scaling vector v which is Hölder continuous with Hölder exponent α for every $0 \le \alpha < -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$. If $\theta^{-1}\,T_0|_W$, $\theta^{-1}\,T_1|_W$ are product bounded with $\theta = \hat{\rho}(T_0|_W, T_1|_W)$ then v is also Hölder continuous with exponent $\alpha = -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$.

4.5.2 Computing a joint spectral radius.

Whereas the usual spectral radius of a single matrix is easily computed from the eigenvalues of that matrix, the joint spectral radius can be difficult to compute exactly. We therefore discuss in this subsection several methods of evaluating or approximating the joint spectral radius.

First, it is easy to see that the joint spectral radius is independent of the choice of basis, i.e., if B is an invertible matrix then $\hat{\rho}(BA_0B^{-1}, BA_1B^{-1}) = \hat{\rho}(A_0, A_1)$. Once a convenient basis is selected, the value of $\hat{\rho}(A_0, A_1)$ can be approximated to any desired degree of accuracy by the computation of the norms and eigenvalues of finitely many matrices. This follows from the next lemma and the fact that $\hat{\rho}(A_0, A_1) = \limsup \hat{\sigma}_m = \limsup \hat{\rho}_m$.

LEMMA 4.6.
$\hat{\sigma}_m \leq \hat{\rho}(A_0, A_1) \leq \hat{\rho}_m$ for every $m > 0$.

PROOF. Given any $d_1, \dots, d_m = 0, 1$ and any $n > 0$ we have

$$\rho(A_{d_1} \cdots A_{d_m})^{1/m} \;=\; \rho\big((A_{d_1} \cdots A_{d_m})^n\big)^{1/mn} \;\leq\; \hat{\sigma}_{mn},$$

whence

$$\hat{\sigma}_m \;=\; \max_{d_j=0,1} \rho(A_{d_1} \cdots A_{d_m})^{1/m} \;\leq\; \limsup_{n\to\infty} \hat{\sigma}_{mn} \;\leq\; \hat{\rho}(A_0, A_1).$$

For the second inequality, let

$$K \;=\; \max\{1, \hat{\rho}_m^{-1}, \dots, \hat{\rho}_m^{-m+1}\} \qquad \text{and} \qquad L \;=\; \max\{1, \hat{\rho}_1, \dots, \hat{\rho}_{m-1}^{m-1}\}.$$

Given any $n > 0$, write $n = mk+r$ with $0 \leq r < m$. Then for any $d_1, \dots, d_n = 0, 1$ we have

$$\|A_{d_1} \cdots A_{d_n}\| \;\leq\; (\hat{\rho}_m^m)^k\, \hat{\rho}_r^r \;=\; \hat{\rho}_m^n\, \hat{\rho}_m^{-r}\, \hat{\rho}_r^r \;\leq\; K\, L\, \hat{\rho}_m^n.$$

Taking the maximum over all $d_j = 0, 1$ we obtain $\hat{\rho}_n \leq (K\,L)^{1/n}\, \hat{\rho}_m$, so $\hat{\rho}(A_0, A_1) = \limsup_{n\to\infty} \hat{\rho}_n \leq \hat{\rho}_m$. $\square$

As a consequence of Lemma 4.6, we have $\sup \hat{\sigma}_m = \hat{\rho}(A_0, A_1) = \lim \hat{\rho}_m = \inf \hat{\rho}_m$. Note that $\sup \hat{\sigma}_m$ is a lower-semicontinuous function of the matrices A_0, A_1 since each $\hat{\sigma}_m$ is a continuous function of those matrices. Similarly, $\inf \hat{\rho}_m$ is upper-semicontinuous, so $\hat{\rho}(A_0, A_1)$ is a continuous function of the matrices A_0, A_1 [HeS].

Unfortunately, the number of matrix products involved in the computation of $\hat{\sigma}_m$ or $\hat{\rho}_m$ grows exponentially with m, and therefore it is usually impractical to achieve a good approximation to $\hat{\rho}(A_0, A_1)$ by simply computing $\hat{\sigma}_m$ and $\hat{\rho}_m$ for various m. On the other hand, if A_0, A_1 satisfy one of the hypotheses of the following lemma then $\hat{\rho}(A_0, A_1)$ can be computed exactly.

LEMMA 4.7.
If A_0, A_1 can be simultaneously upper-triangularized or simultaneously Hermitianized (i.e., there exists an invertible matrix B such that BA_0B^{-1} and BA_1B^{-1} are either both upper-triangular or both Hermitian, respectively) then $\hat{\rho}(A_0, A_1) = \hat{\sigma}_1 = \max\{\rho(A_0), \rho(A_1)\}$.

PROOF. The result for simultaneous upper-triangularization follows immediately from examination of the diagonal entries of the product of two upper-triangular matrices.

To prove the result for simultaneous Hermitianization, let $\|\cdot\|$ be any norm under which $\mathbf{C}^N$ is a Hilbert space, e.g., the Euclidean space norm, for then $\|A\| = \rho(A)$ for any Hermitian matrix A. Given $d_1, \dots, d_m$, we therefore have

$$\begin{aligned} \|A_{d_1} \cdots A_{d_m}\| &= \|B^{-1}(BA_{d_1}B^{-1}) \cdots (BA_{d_m}B^{-1})B\| \\ &\le \|B\| \, \|B^{-1}\| \, \|BA_{d_1}B^{-1}\| \cdots \|BA_{d_m}B^{-1}\| \\ &= \|B\| \, \|B^{-1}\| \, \rho(BA_{d_1}B^{-1}) \cdots \rho(BA_{d_m}B^{-1}) \\ &= \|B\| \, \|B^{-1}\| \, \rho(A_{d_1}) \cdots \rho(A_{d_m}) \\ &\le \|B\| \, \|B^{-1}\| \, \hat{\sigma}_1^m, \end{aligned}$$

since each $BA_{d_j}B^{-1}$ is Hermitian. Thus $\hat{\rho}_m \le \|B\|^{1/m} \, \|B^{-1}\|^{1/m} \, \hat{\sigma}_1$, so $\hat{\rho}(A_0, A_1) = \limsup \hat{\rho}_m \le \hat{\sigma}_1 \le \hat{\rho}(A_0, A_1)$. $\square$

Note that $A_0 = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$, $A_1 = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$ can be simultaneously upper-triangularized but not simultaneously Hermitianized, while $A_0 = \begin{pmatrix} 3 & 0 \\ 3 & -2 \end{pmatrix}$, $A_1 = \begin{pmatrix} -2 & 3 \\ 0 & 3 \end{pmatrix}$ can be simultaneously Hermitianized but not simultaneously upper-triangularized.

The following lemma, a generalization of the first part of Lemma 4.7, can be useful for reducing the computational complexity of the calculation of $\hat{\rho}(A_0, A_1)$ when A_0, A_1 are large. Its proof follows easily using block matrix multiplication.

LEMMA 4.8.
Assume A_0, A_1 can be simultaneously block upper-triangularized, i.e., there exists an invertible matrix B such that

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

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

The above tools are not sufficient to exactly evaluate the joint spectral radius of all pairs of matrices, even when we consider only pairs of 2×2 matrices. For example, one pair which will be of interest in Section 4.8.1 is

$$S_0 = \frac{1}{5}\begin{pmatrix} 3 & 0 \\ 1 & 3 \end{pmatrix} \qquad \text{and} \qquad S_1 = \frac{1}{5}\begin{pmatrix} 3 & -3 \\ 0 & -1 \end{pmatrix}. \tag{4.25}$$

This pair is neither simultaneously upper-triangularizable nor simultaneously Hermitianizable. By direct brute-force calculation of $2^{31}-2$ matrix products we find that

$$\min\{\hat\rho_1,\dots,\hat\rho_{30}\} \;=\; \hat\rho_{30} \;=\; \|S_0^5 S_1 S_0^{11} S_1 S_0^{12}\|^{1/30} \;\approx\; 0.671271 \tag{4.26}$$

and

$$\max\{\hat\sigma_1,\dots,\hat\sigma_{30}\} \;=\; \hat\sigma_{13} \;=\; \rho(S_1 S_0^{12})^{1/13} \;\approx\; 0.659679,$$

where we have chosen the norm $\|u\| = |u_1| + |u_2|$ on $\mathbf{C}^2$ to evaluate $\hat\rho_m$. This enormous calculation does not even specify $\hat\rho(S_0,S_1)$ to an accuracy of two decimal places!

Fortunately, there does exist a recursive algorithm which often will substantially reduce the computational complexity of approximating $\hat\rho(A_0,A_1)$ from above (unfortunately, we know of no similar method for reducing the computational burden of approximating from below). The algorithm is based on the following straightforward generalization of Lemma 4.6 [DL2].

LEMMA 4.9.
Assume that $\{P_j\}$ is a set of building blocks of products of the matrices A_0, A_1. That is,

(a) *each P_j is a product of m_j of the matrices A_0, A_1, and*
(b) *there exists an $r \ge 0$ such that if P is any product of A_0, A_1 then $P = P_{j_1}\cdots P_{j_k} Q$ with Q a product of length at most r.*

Then $\hat\rho(A_0,A_1) \le \max \|P_j\|^{1/m_j}$.

Given an initial "guess" θ for the joint spectral radius, the following algorithm [ColHe1] will construct a set of building blocks $\{P_j\}$ such that $\hat\rho(A_0,A_1) \le \max\|P_j\|^{1/m_j} \le \theta$ (if $\theta < \hat\rho(A_0,A_1)$ then the algorithm will not terminate).

Algorithm 4.10.
Fix $\theta > \hat\rho(A_0,A_1)$. For each of $P = A_0$ and $P = A_1$ in turn, implement the following recursion.

> Let $P = A_{d_1}\cdots A_{d_m}$. If $\|P\|^{1/m} \le \theta$ then keep P as a building block. Otherwise, repeat this step replacing $P = A_{d_1}\cdots A_{d_m}$ by each of $P = A_{d_1}\cdots A_{d_m}A_0$ and $P = A_{d_1}\cdots A_{d_m}A_1$ in turn.

Algorithm 4.10 must terminate if the guess θ exceeds $\hat{\rho}(A_0, A_1)$ since there must be some m such that $\hat{\rho}(A_0, A_1) \le \hat{\rho}_m \le \theta$.

Returning now to the specific case of the matrices S_0, S_1 given in Eq. (4.25), we used Algorithm 4.10 to improve the estimate of $\hat{\rho}(S_0, S_1)$. In particular, Algorithm 4.10 did terminate after an initial guess of $\theta = 0.660025$ was input, so we conclude that $\hat{\rho}(S_0, S_1) \le 0.660025$, a significant improvement over Eq. (4.26). Algorithm 4.10 output a set of 306473 building blocks after computing 612944 matrix products (including those not selected as building blocks). The longest product selected as a building block was of length 139. A direct calculation of $\hat{\rho}_{139}$ (even if it were possible!) could not improve this estimate since, for example,

$$\begin{aligned} &\|S_0^3 S_1 S_0^{12} S_1 S_0^{13} S_1 S_0^{12} S_1 S_0^{11} S_1 S_0^{9} S_1 S_0^{12} S_1 S_0^{10} S_1 S_0^{13} S_1 S_0^{12} S_1 S_0^{11} S_1 S_0^{9} S_1\|^{1/139} \\ &\qquad = \ 0.661276, \end{aligned}$$

whence $\hat{\rho}_{139} \ge 0.661276$. Based on the above numerical evidence, we conjecture that $\hat{\rho}(S_0, S_1) = \rho(S_1 S_0^{12})^{1/13}$.

4.6 Necessary conditions for the existence of continuous scaling vectors.

We devote this section to proving that Theorems 4.3 and 4.5 are sharp, i.e., that we have the following converse [ColHe3].

THEOREM 4.11.
If v is any continuous scaling vector and if W is defined by Eq. (4.19) *then $\hat{\rho}(T_0|_W, T_1|_W) < 1$. Moreover, no Hölder exponent of continuity α for v can exceed $-\log_2 \hat{\rho}(T_0|_W, T_1|_W)$, and can equal $-\log_2 \hat{\rho}(T_0|_W, T_1|_W)$ if and only if $\theta^{-1} T_0|_W$, $\theta^{-1} T_1|_W$ are product bounded with $\theta = \hat{\rho}(T_0|_W, T_1|_W)$.*

To begin the proof, let v be an arbitrary continuous scaling vector; v is then Hölder continuous for some Hölder exponent α in the range $0 \le \alpha \le 1$. Let K be any corresponding Hölder constant.

Choose now any fixed product $T = T_{d_1} \cdots T_{d_m}$ of T_0, T_1, and let $x \in [0, 1]$ be an arbitrary dyadic point (we will impose restrictions on x momentarily). Let $x = .e_1 e_2 \cdots$ be any binary expansion of x. Then by Eq. (4.14) and the continuity of v,

$$T^k v(x) \ = \ v(x_k) \ \to \ v(d) \qquad \text{as } k \to \infty, \tag{4.27}$$

where

$$\begin{aligned} x_1 &= .d_1 \cdots d_m e_1 e_2 \cdots , \\ x_2 &= .d_1 \cdots d_m d_1 \cdots d_m e_1 e_2 \cdots , \\ &\vdots \end{aligned}$$

and $d \in [0,1]$ is the rational (but not necessarily dyadic) point

$$d = .d_1 \cdots d_m d_1 \cdots d_m \cdots .$$

Intuitively, Eq. (4.27) should impose some constraint both on the spectral radius of T and on the Hölder exponent α. To make this precise, let $y_k \in [0,1]$ be the dyadic points

$$\begin{aligned} y_1 &= .d_1 \cdots d_m, \\ y_2 &= .d_1 \cdots d_m d_1 \cdots d_m, \\ &\vdots \end{aligned}$$

We will use the y_k in place of d since d is not necessarily dyadic. Note that

$$v(x_k) - v(y_k) \;=\; T^k\big(v(x) - v(0)\big) \tag{4.28}$$

and

$$|x_k - y_k| \;=\; 2^{-mk}\,|x|. \tag{4.29}$$

If we define W as in Eq. (4.19) then certainly $v(x) - v(0) \in W$. So, let λ be that eigenvalue of $T|_W$ such that $|\lambda| = \rho(T|_W)$. If $v(x) - v(0)$ "has a component" in the λ-eigenspace of $T|_W$, then Eq. (4.28) should imply that $\|v(x_k) - v(y_k)\|$ is on the order of $|\lambda|^k$; combining this with Eq. (4.29) then yields an estimate of the Hölder exponent. In general, the meaning of "component" can be precisely formulated in terms of the Jordan decomposition of W induced by $T|_W$; reference [ColHe3] contains an explicit description of how to do this. In any case, since W is spanned by vectors of the form $v(x) - v(0)$ for dyadic x, there must be *some* dyadic x such that $v(x) - v(0)$ has a component in the λ-eigenspace of $T|_W$. Although we will omit the details, it is then straightforward to prove that there exists a constant $C > 0$ such that

$$\|T^k\big(v(x) - v(0)\big)\| \;\geq\; C\,|\lambda|^k \tag{4.30}$$

for all k [ColHe3]. In particular, since $v(x_k) - v(y_k) \to 0$ as $k \to \infty$, we must have $\rho(T|_W) = |\lambda| < 1$. Combining Eqs. (4.28)–(4.30), we find that

$$\|v(x_k) - v(y_k)\| \;\geq\; K_T\,|x_k - y_k|^{\alpha_T}, \tag{4.31}$$

where $\alpha_T = -\log_2 \rho(T)^{1/m}$ and $K_T = C\,|x|^{-\alpha_T}$. However, v is Hölder continuous with exponent α and constant K, so

$$\|v(x_k) - v(y_k)\| \;\leq\; K\,|x_k - y_k|^{\alpha}. \tag{4.32}$$

Since $|x_k - y_k| \to 0$ as $k \to \infty$, Eqs. (4.31) and (4.32) imply that $\alpha \le \alpha_T$. Taking now the supremum over all $T = T_{d_1} \cdots T_{d_m}$ for $d_j = 0, 1$ we obtain $\hat{\sigma}_m < 1$ and $\alpha \le -\log_2 \hat{\sigma}_m$. Thus $\hat{\rho}(T_0|_W, T_1|_W) = \sup \hat{\sigma}_m \le 1$ and $\alpha \le -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$, i.e., we have proved the statement that the Hölder exponent of v cannot exceed $-\log_2 \hat{\rho}(T_0|_W, T_1|_W)$.

Our analysis has also nearly established that $\hat{\rho}(T_0|_W, T_1|_W) < 1$; in particular, we have shown that $\hat{\rho}(T_0|_W, T_1|_W) = \sup \hat{\sigma}_m \le 1$ with $\hat{\sigma}_m < 1$ for every m. Thus, only the special case

$$\hat{\rho}(T_0|_W, T_1|_W) \;=\; \sup \hat{\sigma}_m \;=\; 1 \qquad \text{and} \qquad \hat{\sigma}_m < 1 \;\text{ for every } m \tag{4.33}$$

remains. It is interesting to note that we do not know if Eq. (4.33) is possible. More generally, we know of no matrices A_0, A_1 for which we can demonstrate that the supremum $\sup \hat{\sigma}_m(A_0, A_1) = \hat{\rho}(A_0, A_1)$ is not achieved for some m. Lagarias and Wang [LW] have conjectured that no such matrices can exist, i.e., it is always the case that $\hat{\rho}(A_0, A_1) = \rho(A_{d_1} \cdots A_{d_m})^{1/m}$ for some finite matrix product $A_{d_1} \cdots A_{d_m}$.

We continue now with the proof of Theorem 4.11; in particular, we show now that the continuity of v implies $\hat{\rho}(T_0|_W, T_1|_W) < 1$ (here we use an argument due to Micchelli and Prautzsch [MiP2] see [ColHe3] for another argument). Let $\{v(x_i) - v(0)\}_{i=1}^J$ be a basis for W, and define a norm on W by $\|w\| = \sum |a_i|$ for $w = \sum a_i \big(v(x_i) - v(0)\big)$. Choose any $\varepsilon < 1$; since v is uniformly continuous there exists a $\delta > 0$ such that $\|v(x) - v(y)\| \le \varepsilon$ whenever $|x - y| < \delta$. Let m be such that $2^{-m} < \delta$, and choose any product $T = T_{d_1} \cdots T_{d_m}$ of length m. Writing the binary expansion of x_i as $x_i = .e_1^i e_2^i \cdots$, define $X_i = .d_1 \cdots d_m e_1^i e_2^i \cdots$ and $Y = .d_1 \cdots d_m$. Then $|X_i - Y| \le 2^{-m}$, so for an arbitrary element $w = \sum a_i \big(v(x_i) - v(0)\big)$ of W we have

$$\begin{aligned}
\|T_{d_1} \cdots T_{d_m} w\| \;&\le\; \sum_{i=1}^{J} |a_i| \; \|T_{d_1} \cdots T_{d_m} \big(v(x_i) - v(0)\big)\| \\
&=\; \sum_{i=1}^{J} |a_i| \; \|v(X_i) - v(Y)\| \\
&\le\; \varepsilon \sum_{i=1}^{J} |a_i| \\
&=\; \varepsilon \, \|w\|.
\end{aligned}$$

Thus $\|T|_W\| \le \varepsilon$, and therefore by taking the supremum over all such matrix products of length m we obtain $\hat{\rho}(T_0|_W, T_1|_W) \le \hat{\rho}_m \le \varepsilon^{1/m} < 1$.

The proof of the final statement of the theorem, that $\alpha = -\log_2 \hat{\rho}(T_0|_W, T_1|_W)$ implies product boundedness, follows easily by using a basis for W as above, and will be omitted [ColHe3].

4.7 Differentiability.

Theorems 4.3, 4.5, and 4.11 characterize those dilation equations which have continuous, compactly supported solutions f. However, if f is a continuously differentiable, compactly supported scaling function for the dilation equation determined by the coefficients $\{c_0, \dots, c_N\}$ then f' is a continuous, compactly supported scaling function for the dilation equation determined by the coefficients $\{2c_0, \dots, 2c_N\}$. Thus, Theorems 4.3, 4.5, and 4.11 implicitly characterize those dilation equations having continuously differentiable, compactly supported solutions. We make this implicit characterization, and its extension to higher derivatives, explicit in this section.

4.7.1 Necessary and sufficient conditions.

To begin, assume that f is an n-times continuously differentiable, compactly supported scaling function for a dilation equation determined by the coefficients $\{c_0, \dots, c_N\}$. In light of Theorem 4.1, it suffices to consider coefficients satisfying the fundamental constraint $\sum c_k = 2$. Let the associated matrices M, T_0, T_1, scaling vector v, and subspace W be constructed as usual. Now, the j^{th} derivative $f^{(j)}$ of f is a continuous, compactly supported scaling function for the dilation equation determined by the coefficients $\{2^j c_0, \dots, 2^j c_N\}$ for each $j = 0, \dots, n$, so let the matrices M^j, T_0^j, T_1^j, scaling vector v^j, and subspace W^j denote the obvious objects constructed with respect to $f^{(j)}$. Clearly, $M^j = 2^j M$, $T_0^j = 2^j T_0$, and $T_1^j = 2^j T_1$; however, the scaling vectors v^j and the subspaces W^j are quite distinct from v and W. In particular, since $1 \in \text{spectrum}(M^j)$ we have $2^{-j} \in \text{spectrum}(M)$ for $j = 0, \dots, n$; this simply amounts to the realization that $(f^{(j)}(1), \dots, f^{(j)}(N-1))^{\text{t}}$ is a right eigenvector for M for the eigenvalue 2^{-j}. Note that this implies $n < N - 1$, and therefore f cannot be infinitely differentiable! From Theorem 4.11 we must have

$$\hat{\rho}(T_0^j|_{W^j}, T_1^j|_{W^j}) \;=\; 2^j\, \hat{\rho}(T_0|_{W^j}, T_1|_{W^j}) \;<\; 1 \qquad \text{for each } j = 0, \dots, n.$$

This supplies a necessary condition for the existence of an n-times differentiable scaling vector.

To make this more explicit, suppose we are given coefficients $\{c_0, \dots, c_N\}$ such that $\sum c_k = 2$ and $1, 2^{-1}, \dots, 2^{-n} \in \text{spectrum}(M)$, and let $a = (a_1, \dots, a_{N-1})^{\text{t}}$ be a right eigenvector for M for the eigenvalue 2^{-n}. Define

$$v^n(0) \;=\; (0, a_1, \dots, a_{N-1})^{\text{t}} \qquad \text{and} \qquad v^n(1) \;=\; (a_1, \dots, a_{N-1}, 0)^{\text{t}}.$$

For dyadic $x = .d_1 \cdots d_m \in [0,1]$ define

$$v^n(x) \;=\; 2^{mn}\, T_{d_1} \cdots T_{d_m} v^n(0),$$

and set

$$W^n \;=\; W^n(a) \;=\; \text{span}\{v^n(x) - v^n(0) : \text{dyadic } x \in [0,1]\}.$$

Then we have the following theorem.

THEOREM 4.12.
Let coefficients $\{c_0, \dots, c_N\}$ *be given such that* $\sum c_k = 2$ *and* $1, 2^{-1}, \dots, 2^{-n} \in \operatorname{spectrum}(M)$. *Then there exists an n-times continuously differentiable, compactly supported scaling function f for the dilation equation determined by the coefficients* $\{c_0, \dots, c_N\}$ *if and only if*

$$\hat{\rho}(T_0|_{W^n}, T_1|_{W^n}) \; < \; 2^{-n} \tag{4.34}$$

for some right eigenvector a for M for the eigenvalue 2^{-n}, *where* $W^n = W^n(a)$ *is defined as above. In this case the* n^{th} *derivative* $f^{(n)}$ *of f is Hölder continuous for each exponent*

$$0 \;\leq\; \alpha \;<\; \alpha_{\max} \;=\; -\log_2\bigl(2^n\, \hat{\rho}(T_0|_{W^n}, T_1|_{W^n})\bigr)$$

but not for any $\alpha > \alpha_{\max}$, *and* $\alpha = \alpha_{\max}$ *is allowed if and only if* $\theta^{-1}\, T_0|_{W^n}$, $\theta^{-1}\, T_1|_{W^n}$ *are product bounded with* $\theta = \hat{\rho}(T_0|_{W^n}, T_1|_{W^n})$.

The sufficient condition in Theorem 4.12 is proved as follows. Assume $\hat{\rho}(T_0|_{W^n}, T_1|_{W^n}) < 2^{-n}$. Then v^n extends to a continuous scaling vector on $[0, 1]$ by Theorem 4.5, and therefore determines a continuous, compactly supported scaling function h for the dilation equation determined by the coefficients $\{2^n c_0, \dots, 2^n c_N\}$. Theorem 4.5 also implies that h is Hölder continuous for at least each $0 \leq \alpha < \alpha_{\max}$, and Theorem 4.11 limits α to at most $\alpha_{\max}$. Together, Theorems 4.5 and 4.11 imply $\alpha = \alpha_{\max}$ is allowed if and only if the stated product boundedness condition holds. Finally, Theorem 4.1 implies that h is the n^{th} derivative of an integrable, compactly supported scaling function f for the dilation equation determined by the coefficients $\{c_0, \dots, c_N\}$.

4.7.2 Imposition of sum rules.

Theorem 4.12 characterizes dilation equations having differentiable solutions without explicit restrictions on the coefficients $\{c_0, \dots, c_N\}$. However, scaling functions useful in applications usually arise from coefficients satisfying a certain number of "equal-opportunity" restrictions on the even- and odd-indexed coefficients; Eq. (4.4) is one example of such a restriction. In particular, in addition to the fundamental constraint $\sum c_k = 2$, it is generally assumed that the coefficients satisfy the following $n + 1$ *sum rules*:

$$\sum_k (-1)^k\, k^j\, c_k \;=\; 0 \qquad \text{for } j = 0, \dots, n \tag{4.35}$$

(note that Eq. (4.35) is just Eq. (4.4) if $n = 0$). These sum rules impose certain conditions on the scaling function f and associated wavelet g which are related to, but distinct from, smoothness. For example, the $n + 1$ sum rules imply that the wavelet g will have $n+1$ vanishing moments. For $j=0$ this follows from the calculation

$$\int g(t)\,dt \;=\; \sum_k (-1)^k\, c_{N-k} \int f(2t-k)\,dt \;=\; 0,$$

and an easy induction then establishes that $\int t^j\, g(t)\,dt = 0$ for $j = 1, \dots, n$. We also have the following.

PROPOSITION 4.13.
Assume the coefficients $\{c_0, \dots, c_N\}$ satisfy $\sum c_k = 2$ and $n+1$ sum rules, and that an associated compactly supported scaling function f exists. Then $\sum k^j\, f(t-k)$ is a polynomial of degree j for each $j = 0, \dots, n$.

In particular, the polynomials $1, \dots, t^n$ can all be reproduced exactly from translates of f. Thus, an arbitrary smooth function can be approximated with error $\mathcal{O}(2^{-mn})$ by combinations of $\{f(2^m t - k)\}_{k \in \mathbf{Z}}$ for each scale 2^{-m} [S]

Proposition 4.13 can be proved by various methods; we sketch one version, restricted for clarity to the case $n = 1$. Assume f is normalized so that $\hat{f}(0) = 1$, and let θ_0 be any function such that

(a) $\hat{\theta}_0(0) = 1$,
(b) $\sum \theta_0(t-k) = 1$, and
(c) $\sum k\, \theta_0(t-k) = t + b$ for some constant b.

For example, this is true if θ_0 is the solution of a dilation equation satisfying two sum rules and such that θ_0 is orthogonal to its integer translates (admittedly, this is circular reasoning, but it is not difficult to prove Proposition 4.13 directly with this extra assumption of orthogonality). An example of such a θ_0 is the Daubechies scaling function D_4 discussed in Section 4.8.1. Now apply the Cascade Algorithm, i.e., define the functions θ_i for $i > 0$ by

$$\theta_i(t) \;=\; \sum_k c_k\, \theta_{i-1}(2t-k). \tag{4.36}$$

Then $\hat{\theta}_i(\gamma) \to \hat{f}(\gamma) = \prod_{j=1}^{\infty} m_0(\gamma/2^j)$ uniformly on compact sets since $\hat{\theta}_0(0) = 1$ and $\hat{\theta}_i(\gamma) = m_0(\gamma/2)\, \hat{\theta}_{i-1}(\gamma/2)$. In the time domain this implies $\theta_i \to f$ weakly in $L^2(\mathbf{R})$. Now, it follows from Eq. (4.36) and the two sum rules that

$$\sum_k \theta_i(t-k) \;=\; \sum_k \theta_{i-1}(2t-k) \tag{4.37}$$

and

$$\sum_k k\, \theta_i(t-k) \;=\; \frac{1}{2} \sum_k k\, \theta_{i-1}(2t-k) \;-\; \frac{A_1}{2} \sum_k \theta_{i-1}(2t-k), \tag{4.38}$$

where

$$A_j = \sum (2k)^j \, c_{2k} = \sum (2k+1)^j \, c_{2k+1}.$$

Applying Eqs. (4.37) and (4.38) recursively and taking the limit as $i \to \infty$ we obtain $\sum f(t-k) = 1$ a.e. and $\sum k\, f(t-k) = t - A_1$ a.e., completing the proof.

As a corollary of Proposition 4.13 we have that if f is n-times differentiable and $n+1$ sum rules are satisfied then $\sum k^j \, f^{(n)}(t-k)$ is identically constant for each $j = 0, \ldots, n$. In the notation of Theorem 4.12, this implies

$$\sum_{k=1}^{N} k^j \left(v_k^n(x) - v_k^n(0) \right) \;=\; 0 \qquad \text{for } j = 0, \ldots, n,$$

where $v_k^n(x)$ is the k^{th} component of $v^n(x)$. Thus

$$W^n \;\subset\; V^n \;=\; \{u \in \mathbf{C}^N : e_j \cdot u = 0 \text{ for } j = 0, \ldots, n\},$$

where $e_j = (1^j, 2^j, \ldots, N^j)$. Moreover, as Daubechies and Lagarias proved [DL2] the assumption of the sum rules implies the following facts about V^n.

LEMMA 4.14.
Assume the coefficients $\{c_0, \ldots, c_N\}$ satisfy $\sum c_k = 2$ and $n+1$ sum rules. Then the following statements hold for each $j = 0, \ldots, n$.

- ***(a)*** $U^j = \operatorname{span}\{e_0, \ldots, e_j\}$ *is left-invariant under both T_0 and T_1 and* $\operatorname{spectrum}(T_0|_{U^j}) = \operatorname{spectrum}(T_1|_{U^j}) = \{1, 2^{-1}, \ldots, 2^{-j}\}$.
- ***(b)*** *V^j is the orthogonal complement of U^j in $\mathbf{C}^N$ and is right-invariant under both T_0 and T_1.*

As a consequence, $1, 2^{-1}, \ldots, 2^{-n} \in \operatorname{spectrum}(M)$.

PROOF. Note that (b) follows immediately from (a) and the definition of V^j. To prove (a), it is sufficient to show that for each $i = 0$, 1 and for $j = 0, \ldots, n$ there exist constants a_{jl}^i such that

$$e_j T_i \;=\; 2^{-j} \, e_j \;+\; \sum_{l=0}^{j-1} a_{jl}^i \, e_l.$$

This is certainly true for $j = 0$ since $e_0 T_0 = e_0 T_1 = e_0$. For simplicity, we prove now only $i = 0$, $j = 1$; the general case is similar.

Since $(T_0)_{pq} = c_{2p-q-1}$, we have

$$(e_1 T_0)_q \;=\; \sum_{p=1}^{N} p \, c_{2p-q-1}.$$

If q is even, say $q = 2r$, then

$$\begin{aligned}(e_1T_0)_{2r} &= \sum_p p\, c_{2p-2r-1} \\ &= \sum_p (r+p+1)\, c_{2p+1} \\ &= \frac{2r+1}{2}\sum_p c_{2p+1} + \frac{1}{2}\sum_p (2p+1)\, c_{2p+1} \\ &= \frac{1}{2}(2r+1+A_1) \\ &= \frac{1}{2}(e_1)_{2r} + \frac{A_1+1}{2}(e_0)_{2r}.\end{aligned}$$

A similar calculation applies if q is odd, so we conclude that

$$e_1T_0 = \frac{1}{2}e_1 + \frac{A_1+1}{2}e_0,$$

as desired.

It follows from (a) that $1, 2^{-1}, \dots, 2^{-n}$ are eigenvalues of both T_0 and T_1, and it remiains only to show that these are also eigenvalues of M. However, given j we can find $u = (u_1, \dots, u_N) \in U^j$ which is a left eigenvector for T_0 for the eigenvalue 2^{-j}. Since u is a linear combination of $e_0, \dots, e_j$ we cannot have $u_2 = \cdots = u_N = 0$. Since $T_0 = \begin{pmatrix} c_0 & 0 \\ * & M \end{pmatrix}$, we conclude that $(u_2, \dots, u_N)$ is a left eigenvector for M for the eigenvalue 2^{-j}. $\square$

As a consequence of Lemma 4.14 and Theorem 4.12, we find that

$$\hat{\rho}(T_0|_{V^n}, T_1|_{V^n}) < 2^{-n} \tag{4.39}$$

is sufficient to ensure the existence of an n-times continuously differentiable, compactly supported scaling function (although it need not be necessary). Since V^n is independent of the coefficients, this is often an easier condition to test than Eq. (4.34).

Note that Eq. (4.39) uses $n+1$ sum rules to conclude n-times differentiability. In practice, however, it is often the case that several more sum rules will be satisfied than the number of derivatives possessed by the scaling function. For example, the Daubechies scaling function D_4 satisfies two sum rules, yet is only continuous and not differentiable. Even so, we can use these "extra" sum rules to obtain extra simplification of the joint spectral radius calculations. This is made precise in the following proposition, which is a slightly weaker version of a result first proved by Daubechies and Lagarias [DL2].

PROPOSITION 4.15.
Assume the coefficients $\{c_0, \dots, c_N\}$ satisfy $\sum c_k = 2$ and $n+1$ sum rules. If

$$\hat{\rho}(T_0|_{V^n}, T_1|_{V^n}) < 2^{-l}$$

for some $0 \le l \le n$*, then an* l*-times continuously differentiable, compactly supported scaling function exists, and its* l^{th} *derivative is Hölder continuous for each exponent*

$$0 \;\le\; \alpha \;<\; \min\bigl\{1,\; -\log_2\bigl(2^l\,\hat\rho(T_0|_{V^n}, T_1|_{V^n})\bigr)\bigr\}.$$

PROOF*. Recall from Lemma 4.14 that $\{e_0,\dots,e_j\}$ is a basis for U^j for each $j = 0,\dots,n$. Since $U^0 \subset \cdots \subset U^n$ and each V^j is the orthogonal complement of U^j in $\mathbf{C}^N$, we have $V^0 \supset \cdots \supset V^n$. Moreover, $\dim(V^n) = N-n-1$, so if $\{\tilde e_{n+1},\dots,\tilde e_{N-1}\}$ is a basis for V^n then $\{e_{j+1},\dots,e_n,\tilde e_{n+1},\dots,\tilde e_{N-1}\}$ is a basis for V^j for each $j = 0,\dots,n$. By considering these complementary bases for U^j and V^j, we can easily construct a change-of-basis matrix B such that

$$BT_iB^{-1} \;=\; \begin{pmatrix} 1 & & & * & \\ & 2^{-1} & & & \\ & & \ddots & & \\ & & & 2^{-n} & \\ & 0 & & & C_i \end{pmatrix}, \qquad i = 0,1,$$

and conclude then from Lemma 4.8 that

$$\hat\rho(T_0|_{V^j}, T_1|_{V^j}) \;=\; \begin{cases} \max\{2^{-j-1},\,\dots,\,2^{-n},\,\hat\rho(C_0,C_1)\}, & j = 0,\dots,n-1,\\ \hat\rho(C_0,C_1), & j = n.\end{cases}$$

Thus,

$$\hat\rho(T_0|_{V^l}, T_1|_{V^l}) \;=\; \max\{2^{-l-1},\; \hat\rho(T_0|_{V^n}, T_1|_{V^n})\},$$

so the hypothesis $\hat\rho(T_0|_{V^n}, T_1|_{V^n}) < 2^{-l}$ implies $\hat\rho(T_0|_{V^l}, T_1|_{V^l}) < 2^{-l}$. The result therefore follows from Theorem 4.12. □

In particular, note that the dimension of V^n can be significantly smaller than the dimension of V^l, resulting in a reduction in the size of the matrices involved in the computation of the joint spectral radius. Because W^l may be strictly smaller than V^l, the upper bound for the Hölder exponent of $f^{(l)}$ in Proposition 4.15 need not be sharp, as is the case for the upper bound in Theorem 4.12.

4.8 Examples.

For simplicity, we will restrict our attention to real-valued coefficients $\{c_0,\dots,c_N\}$. By the uniqueness results of Theorem 4.1, the corresponding scaling functions must then also be real-valued.

*Author's note: This proof is incorrect. A corrected proof can be found in the Errata after the end of this paper.

4.8.1 Four-coefficient dilation equations.

There are no continuous, compactly supported solutions of dilation equations when $N = 0$ or 1. When $N = 2$ the assumptions $\sum c_k = 2$ and $1 \in \text{spectrum}(M)$ imply that Eq. (4.4) holds, whence $W \subset V$ is at most one-dimensional and the computation of $\hat{\rho}(T_0|_W, T_1|_W)$ is trivial. We therefore move directly to the first nontrivial case, $N = 3$. Here we have four coefficients $\{c_0, c_1, c_2, c_3\}$; assume again that they satisfy $\sum c_k = 2$ and $1 \in \text{spectrum}(M)$. It follows then that either

- ***(a)*** $c_0 + c_1 = c_2 + c_3 = 1$, or
- ***(b)*** $c_0 + c_2 = c_1 + c_3 = 1$.

Note that case (b) is simply Eq. (4.4).

In case (a), it is easy to show that $\dim(W) = 3$ if $c_0 \neq c_3$, whence $\hat{\rho}(T_0|_W, T_1|_W) \geq \rho(T_0|_W) = \rho(T_0) \geq 1$. Thus there are no continuous solutions to case (a) when $c_0 \neq c_3$. Since case (a) with $c_0 = c_3$ is a special case of (b), we turn immediately to that case.

Assume, therefore, that Eq. (4.4) holds, and note that this class of dilation equations is a two-parameter family. We select the independent parameters to be c_0 and c_3, and therefore identify this family with the (c_0, c_3)-plane, i.e., each point (c_0, c_3) determines a dilation equation with coefficients $\{c_0, 1-c_3, 1-c_0, c_3\}$ and vice versa. Since Eq. (4.4) implies $W \subset V$ and $\dim(V) = 2$, it follows that either $W = V$ or W is one-dimensional.

The (c_0, c_3)-plane is shown in Figure 4.1 along with several geometrical objects. It has been shown [ColHe4] that integrable scaling functions exist at least for all points in the shaded region, and cannot exist outside the ellipse. We show below that continuous scaling functions are restricted to a subregion of the interior of the triangle, and that differentiable scaling functions occur precisely on the solid portion of the dashed line. The circle $(c_0 - 1/2)^2 + (c_3 - 1/2)^2 = 1/2$ consists of those points which satisfy both Eqs. (4.4) and (4.5), which is a necessary condition for the existence of corresponding multiresolution analyses. In fact, every point on the circle with the single exception of $(c_0, c_3) = (1, 1)$ does determine a multiresolution analysis and hence a wavelet orthonormal basis for $L^2(\mathbf{R})$.

Now, a right eigenvector for M for the eigenvalue 1 is $(c_0, c_3)^{\mathrm{t}}$; therefore $v(0) = (0, c_0, c_3)^{\mathrm{t}}$ and $v(1) = (c_0, c_3, 0)^{\mathrm{t}}$, up to normalization. Computing $v(1/2) = T_1 v(0)$, we find that $v(0)$, $v(1/2)$, and $v(1)$ are not co-linear, and therefore $W = V$, if $1 - c_0 - c_3 \neq 0$. When $1 - c_0 - c_3 = 0$ it is not difficult to show that $\dim(W) = 1$, and therefore $\hat{\rho}(T_0|_W, T_1|_W)$ can be trivially computed in this case. However, $1 - c_0 - c_3 = 0$ is still interesting and we discuss it further below.

When $1 - c_0 - c_3 \neq 0$ we have $\dim(W) = 2$. Using the change-of-basis matrix

$$B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

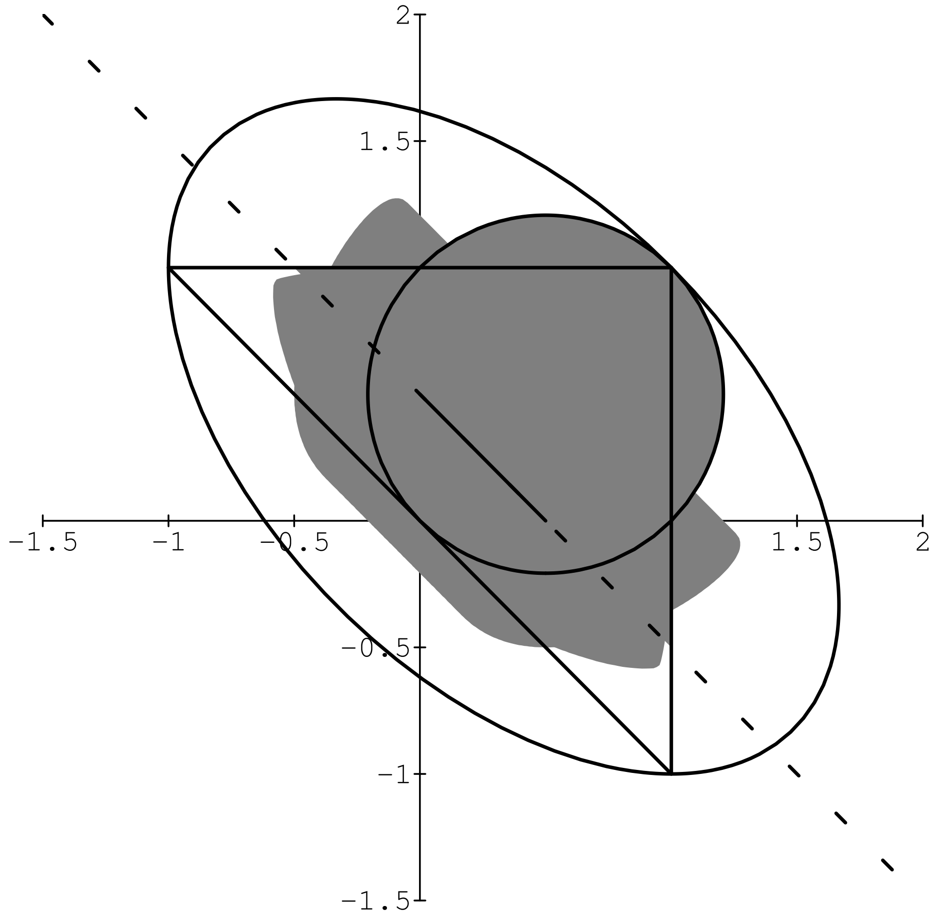

FIGURE 4.1
The (c_0, c_3)-plane, identified with real-valued, four-coefficient dilation equations satisfying Eq. (4.4).

we find that

$$\hat{\rho}(T_0|_W, T_1|_W) \;=\; \hat{\rho}\big((BT_0B^{-1})|_{BW}, (BT_1B^{-1})|_{BW}\big) \;=\; \hat{\rho}(S_0, S_1),$$

where

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

Therefore, a continuous, compactly supported solution to the dilation equation exists if and only if $\hat{\rho}(S_0, S_1) < 1$. The "easy" cases for computing $\hat{\rho}(S_0, S_1)$ are given in the following lemma, whose proof follows from tedious, but elementary, algebra [ColHe2].

LEMMA 4.16.

(a) *S_0, S_1 can be simultaneously upper-triangularized if and only if $1-c_0-c_3 = 0$ or $1-2c_0-2c_3 = 0$.*

(b) *S_0, S_1 can be simultaneously Hermitianized if and only if* $c_0\, c_3\,(1 - c_0 - c_3)\,(1 - 2c_0 - 2c_3) > 0$.

In either of these cases we have $\hat{\rho}(S_0, S_1) = \hat{\sigma}_1 = \max\{|c_0|, |c_3|, |1 - c_0 - c_3|\}$.

Only "half" of the (c_0, c_3)-plane satisfies one of the hypotheses of Lemma 4.16. Those points (c_0, c_3) satisfying one of those hypotheses and resulting in $\hat{\rho}(S_0, S_1) = \hat{\sigma}_1 < 1$ are shown in the shaded region of Figure 4.2. For the remaining points, $\hat{\rho}(S_0, S_1)$ must be approximated by some other method, e.g., by computing $\hat{\sigma}_m$, $\hat{\rho}_m$ or through the use of Algorithm 4.10 (see Section 4.5.2). The triangle shown in Figures 4.1–4.3 is the set $\{(c_0, c_3) : \hat{\sigma}_1 = 1\}$; continuous scaling functions are therefore restricted at least to points in the interior of the triangle. Combining Lemma 4.16 with brute-force calculations of $\hat{\sigma}_{18}$ and $\hat{\rho}_{18}$, we show in Figure 4.3 the following regions.

(a) The shaded region in Figure 4.3 is the set $\{(c_0, c_3) : \hat{\rho}_{18} < 1\}$ combined with the shaded region from Figure 4.2.

(b) The solid curve in Figure 4.3 is the set $\{(c_0, c_3) : \hat{\sigma}_{18} = 1\}$.

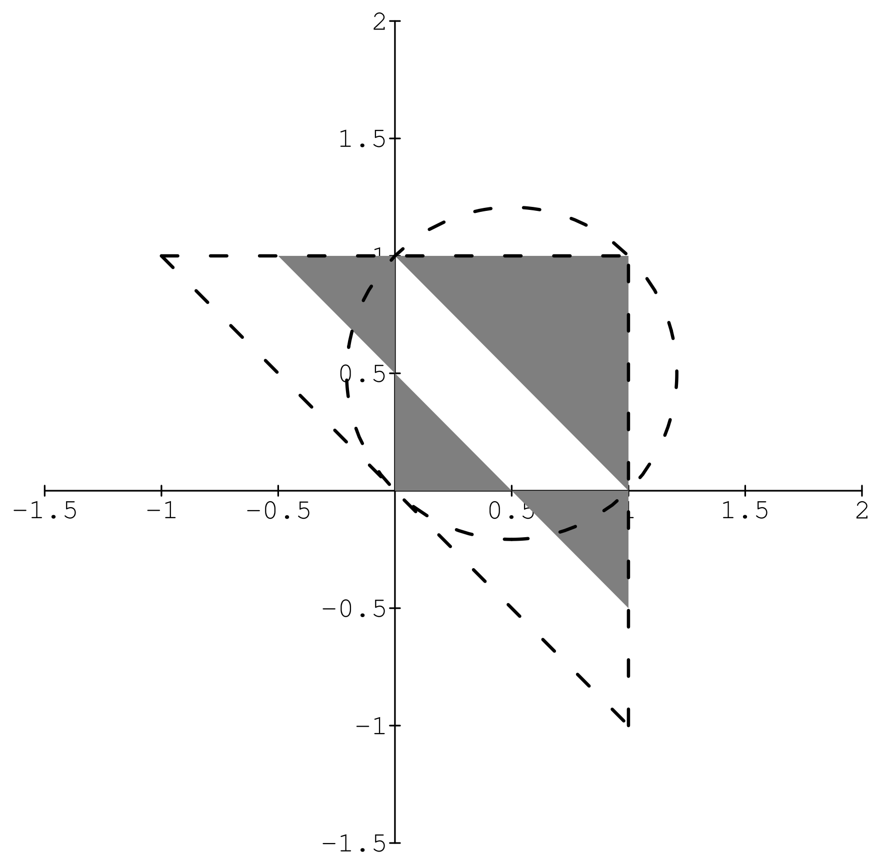

FIGURE 4.2
The subset of the (c_0, c_3)-plane where simultaneous Hermitianization occurs (shaded area).

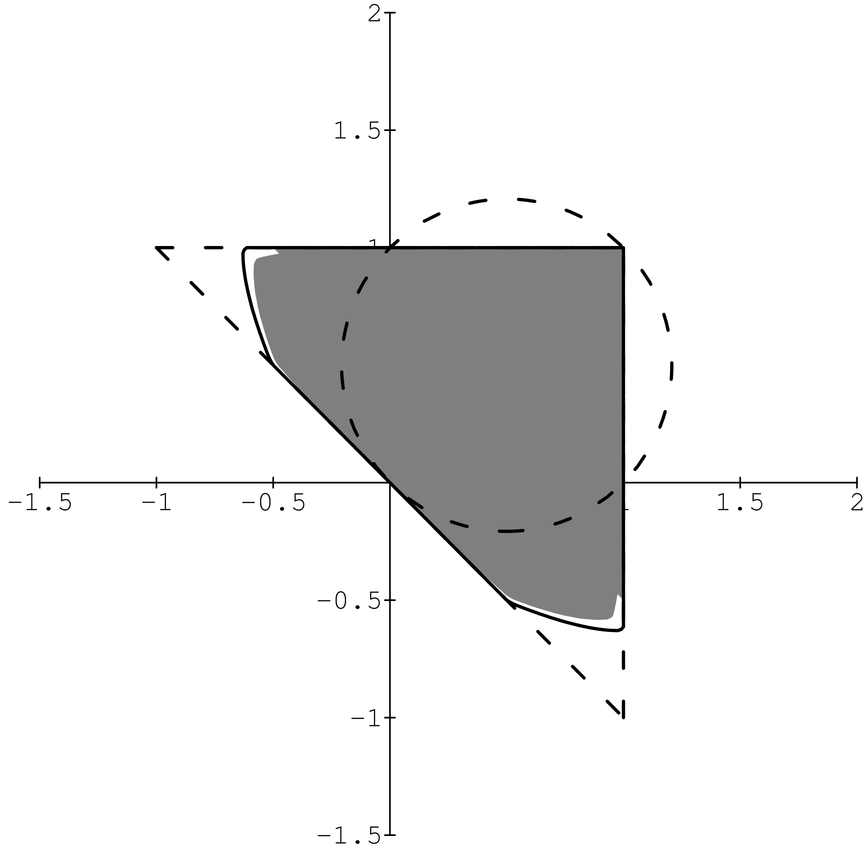

FIGURE 4.3
Numerical approximation of the region in the (c_0, c_3)-plane where continuous, compactly supported scaling functions exist (shaded area).

In particular, continuous, compactly supported scaling functions exist for all (c_0, c_3) lying within the shaded region, and cannot exist for any (c_0, c_3) lying on or outside the solid curve.

We consider now whether any of these scaling functions are differentiable. This requires that both 1 and 1/2 be eigenvalues of M, which implies $1 - c_0 - c_3 = 0$. In this case two sum rules are satisfied, so W^1 equals the one-dimensional space V^1. Therefore the scaling function f is differentiable if and only if $\hat{\rho}(T_0|_{W^1}, T_1|_{W^1}) = \max\{|c_0|, |c_3|\} < 1/2$, i.e., on the solid portion of the dashed line shown in Figure 4.1. No scaling function in the (c_0, c_3)-plane can be twice differentiable since this would require that 1, 1/2, and 1/4 all be eigenvalues of the 2×2 matrix M.

Finally, consider those dilation equations important for constructing multiresolution analyses, i.e., those lying on the circle shown in Figures 4.1–4.3. Only "half" of these have continuous solutions, and none can be differentiable. The "standard" example, the Daubechies scaling function D_4, is that unique point on the circle which satisfies two sum rules (that is, unique up to symmetry about the line $c_0 = c_3$, which corresponds to reversing the order of the coefficients in the dilation equation and time-reversing the corresponding scaling function). This point lies at the intersection of the circle and the dashed line $1 - c_0 - c_3 = 0$. The associated scaling function (shown in Figure 4.4) therefore has the highest regularity on the circle, in the sense of being able to reproduce exactly the greatest number of polynomials. However, it is not the smoothest scaling function on the circle in the sense of greatest Hölder exponent. In particular, D_4 satisfies the hypotheses of Lemma 4.16(a), so its maximum Hölder exponent is $-\log_2 \hat{\rho}(S_0, S_1) = -\log_2 \hat{\sigma}_1 \approx 0.550$ (since D_4 satisfies two sum rules, this can also be obtained more simply by applying Proposition 4.15 with $l = 0$, $n = 1$). On the circle, $-\log_2 \hat{\sigma}_1$ reaches its maximum value of $-\log_2 0.6 \approx 0.737$ at $(c_0, c_3) = (3/5, -1/5)$. Unfortunately, this point does not satisfy either of the hypotheses of Lemma 4.16, and this is not the maximum Hölder exponent of the corresponding scaling function (shown in Figure 4.5). This is, in fact, the example we studied in depth in Section 4.5.2, cf. Eq. (4.25). We determined there after extensive calculation that $0.659679 \le \hat{\rho}(S_0, S_1) \le 0.660025$, and therefore the maximum Hölder exponent of continuity of this scaling function is approximately $-\log_2 0.660 \approx 0.600$. We conjecture that this is the largest Hölder exponent occuring for points on the circle.

4.8.2 Symmetric seven-coefficient dilation equations.

In general, the relationship between the smoothness of scaling functions and the coefficients of the dilation equation is much more complicated than is apparent from the discussion of four-coefficient dilation equations presented in the preceding subsection. This is a result of the fact that when $N = 3$ we have $\hat{\rho}(T_0|_W, T_1|_W) = \hat{\rho}(T_0|_V, T_1|_V)$ even when $W \neq V$. As V is independent of the co-

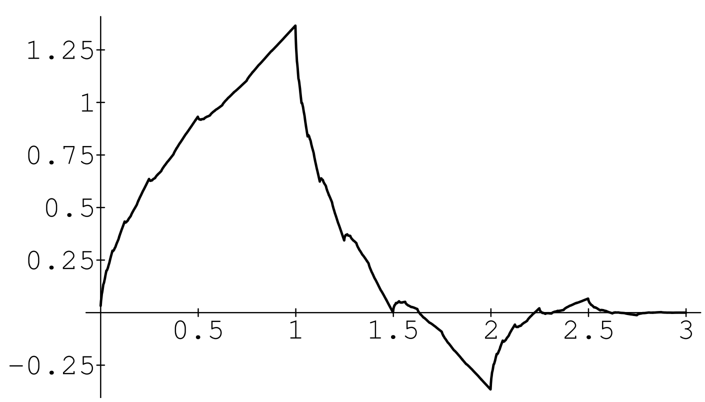

FIGURE 4.4
Daubechies scaling function D_4, corresponding to $(c_0, c_3) = ((1+\sqrt{3})/4, (1-\sqrt{3})/4)$.

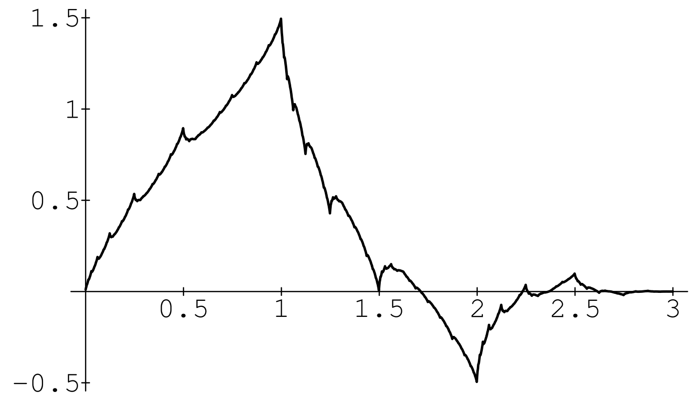

FIGURE 4.5
Scaling function corresponding to $(c_0, c_3) = (3/5, -1/5)$.

efficients, $\hat{\rho}(T_0|_W, T_1|_W)$ and the maximum Hölder exponent $-\log_2 \hat{\rho}(T_0|_W, T_1|_W)$ therefore become continuous as functions of the coefficients.

A markedly different situation can occur when $N > 3$; in particular, the space W and its dimensionality can change dramatically as the coefficients of the dilation equation vary. Since the joint spectral radius is a spectral condition and since each eigenvalue of each product of T_0, T_1 is a continuous function of the coefficients, what becomes important for the determination of $\hat{\rho}(T_0|_W, T_1|_W)$ is which eigenvalues of a product $T_{d_1} \cdots T_{d_m}$ are relevant when the product is restricted to W. In particular, when the dimension of W decreases certain eigenvalues can be rendered irrelevant, thereby changing the joint spectral radius discontinuously. The net effect is that $\hat{\rho}(T_0|_W, T_1|_W)$, and therefore also the maximum Hölder exponent, are "unstable" as functions of the coefficients. We illustrate some these ideas using symmetric seven-coefficient dilation equations.

Set $N = 6$ and suppose that the coefficients $\{c_0, c_1, c_2, c_3, c_4, c_5, c_6\}$ are real, satisfy Eq. (4.4), i.e.,

$$c_0 + c_2 + c_4 + c_6 \;=\; c_1 + c_3 + c_5 \;=\; 1,$$

and are symmetric, i.e.,

$$c_0 \;=\; c_6, \qquad c_1 \;=\; c_5, \qquad c_2 \;=\; c_4.$$

Together, these conditions reduce the number of free parameters to two, which we take as c_0 and c_1. We identify this class of dilation equation with the (c_0, c_1)-plane, i.e., each point (c_0, c_1) determines a dilation equation with coefficients $\{c_0, c_1, 1/2 - c_0, 1 - 2c_1, 1/2 - c_0, c_1, c_0\}$, and conversely. This (c_0, c_1)-plane is shown in Figure 4.6 along with several geometrical objects which we will discuss.

Since we have assumed that the coefficients to the dilation equation are symmetric, any associated scaling function will also be symmetric, i.e., $f(x) = f(6 - x)$.

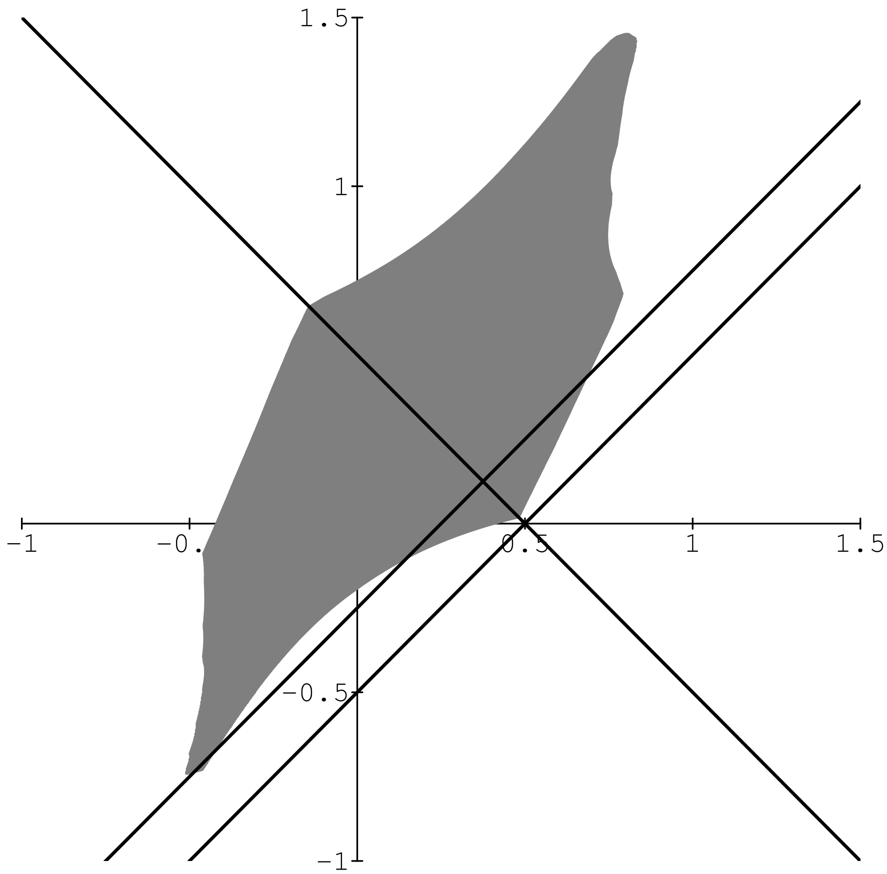

FIGURE 4.6
The (c_0, c_1)-plane, identified with symmetric, real-valued, seven-coefficient dilation equations satisfying Eq. (4.4).

Daubechies [D1] has shown that, for real-valued coefficients, symmetry precludes the possibility of an associated multiresolution analysis, with the single exception of the Haar system. Thus these scaling functions cannot be used to construct wavelet orthonormal bases for $L^2(\mathbf{R})$. However, symmetric functions can be used to construct non-orthogonal wavelet bases [CoDF].

The shaded region in Figure 4.6 is a numerical approximation to the set of points in the (c_0, c_1)-plane for which $\hat{\rho}(T_0|_V, T_1|_V) < 1$; in particular, the shaded region consists of those points (c_0, c_1) such that $\hat{\rho}_{14}(T_0|_V, T_1|_V) < 1$, using the norm $\|u\| = |u_1| + \cdots + |u_6|$. As $\hat{\rho}(T_0|_W, T_1|_W) \le \hat{\rho}(T_0|_V, T_1|_V)$, each point in this region has an associated continuous, compactly supported scaling function.

Consider now the point $(c_0, c_1) = (1/2, 0)$; the coefficients of the associated dilation equation are

$$c_0 = 1/2, \quad c_1 = 0, \quad c_2 = 0, \quad c_3 = 1, \quad c_4 = 0, \quad c_5 = 0, \quad c_6 = 1/2. \tag{4.40}$$

Examination of the matrix M given by Eq. (4.16) shows that 1 is an eigenvalue with multiplicity two. Two linearly independent right eigenvectors are, for example, $(1, 2, 3, 2, 1)^{\mathrm{t}}$ and $(0, 0, 1, 0, 0)^{\mathrm{t}}$. As outlined in Section 4.4, any right eigenvector for M for the eigenvalue 1 is a possible candidate for defining a scaling vector; however, by uniqueness only one can possibly lead to an integrable solution. For example, setting $v(0) = (0, 0, 0, 1, 0, 0)^{\mathrm{t}}$ we find that $W = V$, and therefore $\hat{\rho}(T_0|_W, T_1|_W) \geq 1$ since 1 is a multiple eigenvalue of the 6×6 matrices T_0 and T_1 and $\dim(V) = 5$. Therefore, by Theorem 4.11 this cannot lead to a continuous scaling vector. On the other hand, if we choose instead the vector $(0, 1, 2, 3, 2, 1)^{\mathrm{t}}$ as the candidate for $v(0)$, we find that $v(x) = (x, x+1, x+2, 3-x, 2-x, 1-x)^{\mathrm{t}}$ for $x \in [0, 1]$. Thus W is one-dimensional, from which one can determine that $\hat{\rho}(T_0|_W, T_1|_W) = 1/2$; by comparison, $\hat{\rho}(T_0|_V, T_1|_V) = 1$, and therefore it is not sufficient to examine only V to determine the existence of a continuous scaling vector. This scaling vector is therefore Hölder continuous with exponent 1; note this is obvious since the graph of v is a line segment. The associated scaling function is therefore also Hölder continuous with exponent 1; however, it is not differentiable. The scaling function is simply the triangle function with height 1 supported on the interval $[0, 6]$.

Note that this seven-coefficient dilation equation is simply a "stretched" version of a three-coefficient dilation equation, i.e., we obtain the coefficients in Eq. (4.40) by inserting two zeros between each of the coefficients of the $N = 2$ equation defined by $c_0 = 1/2$, $c_1 = 1$, $c_2 = 1/2$. This procedure effectively stretches the support of the solution of the original three-coefficient equation in the time domain by a factor of three. In particular, the solution to the three-coefficient equation is the symmetric triangle function with support $[0, 2]$ and height 1, so that our continuous solution to Eq. (4.40) is stretched out over the interval $[0, 6]$. Stretched dilation equations have been studied in some detail by Berger and Wang [BW2].

We show now that the stretched dilation equation corresponding to $(1/2, 0)$ lies on the boundary of the region where $\hat{\rho}(T_0|_W, T_1|_W) < 1$; this is due precisely to the fact that $\dim(W)$ changes as one moves away from this point in the (c_0, c_1)-plane. To see this, let us first examine the line ℓ_1 consisting of points of the form $(c_0, c_1) = (1/2 + \delta, -\delta)$ for $\delta \in \mathbf{R}$. This line is plotted in Figure 4.6; note that $\delta = 0$ corresponds to the stretched dilation equation $(1/2, 0)$ and therefore $\dim(W) = 1$ when $\delta = 0$. However, for $\delta \neq 0$ we can use Proposition 4.2 to show that $\dim(W) = 3$. Furthermore, in this case T_0 and T_1 can be simultaneously block upper-triangularized to the form

$$BT_iB^{-1} \;=\; \begin{pmatrix} C_i^1 & * & * & * \\ 0 & 1/2 & * & * \\ 0 & 0 & C_i^2 & * \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where each C_i^k is a 2×2 matrix, where the upper left 3×3 submatrix of BT_iB^{-1} corresponds to $T_i|_W$, and where the upper left 5×5 submatrix of BT_iB^{-1} corresponds to $T_i|_V$. Except for a small range of δ which will not affect us, each pair C_0^k, C_1^k can be simultaneously Hermitianized; in particular, this simultaneous Hermitianization can be carried out for each δ in the interval $[-1/8, 1/8]$. Applying Lemmas 4.7 and 4.8 we find that

$$\hat{\rho}(T_0|_W, T_1|_W) \;=\; \max\left\{\hat{\rho}(C_0^1, C_1^1),\ \tfrac{1}{2}\right\} \;=\; \max\left\{\tfrac{1}{2},\ |\tfrac{1}{2}+\delta|,\ |\tfrac{1}{2}+2\delta|\right\} \qquad (4.41)$$

and

$$\hat{\rho}(T_0|_V, T_1|_V) = \max\left\{\hat{\rho}(T_0|_W, T_1|_W),\ \hat{\rho}(C_0^2, C_1^2)\right\} = \max\left\{\hat{\rho}(T_0|_W, T_1|_W),\ |1+3\delta|\right\}. \qquad (4.42)$$

In particular, note that $\hat{\rho}(T_0|_W, T_1|_W) = 1/2 < \hat{\rho}(T_0|_V, T_1|_V) < 1$ for $\delta \in [-1/8, 0)$ while $1/2 < \hat{\rho}(T_0|_W, T_1|_W) < 1 < \hat{\rho}(T_0|_V, T_1|_V)$ for $\delta \in (0, 1/8]$. Thus, although the points in the segment $[-1/8, 1/8]$ lie completely within the region $\hat{\rho}(T_0|_W, T_1|_W) < 1$ and have W as a proper subset of V, only half of the segment lies in the region $\hat{\rho}(T_0|_V, T_1|_V) < 1$. Note that the only stretched dilation equation along this segment occurs at the point $\delta = 0$.

Consider next the line ℓ_2 consisting of points of the form $(c_0, c_1) = (1/2+\delta, \delta)$ for $\delta \in \mathbf{R}$. This line is plotted in Figure 4.6; we again have that $\delta = 0$ corresponds to the stretched dilation equation $(1/2, 0)$. Along this line, it can be shown [ColHe3] that $W = V$ whenever $\delta \neq 0$. Moreover, $\rho(T_0|_W) > 1$ for all $\delta \neq 0$ sufficiently close to zero. Therefore, there is an $\varepsilon > 0$ such that $\hat{\rho}(T_0|_W, T_1|_W) \geq \rho(T_0|_W) > 1$ for $0 < |\delta| < \varepsilon$, and so no dilation equation for these δ can have a continuous, compactly supported solution. Thus, although the point $(c_0, c_1) = (1/2, 0)$ has a continuous, compactly supported solution, an arbitrarily small change in the coefficients can result in a dilation equation without a continuous solution. In particular, the region where $\hat{\rho}(T_0|_W, T_1|_W) < 1$ is not an open set.

Now consider the line ℓ_3 consisting of those points of the form $(c_0, c_1) = (3/8+\delta, 1/8+\delta)$; this line is again plotted in Figure 4.6. The point $(3/8, 1/8)$, corresponding to $\delta = 0$, lies on the line ℓ_1, and so from Eqs. (4.41) and (4.42) has $\hat{\rho}(T_0|_W, T_1|_W) = 1/2$ and $\hat{\rho}(T_0|_V, T_1|_V) = 5/8$. Again it can be shown [ColHe3] that $W = V$ whenever $\delta \neq 0$. Since $\hat{\rho}(T_0|_V, T_1|_V)$ is a continuous function of the coefficients, there is an interval of points on ℓ_3 near $(3/8, 1/8)$ which have continuous, compactly supported solutions. For $\delta \neq 0$ in this interval, the maximum Hölder exponent of continuity is $-\log_2 \hat{\rho}(T_0|_V, T_1|_V)$. As δ converges to zero, this maximum Hölder exponent converges to the value of $-\log_2 \hat{\rho}(T_0|_V, T_1|_V)$ corresponding to the point $\delta = 0$, i.e., to $-\log_2 5/8$. However, the maximum Hölder exponent at the point $\delta = 0$ is $-\log_2 \hat{\rho}(T_0|_W, T_1|_W) = -\log_2 1/2 = 1$. Hence the maximum Hölder exponent is not continuous as a function of the coefficients even within the region $\hat{\rho}(T_0|_V, T_1|_V) < 1$.

As a final example, consider the point $(-1/16, 0)$, which was studied in detail in an early paper by Dubuc [Du] This is the point in the (c_0, c_1)-plane satisfying the greatest number of sum rules (four), and will serve to illustrate the benefit of extra sum rules. In particular, since $\dim(V^3) = 2$ we can find 2×2 matrices C_0, C_1 such that $\hat{\rho}(T_0|_{V^3}, T_1|_{V^3}) = \hat{\rho}(C_0, C_1)$. Moreover, C_0, C_1 can be simultaneously Hermitianized, with the result that $\hat{\rho}(C_0, C_1) = 1/4$. It therefore follows from

Proposition 4.15 with $l = 1$, $n = 3$ that f is differentiable and f' is Hölder continuous for each exponent $0 \le \alpha < 1$. We could prove this same result without using the two "extra" sum rules by appealing to Theorem 4.12 with $n = 1$, but this would require consideration of 4×4, rather than 2×2, matrices, since $W^1 = V^1$ is four-dimensional in this case. However, this extra complexity is required if we want to determine whether f' is Hölder continuous with exponent $\alpha = 1$. In particular, after some calculation we find $\hat{\rho}(T_0|_{V^1}, T_1|_{V^1}) = 1/4$, and therefore Theorem 4.12 implies that $\alpha = 1$ is allowed if and only if $4\,T_0|_{V^1}$, $4\,T_1|_{V^1}$ are product bounded. However, the eigenvalues of $T_0|_{V^1}$ are $1/4$, $1/8$, and $-1/16$, with $1/4$ a degenerate eigenvalue of multiplicity two. Therefore, by an appropriate change of basis, $T_0|_{V^1}$ can be placed in Jordan canonical form

$$\begin{pmatrix} 1/4 & 1 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/8 & 0 \\ 0 & 0 & 0 & -1/16 \end{pmatrix},$$

whence $\|(4\,T_0|_{V^1})^k\| \to \infty$ as $k \to \infty$ for any choice of norm $\|\cdot\|$. Thus $4\,T_0|_{V^1}$, $4\,T_1|_{V^1}$ are not product bounded, so f' cannot be Hölder continuous for the exponent $\alpha = 1$ (in particular, f is not twice differentiable). This fact cannot be obtained by examining the smaller matrices $T_0|_{V^3}$, $T_1|_{V^3}$ since these ignore the degeneracy of the eigenvalue $1/4$. In fact, $4\,T_0|_{V^3}$, $4\,T_1|_{V^3}$ are product bounded since they can be simultaneously Hermitianized. These results are the most we can obtain from the theorems we have presented. However, Dubuc [Du] proved directly that f' is "almost Lipschitz" in the sense that there exists a constant $C > 0$ such that

$$|f'(x) - f'(y)| \;\le\; C\,|\log|x-y||\,|x-y|.$$

Daubechies and Lagarias [DL2] proved that the same result follows from the product boundedness of the small matrices $4\,T_0|_{V^3}$, $4\,T_1|_{V^3}$. More generally, they proved that if the hypotheses of Proposition 4.15 are fulfilled and if $\hat{\rho}(T_0|_{V^n}, T_1|_{V^n}) = 2^{-l-1}$ with $\theta^{-1}T_0|_{V^n}$, $\theta^{-1}T_1|_{V^n}$ product bounded for $\theta = \hat{\rho}(T_0|_{V^n}, T_1|_{V^n})$, then $f^{(l)}$ is almost Lipschitz. The results of Daubechies and Lagarias [DL2] on the local properties of functions can be used to refine even further the properties satisfied by this scaling function f.

Bibliography.

[B] M.A. Berger, *Lectures on wavelets*, preprint.

[BW1] M.A. Berger and Y. Wang, *Bounded semi-groups of matrices*, Lin. Alg. Appl. **166** (1992), 21–27.

[BW2] M.A. Berger and Y. Wang, *Multi-scale dilation equations and iterated function systems*, Constr. Approx. (to appear).

[CDaMi] A. Cavaretta, W. Dahmen, and C.A. Micchelli, *Stationary Subdivision*, Mem. Amer. Math. Soc. **93** (1991), 1–186.

[CoDF] A. Cohen, I. Daubechies, and J.-C. Feauveau, *Biorthogonal bases of compactly supported wavelets*, Commun. Pure Appl. Math. **65** (1992), 485–560.

[ColHe1] 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), 876–881.

[ColHe2] D. Colella and C. Heil, *On the characterization of continuous scaling functions*, The MITRE Corporation, McLean, Va., Technical Report WP-91W00476 (1992).

[ColHe3] D. Colella and C. Heil, *Characterizations of scaling functions: Continuous solutions*, SIAM J. Matrix Anal. Appl. (to appear).

[ColHe4] D. Colella and C. Heil, *Characterizations of scaling functions, II. Distributional and functional solutions*, (in preparation).

[D1] I. Daubechies, *Orthonormal bases of compactly supported wavelets*, Commun. Pure Appl. Math. **41** (1988), 909–996.

[D2] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Reg. Conf. Ser. Appl. Math., vol. 61, Soc. Ind. Appl. Math., Philadelphia, 1992.

[DL1] I. Daubechies and J. Lagarias, *Two-scale difference equations: I. Existence and global regularity of solutions*, SIAM J. Math. Anal. **22** (1991), 1388–1410.

[DL2] I. Daubechies and J. Lagarias, *Two-scale difference equations: II. Local regularity, infinite products of matrices and fractals*, SIAM J. Math. Anal. **23** (1992), 1031–1079.

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

[Du] S. Dubuc, *Interpolation through an iterative scheme*, J. Math. Anal. Appl. **114** (1986), 185–205.

[DyLe] N. Dyn and D. Levin, *Interpolating subdivision schemes for the generation of curves and surfaces*, International Series of Numerical Mathematics, vol. 94, Birkhäuser, Boston, 1990, pp. 91–106.

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

[E] T. Eirola, *Sobolev characterization of solutions of dilation equations*, SIAM J. Math. Anal. **23** (1992), 1015–1030.

[Ha] A. Haar, *Zur Theorie der orthogonalen Funktionensysteme*, Math. Ann. **69** (1910), 331–371.

[He] C. Heil, *Methods of solving dilation equations*, Proc. of the 1991 NATO ASI on Prob. and Stoch. Methods in Anal. with Appl. (J.S. Byrnes, ed., NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci.,, eds.), Kluwer Academic Publishers, Dordrecht (to appear).

[HeS] C. Heil and G. Strang, *Continuity of the joint spectral radius*, Proc. of the Workshop on Linear Algebra for Signal Processing, Institute for Mathematics and its Applications (to appear).

[LW] J.C. Lagarias and Y. Wang, *The finiteness conjecture for the generalized spectral radius of a set of matrices*, preprint.

[L] W. Lawton, *Tight frames of compactly supported affine wavelets*, J. Math. Phys. **31** (1990), 1898–1901.

[M] S.G. Mallat, *Multiresolution approximations and wavelet orthonormal bases for* $L^2(\mathbf{R})$, Trans. Amer. Math. Soc. **315** (1989), 69–87.

[Me] Y. Meyer, *Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs*, Séminaire Bourbaki **662** (1985–1986).

[MiP1] C.A. Micchelli and H. Prautzsch, *Refinement and subdivision for spaces of integer translates of compactly supported functions*, Numerical Analysis (D.F. Griffith and G.A. Watson, eds.), Academic Press, New York, 1987, pp. 192–222.

[MiP2] C.A. Micchelli and H. Prautzsch, *Uniform refinement of curves*, Lin. Alg. Appl. **114/115** (1989), 841–870.

[R1] O. Rioul, *Simple regularity criteria for subdivision schemes*, SIAM J. Math. Anal. (to appear).

[R2] O. Rioul, *Simple, optimal regularity estimates for wavelets*, preprint.

[RoS] G.C. Rota and G. Strang, *A note on the joint spectral radius*, Kon. Nederl. Akad. Wet. Proc. A **63** (1960), 379–381.

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

[V1] L. Villemoes, *Energy moments in time and frequency for two-scale difference equations*, SIAM J. Math. Anal. (to appear).

[V2] L. Villemoes, *Wavelet analysis of two-scale difference equations*, preprint.

[W] Y. Wang, *On two-scale dilation equations*, preprint.

ERRATA

Note: This errata listing is not included in the published version of this paper.

The proof of Proposition 4.15 given in the published version of this paper is incorrect. Below is a corrected proof.

PROPOSITION 4.15.
Assume the coefficients $\{c_0, \dots, c_N\}$ satisfy $\sum c_k = 2$ and $n+1$ sum rules. If

$$\hat{\rho}(T_0|_{V^n}, T_1|_{V^n}) \;<\; 2^{-l}$$

for some $0 \le l \le n$, then an l-times continuously differentiable, compactly supported scaling function exists, and its l^{th} derivative is Hölder continuous for each exponent

$$0 \;\le\; \alpha \;<\; \min\bigl\{1,\; -\log_2\bigl(2^l\,\hat{\rho}(T_0|_{V^n}, T_1|_{V^n})\bigr)\bigr\}.$$

PROOF. Recall from Lemma 4.14 that $\{e_0, \dots, e_j\}$ is a basis for U^j for each $j = 0, \dots, n$. By Gram-Schmidt, we can find orthonormal vectors $u_0, \dots, u_n$ so that $\{u_0, \dots, u_j\}$ is a basis for U^j for each j. Note that $\dim(V^n) = N - n - 1$; let $\{v_{n+1}, \dots, v_{N-1}\}$ be an orthonormal basis for V^n. Then since each V^j is the orthogonal complement of U^j in $\mathbf{C}^N$, we have that $\{u_{j+1}, \dots, u_n, v_{n+1}, \dots, v_{N-1}\}$ is an orthonormal basis for V^j. Since each U^j is left-invariant under both T_0, T_1, we can write the matrices T_0, T_1 in the basis $\{u_0, \dots, u_n, v_{n+1}, \dots, v_{N-1}\}$ as

$$BT_iB^{-1} \;=\; \begin{pmatrix} 1 & & & 0 & \\ & 2^{-1} & & & \\ & & \ddots & & \\ & & & 2^{-n} & \\ & * & & & C_i \end{pmatrix}, \qquad i = 0, 1,$$

for an appropriate change-of-basis matrix B. Because V^j is the orthogonal complement of U^j, and because $\{u_0, \dots, u_n, v_{n+1}, \dots, v_{N-1}\}$ is an orthonormal basis for $\mathbf{C}^N$, it is the case that the lower right submatrix

$$\begin{pmatrix} 2^{-j-1} & & 0 & \\ & \ddots & & \\ & & 2^{-n} & \\ & * & & C_i \end{pmatrix}$$

of BT_iB^{-1} is the matrix for $T_i|_{V^j}$ in the basis $\{u_{j+1}, \dots, u_n, v_{n+1}, \dots, v_{N-1}\}$. Therefore, by Lemma 4.7,

$$\hat{\rho}(T_0|_{V^j}, T_1|_{V^j}) \;=\; \begin{cases} \max\{2^{-j-1},\, \dots,\, 2^{-n},\, \hat{\rho}(C_0, C_1)\}, & j = 0, \dots, n-1, \\ \hat{\rho}(C_0, C_1), & j = n. \end{cases}$$

Thus,

$$\hat{\rho}(T_0|_{V^l}, T_1|_{V^l}) \;=\; \max\{2^{-l-1},\; \hat{\rho}(T_0|_{V^n}, T_1|_{V^n})\},$$

so the hypothesis $\hat{\rho}(T_0|_{V^n}, T_1|_{V^n}) < 2^{-l}$ implies $\hat{\rho}(T_0|_{V^l}, T_1|_{V^l}) < 2^{-l}$. The result therefore follows from Theorem 4.12. $\square$

Acknowledgment. We thank Gustaf Grippenberg for pointing out that the published proof of Proposition 4.15 was incorrect.