SIAM J. MATH. ANAL.
Vol. 30, No. 3, pp. 678–697

ARBITRARILY SMOOTH ORTHOGONAL NONSEPARABLE WAVELETS IN $\mathbb{R}^2$*

EUGENE BELOGAY† AND YANG WANG†

Abstract. For each $r \in \mathbb{N}$, we construct a family of bivariate orthogonal wavelets with compact support that are nonseparable and have vanishing moments of order r or less. The starting point of the construction is a scaling function that satisfies a dilation equation with special coefficients and a special dilation matrix M: the coefficients are aligned along two adjacent rows, and $|\det(M)| = 2$. We prove that if $M^2 = \pm 2I$, e. g., $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ or $M = \left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$, then the smoothness of the wavelets improves asymptotically by $1 - \frac{1}{2}\log_2 3 \approx 0.2075$ when r is incremented by 1. Hence they can be made arbitrarily smooth by choosing r large enough.

Key words. nonseparable wavelets, smooth orthogonal scaling function, regularity

AMS subject classifications. Primary, 42C15; Secondary, 26B35, 41A25, 41A63, 65D20

PII. S0036141097327732

1. Introduction.

Since the introduction by Daubechies [7] of compactly supported orthogonal wavelet bases in $\mathbb{R}^1$ with arbitrarily high smoothness, various new wavelet bases (often with specially tailored properties) have been constructed and applied successfully in image processing, numerical computation, statistics, etc. Many of these applications, such as image compression, employ wavelet bases in $\mathbb{R}^2$. Virtually all of these bases are *separable*; that is, the bivariate basis functions are simply tensor products of univariate basis functions. A separable wavelet basis is easy to construct and simple to study, for it inherits features of the corresponding wavelet basis in $\mathbb{R}^1$, such as smoothness and support size. Separable wavelet transforms are easy to implement.

Nevertheless, separable bases have a number of drawbacks. Because they are so special, they have very little design freedom. Furthermore, separability imposes an unnecessary product structure on the plane, which is artificial for natural images. For example, the zero set of a separable scaling function contains horizontal and vertical lines. This "preferred directions" effect can create unpleasant artifacts that become obvious at high image compression ratios. Nonseparable wavelet bases offer the hope of a more isotropic analysis [6, 12, 15].

Despite the success in constructing univariate orthogonal and multivariate biorthogonal wavelet bases with arbitrarily high smoothness, a general theory of smooth multivariate orthogonal nonseparable wavelets is not currently available, and only a few such constructions have been published. Gröchenig and Madych [9] constructed several nonseparable Haar-type scaling functions in $\mathbb{R}^n$, which are discontinuous indicator functions of (often fractal-like) compact sets. Cohen and Daubechies [6] used the univariate construction [7] to produce nonseparable scaling functions with higher accuracy, which are not continuous, as proved by Villemoes [16]. Continuous nonseparable scaling functions were constructed by Kovačević and Vetterli [12], and recently by He and Lai [10], but none of these functions is differentiable.

*Received by the editors September 22, 1997; accepted for publication (in revised form) May 26, 1998; published electronically April 9, 1999.
http://www.siam.org/journals/sima/30-3/32773.html
†School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160 (ebelogay@math.gatech.edu, wang@math.gatech.edu).

In this paper we present a new family of arbitrarily smooth, orthogonal, nonseparable wavelet bases in $\mathbb{R}^2$. Our construction follows the standard multiresolution analysis (MRA) approach [8, 13, 14]: it focuses on a scaling function that solves a dilation (or refinement) equation with special coefficients and a special dilation matrix M with $|\det(M)| = 2$. Such dilation matrices make a popular "laboratory case," partly because the MRA involves only one wavelet [4, 6, 12, 15]. The wavelet is easy to construct from the scaling function and has the same smoothness, so we deal mainly with the scaling function.

The coefficients in the dilation equation, called scaling coefficients, determine the properties of the scaling function. To construct a scaling function usually means to find its scaling coefficients. We characterize completely the set of two-row scaling coefficients that produce nonseparable orthogonal scaling functions with arbitrarily high accuracy. We show that our construction can produce scaling functions of any desired smoothness for the special dilation matrix $\left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$.

The paper is organized as follows. In section 2 we introduce some basic notions and assumptions and state our two main results: the first describes the scaling coefficients; the second determines the smoothness of the scaling function. In section 3 we formulate and solve the equations that the scaling coefficients must satisfy in order for the scaling function to be orthogonal and accurate. In section 4 we prove the smoothness result. In section 5 we plot some of the new scaling functions and explain how our results can be formulated for dilation matrices other than $\left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$.

Note. After submitting this paper, the authors learned about related independent results by Ayache [1], who recently constructed arbitrarily smooth nonseparable orthogonal wavelets by perturbing the separable wavelets for dilation $\left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right)$.

2. Preliminaries and main results. Let M be an expanding 2×2 matrix with integer entries such that $|\det(M)| = 2$. The key ingredients to an MRA with such a *dilation matrix* M are two functions: a *scaling function* ϕ and a *wavelet* ψ. The scaling function $\phi : \mathbb{R}^2 \to \mathbb{R}$ satisfies a *dilation equation* (*refinement equation*) of the form

$$\phi(\mathbf{x}) = 2 \sum_{\mathbf{n} \in \mathbb{Z}^2} c_{\mathbf{n}} \phi(M\mathbf{x} - \mathbf{n}), \qquad \text{where} \quad \sum_{\mathbf{n} \in \mathbb{Z}^2} c_{\mathbf{n}} = 1. \tag{2.1}$$

The numbers $c_{\mathbf{n}}$ are the *scaling coefficients* (*low-pass filter coefficients*) of $\phi(\mathbf{x})$. We assume that they are real and that $c_{\mathbf{n}} \neq 0$ for only finitely many $\mathbf{n} \in \mathbb{Z}^2$ (ensuring that $\phi(\mathbf{x})$ has compact support). A convenient way to work with the scaling coefficients as a whole is to consider the *coefficient mask* (*z-transform*)

$$C(z, w) := \sum_{(m,n) \in \mathbb{Z}^2} c_{(m,n)} z^m w^n, \quad \text{where } (z, w) \in \mathbb{C}^2.$$

Note that $C(1, 1) = 1$. The *symbol* (*frequency response*) of (2.1) is

$$m(\omega_1, \omega_2) := C(e^{i\omega_1}, e^{i\omega_2}).$$

Denote M^{T} by W. The mask is said to satisfy *Cohen's criterion* [16, p. 181] if there exists a compact fundamental domain Ω of the lattice $2\pi\mathbb{Z}^2$ with the property

$$m(W^{-j}\omega) \neq 0 \quad \text{for all } j \geq 1 \text{ and for all } \omega \in \Omega. \tag{2.2}$$

An important task in wavelet theory is to relate the properties of the scaling function $\phi(\mathbf{x})$ to the properties of the coefficient mask $C(z, w)$. Our goal is to find coefficients in (2.1) that produce a scaling function with two important properties: orthogonality and accuracy.

A scaling function $\phi(\mathbf{x}) \in L^2(\mathbb{R}^2)$ is called *orthogonal* if the set of its lattice translates $\{\phi(\mathbf{x}-\mathbf{k}) : \mathbf{k} \in \mathbb{Z}^2\}$ is orthogonal. The following *orthogonal coefficients condition* is necessary for $\phi(\mathbf{x})$ to be orthogonal:

$$2 \sum_{\mathbf{n}\in\mathbb{Z}^2} c_{\mathbf{n}} c_{\mathbf{n}+M\mathbf{k}} = \delta_{\mathbf{0},\mathbf{k}} \qquad \text{for all } \mathbf{k} \in \mathbb{Z}^2, \tag{2.3}$$

where $\delta_{\mathbf{m},\mathbf{k}}$ is the Kronecker symbol. The condition (2.3) becomes sufficient if, in addition, the scaling coefficients satisfy Cohen's criterion (2.2). The coefficient mask $C(z,w)$ is called *orthogonal* if (2.3) holds.

A scaling function $\phi(\mathbf{x})$ is said to have *accuracy* (*regularity*) r if the space of the infinite linear combinations of $\{\phi(\mathbf{x}-\mathbf{k}) : \mathbf{k} \in \mathbb{Z}^2\}$ contains all polynomials of degree $r-1$ or less. In this case the coefficient mask $C(z,w)$ is said to have accuracy r as well. (How the accuracy of ϕ relates to $C(z,w)$ is explained in the next section.) Accuracy is a desirable property in many applications; for example, in image processing it implies that the polynomial components of the filtered signal will not "leak" into the high-pass band, which improves compression.

In this paper we measure the smoothness of a real bivariate function ϕ by its Hölder exponent (there are alternative regularity measures, such as the Sobolev exponent). Let $s = m + \gamma$, where $0 \le \gamma < 1$ and m is a nonnegative integer. Then we say that $\phi(\mathbf{x}) \in C^s$ if for each partial derivative $D^\alpha \phi(\mathbf{x})$, where $\alpha = (\alpha_1, \alpha_2)$ and $\alpha_1 + \alpha_2 \le m$, there is a constant $c > 0$ such that $|D^\alpha\phi(\mathbf{x}) - D^\alpha\phi(\mathbf{y})| \le c|\mathbf{x}-\mathbf{y}|^\gamma$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$. If the Fourier transform $\widehat{\phi}(\omega)$ of $\phi(\mathbf{x})$ satisfies

$$|\widehat{\phi}(\omega)| \le c \cdot (1 + |\omega_1|)^{-s-1-\epsilon}(1 + |\omega_2|)^{-s-1-\epsilon}$$

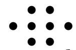
for some constants $c > 0$ and $\epsilon > 0$, then $\phi(\mathbf{x}) \in C^s$. It is usually harder to estimate the smoothness of a scaling function than its accuracy.

The *autocorrelation* (*product filter*) of a real polynomial $F(z_1, z_2)$ is defined to be

$$\mathcal{P}_F(z_1, z_2) := F(z_1, z_2)F(z_1^{-1}, z_2^{-1}).$$

The polynomial F is called a *spectral factor* of $\mathcal{P}_F$. Note that $\mathcal{P}_F(e^{i\omega_1}, e^{i\omega_2}) = |F(e^{i\omega_1}, e^{i\omega_2})|^2$ and is therefore always real and nonnegative. Conversely, the Fèjer–Riesz theorem [7] ensures that every real-valued nonnegative univariate trigonometric polynomial $P(e^{i\omega_1})$ can be factored (nonuniquely) as $|F(e^{i\omega_1})|^2$. In most cases the spectral factor F can only be computed numerically. (Strang and Nguyen [15, p. 157] review several spectral factorization methods.)

Spectral factorization is a key technique in univariate orthogonal wavelet theory. Daubechies [7] constructed orthogonal univariate wavelets of arbitrarily high accuracy by deriving an analytic formula for the autocorrelation of the coefficient mask; the scaling coefficients themselves must be computed numerically. Theorem 2.1 is a similar result in a special bivariate setting.

The main difficulty in constructing bivariate nonseparable smooth wavelets is that some key univariate techniques, such as polynomial factorization, do not generalize to the bivariate case. The Fèjer–Riesz theorem is one of them: e.g., $2+\cos(\omega_1)+\cos(\omega_2)$ is real and nonnegative; yet it cannot be factored as $|F(e^{i\omega_1}, e^{i\omega_2})|^2$. One approach to nonseparable orthogonal wavelets is to solve the accuracy and orthogonality conditions for a specific arrangement of unknown scaling coefficients. Unfortunately, the resulting system of linear and quadratic equations can only be solved (even numerically) when the number of coefficients is rather small. (Kovačević and Vetterli [12] placed the 8 nonzero coefficients in the  pattern; He and Lai [10] used a 4×4 mask.) Another approach to nonseparable wavelets, which we adopt here, is to study special cases of (2.1) and derive the scaling coefficients explicitly.

Our construction employs a special coefficient mask for a particular dilation matrix (generalizations are discussed in section 5). First, we fix the dilation matrix

$$M = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}.$$

The property $M^2 = 2I$ will be useful in the proof of Theorem 2.2. Next, we restrict the placement of the nonzero scaling coefficients in (2.1) to two adjacent rows in the first quadrant. That is, $\operatorname{supp} c := \{\mathbf{n} \in \mathbb{Z}^2 : c_{\mathbf{n}} \neq 0\} \subseteq \{0, \ldots, N\} \times \{0, 1\}$ for some $N \in \mathbb{N}$. As a result, the coefficient mask has the form

$$C(z, w) = A(z) + wB(z), \tag{2.4}$$

where $A(z)$ and $B(z)$ are polynomials of one complex variable. The theorems below do not generalize easily to masks with more than two rows.

Observe that $A(1)+B(1) = C(1, 1) = 1$, so it is impossible that $A(1) = B(1) = 0$. Therefore, we assume that $A(1) \neq 0$, for we can always achieve it by a suitable change of variables in (2.1), as explained at the end of section 3. Furthermore, we assume that $B(0) \neq 0$. Cohen and Daubechies [6] noted that if $B(z) \equiv 0$ (that is, if the scaling coefficients are aligned along a single horizontal line), then the scaling function is separable. A simple but important example of such a one-row mask is the *Haar coefficient mask*

$$H(z) := \frac{1}{2}(1 + z). \tag{2.5}$$

The corresponding separable scaling function is the indicator of the unit square.

Our first theorem gives the necessary conditions for a two-row coefficient mask to produce a nonseparable orthogonal scaling function of arbitrarily high accuracy. The proof is in section 3.

THEOREM 2.1. *Let $\phi(\mathbf{x})$ satisfy the dilation equation (2.1) with dilation $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ and coefficient mask $C(z, w) = A(z) + wB(z)$, where $A(1) \neq 0$ and $B(0) \neq 0$. Let $r \in \mathbb{N}$ and let ν be an odd integer with $\nu \geq \deg A$. If the scaling function $\phi(\mathbf{x})$ is orthogonal and has accuracy $r+1$, then the polynomials $A(z)$ and $B(z)$ have the form*

$$z^{\nu} A(z^{-1}) = H^r(z) L(z) S(z^2), \tag{2.6}$$
$$B(z) = H^r(z) L(-z) Q(z^2) H^{2r}(-z), \tag{2.7}$$

where $L(z)$, $S(z)$, and $Q(z)$ are any polynomials that satisfy

$$\mathcal{P}_S(z^2) = 1 - \left(\frac{1 - z^2}{4}\right)^r \left(\frac{1 - z^{-2}}{4}\right)^r \mathcal{P}_Q(z^2), \tag{2.8}$$

$$\mathcal{P}_L(z) = \sum_{j=0}^{r-1} \binom{r+j-1}{j} \left(\frac{1-u}{2}\right)^j + (1-u)^r u\, R(u^2), \quad u := \frac{1}{2}(z + z^{-1}), \tag{2.9}$$

$$S(1) = L(1) = 1, Q(1) = (-1)^r L(-1), \quad \text{and} \quad L(0)Q(0) \neq 0 \tag{2.10}$$

and $R(z)$ is an arbitrary polynomial.

Remark. The converse of Theorem 2.1 holds if, in addition to (2.6)–(2.10), the mask $C(z,w)$ satisfies Cohen's criterion (2.2).

Theorem 2.1 suggests the following procedure for obtaining the coefficients of $C(z,w)$:

(i) Choose polynomials Q and R so that the right-hand sides of (2.8)–(2.9) are nonnegative along the unit circle $|z|=1$ (thus ensuring that the next step can be performed).

(ii) Using some spectral factorization method, make a list of all polynomials S and L that satisfy (2.8)–(2.10). Choose a specific pair. (Observe that if $R=0$, then the polynomial L is one of the univariate orthogonal coefficients masks with accuracy r constructed by Daubechies [7].)

(iii) Substitute L, S, and Q in (2.6)–(2.7) and expand. Choose an odd $\nu \geq \deg A$.

Note that the choices in step (ii) are not unique. Unlike the univariate case, there is no obvious way to single out a "minimal phase" coefficient mask $C(z,w)$.

The minimal degree of $A(z)$ and $B(z)$ in Theorem 2.1 is $4r-1$ and is achieved if

$$Q(z) = \text{const} = (-1)^r L(-1), \quad R(z) = \text{const} = 0, \quad \text{and} \quad \nu = 4r-1. \tag{2.11}$$

There are $2^{1+2\lfloor r/2 \rfloor}$ coefficient masks $C(z,w)$ that satisfy the conditions (2.11) and $A(0) \neq 0$ in addition to the conditions in Theorem 2.1. Denote the set of those masks by $\mathcal{C}_{r+1}$. Every $C(z,w)$ in $\mathcal{C}_{r+1}$ produces a scaling function $\phi(\mathbf{x})$ with accuracy $r+1$. Denote the set of those scaling functions by Φ_{r+1}.

Our second theorem states that all functions in Φ_{r+1} are compactly supported orthogonal nonseparable scaling functions that can be made arbitrarily smooth by choosing the accuracy $r+1$ large enough. The proof is in section 4.

THEOREM 2.2. *Let $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ and let $\phi \in \Phi_{r+1}$. Then*

(i) *ϕ is an orthogonal scaling function with accuracy $r+1$;*

(ii) *$\operatorname{supp}\phi \subseteq [0, 4r+1] \times [0, 4r]$;*

(iii) *ϕ is nonseparable;*

(iv) *if $r \geq 5$, then $\phi \in C^{r-\mu_r-2}$, where*

$$\mu_r := \frac{1}{2}\log_2\left(\sum_{j=0}^{r-1}\binom{r+j-1}{j}\left(\frac{3}{4}\right)^j\right).$$

Furthermore,

$$r - \mu_r > \left(1 - \frac{1}{2}\log_2 3\right) r + \frac{1}{2}\log_2 3 \tag{2.12}$$

and

$$\lim_{r\to\infty} \frac{r-\mu_r-2}{r} = 1 - \frac{1}{2}\log_2 3 \approx 0.2075. \tag{2.13}$$

Remark. As in the univariate case, the statements (2.12)–(2.13) guarantee the existence of orthogonal wavelets of any desired smoothness. In particular, $\Phi_{5+1} \subset C^0$, $\Phi_{9+1} \subset C^1$, and $\Phi_{13+1} \subset C^2$. The smoothness estimate $\phi \in C^{r-\mu_r-2}$ applies to all ϕ in Φ_{r+1} uniformly and is not sharp. Lemma 5.1 provides sharper smoothness estimates for individual functions in Φ_{r+1}, involving numerical computations.

If $\phi(\mathbf{x})$ is an orthogonal scaling function, then its associated *wavelet* $\psi(\mathbf{x})$ has the form

$$\psi(\mathbf{x}) = 2\sum_{\mathbf{n}\in\mathbb{Z}^2} d_{\mathbf{n}}\phi(M\mathbf{x}-\mathbf{n}).$$

FIG. 2.1. *The quincunx and the column sublattice.*

Obviously, the wavelet ϕ has the same smoothness as the scaling function ϕ. It is known that if the *wavelet coefficients* (*high-pass filter coefficients*) $d_{\mathbf{n}}$ are given by

$$d_{\mathbf{n}} = (-1)^{n_1} c_{\mathbf{e}-\mathbf{n}}, \quad \text{where } \mathbf{n} = (n_1, n_2) \text{ and } \mathbf{e} = (1, 0),$$

then the wavelet $\psi(\mathbf{x})$ is orthogonal to $\{\phi(\mathbf{x} - \mathbf{k}) : \mathbf{k} \in \mathbb{Z}^2\}$, and its lattice translates and dilates form an orthogonal basis of $L^2(\mathbb{R}^2)$ [13, 14].

Although Theorems 2.1 and 2.2 are formulated for $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$, they can be generalized [2] to other expanding matrices with integer entries and $|\det(M)| = 2$. Each such matrix transforms the lattice $\mathbb{Z}^2$ into three types of sublattices: the *quincunx*, the *column* sublattice (see Figure 2.1), and the *row* sublattice (the row sublattice is merely a transpose of the column sublattice).

The accuracy and the orthogonality conditions depend only on the sublattice type and not on the dilation matrix, as we shall see in section 3. Therefore, Theorem 2.1 still holds if $\left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ is replaced by any other expanding integral matrix M that generates the column sublattice $2\mathbb{Z} \times \mathbb{Z}$, such as $\left(\begin{smallmatrix} 0 & -2 \\ 1 & 0 \end{smallmatrix}\right)$. If $\left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ is replaced by the quincunx matrix $\left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$, or any other matrix that generates the quincunx lattice, such as $\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$, then $A(z) + wB(z)$ must be replaced by $A(z) + wzB(z)$, as we shall see at the end of section 5.

In contrast, the smoothness of the scaling function does depend on the dilation matrix. Theorem 2.2 can be extended to any dilation M with the properties $M^2 = \pm 2I$, such as $\left(\begin{smallmatrix} 0 & -2 \\ 1 & 0 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$. Although the quincunx matrix $\left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ has often been used in nonseparable constructions, the authors know of no scaling functions that are both differentiable and orthogonal with respect to the quincunx matrix; it is quite possible that $\left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ admits no such scaling functions at all.

3. Accuracy and orthogonality conditions. In this section we prove Theorem 2.1. First, we study how the accuracy and the orthogonality of the scaling function restrict the scaling coefficients. Then, we derive the formulas (2.6)–(2.10). Recall that $H(z) = \frac{1}{2}(1 + z)$.

Accuracy. The accuracy of the scaling function ϕ has many equivalent formulations, such as the vanishing moments condition for the wavelet ($\int \mathbf{x}^\alpha \psi(\mathbf{x})\, d\mathbf{x} = 0$, $|\alpha| < r$) or the "sum rules" on the coefficients $c_{\mathbf{n}}$ [4]. Here, we prefer to work with the sum rules. Applied to our dilation matrix $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$, a general result by Cabrelli, Heil, and Molter [4] states that an orthogonal scaling function ϕ has accuracy r if and only if its coefficient mask $C(z, w)$ satisfies the following *accuracy condition*:

$$\frac{\partial^{p+q}}{\partial z^p \partial w^q} C(-1, 1) = 0 \quad \text{for all } p, q \geq 0 \text{ with } p + q < r. \tag{3.1}$$

Observe that the analogous condition for a univariate polynomial $C(z)$ implies the factorization $C(z) = (1+z)^r Q(z)$ for some $Q(z)$. Unfortunately, bivariate polynomials $C(z, w)$ cannot be described in such a simple way. (This fact is one of the main obstacles in studying nonseparable bivariate wavelets.) Nevertheless, the accuracy condition (3.1) takes on a simple form for our special two-row mask $C(z,w)$.

LEMMA 3.1. *Let* $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$. *The dilation equation* (2.1) *with coefficient mask* $C(z,w) = A(z) + wB(z)$ *has accuracy* $r+1$ *if and only if*

$$A(z) = H^r(z)A_0(z), \quad B(z) = H^r(z)B_0(z), \tag{3.2}$$

$$A_0(-1) + B_0(-1) = 0. \tag{3.3}$$

Proof. It is easy to check that (3.2)–(3.3) imply (3.1). We now prove the converse. By taking $q=0$ in (3.1), we obtain $A^{(p)}(-1) + B^{(p)}(-1) = 0$ for all $0 \le p \le r$. By taking $q=1$, we obtain $B^{(p)}(-1) = 0$ for all $0 \le p \le r-1$. These conditions imply $A^{(p)}(-1) = 0$ for all $0 \le p \le r-1$. Hence $A(z) = H^r(z)A_0(z)$ and $B(z) = H^r(z)B_0(z)$. Finally, $A^{(r)}(-1) + B^{(r)}(-1) = 0$ yields $A_0(-1) + B_0(-1) = 0$. ∎

Note that, unlike the univariate case, here accuracy $r+1$ corresponds to r Haar factors plus one extra condition (3.3).

Orthogonality. If $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$, the orthogonality condition (2.3) is equivalent to

$$\mathcal{P}_C(z,w) + \mathcal{P}_C(-z,w) = 1.$$

For the special two-row coefficient mask (2.4), this condition splits further into the following two conditions imposed on the polynomials A and B:

$$\mathcal{P}_A(z) + \mathcal{P}_A(-z) + \mathcal{P}_B(z) + \mathcal{P}_B(-z) = 1, \tag{3.4}$$

$$A(z^{-1})B(z) + A(-z^{-1})B(-z) = 0. \tag{3.5}$$

Remark. The second condition means that $A(z^{-1})B(z)$ contains no even powers of z. Therefore, $\deg A$ must be odd if $B(0) \ne 0$.

The main task in the remainder of this section is to solve (3.2)–(3.5) for the unknown polynomials A and B. Observe that (3.2)–(3.3) are linear and that (3.4)–(3.5) are quadratic.

First, we investigate (3.5).

LEMMA 3.2. *For every polynomial* $p(z) = z^{2n}p_1(z)$ *with* $p_1(0) \ne 0$ *and* $n \ge 0$, *there exist polynomials* q *and* ℓ *such that* $p(z) = q(z^2)\ell(z)$ *and* $\gcd(\ell(z), \ell(-z)) = 1$.

Proof. Consider $t(z) := \gcd(p(z), p(-z))$. Clearly, $t(z) = t(-z) = q(z^2)$ for some polynomial $q(z)$. Let $\ell(z) := p(z)/q(z^2)$, and observe that $\gcd(\ell(z), \ell(-z)) = 1$. ∎

LEMMA 3.3. *The following are equivalent:*

(i) *The polynomials* A *and* B *satisfy condition* (3.5) *and* $B(0) \ne 0$.

(ii) *There exist an odd integer* $\nu \ge \deg A$ *and real polynomials* s, q, *and* ℓ *such that* $z^\nu A(z^{-1}) = s(z^2)\ell(z)$, $B(z) = q(z^2)\ell(-z)$, *and* $\gcd(\ell(z), \ell(-z)) = 1$.

Proof. The implication (ii) ⇒ (i) follows from

$$z^\nu s(z^2)\ell(z)q(z^2)\ell(-z) + (-z)^\nu s(z^2)\ell(-z)q(z^2)\ell(z) = 0.$$

Assume that (i) holds, and fix an odd $\nu \ge \deg A$. Lemma 3.2 allows us to write $z^\nu A(z^{-1}) = s(z^2)\ell(z)$ and $B(z) = q(z^2)m(z)$ for some polynomials s, q, ℓ, and m. Substituting $A(z^{-1})$ and $B(z)$ in (3.5) yields

$$m(z)\ell(z) = m(-z)\ell(-z).$$

Since $\gcd(m(z), m(-z)) = \gcd(\ell(z), \ell(-z)) = 1$, it follows that $m(z) = \ell(-z)$. ∎

Next, we take into account the accuracy of $C(z,w)$.

LEMMA 3.4. *Let A and B be polynomials with $A(1) \neq 0$ and $B(0) \neq 0$. Then the following are equivalent:*

(i) *The two-row coefficient mask $C(z,w) = A(z) + wB(z)$ satisfies* (3.5) *and has accuracy $r+1$.*

(ii) *There exists an odd integer $\nu \geq \deg A$ and real polynomials S, Q, and L with $L(0)Q(0) \neq 0$, $S(1) = L(1) = 1$, and $Q(1) = (-1)^r L(-1)$ such that*

$$z^\nu A(z^{-1}) = H^r(z)L(z)S(z^2),$$
$$B(z) = H^r(z)L(-z)Q(z^2)H^{2r}(-z).$$

Proof. The implication (ii)⇒(i) follows from Lemmas 3.1 and 3.3 by letting $s(z^2) = S(z^2)$, $q(z^2) = Q(z^2)\big(\frac{1-z^2}{2}\big)^r$, and $\ell(z) = H^r(z)L(z)$. We now prove that (i)⇒(ii). By Lemma 3.1, $H^r(z)$ divides both $A(z)$ and $B(z)$. Therefore, $z=-1$ is a root of multiplicity $2r$ of $P(z) := z^\nu A(z^{-1})B(z)$. On the other hand, $P(z) = P(-z)$ by (3.5); hence $z = 1$ is another root of multiplicity $2r$ of $P(z)$. Since $A(1) \neq 0$ by assumption, $H^{2r}(-z)$ must divide $B(z)$. Define

$$A_1(z) := A(z)z^r/H^r(z), \quad \text{and} \quad B_1(z) := B(z)/\big(H^r(z)H^{2r}(-z)\big).$$

Noting that $H(z^{-1}) = z^{-1}H(z)$, we rewrite

$$A(z^{-1})B(z) = H(z)^{2r}H(-z)^{2r}A_1(z^{-1})B_1(z)$$

and substitute in (3.5). Factoring out $H(z)^{2r}H(-z)^{2r}$ yields the equation

$$A_1(z^{-1})B_1(z) + A_1(-z^{-1})B_1(-z) = 0.$$

Therefore, Lemma 3.2 applies to A_1 and B_1, and we obtain the desired factorization in (ii). Since $B(1) = 0$, and therefore $A(1) = 1$, we can normalize L and S so that $L(1) = S(1) = 1$. The only restriction on Q, that $Q(1) = (-1)^r L(-1)$, then follows directly from the extra accuracy condition (3.3). ∎

Finally, to prove Theorem 2.1 we need to solve the remaining equation, (3.4).

Proof of Theorem 2.1. We need only show that the polynomials S, Q and L in Lemma 3.4 have the autocorrelations given by (2.8) and (2.9). By Lemma 3.4,

$$\mathcal{P}_A(z) = \mathcal{P}_H^r(z)\mathcal{P}_S(z^2)\mathcal{P}_L(z), \quad \text{and} \quad \mathcal{P}_B(z) = \mathcal{P}_H^r(z)\mathcal{P}_H^{2r}(-z)\mathcal{P}_L(-z)\mathcal{P}_Q(z^2).$$

Hence $\mathcal{P}_A(z) + \mathcal{P}_B(-z) = \mathcal{P}_H^r(z)\mathcal{P}_L(z)U(z^2)$, where

$$U(z^2) := \mathcal{P}_S(z^2) + \mathcal{P}_H^r(-z)\mathcal{P}_H^r(z)\mathcal{P}_Q(z^2).$$

Therefore, the orthogonality condition (3.4) can be factored as follows:

$$\big(\mathcal{P}_H^r(z)\mathcal{P}_L(z) + \mathcal{P}_H^r(-z)\mathcal{P}_L(-z)\big)U(z^2) = 1.$$

It follows that $U(z^2)$ must be a constant. In particular, $U(z^2) = U(1) = \mathcal{P}_S(1) = 1$, which immediately yields (2.8). Finally, the equation

$$\mathcal{P}_H^r(z)\mathcal{P}_L(z) + \mathcal{P}_H^r(-z)\mathcal{P}_L(-z) = 1 \tag{3.6}$$

characterizes all univariate orthogonal masks with accuracy r and was solved by Daubechies [7]. The solution to (3.6) is given by (2.9). ∎

Remark. Although the results in this section are formulated for $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$, the accuracy and orthogonality conditions (2.3) and (3.1) depend only on the sublattice $M\mathbb{Z}^2$ (in our case, the column sublattice), not on the particular dilation matrix. Therefore, Theorem 2.1 and all results in this section apply to any other integer expanding matrix M that satisfies $M\mathbb{Z}^2 = 2\mathbb{Z} \times \mathbb{Z}$. Corollary 5.3 and Lemma 5.4 explain how to apply Theorem 2.1 to other dilation matrices, such as $\left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right)$.

The statement of Theorem 2.1 contains two assumptions: $B(0) \neq 0$ and $A(1) \neq 0$. Now we explain why these assumptions impose no loss of generality.

First, by shifting the variables in the dilation equation (2.1), one can check that a shift of the coefficient mask by a vector $\mathbf{s}$ results in a shift of the scaling function by $(M - I)\mathbf{s}$. Therefore, without loss of generality, the coefficients of any mask $C(z, w)$ can always be shifted so that $C(z, w)$ contains no negative powers of w, and so that the smallest power of z in $B(z)$ is zero. (The Laurent polynomial $A(z)$ may contain negative powers of z; nevertheless, Theorem 2.1 holds.)

Second, suppose that $A(1) = 0$ and $B(1) \neq 0$ (recall that $A(1) + B(1) = 1$). Change the variables $\mathbf{x} \mapsto -\mathbf{x}$ and $\mathbf{n} \mapsto -\mathbf{n}$ in (2.1), and note (cf. Lemma 5.2) that if $\phi(\mathbf{x})$ solves (2.1) with the coefficient mask $C(z, w)$, then $\phi(-\mathbf{x})$ solves (2.1) with the coefficient mask $\tilde{C}(z, w) = C(z^{-1}, w^{-1})$. Shift the coefficients of $\tilde{C}(z, w)$ to the first quadrant (that is, multiply by $z^N w$, where $N = \max(\deg A, \deg B)$) and observe that $z^N w\, \tilde{C}(z, w) = \tilde{A}(z) + w\tilde{B}(z)$, where $\tilde{A}(1) = B(1) \neq 0$. The shift of $\phi(-\mathbf{x})$ is orthogonal and has accuracy r if and only if $\phi(\mathbf{x})$ is orthogonal and has accuracy r. So the assumption $A(1) \neq 0$ causes no loss of generality in Theorem 2.1.

4. Smoothness. In this section we prove Theorem 2.2. To prove the orthogonality of a scaling function $\phi \in \Phi_{r+1}$, we check Cohen's criterion (4.6). We establish the smoothness of ϕ by estimating the decay of its Fourier transform $\widehat{\phi}$ in a series of lemmas. In this section, $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ and $W = M^{\mathrm{T}}$.

The *symbol* of the dilation equation (2.1) with coefficient mask $C(z, w)$ is the trigonometric polynomial $m(\omega_1, \omega_2) := C(e^{i\omega_1}, e^{i\omega_2})$. Taking the Fourier transform of (2.1) and iterating, we obtain

$$\widehat{\phi}(\omega) = \widehat{\phi}(0) \cdot \prod_{j=1}^{\infty} m(W^{-j}\omega). \tag{4.1}$$

Since $W^{-2} = \frac{1}{2}I$, the infinite product (4.1) breaks into two parts:

$$\prod_{j=1}^{\infty} m(W^{-j}\omega) = \left(\prod_{j=1}^{\infty} m(2^{-j}\omega)\right)\left(\prod_{j=1}^{\infty} m(2^{-j}W\omega)\right). \tag{4.2}$$

Our goal is to derive an estimate for $\prod_{j=1}^{\infty} m(2^{-j}\omega)$ when $C(z, w) \in \mathcal{C}_{r+1}$ by using only the first frequency ω_1.

First, let $C(z, w) = H(z) = \frac{1}{2}(1 + z)$ in (2.1). It can be checked that (2.1) is solved by the Haar scaling function $\chi_{[0,1)^2}$ and that

$$\prod_{j=1}^{\infty} m_H(W^{-j}\omega) = \widehat{\chi_{[0,1)}}(\omega_1) \cdot \widehat{\chi_{[0,1)}}(\omega_2) = \frac{(e^{i\omega_1} - 1)}{i\omega_1} \cdot \frac{(e^{i\omega_2} - 1)}{i\omega_2},$$

where $m_H(\omega_1, \omega_2) := H(e^{i\omega_1})$. Therefore, there is a constant $c > 0$ such that

$$\Bigl|\prod_{j=1}^{\infty} m_H(W^{-j}\omega)\Bigr| \leq c \cdot \bigl(1 + |\omega_1|\bigr)^{-1}\bigl(1 + |\omega_2|\bigr)^{-1}. \tag{4.3}$$

Next, fix $r \in \mathbb{N}$, and consider any coefficient mask $C(z,w) \in \mathcal{C}_{r+1}$, as described in Theorem 2.1 and (2.11). The corresponding symbol has the form

$$m(\omega_1,\omega_2) = C(e^{i\omega_1}, e^{i\omega_2}) = H^r(e^{i\omega_1})\big(A_0(e^{i\omega_1}) + e^{i\omega_2}B_0(e^{i\omega_1})\big).$$

Define

$$q(\omega_1,\omega_2) := A_0(e^{i\omega_1}) + e^{i\omega_2}B_0(e^{i\omega_1}) \quad \text{and} \quad \ell(\omega_1) := |A_0(e^{i\omega_1})| + |B_0(e^{i\omega_1})|. \tag{4.4}$$

Clearly, $|q(\omega_1,\omega_2)| \le \ell(\omega_1)$. We estimate $\ell(\omega_1)$ in a series of lemmas. Define

$$T_r(y) := \sum_{j=0}^{r-1} \binom{r+j-1}{j} y^j. \tag{4.5}$$

LEMMA 4.1. *Let $L(z)$ satisfy* (2.9). *Assume $r \ge 1$. Then*

$$\mathcal{P}_H^r(-e^{i\omega_1})L^2(-1) \le T_r\left(\sin^2\frac{\omega_1}{2}\right),$$

and equality holds only if $\omega_1 \equiv \pi \pmod{2\pi}$.

Proof. From (2.9) we obtain $L^2(-1) = \mathcal{P}_L(-1) = T_r(1) = \sum_{j=0}^{r-1}\binom{r+j-1}{j}$. By the definition of H in (2.5),

$$\mathcal{P}_H^r(-e^{i\omega_1}) = \left(\frac{1-\cos\omega_1}{2}\right)^r = \left(\sin\frac{\omega_1}{2}\right)^{2r} \le \left(\sin\frac{\omega_1}{2}\right)^{2j} \quad \text{for all } j \le r.$$

Hence

$$\begin{aligned} L^2(-1)\mathcal{P}_H^r(-e^{i\omega_1}) &= \sum_{j=0}^{r-1}\binom{r+j-1}{j}\left(\sin\frac{\omega_1}{2}\right)^{2r} \\ &\le \sum_{j=0}^{r-1}\binom{r+j-1}{j}\left(\sin\frac{\omega_1}{2}\right)^{2j} = T_r\left(\sin^2\frac{\omega_1}{2}\right). \end{aligned}$$

Because the coefficients of T_r are positive, equality holds only when $\sin^2\frac{\omega_1}{2} = 1$, that is, when $\omega_1 \equiv \pi \pmod{2\pi}$. ∎

COROLLARY 4.2. *There exist real polynomials S and L that satisfy* (2.8)–(2.11).

Proof. We need only show that (2.8) defines a valid autocorrelation; that is,

$$1 - \mathcal{P}_H^r(z)\mathcal{P}_H^r(-z)L^2(-1) \ge 0 \quad \text{for all} \quad |z| = 1.$$

Let $|z| = 1$. From Lemma 4.1 and (3.6) we obtain

$$\mathcal{P}_H^r(z)\mathcal{P}_H^r(-z)L^2(-1) \le \mathcal{P}_H^r(z)\mathcal{P}_L(z) = 1 - \mathcal{P}_H^r(-z)\mathcal{P}_L(-z) \le 1.$$

The last inequality holds because $\mathcal{P}_H(z) \ge 0$ and $\mathcal{P}_L(z) \ge 0$ for $|z| = 1$. ∎

LEMMA 4.3. $|A_0(e^{i\omega_1})|^2 \le T_r(\sin^2\frac{\omega_1}{2})$ *and* $|B_0(e^{i\omega_1})|^2 \le T_r(\sin^2\frac{\omega_1}{2})$. *Each equality holds only if $\omega_1 \equiv \pi \pmod{2\pi}$.*

Proof. By the construction in Theorem 2.1 and (2.11) we have

$$\begin{aligned} \mathcal{P}_{A_0}(z) &= \mathcal{P}_L(z)\mathcal{P}_S(z^2) = \mathcal{P}_L(z)\big(1 - \mathcal{P}_H^r(z)\mathcal{P}_H^r(-z)L^2(-1)\big), \\ \mathcal{P}_{B_0}(z) &= \mathcal{P}_H^r(-z)\mathcal{P}_H^r(-z)\mathcal{P}_L(-z)L^2(-1). \end{aligned}$$

Since $\mathcal{P}_H^r(e^{i\omega}) \geq 0$ for all ω,

$$|A_0(e^{i\omega_1})|^2 = \mathcal{P}_{A_0}(e^{i\omega_1}) \leq \mathcal{P}_L(e^{i\omega_1}) \cdot 1 = T_r(\sin^2 \tfrac{\omega_1}{2}).$$

Observe that (3.6) yields $0 \leq \mathcal{P}_H^r(-z)\mathcal{P}_L(-z) \leq 1$ for $z = e^{i\omega_1}$. Therefore,

$$|B_0(e^{i\omega_1})|^2 = \mathcal{P}_{B_0}(e^{i\omega_1}) \leq \mathcal{P}_H^r(-e^{i\omega_1}) \cdot 1 \cdot L^2(-1).$$

Lemma 4.1 completes the inequality for B_0. ∎

Before we proceed with the estimate of the infinite product (4.2), we recall the following lemma (without proof) from Daubechies [8, pp. 220–226]:

Lemma 4.4 (Cohen and Conze [5]). *Let* $\mu_r := \frac{1}{2}\log_2 T_r(\frac{3}{4})$. *Then*

$$T_r\big(\sin^2 \tfrac{\omega_1}{2}\big) \leq T_r\big(\tfrac{3}{4}\big) \quad \text{if } |\omega_1| \leq \tfrac{2\pi}{3},$$
$$T_r(\sin^2 \omega_1)T_r\big(\sin^2 \tfrac{\omega_1}{2}\big) \leq T_r^2\big(\tfrac{3}{4}\big) \quad \text{if } \tfrac{2\pi}{3} < |\omega_1| \leq \pi.$$

Furthermore,

$$r - \mu_r > \left(1 - \frac{1}{2}\log_2 3\right) r + \frac{1}{2}\log_2 3,$$
$$\lim_{r\to\infty} \frac{r - \mu_r - 2}{r} = 1 - \frac{1}{2}\log_2 3 \approx 0.2075. \qquad ∎$$

Lemma 4.5. *There exist constants* $c > 0$ *and* $\epsilon > 0$ *such that*

$$\prod_{j=1}^{\infty} \ell(2^{-j}\omega_1) \leq c \cdot |\omega_1|^{\mu_r + 1 - \epsilon}, \quad \text{where } \mu_r := \tfrac{1}{2}\log_2 T_r(\tfrac{3}{4}).$$

Proof. Lemma 4.3 implies that $\ell^2(\omega_1) = \big(|A_0(e^{i\omega_1})| + |B_0(e^{i\omega_1})|\big)^2 \leq 4\,T_r(\sin^2 \frac{\omega_1}{2})$ and that equality holds only when $\omega_1 \equiv \pi \pmod{2\pi}$. By combining these statements with Lemma 4.4, we obtain the following strict inequalities:

$$\ell^2(\omega_1) < 4\,T_r\big(\tfrac{3}{4}\big) \quad \text{if } |\omega_1| \leq \tfrac{2\pi}{3},$$
$$\ell^2(\omega_1)\ell^2(2\omega_1) < 4^2\,T_r^2\big(\tfrac{3}{4}\big) \quad \text{if } \tfrac{2\pi}{3} < |\omega_1| \leq \pi.$$

Since ℓ and T_r are continuous, we can preserve these inequalities even after we replace 4 by $(4-\delta)$ for some small $\delta > 0$. By Theorem 2.3 of Cohen and Daubechies [6, p. 69], there exists a constant $c > 0$ such that

$$\prod_{j=1}^{\infty} \ell(2^{-j}\omega_1) \leq c \cdot |\omega_1|^b,$$

where $b := \frac{1}{2}\log_2\big((4-\delta)T_r(\frac{3}{4})\big)$. Observe that $b = 1 - \epsilon + \mu_r$ for some $\epsilon > 0$. ∎

Corollary 4.6. *Let* $W = M^{\mathrm{T}}$. *There exist constants* $c > 0$ *and* $\epsilon > 0$ *such that*

$$\left|\prod_{j=1}^{\infty} q(W^{-j}\omega)\right| \leq c \cdot |\omega_1|^{\mu_r+1-\epsilon}\,|\omega_2|^{\mu_r+1-\epsilon},$$

where $\mu_r := \frac{1}{2}\log_2 T_r(\frac{3}{4})$, *and* $q(\omega)$ *is defined in* (4.4).

Proof. Note that $W(\omega_1,\omega_2)=(\omega_2,2\omega_1)$. Therefore, (4.2) and (4.4) yield

$$\begin{aligned}\left|\prod_{j=1}^{\infty} q(W^{-j}\omega)\right| &= \prod_{j=1}^{\infty}\left|q(2^{-j}\omega)\right| \prod_{j=1}^{\infty}\left|q(2^{-j}W\omega)\right| \\ &\le \prod_{j=1}^{\infty}\ell(2^{-j}\omega_1)\prod_{j=1}^{\infty}\ell(2^{-j}\omega_2) \\ &\le c\cdot|\omega_1|^{\mu_r+1-\epsilon}\,|\omega_2|^{\mu_r+1-\epsilon}. \qquad \square\end{aligned}$$

The following lemma establishes a simple fact about the zeros of $m(\omega_1,\omega_2)$:

LEMMA 4.7. *Suppose that* $m(\omega_1,\omega_2)=0$. *Then* $\omega_1\equiv\pi \pmod{2\pi}$.

Proof. If $m(\omega_1,\omega_2)=0$, then $|A(e^{i\omega_1})|=|B(e^{i\omega_1})|$, and therefore $\mathcal{P}_A(z)=\mathcal{P}_B(z)$ for some $z=e^{i\omega_1}$. Thus, either $H(z)=0$, and therefore $z=-1$, or $\mathcal{P}_{A_0}(z)=\mathcal{P}_{B_0}(z)$; that is,

$$\left(1-\mathcal{P}_H^r(-z)\mathcal{P}_H^r(z)L^2(-1)\right)\mathcal{P}_L(z)=\mathcal{P}_H^{2r}(-z)\mathcal{P}_L(-z)L^2(-1),$$

and therefore

$$\mathcal{P}_L(z)=\mathcal{P}_H^r(-z)\mathcal{P}_H^r(z)\mathcal{P}_L(z)L^2(-1)+\mathcal{P}_H^r(-z)\mathcal{P}_H^r(-z)\mathcal{P}_L(-z)L^2(-1).$$

Due to (3.6),

$$\mathcal{P}_L(z)=\mathcal{P}_H^r(-z)L^2(-1).$$

By Lemma 4.1, this can happen only when $\omega_1\equiv\pi \pmod{2\pi}$. $\square$

Proof of Theorem 2.2.

(i) Since the conditions for accuracy (3.2)–(3.3) and orthogonality (2.3) hold by Theorem 2.1, to ensure that ϕ is an orthogonal scaling function, we need only verify Cohen's criterion by finding a compact fundamental domain Ω of the lattice $2\pi\mathbb{Z}^2$ with the property

$$m(W^{-j}\omega)\neq 0 \quad \text{for all } j\ge 1 \text{ and for all } \omega\in\Omega. \tag{4.6}$$

Let $\Omega=[-\pi,\pi]^2$. Clearly, Ω is a compact fundamental domain of $2\pi\mathbb{Z}^2$. Observe that $W^{-1}\Omega=[-\frac{\pi}{2},\frac{\pi}{2}]\times[-\pi,\pi]$. Therefore, Lemma 4.7 guarantees that $m(\omega)\neq 0$ for $\omega\in W^{-1}\Omega$. Finally, $W^{-j-1}\Omega\subset W^{-j}\Omega$ for all $j\ge 1$, which proves (4.6).

(ii) Berger and Wang [3] showed that the support of the solution to the dilation equation (2.1) is the attractor of the iterated function system (IFS)

$$\left\{f_{\mathbf{n}}(\mathbf{x}):=M^{-1}(\mathbf{x}+\mathbf{n}) \;:\; c_{\mathbf{n}}\neq 0\right\}.$$

If $\operatorname{supp} c=[0,N_x]\times[0,N_y]$, it can be checked that $\operatorname{supp}\phi\subseteq[0,N_x+2N_y]\times[0,N_x+N_y]$. For all ϕ in Φ_{r+1}, we have $N_x=4r-1$ and $N_y=1$, which proves part (ii).

(iii) Define $m_1(\omega):=m(\omega)\,m(W^{-1}\omega)$. By (4.1), we have $\widehat{\phi}(\omega)=m_1(\omega/2)\,\widehat{\phi}(\omega/2)$. Assume that $\phi(\mathbf{x})$ is separable. Then $\widehat{\phi}(\omega)$ and $m_1(\omega/2)$ are separable. This is impossible, because the coefficient mask $C(z,w)$ is nonseparable by (2.6)–(2.7).

(iv) Let $\mu_r:=\frac{1}{2}\log_2 T_r(\frac{3}{4})$. Combining (4.1)–(4.3) with Corollary 4.6, we obtain the following estimate for the decay of the Fourier transform of any ϕ in Φ_{r+1}:

$$|\widehat{\phi}(\omega)|\le c\cdot(1+|\omega_1|)^{\mu_r+1-r-\epsilon}(1+|\omega_2|)^{\mu_r+1-r-\epsilon}$$

for some constants $c > 0$ and $\epsilon > 0$. As a result, $\phi \in C^{r-\mu_r-2}(\mathbb{R}^2)$. Lemma 4.4 provides the asymptotics for $r - \mu_r$. ∎

Remark. Observe that, aside from the claim on the size of supp ϕ in part (ii), in the proof of Theorem 2.2 and Corollary 4.6 we used only the following properties of the dilation matrix $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$:

$$M^2 = \pm 2I \quad \text{and} \quad M\mathbb{Z}^2 = 2\mathbb{Z} \times \mathbb{Z}. \tag{4.7}$$

Therefore, aside from part (ii), Theorem 2.2 holds for any dilation matrix M satisfying (4.7), such as $M = \left(\begin{smallmatrix} 0 & -2 \\ 1 & 0 \end{smallmatrix}\right)$. Lemma 5.2 and a result by Lagarias and Wang [11] extend Theorem 2.2 to all dilation matrices with $M^2 = \pm 2I$ [2].

5. Examples. In this section we consider a few of the coefficient masks in Theorem 2.1 for $r = 1$, 2, and 6. We plot the corresponding scaling functions and study their smoothness. Finally, we discuss how to modify these masks for other dilation matrices. Additional coefficient values can be found in the appendix of [2].

Devil's Tower. The simplest case of Theorem 2.1 is when $r = 1$ and (2.11) holds. In this case $L(z) = 1$, and the autocorrelation

$$\mathcal{P}_S(z^2) = 1 - \Big(\frac{1-z^2}{4}\Big)\Big(\frac{1-z^{-2}}{4}\Big)$$

has only two spectral factors, which can be computed explicitly:

$$S^{(1)}(z^2) = \frac{2-\sqrt{3}}{4} + \frac{2+\sqrt{3}}{4}z^2 \quad \text{and} \quad S^{(2)}(z^2) = \frac{2+\sqrt{3}}{4} + \frac{2-\sqrt{3}}{4}z^2.$$

The corresponding coefficient masks are given below. For brevity we omit the powers of z and w; the constant term c_{00} is anchored at the lower-left corner:

$$c^{(1)} = \frac{1}{8}\begin{bmatrix} -1 & 1 & 1 & -1 \\ 2-\sqrt{3} & 2-\sqrt{3} & 2+\sqrt{3} & 2+\sqrt{3} \end{bmatrix},$$

$$c^{(2)} = \frac{1}{8}\begin{bmatrix} -1 & 1 & 1 & -1 \\ 2+\sqrt{3} & 2+\sqrt{3} & 2-\sqrt{3} & 2-\sqrt{3} \end{bmatrix}.$$

The corresponding scaling functions $\phi^{(1)}$ and $\phi^{(2)}$ are supported on $[0, 5] \times [0, 4]$, have accuracy 2, and are discontinuous. The mesh plot of $\phi^{(1)}$ resembles the famous Wyoming mountain, as depicted in Figure 5.1.

Continuous scaling function. In the case $r = 2$, four spectral factors of $\mathcal{P}_S$ are combined with two spectral factors of $\mathcal{P}_L$ to produce eight coefficient masks in $\mathcal{C}_{2+1}$ (see Table A.1 in the Appendix). The second mask has the following coefficients:

$$c = \begin{bmatrix} 0.03697 & -0.06403 & -0.05678 & 0.13797 & 0.00265 & -0.08384 & 0.01716 & 0.00991 \\ 0.00790 & -0.01369 & -0.13808 & 0.12118 & 0.54993 & 0.34617 & 0.08025 & 0.04633 \end{bmatrix}. \tag{5.1}$$

The corresponding scaling function (Figure 5.2) has accuracy $2 + 1 = 3$ and is continuous. The continuity does not follow from Theorem 2.2, because $2 - \mu_2 - 2 \approx 1.339 - 2 < 0$. Instead, the continuity of the scaling function in Figure 5.2 follows from Lemma 5.1, which is a straightforward modification of Lemma 7.1.2 by Daubechies [8, p. 217]; the proof is essentially the same and is therefore omitted here.

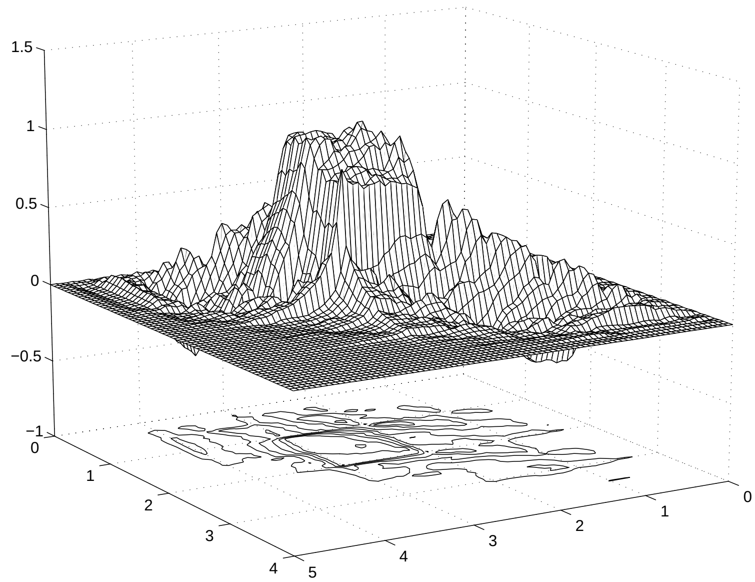


FIG. 5.1. *"Devil's Tower": a discontinuous scaling function with accuracy* $1+1=2$ *and support* $[0,5]\times[0,4]$.

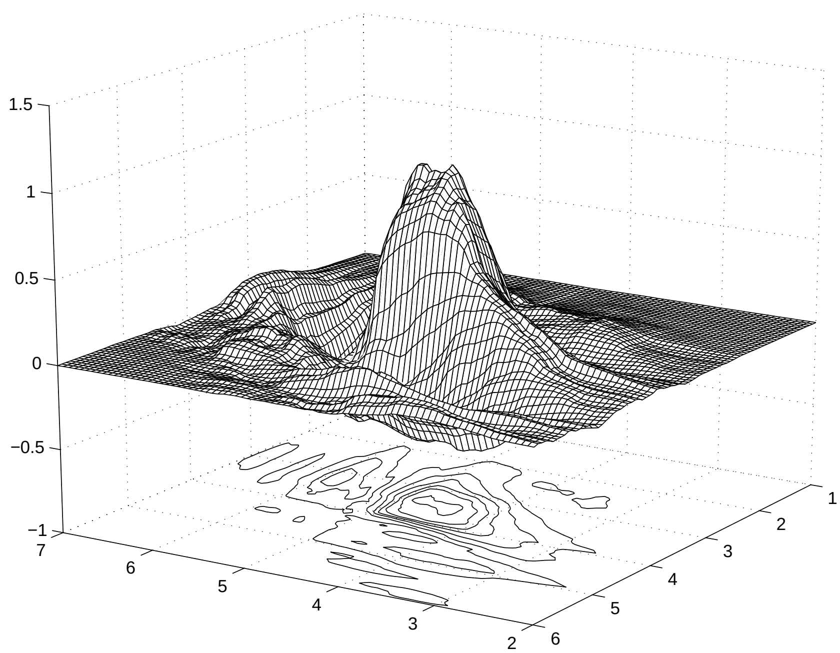


FIG. 5.2. *"Resting Dog": a continuous nonseparable scaling function with accuracy* $2+1=3$ *and support* $[0,9]\times[0,8]$.

LEMMA 5.1. *Let $W = M^{\mathrm{T}}$. Suppose that there exist $p \in \mathbb{N}$ and $\lambda > 0$ such that*

$$\sup_{\omega} \prod_{j=0}^{p-1} \left| q(2^{-j}\omega)\, q(2^{-j}W^{-1}\omega) \right| < 2^{\lambda p}, \tag{5.2}$$

where $q(\omega)$ is defined in (4.4). *Then there exist constants c' and c'' such that*

$$\prod_{j=1}^{\infty} \left| q(W^{-j}\omega) \right| < c' \cdot (1+|\omega|)^{\lambda} < c'' \cdot (1+|\omega_1|)^{\lambda}(1+|\omega_2|)^{\lambda}.$$

Hence $\phi \in C^{r-1-\lambda}$. □

The authors evaluated the trigonometric polynomial in (5.2) numerically for 120 values of ω within a single period and verified that the coefficient mask (5.1) satisfies (5.2) for $p = 4$ and $\lambda = 1$. Therefore, the scaling function in Figure 5.2 is continuous.

While Theorem 2.2 claims only that $\Phi_{5+1} \subset C^0$, it can be verified numerically that all masks in $\mathcal{C}_{3+1}$ (see Table A.3 in the Appendix) satisfy (5.2) for $p = 4$ and $\lambda = 2$; hence $\Phi_{3+1} \subset C^0$.

Differentiable scaling functions. About half of the 128 coefficient masks in $\mathcal{C}_{6+1}$ satisfy (5.2) for $p = 4$ and $\lambda = 4$ and therefore produce differentiable scaling functions. To the best of the authors' knowledge, these are the first nonseparable orthogonal scaling functions (for a dilation matrix with $|\det(M)| = 2$) that are differentiable. The 2×24 scaling coefficients of the differentiable scaling function in Figure 5.3 are displayed in Table A.2 in the Appendix. Numerical computations confirm also that $\Phi_{7+1} \subset C^1$.

Other dilations. The following lemma explains what happens when we manipulate the coefficient mask in the dilation equation (2.1). Let J be any lattice-preserving matrix, that is, any matrix J such that both J and J^{-1} have integer entries. An example is $J = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. For any function $f : \mathbb{R}^2 \to \mathbb{R}$ define its "upsampling by J" as $(J\!\uparrow\! f)(\mathbf{x}) := f(J^{-1}\mathbf{x})$. Similarly, define the "upsampled by J" scaling coefficients as $(J\!\uparrow\! c)_{\mathbf{n}} := c_{J^{-1}\mathbf{n}}$.

LEMMA 5.2. *Let ϕ be a solution to the dilation equation* (2.1) *with coefficients c and dilation matrix M. Then $J\!\uparrow\!\phi$ is a solution to the dilation equation* (2.1) *with coefficients $J\!\uparrow\! c$ and dilation matrix JMJ^{-1}.*

Proof. Make the substitution $\mathbf{x} \mapsto J^{-1}\mathbf{y}$ in the dilation equation. □

A surprising example is $J = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, the matrix of the reflection about the y-axis. Let $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$. Then $JMJ^{-1} = -M$. Therefore, the scaling function with coefficient mask $w^{-1}(B(z) + wA(z))$ and dilation M is a reflection of the scaling function with coefficient mask $A(z) + wB(z)$ and dilation $-M$, but it is *not* a reflection of the scaling function with coefficient mask $A(z) + wB(z)$ and dilation M. The difference is obvious in Figure 5.4, which shows the contour plots of two scaling functions with the same dilation matrix $M = \left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ and x-reflected coefficients. (The one on the left was plotted in Figure 5.2.)

The plots exhibit a peculiar feature: all scaling functions appear so far to be "almost symmetric" with respect to the bisectrix $x = y$. This is not surprising, since the dilation matrix $\left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ swaps the axes and stretches along x by a factor of two. Unfortunately, the case $|\det(M)| = 2$ is similar to the univariate case $M = 2$—the only symmetric orthogonal scaling function is the Haar function.

Lemma 5.2 allows us to adapt the coefficient masks described in Theorem 2.1 to dilation matrices that generate the quincunx sublattice in the following way.

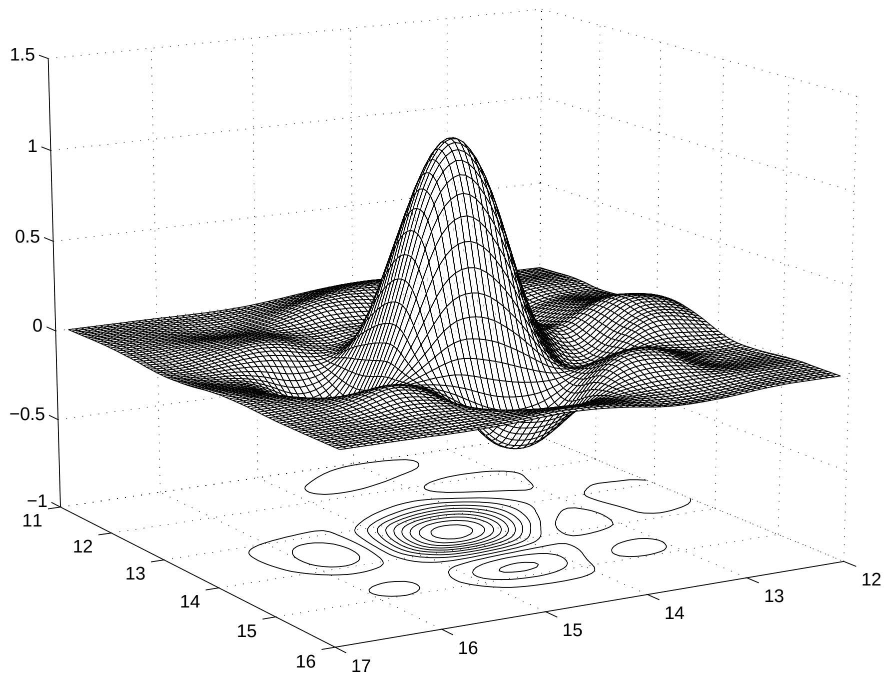


FIG. 5.3. *A differentiable nonseparable scaling function with accuracy* $6+1=7$.

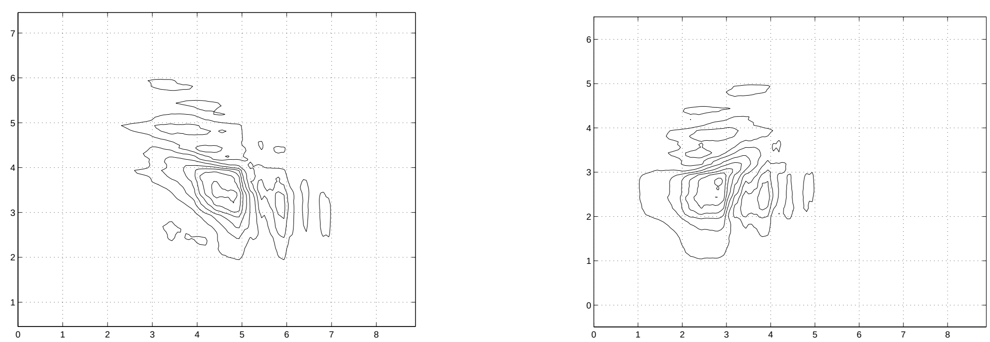


FIG. 5.4. *Level sets of scaling functions with* x*-reflected scaling coefficients: the functions are not similar.*

COROLLARY 5.3. *Let* ν *be odd. If* $A(z)+wB(z)$ *is a two-row orthogonal coefficient mask with accuracy* r *for the column lattice, then* $A(z)+z^\nu wB(z)$ *is a two-row orthogonal coefficient mask with accuracy* r *for the quincunx lattice.*

Proof. Set $M=\left(\begin{smallmatrix}0&2\\1&0\end{smallmatrix}\right)$ and $J=\left(\begin{smallmatrix}1&\nu\\0&1\end{smallmatrix}\right)$ in Lemma 5.2. Observe that the J-upsampling of $A(z)+wB(z)$ is $A(z)+z^\nu wB(z)$ and that the matrix $J\left(\begin{smallmatrix}0&2\\1&0\end{smallmatrix}\right)J^{-1}$ generates the quincunx sublattice. ∎

If we shift the top row of coefficients in (5.1) one position to the left and use the dilation matrix $\left(\begin{smallmatrix}1&-1\\1&1\end{smallmatrix}\right)$ in (2.1), then we obtain an orthogonal nonseparable quincunx scaling function with accuracy $2+1=3$, plotted in Figure 5.5 (cf. Kovačević and Vetterli's scaling function with accuracy 2 [12]).

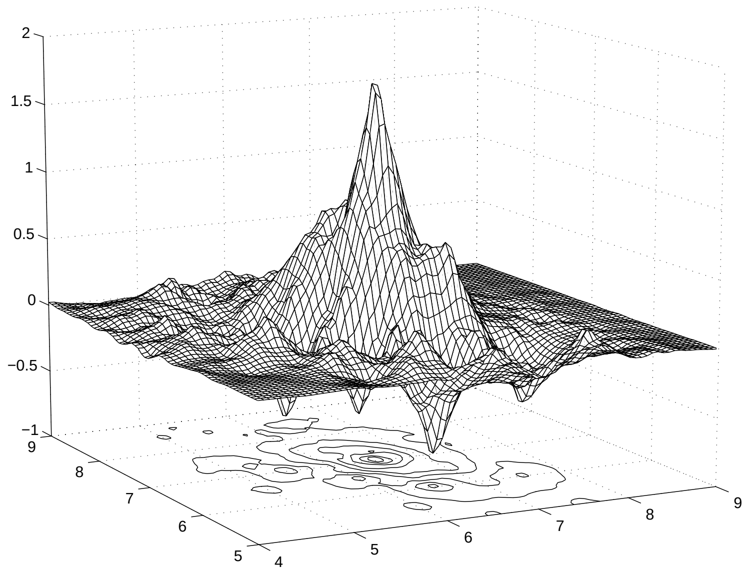


FIG. 5.5. *An orthogonal nonseparable scaling function with accuracy* $2+1=3$ *for the quincunx dilation matrix.*

Every 2×2 dilation matrix M with $M^2 = 2I$ can be factored in the form $J\left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)J^{-1}$ [11]. Therefore, Lemma 5.2 and Theorem 2.2 provide arbitrarily smooth nonseparable wavelet bases for such dilation matrices.

The convolution of two sets of scaling coefficients $a_{\mathbf{n}}$ and $b_{\mathbf{n}}$ is defined by $(a*b)_{\mathbf{n}} := \sum_{\mathbf{k}\in\mathbb{Z}^2} a_{\mathbf{k}}\, b_{\mathbf{n}-\mathbf{k}}$. The following statement combines two levels of the MRA into one.

LEMMA 5.4. *Let* ϕ *be a solution to the dilation equation* (2.1) *with coefficients* c *and dilation matrix* M*. Then* ϕ *is also a solution to the dilation equation* (2.1) *with coefficients* $c * (M \uparrow c)$ *and dilation matrix* M^2*.*

Proof. Iterate the dilation equation (2.1); that is, replace each ϕ on the right by the sum of its dilates. ∎

Since $\left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)^2 = 2I$, by Theorem 2.2 and Lemma 5.4, we obtain the following.

COROLLARY 5.5. *Aside from the separable Daubechies MRA, there exists a bivariate nonseparable orthogonal MRA of any given smoothness for the dilation* $\left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right)$*.*

Ayache [1] obtained this result independently by perturbing the separable Daubechies basis.

6. Summary. We showed how to obtain orthogonal two-row coefficient masks (low-pass filter coefficients) with arbitrarily high accuracy for dilation matrices M with $|\det(M)| = 2$, such as $\left(\begin{smallmatrix} 0 & 2 \\ 1 & 0 \end{smallmatrix}\right)$ or $\left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$. We proved that if $M^2 = \pm 2I$, then the smoothness of the scaling functions corresponding to those coefficient masks increases asymptotically with the accuracy and can be made arbitrarily high.

Appendix. Scaling coefficients. The following tables contain the coefficients of $A(z)$ and $B(z)$ that satisfy the conditions (2.6)–(2.11) in Theorem 2.1, that is, the coefficients of $C(z,w)$ in $\mathcal{C}_{r+1}$ for some values of r. Generally speaking, the scaling coefficients in the beginning of each table correspond to the "least symmetric" scaling functions; the scaling coefficients towards the end of the table correspond to the "most symmetric" scaling functions.

Table A.1
Scaling coefficients (first half of $\mathcal{C}_{2+1}$) of accuracy $2+1$.

n	a_n	b_n	a_n	b_n
	Solution 1		Solution 2	
0	0.001113697631	0.036969146934	0.007903570868	0.036969146934
1	-0.001928980881	-0.064032440802	-0.013689386305	-0.064032440802
2	-0.011710875083	-0.056780853066	-0.138084015734	-0.056780853066
3	0.003658326071	0.137970734671	0.121182418797	0.137970734671
4	-0.058943362439	0.002654265329	0.549926093940	0.002654265329
5	0.169446270789	-0.083844146934	0.346172096397	-0.083844146934
6	0.569540539891	0.017157440802	0.080254350926	0.017157440802
7	0.328824384020	0.009905853066	0.046334871111	0.009905853066
	Solution 3		Solution 4	
0	-0.012415391296	0.036969146934	-0.088108228150	0.036969146934
1	0.021504088520	-0.064032440802	0.152607927721	-0.064032440802
2	-0.006740179593	-0.056780853066	0.565028719464	-0.056780853066
3	0.197013817950	0.137970734671	0.336639086235	0.137970734671
4	0.570245056104	0.002654265329	0.030278563342	0.002654265329
5	0.310978621572	-0.083844146934	0.014909362188	-0.083844146934
6	-0.051089485215	0.017157440802	-0.007199054655	0.017157440802
7	-0.029496528042	0.009905853066	-0.004156376143	0.009905853066

Table A.2
The scaling coefficients of the differentiable scaling function in Figure 5.3.

n	a_n	b_n
0	0.000143412339	0.000082104697
1	-0.000323736348	-0.000185341617
2	-0.001920307470	-0.000431083617
3	0.003595546890	0.000549853896
4	0.008522556302	0.003143374464
5	-0.004408987885	-0.002050950686
6	-0.013081025520	-0.013941778594
7	-0.060122740069	0.012068826968
8	-0.098853166909	0.031125316212
9	0.162212221109	-0.036965110063
10	0.525503586261	-0.036331876410
11	0.406747601563	0.060643179987
12	0.037817366952	0.019734738252
13	-0.036595246129	-0.057961647081
14	0.043460344536	0.000210716144
15	0.031894654283	0.033234701933
16	-0.002789322971	-0.006059880986
17	-0.003862049736	-0.011382637768
18	0.001292258468	0.003005793635
19	0.000931477590	0.002333217232
20	-0.000121744397	-0.000582912091
21	-0.000080277824	-0.000304243713
22	0.000026042408	0.000045488295
23	0.000011536556	0.000020150912

Table A.3

Scaling coefficients (first half of $\mathcal{C}_{3+1}$) of accuracy $3+1$.

n	a_n	b_n	a_n	b_n
	Solution 1		Solution 2	
0	0.000062409389	-0.011623030789	0.000326746845	-0.011623030789
1	-0.000151373802	0.028191629145	-0.000792523573	0.028191629145
2	-0.000616099021	0.018801633679	-0.004873897732	0.018801633679
3	0.001728930996	-0.089291978349	0.013049802109	-0.089291978349
4	0.003825989061	0.016318478955	0.044932505092	0.016318478955
5	-0.006639890465	0.099956916164	-0.101357239374	0.099956916164
6	0.011989717589	-0.045534931049	-0.168647288127	-0.045534931049
7	-0.050002611258	-0.046035169850	0.262229312212	-0.046035169850
8	-0.071220291676	0.025023044461	0.522072641289	0.025023044461
9	0.325850800478	0.008409358878	0.283090226735	0.008409358878
10	0.555958274659	-0.002985195258	0.106189292633	-0.002985195258
11	0.229214144050	-0.001230755988	0.043780421891	-0.001230755988
	Solution 3		Solution 4	
0	0.004635883479	-0.011623030789	0.024271352757	-0.011623030789
1	-0.011244322602	0.028191629145	-0.058870099219	0.028191629145
2	0.005799855170	0.018801633679	-0.092072378476	0.018801633679
3	0.003357833887	-0.089291978349	0.314552501801	-0.089291978349
4	0.012240306723	0.016318478955	0.551933076802	0.016318478955
5	0.360239746045	0.099956916164	0.242555222384	0.099956916164
6	0.554047592739	-0.045534931049	0.023310939204	-0.045534931049
7	0.174496893426	-0.046035169850	0.003917651739	-0.046035169850
8	-0.084208083428	0.025023044461	-0.008872536334	0.025023044461
9	-0.029935887232	0.008409358878	-0.002744659377	0.008409358878
10	0.007484445316	-0.002985195258	0.001429546047	-0.002985195258
11	0.003085736475	-0.001230755988	0.000589382672	-0.001230755988

Note. To save space, we displayed only the *first half* of $\mathcal{C}_{r+1}$. To obtain the other half, simply take the coefficients (both a_n and b_n) in reverse order. Recall that the scaling functions with scaling coefficients in the second half of the family $\mathcal{C}_{r+1}$ are different than, and are not mere flips of, the scaling functions with scaling coefficients in the first (tabulated) half (see the discussion after Lemma 5.2).

Acknowledgments. The authors thank the referees for their useful suggestions, and Albert Cohen for the helpful discussions and for communicating Ayache's work. Special thanks go to Chris Heil, for providing the preprint [4] of the accuracy condition and for carefully revising the early drafts, and to Tim Flaherty for his meticulous work in discovering several inaccuracies in the manuscript. Thanks to Huibin Zhou for his comments. Wang acknowledges the support of the Mathematics Department at Cornell University, where he was a visitor during the completion of this paper. The fruitful blend of theoretical mathematics and computer numerics was essential to our work; the plots and the numerical computations were aided by Mathematica and MATLAB.

REFERENCES

[1] A. Ayache, *Construction of nonseparable dyadic compactly supported orthonormal wavelet bases for $L^2(\mathbb{R}^2)$ of arbitrarily high regularity*, Rev. Mat. Iberoamericana, to appear.

[2] E. Belogay, *Construction of Smooth Orthogonal Wavelets with Compact Support in $\mathbb{R}^d$*, Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA, 1998.

[3] M. A. Berger and Y. Wang, *Multidimensional two-scale dilation equations*, Wavelets: A Tutorial in Theory and Applications, Wavelet Anal. Appl. 2, C. K. Chui, ed., Academic Press, New York, (1992), pp. 295–323.

[4] C. Cabrelli, C. Heil, and U. Molter, *Accuracy of lattice translates of several multidimensional refinable functions*, J. Approx. Theory, 95 (1998), pp. 5–52.

[5] A. Cohen and J. P. Conze, *Régularité des bases d'ondelettes et mesures ergodiques*, Rev. Mat. Iberoamericana, 8 (1992), pp. 351–365.

[6] A. Cohen and I. Daubechies, *Nonseparable bidimensional wavelet bases*, Rev. Mat. Iberoamericana, 9 (1993), pp. 51–137.

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

[8] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics 61, SIAM, Philadelphia, PA, 1992.

[9] K. Gröchenig and W. Madych, *Multiresolution analysis, Haar bases, and self-similar tilings,* IEEE Trans. Inform. Theory, 38 (1992), pp. 558–568.

[10] W. He and M. J. Lai, *Examples of bivariate nonseparable compactly supported orthonormal continuous wavelets*, in Wavelet Applications in Signal and Image Processing IV, Proceedings, SPIE 3169, 1997, pp. 303–314.

[11] J. C. Lagarias and Y. Wang, *Haar type orthonormal wavelet basis in $\mathbb{R}^2$*, J. Fourier Anal. Appl., 2 (1995), pp. 1–14.

[12] J. Kovačević and M. Vetterli, *Nonseparable multidimensional perfect reconstruction filterbanks,* IEEE Trans. Inform. Theory, 38 (1992), pp. 533–555.

[13] S. Mallat, *Multiresolution analysis and wavelets*, Trans. Amer. Math. Soc., 315 (1989), pp. 69–88.

[14] Y. Meyer, *Ondelettes et Opérateurs*, Hermann, Paris, 1990.

[15] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, 1996.

[16] L. Villemoes, *Continuity of nonseparable quincunx wavelets,* Appl. Comput. Harmon. Anal., 1 (1994), pp. 180–187.