Scaling Functions Robust to Translations

Steven A. Benno, *Member, IEEE,* and José M. F. Moura, *Fellow, IEEE*

Abstract— The discrete wavelet transform (DWT) is popular in a wide variety of applications. Its sparse sampling eliminates redundancy in the representation of signals and leads to efficient processing. However, the DWT lacks translation invariance. This makes it ill suited for many problems where the received signal is the superposition of arbitrarily shifted replicas of a transmitted signal as when multipath occurs, for example.

The paper develops algorithms for the design of orthogonal and biorthogonal compact support scaling functions that are robust to translations. Our approach is to maintain the critical sampling of the DWT while designing multiresolution representations for which the coefficient energy redistributes itself mostly within each subband and not across the entire time-scale plane. We obtain expedite algorithms by decoupling the optimization from the constraints on the scaling function. Examples illustrate that the designed scaling function significantly improves the robustness of the representation.

Index Terms— Multiresolution, robust representations, shift-invariant scaling functions, translation invariant, wavelets.

I. INTRODUCTION

THE DISCRETE wavelet transform (DWT) is popular in a wide variety of applications. Its discrete subband decomposition allows for a coarse-to-fine multiresolution analysis (MRA) of signals, and its sparse critically sampled dyadic grid is computationally more efficient to compute than the fast Fourier transform (FFT). It is precisely this sparse grid, however, that causes the DWT to fail to be invariant even to translations of the input by multiples of the sampling period. As the input is translated in time, coefficient energy from one subband of the DWT escapes into other subbands, even though the spectral content of the signal does not change.

Strang [1] has commented on the DWT's lack of invariance as a major drawback in using orthogonal wavelet transforms in pattern recognition applications. Several approaches have been developed to address this problem. One approach is to use the translation-invariant continuous wavelet transform [2], but this representation is highly redundant and computationally expensive. Another approach is to compute the dyadic wavelet transform at the full sampling density and to extract a subset of samples from the overcomplete representation to form an orthonormal representation [3]. In [4] and [5], a library of shifted basis functions is used to achieve translation invariance. These approaches, however, are translation invariant only

when the entire input signal is uniformly shifted. If there are components within the input signal that are translated relative to each other, these representations do not provide a superposition of translated wavelet coefficients.

In [6] and [7], the authors present a method for choosing wavelets that minimize the worst-case approximation error for representing all signals in a class at some prescribed scale. Specifically, they consider the frequency domain L^p class. To obtain more explicit results, a crucial assumption made in [6] and [7] is that the signals being analyzed are bandlimited. The paradigm of our work is related to [6] and [7] in that we are also minimizing the approximation error in representing a class of functions. In our case, however, we consider the class of functions generated by the translations of a mother function. We address the issue of translation invariance explicitly. In a sense, as we look for robustness to shifts of the same signal, our problem is more restricted than the more general problem studied in [7]. However, by considering this more restricted framework and attacking directly the shift invariance, we do not require the bandlimited assumption to get explicit design algorithms. Indeed, bandlimited constraints represent special cases of our work and are discussed in the Appendix. We obtain explicit algorithms by minimizing a translation error measure at a midpoint.

Our approach to the translation-invariance problem is to apply the concept of *shiftability* to the DWT. A signal g is *shiftable* [8] if and only if, given τ, there exists a set of functions $a_k(\tau)$ such that

$$\forall \tau \in \mathbb{R}: \quad g(t-\tau) = \sum_k a_k(\tau)g(t-k). \tag{1}$$

Note that the coefficients $a_k(\tau)$ are a function of τ and are computed by inner products with $\widehat{g}_k(t)$, which is a dual function of $g(t)$, so that $a_k(\tau) = \langle g(t-\tau), \widehat{g}_k(t) \rangle$, where $\langle \cdot, \cdot \rangle$ is the inner product. In other words, a function $g(t)$ is said to be shiftable if an arbitrary shift of $g(t)$ by τ can be represented by a linear superposition of its integer shifts $\{g(t-k)\}_{k \in \mathbb{Z}}$. When (1) holds, the set of functions $\{g(t-k)\}_{k \in \mathbb{Z}}$ generates a *shiftable representation*.

By designing a shiftable MRA, the efficient sparseness of the critically sampled dyadic grid is maintained, whereas the coefficient energy in each subband is invariant to translations of the input signal. It is to say that the *subspaces* generated by the MRA are translation invariant, e.g., if $h(t) \in V_0$, then for $\tau \in \mathbb{R}$, $h(t-\tau) \in V_0$. Furthermore, using a shiftable MRA allows us to interpolate between grid points within a subband by interpolating with grid points within that subband alone. In other words, interpolating in time between grid points is accomplished with a one-dimensional (1-D) interpolation

instead of a two-dimensional (2–D) interpolation in the time-scale plane.

In general, however, (1) will not be satisfied, i.e., a function will not be shiftable. For example, it is not possible for a function to be shiftable and to have compact support [6]. The goal of shiftable MRA's is realized under appropriate bandlimited conditions. We discuss these conditions, and, for the cases where these conditions are violated, we relax the hard constraint of shiftability and consider instead the design of *robust* signals for which, in (1), we have approximate equality.

The goal becomes the design of signals g for which the mean square error in representing their arbitrary shifts by its integer translates is small, i.e., for small arbitrarily chosen $\delta > 0$, *robust* signals satisfy

$$\sup_{0 \leq \tau \leq 1} \mathcal{E}(\tau) = \sup_{0 \leq \tau \leq 1} \left\| g(t - \tau) - \sum_k a_k(\tau) g(t - k) \right\|^2 \leq \delta. \tag{2}$$

This translation error is a measure of the representation's *robustness* to continuous translations. The design of *scaling* functions that minimize $\mathcal{E}(\tau)$ is a constrained minimization, for example, scaling functions satisfy the two-scale equation constraint [9]. This constrained optimization problem is, in general, difficult. To obtain explicit algorithms for designing *robust* orthonormal and biorthogonal scaling functions, rather than solving directly the problem in (2), we minimize $\mathcal{E}(\frac{1}{2})$, i.e., $\mathcal{E}(\tau)$ at $\tau = 1/2$. In our experience, and as also reported in the literature [3], [10], with many lowpass signals and many scaling functions, $\mathcal{E}(\frac{1}{2})$ is the maximum of $\mathcal{E}(\tau)$ for $0 \leq \tau \leq 1$ so that in these cases, minimizing $\mathcal{E}(\frac{1}{2})$ does solve (2). In general, we need to check that indeed, this holds; when it does not, our algorithms only design the scaling function that minimizes $\mathcal{E}(\frac{1}{2})$. In the sequel, unless otherwise specified, when we refer to the design of *robust* signals, we refer to the minimization of $\mathcal{E}(\frac{1}{2})$.

We will consider the design of robust scaling functions (usually represented in the literature by the symbol ϕ), namely, the design of robust orthonormal and robust biorthogonal scaling functions that minimize $\mathcal{E}(\frac{1}{2})$. In the orthonormal case, we seek functions $\phi(t)$ that are as robust as possible with respect to (w.r.t.) $\mathcal{E}(\frac{1}{2})$ while satisfying the constraints of orthonormality and the two-scale equation. By imposing the additional constraint of compact support, we present an elegant strategy to take advantage of the parameterization by Zou and Tewfik [11] to find that scaling function whose corresponding translation error $\mathcal{E}(\frac{1}{2})$ is smaller than any other for a given length of support. In this sense, our algorithm finds the orthonormal scaling function with compact support that is optimally robust with respect to (w.r.t.) $\mathcal{E}(\frac{1}{2})$.

In the biorthogonal case, we parallel the orthogonal case but exchange the orthonormality constraint for the less strict biorthogonality constraint. Once again, we utilize a characterization of biorthogonal scaling functions with compact support [12] to find the pairs $\phi(t)$ and $\tilde{\phi}(t)$ that are optimally robust wrt $\mathcal{E}(\frac{1}{2})$. We show that the extra freedom gained by going from the orthonormal case to the biorthogonal case leads to more robust representations.

We note that in this paper, the *length of support* of the scaling functions is kept fixed. In [13]–[15], we design robust representations with an arbitrarily small translation error $\mathcal{E}(\tau)$ at the expense of *increasing the support* of the scaling functions.

We now provide a brief summary of the paper. In Section II, we formally describe the problem of translation invariance and robust representations. The section presents a general expression for $\mathcal{E}(\tau)$ that is valid for unconstrained $g(t)$. In the Appendix, we gain insight into translation invariance by looking at special cases that involve bandlimited assumptions. The general translation error expression presented in Section II is used as the basis for designing robust representations in the remainder of the paper. In Sections III and IV, we apply the notion of robust representations to the DWT for the orthogonal case and biorthogonal case, respectively. For each case, we present an algorithm for designing robust scaling functions that are optimal w.r.t. $\mathcal{E}(\frac{1}{2})$. Finally, Section V concludes the paper.

II. ROBUST REPRESENTATIONS

We develop the notions of shiftability and of robust representations. In the sequel, we will deal with $\mathbb{Z}$, which is the set of integers, $L^2(\mathbb{R})$, which is the space of finite energy functions defined on the real line, and $\ell^2(\mathbb{Z})$, which is the set of finite square summable sequences $(f_k \colon k \in \mathbb{Z})$ indexed by the integers.

1) Representation Subspace: Let $g \in L^2(\mathbb{R})$. We associate with this g the following subspace $\mathcal{G}$ that we refer to as its *representation* subspace. The integer shifted versions of g are written as $g(\cdot - k)$ to emphasize that the integer k is fixed. The representation subspace $\mathcal{G}$ is the closed linear span of all integer shifted replicas of g, i.e.,

$$\mathcal{G} = \overline{\text{span}\{g(\cdot - k)\}} \tag{3}$$

$$= \left\{ f \colon f = \sum_k f_k g(\cdot - k) \text{ with the sequence} \right.$$

$$\left. (f_k \colon k \in \mathbb{Z}) \in \ell^2(\mathbb{Z}) \right\}. \tag{4}$$

In (3) and (4), we have used the common notation of omitting the limits of the summation when summing over the set of integers $\mathbb{Z}$.

When g is a scaling function ϕ, in multiresolution analysis (MRA) parlance, $\mathcal{G}$ is the central space of the MRA usually represented by V_0.

Clearly, the subspace $\mathcal{G}$ is invariant to integer translations, i.e.,

$$\text{if } f \in \mathcal{G}, \qquad \text{then } \forall k \in \mathbb{Z}, \ f(\cdot - k) \in \mathcal{G}.$$

However, for general $g \in L^2(\mathbb{R})$ and arbitrary real valued noninteger $\tau \in \mathbb{R}$, we will usually have that the arbitrarily shifted replica $g(\cdot - \tau) \notin \mathcal{G}$. When, for arbitrary real valued τ, $g(\cdot - \tau) \in \mathcal{G}$, the function g is said to be shiftable. The next paragraph introduces more formally the notion of shiftability.

2) Shiftability: A function $g \in L^2(\mathbb{R})$ is shiftable [8] if and only if

$$\forall \tau \in \mathbb{R}, \qquad g(t - \tau) = \sum_k a_k(\tau) g(t - k). \qquad (5)$$

Equation (5) states that the arbitrarily shifted replica $g(\cdot - \tau)$ is in the representation subspace $\mathcal{G}$.

We note that the integer translates of g span the $\mathcal{G}$ but are not necessarily orthogonal. Therefore, the coefficients $a_k(\tau)$ in the expansion (5) are usually computed with the help of a biorthogonal function $\tilde{g}$ [9], which is associated with g. In particular, $a_k(\tau) = \langle g_\tau, \tilde{g}_k \rangle$, where $\langle \cdot, \cdot \rangle$ is the inner product defined in $\mathcal{G}$, $g_\tau(\cdot) = g(\cdot - \tau)$, and $\tilde{g}_k(\cdot) = \tilde{g}(\cdot - k)$ is the kth translate of the biorthoghonal $\tilde{g}$.

In summary, the definition in (5) states that a function is shiftable iff its representation subspace $\mathcal{G}$ is invariant to *arbitrary* real valued translations $\tau \in \mathbb{R}$, i.e.,

for any $f \in \mathcal{G}$ and for any $\tau \in \mathbb{R}$: $f(\cdot - \tau) \in \mathcal{G}$.

3) Robust Representations: There may be extra requirements on g besides finite energy that prevent (5) from being satisfied with equality, i.e., g from being shiftable. We introduce the *translation error*

$$\mathcal{E}(\tau) = \left\| g(t - \tau) - \sum_k a_k g(t - k) \right\|^2 \qquad (6)$$

where the explicit dependence of a_k on τ has been dropped for convenience. When this mean square error is bounded above by a small threshold $\delta > 0$, the representation is robust to translations. A reasonable design criterion is then to minimize $\sup_{0 \leq \tau \leq 1} \mathcal{E}(\tau)$. This, however, does not lead in general to explicit algorithms without introducing additional constraints on the function g, such as restricting g to be bandlimited; see the Appendix as well as [6] and [7]. In order to obtain explicit algorithms for the design of robust orthonormal and biorthogonal scaling functions, we consider a more limited version of this problem. More precisely, the explicit algorithms developed in the next two sections design scaling functions ϕ that minimize $\mathcal{E}(\frac{1}{2})$. For many lowpass functions and many scaling functions, we have experimentally observed, and other authors have noted [3], [10], that the minimum mean square translation error $\mathcal{E}(\tau)$ is maximum at $\tau = 1/2$. In other words, the leakage of the signal energy to higher scale levels is a maximum when a signal that is an element of the MRA subspace $V_0 = \mathcal{G}$ is shifted by 1/2. Thus, for classes of scaling functions for which the maximum of $\mathcal{E}(\tau)$ does occur at 1/2, it is sufficient to work with $\mathcal{E}(\frac{1}{2})$ to characterize the behavior of the scaling functions under fractional translations, and our algorithms do lead to the *optimal* robust scaling function that minimizes the supremum of $\mathcal{E}(\tau)$. If, for the class of scaling functions, this is not true, then our algorithms only provide the scaling function that minimizes $\mathcal{E}(\frac{1}{2})$. In practice, once we have designed a ϕ, we plot the value of $\mathcal{E}(\tau)$ for that ϕ and verify that the maximum of $\mathcal{E}(\tau)$ for $0 \leq \tau \leq 1$ stays below a desired threshold δ. In work published elsewhere, we consider the design of $\mathcal{G}$ under different setups [13]–[15] and

different metrics, in particular the gap metric [16], [17] or its generalizations [18].

Before we address the problem of designing robust representations, we compute, in the next subsection, an alternative form for the translation error, which will be useful in later sections.

A. Translation Error

We rewrite the translation error given by (6) in the Fourier domain

$$\mathcal{E}(\tau) = \left\| G(f) e^{-j2\pi f \tau} - \left(\sum_k a_k e^{-j2\pi fk} \right) G(f) \right\|^2 \qquad (7)$$

where G is the Fourier transform of g. If the integer translates $\{ g(t - k) \}_{k \in \mathbb{Z}}$ are not an orthonormal basis of $\mathcal{G}$, it is well known that the coefficients $\{ a_k \}$ of the representation that minimizes the translation error are computed by inner products with the integer translates of a dual function $\tilde{g}$ of g, i.e.,

$$a_k = \int_{-\infty}^{\infty} g(t - \tau) \tilde{g}(t - k) \, dt \qquad (8)$$
$$= \langle T_\tau g, T_k \tilde{g} \rangle \qquad (9)$$
$$= \langle E_\tau G, E_k \tilde{G} \rangle. \qquad (10)$$

In (9), T_τ is the translation operator $T_\tau g = g(t - \tau)$. In (10), $\tilde{G}$ is the Fourier transform of the dual $\tilde{g}$, , and E_τ is the modulation operator $E_\tau G = G(f) e^{-j2\pi f \tau}$. Writing (10) explicitly, and then invoking the Poisson summation formula, we obtain successively

$$\sum_k a_k e^{-j2\pi fk}$$
$$= \int_{-\infty}^{+\infty} G(\omega) \overline{\tilde{G}(\omega)} e^{-j2\pi \omega \tau} \left(\sum_k e^{-j2\pi(f - \omega)k} \right) d\omega$$
$$= \int_{-\infty}^{+\infty} G(\omega) \overline{\tilde{G}(\omega)} e^{-j2\pi \omega \tau} \left(\sum_k \delta(f - \omega + k) \right) d\omega$$
$$= e^{-j2\pi f \tau} \sum_k G(f + k) \overline{\tilde{G}(f + k)} e^{-j2\pi \tau k}. \qquad (11)$$

In (11), the sum on the right-hand side is identified as the Zak transform $Z_{G\overline{\tilde{G}}}(f, \tau)$ of the product $G(f) \overline{\tilde{G}(f)}$. In general, the Zak transform of a function $h(x)$ is given by (see [19])

$$Z_h(x, y) = \sum_k h(x + k) e^{-j2\pi yk}. \qquad (12)$$

A similar object is also referred to as the Weil–Brezin mapping [20].

Substituting (11) and (12) into (7), we obtain, after some manipulation

$$\mathcal{E}(\tau) = \|g\|^2 + \int_0^1 \left| Z_{G\overline{\tilde{G}}}(f, \tau) \right|^2 Z_{|G|^2}(f, 0) \, df$$
$$- 2 \int_0^1 \overline{Z_{|G|^2}(f, \tau)} Z_{G\overline{\tilde{G}}}(f, \tau) \, df. \qquad (13)$$

Equation (13) expresses, via the Zak transform, the translation error in terms of the Fourier transforms of the autocorrelation
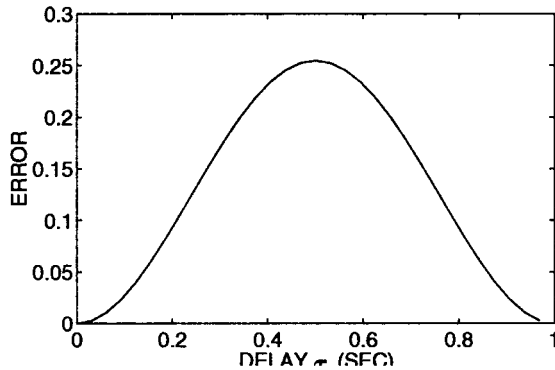
Fig. 1.   Translation error $\mathcal{E}(\tau)$ for the Daubechies D8 scaling function.

function of g and of the cross-correlation of g with its dual $\tilde{g}$. In deriving (13), we made no assumptions regarding the dual except that it exists. Furthermore, we see that the Zak transform arises naturally in this error expression.

In the special case where $\{g(t-k)\}_{k \in \mathbb{Z}}$ is orthonormal, $g = \tilde{g}$, and

$$Z_{|G|^2}(f, 0) = \sum_k |G(f+k)|^2 = 1.$$

Equation (13) reduces to

$$\mathcal{E}(\tau) = 1 - \int_0^1 \left| Z_{|G|^2}(f, \tau) \right|^2 df. \qquad (14)$$

Equation (14) shows that for orthogonal g, for each τ, $\mathcal{E}(\tau)$ is given in terms of the integral along the frequency axis of the magnitude square of the Zak transform of the energy spectral density of g. Intuitively, from (14), we see that the error $\mathcal{E}(\tau)$ is smaller when $|Z_{|G|^2}(f, \tau)|$ is less dependent on τ. In [15], we have developed algorithms that iteratively reshape $|Z_{|G|^2}(f, \tau)|$ to design g with small $\mathcal{E}(\tau)$ and present figures that illustrate that $|Z_{|G|^2}(f, \tau)|$ tends to become independent of τ.

Fig. 1 is a plot of the translation error for the Daubechies D8 scaling function. The error is zero when $\tau = 0, 1$, which corresponds to the scaling function being perfectly aligned with the lattice. When the scaling function is shifted halfway between grid points, the translation error attains its maximum value of 0.255, i.e., 25.5% of the coefficient energy escapes the representation subspace $\mathcal{G}$ for D8 when $\tau = 1/2$.

In the next two sections, we reduce this error by appropriately choosing a scaling function that minimizes the translation error $\mathcal{E}(\frac{1}{2})$ for a given filter length. In Section III, we design optimally robust with respect to (w.r.t.) $\mathcal{E}(\frac{1}{2})$ orthonormal scaling functions. In Section IV, we relax the orthonormality condition and look for optimally robust w.r.t. $\mathcal{E}(\frac{1}{2})$ biorthogonal representations.

III. ROBUST ORTHONORMAL SCALING FUNCTIONS

We now consider discrete wavelet transforms that are as robust as possible or *optimally robust* w.r.t. $\mathcal{E}(\frac{1}{2})$. We first derive an appropriate expression for the translation error for orthonormal scaling functions at $\tau = 1/2$ and then minimize this translation error value within the class of orthonormal scaling functions with compact support.

A. Minimax Design of Robust Orthonormal Scaling Functions

Let $\phi(t)$ be a scaling function[1] and V_0 its representation subspace, i.e.,

$$V_0 = \mathrm{span}\{\phi(t-k)\}_{k \in \mathbb{Z}}. \qquad (15)$$

Because $\phi(t)$ is a scaling function, it satisfies the two-scale equation

$$\phi(t) = \sum_k h_k \phi(2t - k). \qquad (16)$$

Designing functions $\phi(t)$ to be optimally robust w.r.t. $\mathcal{E}(\frac{1}{2})$ to translations is now constrained by the two-scale equation. In other words, the design of $\phi(t)$ that minimizes the translation error is constrained by (16).

Optimization problems with functional constraints are hard. Specifying $\{h_k\}$ completely determines its corresponding scaling function $\phi(t)$. References [11], [21], and [22] discuss unconstrained parameterizations Θ for *all* possible sequences $\{h_k(\Theta)\}$ that generate orthonormal scaling functions with compact support. We follow Zou and Tewfik [11]. By expressing the cost function in terms of $\{h_k(\Theta)\}$ instead of $\phi(t)$, we can solve the optimization problem directly in terms of the free parameters $\Theta = \{\theta_i\}$ and drop the constraint imposed by (16). Other authors [23], [24] who have addressed similar design issues regarding scaling functions have found limited success in expressing their cost functions directly in terms of only the two-scale coefficients $\{h_k\}$.

1) Minimization Strategy: In the example seen in Section II, the translation error $\mathcal{E}(\tau)$ in the interval $\tau \in [0, 1]$ is a concave downward function symmetric about $\tau = 1/2$, where it attains its maximum value. This suggests that we develop a strategy to minimize the translation error $\mathcal{E}(\tau)$ at $\tau = 1/2$. By restricting attention to $\mathcal{E}(\frac{1}{2})$, we can express $\mathcal{E}(\frac{1}{2})$ in terms of the coefficients of the two-scale equation only. We can then take advantage of the unconstrained parameterization mentioned above to develop an algorithm that designs orthonormal scaling functions that minimize $\mathcal{E}(\frac{1}{2})$.

2) Cost $\mathcal{E}(\frac{1}{2})$: We first derive an expression for the translation error $\mathcal{E}(\tau)$ at $\tau = 1/2$ for orthonormal scaling functions. Let $\phi(t)$ be the orthonormal scaling function and $\Phi(f)$ its Fourier transform. From (14), the translation error at $\tau = 1/2$ is

$$\mathcal{E}\left(\tfrac{1}{2}\right) = 1 - \int_0^1 \left| Z_{|\Phi|^2}\left(f, \tfrac{1}{2}\right) \right|^2 df. \qquad (17)$$

The Zak transform of a Fourier pair $g(t)$ and $G(f)$ satisfies

$$Z_G(f, \tau) = e^{j2\pi \tau f} Z_g(-\tau, f). \qquad (18)$$

This property enables us to relate the Zak transform of $|\Phi(t)|^2$ to the Zak transform of its Fourier pair $R_\phi(t)$, which is the autocorrelation of $\phi(t)$. Using (18) and the fact that R_ϕ is real and even, we obtain, for $\tau = 1/2$

$$\left| Z_{R_\phi}\left(\tfrac{1}{2}, f\right) \right|^2 = \left| Z_{|\Phi|^2}\left(f, \tfrac{1}{2}\right) \right|^2.$$

[1] Scaling functions are usually represented in the literature by ϕ, which is why we use in this section ϕ rather than g.

Substituting this into (17), $\mathcal{E}(\frac{1}{2})$ becomes

$$\mathcal{E}\left(\tfrac{1}{2}\right) = 1 - \int_0^1 \left| \sum_k R_\phi\left(k + \tfrac{1}{2}\right) e^{-j2\pi fk} \right|^2 \, df \quad (19)$$

$$= 1 - \sum_k \left| R_\phi\left(k + \tfrac{1}{2}\right) \right|^2. \quad (20)$$

Equation (20) expresses the translation error $\mathcal{E}(\frac{1}{2})$ in terms of the samples $\{R_\phi(k + \frac{1}{2})\}$.

We now relate the samples of the autocorrelation function $R_\phi(k + \frac{1}{2})$ to the coefficients of the two-scale equation. Since $R_\phi(t)$ is the autocorrelation function of the scaling function $\phi(t)$, it also solves the two-scale equation

$$R_\phi(t) = \tfrac{1}{2} \sum_n h_n' R_\phi(2t - n) \quad (21)$$

where $\{h_n'\}$ is the autocorrelation sequence of the coefficients from the original two-scale equation that determines $\phi(t)$, i.e.,

$$h_n' = \sum_k h_{n+k} h_k.$$

Using (21) to express $R_\phi(k + \frac{1}{2})$, we have

$$R_\phi\left(k + \tfrac{1}{2}\right) = \tfrac{1}{2} \sum_n h_n' R_\phi(2k + 1 - n).$$

Since $\{\phi(t - k)\}_{k \in \mathbb{Z}}$ are orthonormal, we have successively

$$R_\phi(2k + 1 - n) = \delta_{2k+1, n}$$
$$R_\phi\left(k + \tfrac{1}{2}\right) = \tfrac{1}{2} h_{2k+1}'. \quad (22)$$

In other words, the samples $R_\phi(k + \frac{1}{2})$ are equal to the odd samples of the autocorrelation sequence of $\{h_k\}$ divided by 2. Substituting (22) in (20), we obtain, for the translation error

$$\mathcal{E}\left(\tfrac{1}{2}\right) = 1 - \tfrac{1}{4} \sum_k \left| h_{2k+1}' \right|^2. \quad (23)$$

The error is expressed in terms of the energy of the *odd* samples of the autocorrelation sequence of the coefficients $\{h_k\}$ of the original two-scale equation.

We now investigate what constraints, if any, the two-scale equation (16) imposes on the odd samples of the autocorrelation sequence $\{h_{2k+1}'\}$. Let $H(z) = \sum_k h_k z^{-k}$ be the Z-transform of the sequence $\{h_k\}$. Similarly, let $H'(z) = H(z)H(z^{-1})$ be the Z-transform of $\{h_k'\}$. Recall from [25] that $\{\phi(t - k)\}_{k \in \mathbb{Z}}$ being orthonormal is equivalent to $H'(z)$ satisfying

$$H'(z) + H'(-z) = 2. \quad (24)$$

The left-hand side of (24) is an even polynomial in z, which is constrained to equal a constant equal to 2. Since (24) involves only the even samples $\{h_{2k}'\}$, the orthogonality of the scaling function is equivalent to constraining all of the even lags of the autocorrelation sequence $\{h_k'\}$ to be zero except for the zero lag, which equals the energy of $\phi(t)$. On the other hand, the two-scale equation does not constrain the odd lags $\{h_{2k+1}'\}$, which, by (20), completely determine the translation error $\mathcal{E}(\frac{1}{2})$. This states that there is no fundamental conflict in designing orthonormal functions minimizing $\mathcal{E}(\frac{1}{2})$

that simultaneously satisfy the two-scale equation. All of the degrees of freedom available in the minimax strategy for designing the scaling function can be dedicated to reducing $\mathcal{E}(\frac{1}{2})$. We show how to do this next.

B. Robust O.N. Scaling Functions with Compact Support

MRA's based on compactly supported wavelets are computationally efficient. Daubechies [25] was the first to provide a characterization of all orthonormal compactly supported scaling functions in terms of the coefficients of the two-scale equation $\{h_k\}$. Alternative parameterizations are in [11], [21], and [22]. We use the results in [11] that provide an unconstrained parameterization of the $\{h_k\}$ in terms of a set of angles $\Theta = \{\theta_i\}$. We apply this parameterization to the expression found in (23) for $\mathcal{E}(\frac{1}{2})$. In this way, the difficult problem of finding robust functions that minimize $\mathcal{E}(\frac{1}{2})$ subject to the constraint of satisfying the two-scale equation becomes an unconstrained parameter optimization problem.

1) Robust Compact Support O.N. Scaling Functions: Zou and Tewfik [11] provide a parameterization for all possible orthonormal scaling functions and wavelets with a given length $2N$. The coefficients h_k are functions of sines and cosines of angles $\Theta = \{\theta_k\}$. It is easy to express the autocorrelation sequence $h_k'(\Theta)$ as a function of Θ. In turn, the maximum translation error $\mathcal{E}(\frac{1}{2})$ in (23) becomes

$$\mathcal{E}\left(\tfrac{1}{2}, \Theta\right) = 1 - \tfrac{1}{4} \sum_k \left| h_{2k+1}'(\Theta) \right|^2 \quad (25)$$

where we have indicated explicitly the dependence on θ. The gradient with respect to Θ is

$$\nabla_\Theta \mathcal{E}\left(\tfrac{1}{2}, \Theta\right) = -\tfrac{1}{2} \sum_k h_{2k+1}'(\Theta) \nabla_\Theta h_{2k+1}'(\Theta). \quad (26)$$

The examples that follow show that for a given N, it is relatively straightforward to find closed-form expressions for (25) and (26). These are then used in optimization algorithms, like gradient descent, to find the minimal value of $\mathcal{E}(\frac{1}{2}, \Theta)$. In general, $\mathcal{E}(\frac{1}{2}, \Theta)$ may have multiple local minima, and the usual techniques (e.g., random restarts) may be needed to ensure that the optimization algorithm finds the absolute minima.

Finally, it should be noted that since the valid sequences of length $2N$ are a subset of the valid sequences of length $2(N + 1)$ (recall that all valid sequences are of even length), then in a trivial way, the minimal value $\mathcal{E}_{2(N+1)}^\star(\frac{1}{2}) \leq \mathcal{E}_{2N}^\star(\frac{1}{2})$. In other words, the minimum error $\mathcal{E}^\star(\frac{1}{2})$ is monotonically decreasing with the number of coefficients. Consequently, we expect that the minimum error attainable by scaling functions with infinite extent in the time domain will have a smaller error than is possible by any scaling function with compact support. Indeed, we have demonstrated that the sinc function is perfectly shiftable in the Appendix.

C. Design Examples

In this section, we present results of the theory developed in the previous subsections. We consider first scaling functions
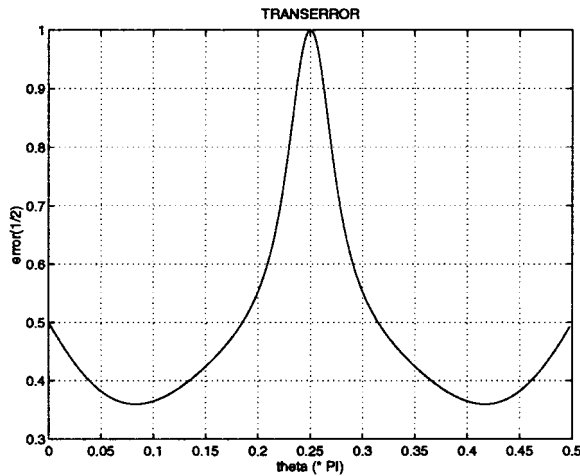
Fig. 2. Maximum translation error $\mathcal{E}(\frac{1}{2}, \theta)$ for the single parameter case.

with $N = 2$, which are parameterized by one parameter and then with $N = 3$, which are parameterized by two parameters.

1) Single Parameter Case: In this example, we use the parameterization for the four coefficient scaling function. From [11], all length-4 sequences are given by

$$h_{0,1} = \sqrt{2} \sin\theta \sin(\theta \mp \pi/4)$$
$$h_{2,3} = \sqrt{2} \cos\theta \sin(\theta \pm \pi/4)$$

where θ is the unconstrained free parameter, and the coefficients are normalized so that $\sum_k h_k = 2$. $\mathcal{E}(\frac{1}{2})$ is given by the energy in the odd coefficients of the autocorrelation sequence of $\{h_k\}$. Taking advantage of the autocorrelation's even symmetry, we calculate only the positive odd terms

$$h_1' = \frac{3}{8} + \frac{1}{4}\sin(2\theta) + \frac{1}{8}\cos(4\theta)$$
$$h_3' = \frac{1}{8} - \frac{1}{4}\sin(2\theta) - \frac{1}{8}\cos(4\theta).$$

The error is given by the trigonometric polynomial

$$\mathcal{E}\left(\tfrac{1}{2}, \theta\right) = 1 - 2\big[|h_1'(\theta)|^2 + |h_3'(\theta)|^2\big]$$
$$= \tfrac{17}{32} - \tfrac{1}{8}\sin(2\theta) - \tfrac{1}{8}\sin(6\theta) - \tfrac{1}{32}\cos(8\theta)$$

and is shown in Fig. 2 for $\theta \in [0, \pi/2]$. The scaling functions for this range of θ are smooth. We provide in Figs. 3 and 4 several scaling functions and their corresponding translation functions $\mathcal{E}(\tau)$. These $\mathcal{E}(\tau)$ all exhibit a concave behavior with the maximum occurring at $\mathcal{E}(\frac{1}{2})$. We are not stating, however, that this is true for every scaling function. A closed-form expression for the derivative of the error is readily found to be

$$\frac{d\mathcal{E}\left(\tfrac{1}{2}, \theta\right)}{d\theta} = -\frac{1}{4}\cos(2\theta) - \frac{3}{4}\cos(6\theta) + \frac{1}{4}\sin(8\theta).$$

Its zero crossings are candidates for local minima of the shiftability error $\mathcal{E}(\frac{1}{2}, \theta)$. Alternatively, the gradient can be used in numerical methods to find the value of θ that minimizes $\mathcal{E}(\frac{1}{2})$. First, notice that $\mathcal{E}(\frac{1}{2}, \theta)$ has multiple maxima and minima, and care must be taken to ensure that an absolute minima is found. Second, the absolute minima occurs at two values ($\theta = \pi/12$ and $5\pi/12$) that produce time-reversed

versions of the same filter because they have the same autocorrelation sequence. The value of $\theta = 5\pi/12$ coincides with the Daubechies D2 scaling function [25].

2) Two Parameter Case: The length-6 sequence is parameterized by $\Theta = \{\theta_1, \theta_2\}$, which is given by

$$
\begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \end{bmatrix} = \frac{1}{4}
\begin{bmatrix}
1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & -1 \\
2 & 0 & 0 & 2 & 0 & 2 & 0 & -2 & 2 \\
2 & 0 & 0 & 2 & 0 & -2 & 0 & 2 & 2 \\
1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 \\
1 & 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1
\end{bmatrix}
$$
$$
\cdot
\begin{bmatrix}
1 \\
\cos(\theta_1) \\
\cos(\theta_2) \\
\cos(\theta_1)\cos(\theta_2) \\
\sin(\theta_1) \\
\sin(\theta_1)\cos(\theta_2) \\
\sin(\theta_2) \\
\cos(\theta_1)\sin(\theta_2) \\
\sin(\theta_1)\sin(\theta_2)
\end{bmatrix}
$$

and the odd terms of its autocorrelation sequence are

$$
\begin{bmatrix} h_1' \\ h_3' \\ h_5' \end{bmatrix} = \frac{1}{16}
\begin{bmatrix}
10 & 0 & 4 & 4 & 4 & -2 & -4 & -2 & 2 \\
5 & -4 & -2 & -2 & -2 & 1 & 6 & 1 & -3 \\
1 & -2 & 4 & -2 & -2 & 1 & -2 & 1 & 1
\end{bmatrix}
$$
$$
\cdot
\begin{bmatrix}
1 \\
\cos(\theta_1) \\
\cos(\theta_2) \\
\cos(\theta_1 - \theta_2) \\
\cos(\theta_1 + \theta_2) \\
\cos(2\theta_1) \\
\cos(2\theta_1 - \theta_2) \\
\cos(2\theta_2) \\
\cos(2\theta_1 - 2\theta_2)
\end{bmatrix}
$$

from which its gradient can be calculated. Fig. 5 shows the error surface for this example. Black and white correspond to large and small values of $\mathcal{E}(\frac{1}{2})$, respectively. The minimizing sequences are for $\Theta = (0.38, 1.64)$ and $(1.64, 0.38)$, which correspond to the light spots in the upper-left and lower-right corners of Fig. 5, respectively. These correspond to time-reversed versions of the same sequence. The minimum value for the two-parameter case is $\mathcal{E}_6^\star(\frac{1}{2}) = 0.226$, which is smaller than $\mathcal{E}_4^\star(\frac{1}{2}) = 0.359$ from the single parameter case.

Clearly, as N increases, the closed-form expressions for $\{h_k'\}$ and $\mathcal{E}(\frac{1}{2})$ quickly become tedious but are otherwise straightforward and easily computed numerically by software packages like Matlab or symbolically with the help of Maple or Mathematica.

IV. ROBUST COMPACT SUPPORT BIORTHOGONAL SCALING FUNCTIONS

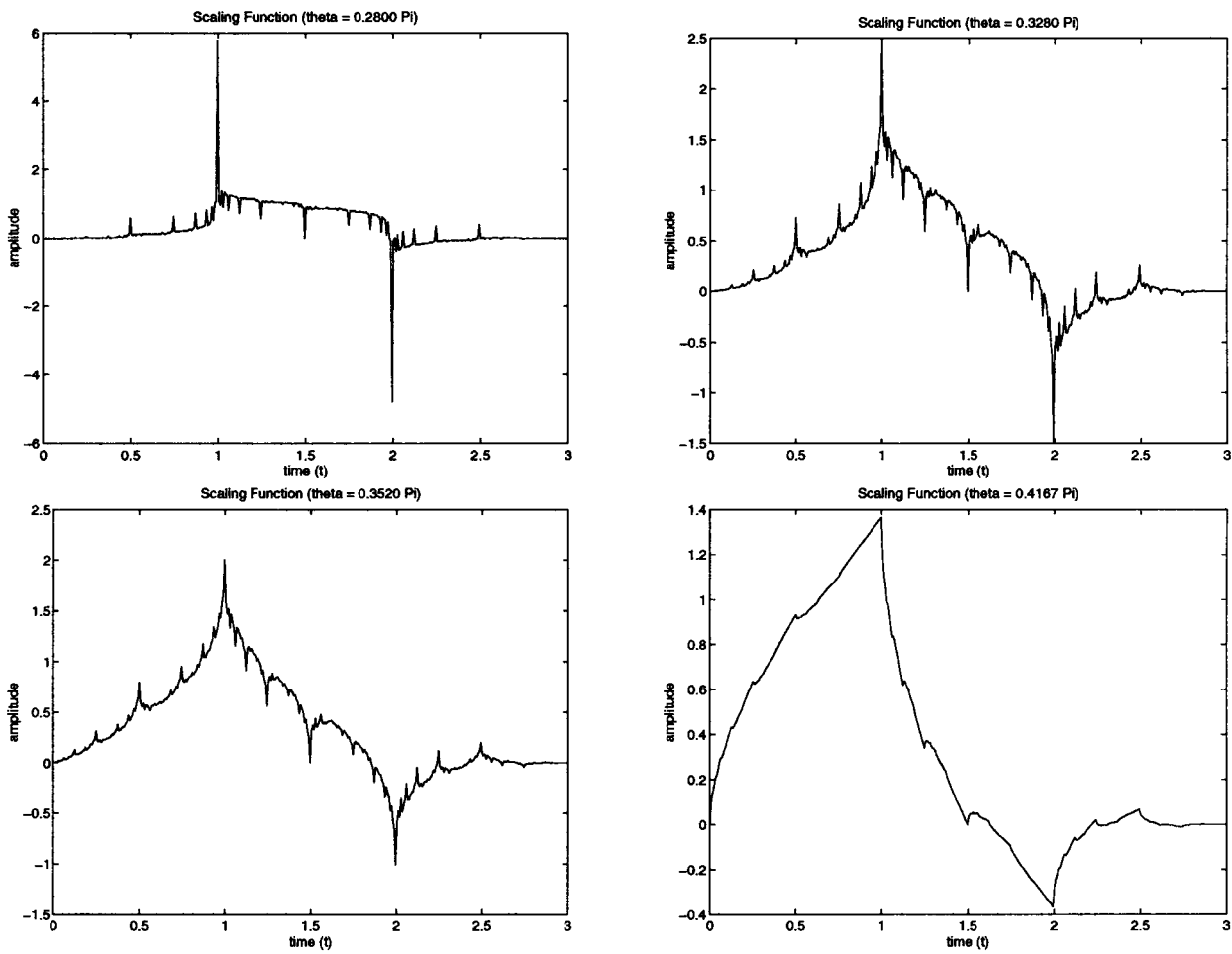In the previous section, we used the parameterization of compactly supported o.n. scaling functions to design o.n.
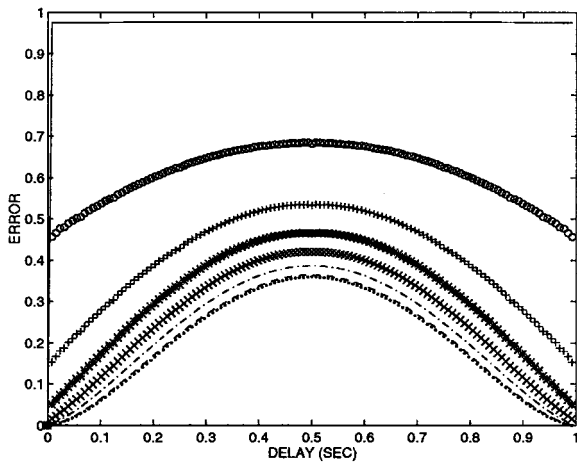
Fig. 3. Scaling functions for several values of θ.

Fig. 4. Eight $\epsilon(\tau)$ curves for scaling functions corresponding to different θ values. From top to bottom: $\theta = 0.256\pi$; $\theta = 0.28\pi$; $\theta = 0.304\pi$; $\theta = 0.328\pi$; $\theta = 0.352\pi$; $\theta = 0.376\pi$; $\theta = 0.4\pi$; $\theta = (5/12)\pi$.

Fig. 5. Maximum translation error $\mathcal{E}(\frac{1}{2}, \{\theta_1, \theta_2\})$ for the two parameter case.

scaling functions with compact support that minimize $\mathcal{E}(\frac{1}{2})$. Key to our development was the strategy of expressing the translation error $\mathcal{E}(\frac{1}{2})$ in terms of the odd samples of the autocorrelation of the coefficients of the two-scale equation. In this section, we extend these results to the case of biorthogonal scaling functions.
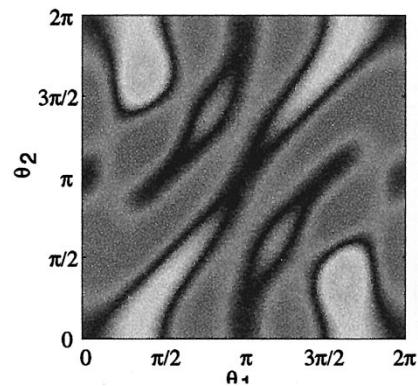
Requiring a compactly supported scaling function to be orthonormal is highly constraining. For example, it is not possible to have one that is symmetric. Dropping the orthonormality constraint, we have greater freedom in designing a scaling function and its nonunique biorthogonal dual to satisfy multiple design criteria. Because nonorthogonal representations are redundant, the redundancy will allow for more robust representations. This is supported by the following trivial argument. Since o.n. scaling functions are contained within the class of biorthogonal scaling functions, the minimum

translation error obtainable by a biorthogonal representation is at least as small as the minimal value obtained by an orthonormal representation, i.e., $\mathcal{E}_{\mathrm{bi}}^{\star}(\frac{1}{2}) \leq \mathcal{E}_{\mathrm{on}}^{\star}(\frac{1}{2})$, where $\mathcal{E}^{\star}$ represents the minimal value of $\mathcal{E}$ for a given length of support.

A. Maximum Translation Error for Biorthogonal Scaling Functions

The approach for the biorthogonal case parallels the minimax strategy for the o.n. case. First, we derive an expression for the error at $\tau = \frac{1}{2}$ and then demonstrate how each term in that expression can be expressed directly in terms of the two-scale coefficients that determine the scaling function and its dual. Without loss of generality, we restrict the discussion to the space V_0 given in (15) generated by the integer translates of $\phi(t)$. Starting with the definition for $\mathcal{E}(\frac{1}{2})$, we have

$$\mathcal{E}\left(\tfrac{1}{2}\right) = \left\| \phi\left(t - \tfrac{1}{2}\right) - \sum_k c_k \phi(t - k) \right\|^2.$$

The $\{c_k\}$ are found by inner products with the biorthogonal dual $\{\tilde{\phi}(t - k)\}$

$$c_k = \left\langle \phi\left(t - \tfrac{1}{2}\right), \tilde{\phi}(t - k) \right\rangle = R_{\phi\tilde{\phi}}\left(k - \tfrac{1}{2}\right).$$

The function $R_{\phi\tilde{\phi}}(\tau)$ is the cross correlation between ϕ and its biorthogonal dual $\tilde{\phi}$. Substituting this into the expression for $\mathcal{E}(\frac{1}{2})$ and expanding, we obtain

$$\mathcal{E}\left(\tfrac{1}{2}\right) = \|\phi\|^2 + \sum_{k_1} \sum_{k_2} R_{\phi\tilde{\phi}}\left(k_1 - \tfrac{1}{2}\right) R_{\phi\tilde{\phi}}\left(k_2 - \tfrac{1}{2}\right)$$
$$\cdot R_{\phi}(k_1 - k_2) - 2 \sum_k R_{\phi\tilde{\phi}}\left(k - \tfrac{1}{2}\right) R_{\phi}\left(k - \tfrac{1}{2}\right). \quad (27)$$

In (27), R_{ϕ} is, like before, the autocorrelation function of $\phi(t)$. In the special case of the previous section, where $\{\phi(t - k)\}_{k \in \mathbb{Z}}$ are orthonormal, $R_{\phi\tilde{\phi}} = R_{\phi}$, $R_{\phi}(k_1 - k_2) = \delta_{k_1, k_2}$, and the above expression becomes the same as that derived in (20).

Next, we take advantage of the fact that ϕ and $\tilde{\phi}$ each satisfy a two-scale equation

$$\phi(t) = \sum_k h_k \phi(2t - k)$$
$$\tilde{\phi}(t) = \sum_k \tilde{h}_k \tilde{\phi}(2t - k). \quad (28)$$

We use these equations to relate $\mathcal{E}(\frac{1}{2})$ to the coefficients $\{h_k\}$ and $\{\tilde{h}_k\}$.

1) $R_{\phi\tilde{\phi}}(k_2 - \frac{1}{2})$ and the Two-Scale Coefficients: Using the definition of a cross-correlation function, the two-scale equations satisfied by ϕ and $\tilde{\phi}$, and the fact that ϕ and $\tilde{\phi}$ are biorthogonal, $R_{\phi\tilde{\phi}}(k_2 - \frac{1}{2})$ becomes

$$R_{\phi\tilde{\phi}}\left(k - \tfrac{1}{2}\right) = \int_{-\infty}^{\infty} \phi\left(t + k - \tfrac{1}{2}\right) \tilde{\phi}(t)\, dt \quad (29)$$

$$= \sum_{k_1} \sum_{k_2} h_{k_1} \tilde{h}_{k_2} \int \phi(2t + 2k - 1 - k_1)$$
$$\cdot \tilde{\phi}(2t - k_2)\, dt \quad (30)$$

$$= \tfrac{1}{2} \sum_{k_1} h_{k_1 + 2k - 1} \tilde{h}_{k_1} \quad (31)$$

$$= \tfrac{1}{2} p_{2k-1} \quad (32)$$

where $\{p_{2k-1}\}$ are the odd coefficients of the cross-correlation sequence between $\{h_k\}$ and $\{\tilde{h}_k\}$. Hence, $R_{\phi\tilde{\phi}}(k - \frac{1}{2})$ is readily expressed in terms of $\{h_k\}$ and $\{\tilde{h}_k\}$.

2) $R_{\phi}(k)$ and $R_{\phi}(k - \frac{1}{2})$ and the Two-Scale Coefficients: $R_{\phi}(k)$ and $R_{\phi}(k - \frac{1}{2})$ can also be calculated in terms of $\{h_k\}$ in a straightforward way. We assume that $\{h_k\}$ is a finite sequence. As seen before in (21), the correlation function $R_{\phi}(t)$ satisfies the two-scale equation. We find the integer samples $R_{\phi}(n)$ following the algorithm outlined by Strang [1]. A scaling function $\phi(t)$ with support on $t \in (0, 2N - 1)$ is defined via the two-scale equation by a sequence $\{h_k\}$ of length $2N$. The corresponding autocorrelation function $R_{\phi}(t)$ has support on $t \in (-(2N - 1), 2N - 1)$, and $\{h'_k\}$ has length $4N - 1$. Define $\mathbf{R}_{\phi} = [R_{\phi}(2N - 2) \cdots R_{\phi}(1) R_{\phi}(0) R_{\phi}(1) \cdots R_{\phi}(2N - 2)]^T$ and

$$\mathbf{H}' = \tfrac{1}{2} \begin{bmatrix} & & & \cdots & & & & \\ \cdots & h'_2 & h'_1 & h'_0 & 0 & 0 & 0 & \cdots \\ & \cdots & h'_2 & h'_1 & h'_0 & 0 & 0 & 0 & \cdots \\ & & \cdots & h'_2 & h'_1 & h'_0 & 0 & \cdots \\ & & & \cdots & & & & \end{bmatrix}.$$

The integer values of $R_{\phi}(t)$ can be expressed as

$$\mathbf{R}_{\phi} = \mathbf{H}' \mathbf{R}_{\phi}. \quad (33)$$

Matrix $\mathbf{H}'$ is $(2(2N-2)+1) \times (2(2N-2)+1)$. The equation shows that $\mathbf{R}_{\phi}$ is the eigenvector of $\mathbf{H}'$ corresponding to the eigenvalue 1. Once the integer samples of the autocorrelation function are known, the half integer samples are found by iterating the two-scale equation once and subsampling to keep only the odd samples

$$R_{\phi}\left(n - \tfrac{1}{2}\right) = \tfrac{1}{2} \sum_k h'_k R_{\phi}(2n - 1 - k). \quad (34)$$

Hence, by substituting the results of (32)–(34) into (27), the translation error at the midpoint $\mathcal{E}(\frac{1}{2})$ is expressed directly in terms of the autocorrelation and cross-correlation sequences of the coefficients of the two-scale equations for the scaling function and its dual.

B. Review of Perfect Reconstruction Filter Banks

Now that we know how to express $\mathcal{E}(\frac{1}{2})$ in terms of $\{h_k\}$ and $\{\tilde{h}_k\}$, we need a characterization of all valid sequences $\{h_k\}$ and $\{\tilde{h}_k\}$ that lead to a biorthogonal multiresolution analysis. Extensive work has been done on this subject by connecting MRA's to perfect reconstruction filter banks (PRFB's) [12], [25], [26]. The following is a summary of the relevant results.

In order for two sequences $\{h_k\}$ and $\{\tilde{h}_k\}$ to generate a biorthogonal multiresolution, they must satisfy two conditions. The first is the regularity condition that each sequence must

satisfy in order to guarantee that the infinite products

$$\Phi(\omega) = \lim_{i \to \infty} \prod_{k=1}^{i-1} \frac{1}{2^k} H(e^{j\omega/2^k})$$

$$\tilde{\Phi}(\omega) = \lim_{i \to \infty} \prod_{k=1}^{i-1} \frac{1}{2^k} \tilde{H}(e^{j\omega/2^k}) \tag{35}$$

both converge pointwise to continuous functions. The regularity condition given in [25] for orthonormal scaling functions is applicable to the biorthogonal case. Reference [27] provides two different regularity conditions adapted for biorthogonal scaling functions. Convergence of (35) depends, in part, on both filters $H(z)$ and $\tilde{H}(z)$ having a sufficient number of zeros at $z = -1$.

Assuming $H(z)$ and $\tilde{H}(z)$ are sufficiently regular, the second condition we impose is that the limiting functions of (35) are biorthogonal to each other, which is equivalent to requiring

$$H(z)\tilde{H}(z^{-1}) + H(-z)\tilde{H}(-z^{-1}) = 2. \tag{36}$$

The sequence $\{p_k\}$ corresponding to $P(z) = H(z)\tilde{H}(z^{-1})$ is the cross-correlation sequence between $\{h_k\}$ and $\{\tilde{h}_k\}$

$$p_k = \sum_n h_{k+n}\tilde{h}_n$$

whose odd terms appear in (32). Analogous to (24) for the orthonormal case, (36) constrains only the coefficients of the *even* powers in $P(z)$ to equal zero except for the zeroth lag, which is constrained to equal 2. Since the expression in (27) for $\mathcal{E}(\frac{1}{2})$ depends on $R_{\phi\tilde{\phi}}(k - \frac{1}{2}) = \frac{1}{2}p_{2k-1}$, i.e., the *odd* coefficients of $\{p_k\}$, there is no conflict in designing optimally robust w.r.t. $\mathcal{E}(\frac{1}{2})$ functions and satisfying the constraints for constructing biorthogonal scaling functions.

1) Parameterization of Biorthogonal Scaling Functions: Equation (36) is a special form of a Bezout identity. Given one of the filters, say, $H(z)$, the minimal length complementary filter can be found by continued fraction expansions (Euclid's algorithm) or by solving a system of linear equations [28]. The complementary filter is not unique. After solving for a valid minimal length complementary filter, we can parameterize all other complementary filters with increased length using the procedure presented in [12], which we outline now.

If there exists a polynomial $\tilde{A}(z)$ such that

$$H(z)\tilde{A}(z^{-1}) + H(-z)\tilde{A}(-z^{-1}) = 0$$

then we can define a new complementary filter by

$$\tilde{H}^\star(z) = \tilde{H}(z) + \tilde{A}(z)$$

that still satisfies (36). The following proposition adapted from [12] gives a characterization for $\tilde{A}(z)$.

Proposition IV.1: All filters of length $N + 2m - 2$ that are complementary to a length-N filter $H(z)$ have the form

$$\tilde{H}^\star(z) = z^{-2k}\tilde{H}(z) + z^{-1}E(z)H(-z) \tag{37}$$

where $k \in \{0, 1, \cdots, m\}$, $\tilde{H}(z)$ is a length-$N - 2$ complementary filter, and $E(z) = E(-z)$ is a polynomial of degree

$2(m - 1)$ of the form

$$E(z) = \sum_{i=0}^{m-1} \beta_i z^{-2i}. \tag{38}$$

The parameter k changes the location of the single nonzero even coefficient, whereas increasing m increases the length of the new complementary filter $\tilde{H}^\star(z)$. Hence, Proposition IV.1 provides a characterization of all possible higher order complementary filters.

From (36), it follows that any valid polynomial $P(z)$ that satisfies (36) can be factored into $H(z)$ and its complementary filter $\tilde{H}(z)$ to produce a PRFB. Equation (36) depends only on the product of the two filters; therefore, we have additional design freedom in the way the zeros of $P(z)$ are assigned to either $H(z)$ or $\tilde{H}(z)$, as long as our factorization yields two filters that are sufficiently regular to allow (35) to converge.

C. Robust Biorthogonal Scaling Functions with Compact Support

Using (27) and (32)–(34) to express the error in terms of the coefficients of the two-scale equations in (28) and with the above characterization of the biorthogonal two-scale coefficients, we are ready to design representations that are optimally robust w.r.t. $\mathcal{E}(\frac{1}{2})$.

The previous subsection provides a method for finding complementary filters when we are *given* $H(z)$, but given a clean slate, how should we commence? Since we require both $H(z)$ and $\tilde{H}(z)$ to converge to a scaling function according to (35), a minimum number of zeros at $z = -1$ is required. Hence, by choosing $H(z) = H_b(z) = 2((1 + z^{-1})/2)^N$, i.e., the binomial filter, we are guaranteed that our system will contain a minimum number of zeros at $z = -1$. Because we are free to assign the zeros of $P(z) = H(z)\tilde{H}(z^{-1})$ to $H(z)$ and $\tilde{H}(z^{-1})$ as desired, without fear of destroying the PRFB property, we can divide the zeros at $z = -1$ between $H(z)$ and $\tilde{H}(z)$ to satisfy the regularity condition. Proposition 4.4 in [12] guarantees the existence of a complementary filter to the binomial filter so that this starting point is guaranteed to produce a valid PRFB.

1) Robust Biorthogonal Scaling Function Design: The procedure for designing $H(z)$ and $\tilde{H}(z)$ in order to minimize $\mathcal{E}(\frac{1}{2})$ is as follows.

1) Define an initial binomial filter $H_b(z)$ with a specified number of zeros N at $z = -1$.
2) Find its minimal-length complementary filter $\tilde{H}_b(z)$ by any of the techniques available. See [9] for a closed-form expression to the filter complementary to the binomial filter. Once $\tilde{H}_b(z)$ is found, determine $P(z) = H_b(z)\tilde{H}_b(z)$.
3) Calculate $\mathcal{E}(\frac{1}{2})$ for all possible factorizations of $P(z) = H(z)\tilde{H}(z)$ such that $H(z)$ and $\tilde{H}(z)$ both satisfy the regularity condition. If the minimum value is suitable, stop here.
4) If a smaller error is desired, increase the degrees of freedom via Proposition IV.1. Starting with $m = 1$, optimize $\mathcal{E}(\frac{1}{2})$ w.r.t. $\{\beta_i\}$, and factor $P(z)$ subject to the constraint of the regularity condition. If the
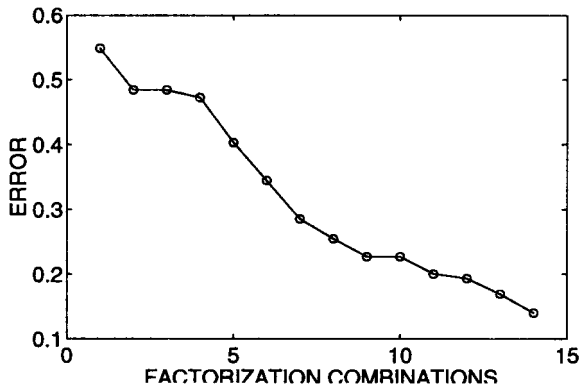
Fig. 6. Translation error $\mathcal{E}(\frac{1}{2})$ for all valid factorizations of the D8 autocorrelation sequence.

resulting minimal value of $\mathcal{E}(\frac{1}{2})$ for this value of m is not acceptable, increment m by one, and repeat the optimization and factorization procedure until the translation error is acceptably small.

Although the PRFB constraint is independent of how the zeros of $P(z)$ are factored, the expression for $\mathcal{E}(\frac{1}{2})$ is not. On the contrary, experimental results have shown considerable differences in the translation error for different factorizations. This is demonstrated in Example 1 below.

This factorization problem, as well as the eigenvector problem in (33), make it difficult to express R_ϕ and, hence, $\mathcal{E}(\frac{1}{2})$ directly in terms of the parameters $\{\beta_i\}$. If, however, the problem statement were different and we were given $H(z)$, or if $H(z)$ was selected according to some other criteria, as in [11], then the problem would be reduced to finding only $\tilde{H}(z)$ to minimize $\mathcal{E}(\frac{1}{2})$. In this case, the combinatorics of factoring the zeros of $P(z)$ between $H(z)$ and $\tilde{H}(z)$ and the eigenvector problem are avoided because $H(z)$ is now fixed, and the translation error and its gradient can be expressed explicitly in terms of the unknown parameters $\{\beta_i\}$.

D. Design Examples

Example 1: We demonstrate the effect of assigning factors to $H(z)$ and $\tilde{H}(z^{-1})$ on $\mathcal{E}(\frac{1}{2})$. The translation error $\mathcal{E}(\frac{1}{2})$ is affected by how the factors of $P(z)$ are assigned to $H(z)$ and $\tilde{H}(z^{-1})$. We demonstrate this effect by starting with $P(z) = H_{D8}(z)H_{D8}(z^{-1})$, which is the autocorrelation sequence of the Daubechies D8 coefficients. We calculate $\mathcal{E}(\frac{1}{2})$ for each valid factorization of $P(z)$ such that $H(z)$ and $\tilde{H}(z)$ satisfy the regularity condition in [25, Prop. 3.3]. Fig. 6 shows the translation error $\mathcal{E}(\frac{1}{2})$ for each valid factorization of $P(z)$ plotted in descending order. Factorizations 9 and 10 correspond to the orthogonal cases where $H(z) = \tilde{H}(z) = H_{D8}(z)$ with $\mathcal{E}_{D8}(\frac{1}{2}) = 0.255$, whereas the optimal factorization w.r.t. $\mathcal{E}(\frac{1}{2})$ has a translation error of $\mathcal{E}^\star(\frac{1}{2}) = 0.140$; this is a 2-dB improvement over the D8 scaling function. There are also valid factorizations that produce very large translation errors, where the worst is nearly 6 dB above $\mathcal{E}^\star(\frac{1}{2})$. This example demonstrates that the factorization of $P(z)$ into $H(z)$ and $\tilde{H}(z^{-1})$ has a profound effect on the resulting translation error $\mathcal{E}(\frac{1}{2})$.

Example 2: This example illustrates the design procedure outlined in the previous subsection. We start with the same
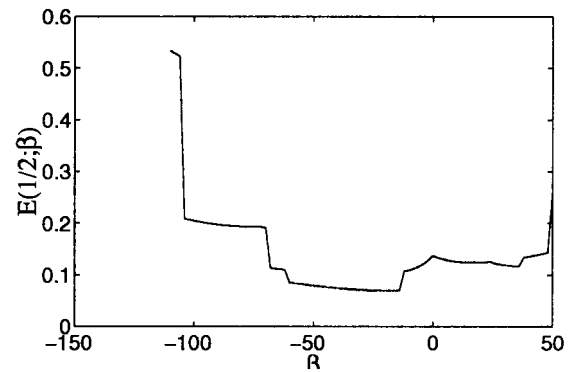

Fig. 7. Translation error $\mathcal{E}(\frac{1}{2}; \beta)$ for the optimal factorizations of $P(z; \beta)$ as a function of β.
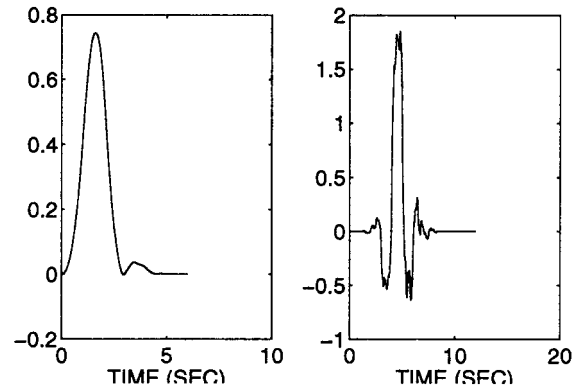

Fig. 8. Optimally robust w.r.t. $\mathcal{E}(\frac{1}{2})$ scaling function (left) and its biorthogonal dual (right) from Example 2.

$P(z)$ as in the previous example, which is the same as starting with the binomial filter

$$H_b(z) = 2\left(\frac{1+z^{-1}}{2}\right)^8$$

and its minimal length complementary filter $\tilde{H}_b(z)$. Taking $k = 1$ and $E(z) = (1 + \beta z^{-2})$ in (37) of Proposition IV.1, we extend the length of the minimal length complementary filter

$$\tilde{H}_b^\star(z; \beta) = z^{-2}\tilde{H}_b(z) + z^{-1}(1 + \beta z^{-2})H_b(-z).$$

We determine the factorization that minimizes $\mathcal{E}(\frac{1}{2})$

$$P'(z; \beta) = H_b(z)\tilde{H}_b'(z^{-1}; \beta)$$

into $H(z; \beta)$ and $\tilde{H}(z; \beta)$ for each value of β.

Fig. 7 is a plot of $\mathcal{E}(\frac{1}{2}; \beta)$ as a function of β. The optimal value $\mathcal{E}^\star(\frac{1}{2}; \beta) = 0.0694$ occurs at $\beta = -18.56$, and the corresponding optimal scaling function and its biorthogonal dual are plotted in Fig. 8. The discontinuities in $\mathcal{E}(\frac{1}{2}; \beta)$ are due to two phenomena. First, continuously varying β does not always correspond to a continuous trajectory of the roots of $P'(z; \beta)$ in the complex plane. Abrupt changes in the zeros of $P'(z; \beta)$ as a function of β lead to discontinuities in $\mathcal{E}(\frac{1}{2}; \beta)$. The second is due to the fact that for each value of β, we are minimizing $\mathcal{E}(\frac{1}{2}; \beta)$ with respect to the factorization of $P'(z; \beta)$. Different factorizations can cause abrupt changes in $\mathcal{E}(\frac{1}{2}; \beta)$, as demonstrated in the previous example.

V. CONCLUSION

We have described the issue of subspace representations that are robust to translations of the input signal. We measured robustness in terms of the mean square error $\mathcal{E}(\tau)$ in representing a signal and its arbitrary delays in terms of the integer shifts of a mother function. We derived, in the Appendix, expressions for this translation error under various bandlimited assumptions, and for each case, we determined under what conditions the translation error goes to zero. We also derived a general expression for $\mathcal{E}(\tau)$ without imposing bandlimited constraints.

We addressed the design of scaling functions with compact support that are robust to translations. We are motivated by the lack of translation invariance of critically sampled dyadic wavelet transforms. We define and design the robust scaling functions as the minimizers of $\mathcal{E}(\frac{1}{2})$. The strategy of minimizing $\mathcal{E}(\frac{1}{2})$ enabled us to express the error in terms only of the coefficients of the two-scale equation. Taking advantage of the parameterizations for orthonormal and biorthogonal scaling functions with compact support, we proposed efficient algorithms for finding scaling functions that are optimally robust with respect to $\mathcal{E}(\frac{1}{2})$ for a given length of support. In future work, it will be important to determine under what conditions on the scaling function ϕ, $\mathcal{E}(\frac{1}{2})$ maximizes $\mathcal{E}(\tau)$ and to extend our algorithms to the cases where the maximum may occur at other values of the translation parameter τ. An alternative is to extend the algorithms presented in the paper to a different translation error metric like the integral of $\mathcal{E}(\tau)$ in the interval $[0, 1]$.

APPENDIX
BANDLIMITED ASSUMPTIONS

In this Appendix, we consider the effect of bandlimited constraints on robust representations. Because bandlimited constraints simplify the expression for the translation error, we explore in this Appendix a more general version of the problem presented in the main body of this paper.

Let $h \in L^2(\mathbb{R})$ be given. We consider the representation of h by translates of another function g to be designed. In general, we look for robust representations, i.e., for a g such that the translation error of the representation

$$\mathcal{E}(\tau) = \left\| h(t-\tau) - \sum_k a_k g(t-k) \right\|^2 \tag{39}$$

is small. We consider the case where h, g, or both are bandlimited.

1) Bandlimited Subspace $\mathcal{B}$: Let $\mathcal{B}$ be the subspace of $L^2(\mathbb{R})$ of bandlimited functions $x(t)$ with bandwidth W. We normalize its bandwidth to the interval $[-\frac{1}{2}, \frac{1}{2}]$, i.e.,

$$\mathcal{B} = \left\{ x(t) : X(f) = 0 \quad \text{for } |f| > \tfrac{1}{2} \right\}$$

where $X(f)$ is the Fourier transform of $x(t)$.

In the discussions that follow, we examine the three combinations of bandlimited assumptions: $h, g \in \mathcal{B}$; $g \in \mathcal{B}$, $h \notin \mathcal{B}$; and $h \in \mathcal{B}$, $g \notin \mathcal{B}$. For each scenario, we derive an expression for the translation error and consider conditions on g that make $\mathcal{E}(\tau)$ as small as possible.

A. g Bandlimited and h Bandlimited

We consider $h, g \in \mathcal{B}$. We seek to see when

$$\forall \tau \in \mathbb{R}, \qquad h(t-\tau) = \sum_k a_k g(t-k). \tag{40}$$

The Fourier transform of (40)

$$H(f)e^{-j2\pi f\tau} = \left(\sum_k a_k e^{-j2\pi fk} \right) G(f)$$

needs to be satisfied only for $|f| < \frac{1}{2}$ because of the bandlimited assumptions on h and g. Hence, for a given $h \in \mathcal{B}$, we can represent the arbitrary translates $h(t-\tau)$ by $\{g(t-k)\}_{k \in \mathbb{Z}}$ if the ratio $H(f)/G(f)$ can be expressed as the fundamental period of a Fourier series

$$|f| < \frac{1}{2}, \qquad \sum_k a_k e^{-j2\pi fk} = \frac{H(f)e^{-j2\pi f\tau}}{G(f)}. \tag{41}$$

If (41) holds, the error $\mathcal{E}(\tau)$ is zero. If we can choose g, an obvious choice is to pick $g = h$. This shows that bandlimited functions are translation invariant. A second possibility is $g(t) = \text{sinc}(t)$. In this case, $G(f) \equiv 1$, $|f| < 1/2$; therefore

$$\forall f \in \mathbb{R}, \qquad \sum_k a_k e^{-j2\pi fk} = H(f)e^{-j2\pi f\tau} * \sum_k \delta(f+k)$$

or $a_k(\tau) = h(k-\tau)$, which, due to the delay parameter τ, is an extension of the sampling theorem [29]. Because of the bandlimited assumptions on g and h, there is no aliasing, and the general shiftability condition in (40) is satisfied.

Although the discussion is based on bandlimited lowpass functions, the results can be extended to include bandpass functions that are modulations of bandlimited lowpass functions.

B. g Bandlimited and h Nonbandlimited

Let $g \in \mathcal{B}$ and $h \notin \mathcal{B}$. We derive an appropriate expression for $\mathcal{E}(\tau)$ given by (39) and determine conditions on g that minimize the translation error for a given h. Applying Parseval's Identity to (39), we have

$$\mathcal{E} = \int_{-\infty}^{\infty} \left| H(f)e^{-j2\pi f\tau} - \left(\sum_k a_k e^{-j2\pi fk} \right) G(f) \right|^2 df$$

$$= \int_{|f|<1/2} |H(f)e^{-j2\pi f\tau} - A(f)G(f)|^2 df$$

$$+ \int_{|f|>1/2} |H(f)|^2 df \tag{42}$$

where

$$A(f) = \sum_k a_k e^{-j2\pi fk}. \tag{43}$$

The last term in (42) is the out-of-band energy in h and is independent of $G(f)$. It is fixed for a given h, regardless of the choice of g. The first term in (42) is made zero by choosing, like before in (41)

$$|f| < \tfrac{1}{2}, \qquad A(f)G(f) = H(f)e^{-j2\pi f\tau}. \tag{44}$$

If g is given, this equation determines $A(f)$. If g is to be determined, a choice is

$$|f| < \tfrac{1}{2}, \qquad G(f) = 1 \quad \text{and} \quad A(f) = H(f)e^{-j2\pi f\tau}.$$

This corresponds to $g(t) = \text{sinc}(t)$. A second possibility is

$$|f| < \tfrac{1}{2}, \qquad G(f) = H(f) \quad \text{and} \quad A(f) = e^{-j2\pi f\tau}.$$

The minimum obtainable $\mathcal{E}(\tau)$ is

$$\mathcal{E}_{\min}(\tau) = \int_{|f|>1/2} |H(f)|^2 \, df. \tag{45}$$

Hence, for the case when h is not bandlimited but g is bandlimited, there is a minimum error for the representation of arbitrary translates of h by the integer translates of g. The minimum error is given by (45), and the alternatives of g given by (44) that achieve it are well understood.

C. g Nonbandlimited and h Bandlimited

Let $g \notin \mathcal{B}$ and $h \in \mathcal{B}$. The translation error $\mathcal{E}(\tau)$ becomes

$$\mathcal{E}(\tau) = \int_{|f|<1/2} |H(f)e^{-j2\pi f\tau} - A(f)G(f)|^2 \, df$$
$$+ \int_{|f|>1/2} |A(f)|^2 |G(f)|^2 \, df \tag{46}$$

where $A(f)$ is given in (43). The first integral is made zero by choosing $A(f)$ to satisfy (44). Unlike the previous case, the choice of $A(f)$ now affects the second term as well.

Define the fundamental period $A_p(f)$ of the periodic $A(f)$ and the tail $G_t(f)$ of $G(f)$ as

$$A_p(f) = \begin{cases} A(f), & |f| \leq \tfrac{1}{2} \\ 0, & |f| > \tfrac{1}{2} \end{cases}$$

$$G_t(f) = \begin{cases} 0, & |f| \leq \tfrac{1}{2} \\ G(f), & |f| > \tfrac{1}{2}. \end{cases}$$

Define the energy spectral densities

$$E_{A_p}(f) = |A_p(f)|^2$$
$$E_G(f) = |G(f)|^2$$

and

$$E_{G_t}(f) = |G_t(f)|^2. \tag{47}$$

Then, by application of the Schwartz inequality, we bound the second integral in (46) as

$$\int_{|f|>1/2} |A(f)|^2 |G(f)|^2 \, df$$
$$= \int_{|f|>1/2} E_{A_p}(f) E_{G_t}(f) \, df$$
$$= \sum_{k=1}^{\infty} \int_{k-1/2}^{k+1/2} E_{A_p}(f) E_{G_t}(f) \, df \tag{48}$$
$$\leq \sum_{k=1}^{\infty} \left(\int_{k-1/2}^{k+1/2} E_{A_p}(f) \, df \right)^{1/2} \left(\int_{k-1/2}^{k+1/2} E_{A_p}(f) \, df \right)^{1/2}$$
$$= \|E_{A_p}(f)\| \, \|E_{G_t}(f)\|$$
$$= \|A_p(f)\|_4^2 \|G_t(f)\|_4^2 \tag{49}$$

where $\|\cdot\|_4$ is the L_4-norm, and $\|\cdot\|$ is the L_2-norm. A strategy to minimize (46) is to choose $A(f)$ and $G(f)$ so that (44) is satisfied $A(f)$ is smooth, and the bound (48) is minimized. Equation (48) goes to zero as the tails $G_t(f)$ go to zero, which implies that ideally, we want g to be bandlimited. If g is constrained to be in some class of signals, for example, signals of finite duration T, then minimizing $\mathcal{E}(\tau)$ is a constrained optimization problem. The desired $g(t)$ should be the element within the allowable class that is as close as possible to being bandlimited.

REFERENCES

[1] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM Rev.*, vol. 31, pp. 614–627, Dec. 1989.

[2] R. Kronland-Martinet, J. Morlet, and A. Grossman, "Analysis of sound patterns through wavelet transforms," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 1, pp. 97–126, 1987.

[3] F. Bao and N. Erdöl, "Optimal initial phase wavelet transform," in *Proc. 6th IEEE Digital Signal Process. Workshop*, Adelaide, Australia, 1994, pp. 187–190.

[4] I. Cohen, S. Raz, and D. Malah, "Shift invariant wavelet packet bases," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Detroit, MI, May 1995, pp. II-1081–II-1084.

[5] S. D. Marco, P. Heller, and J. Weiss, "An M-band, 2-dimensional translation invariant wavelet transform and applications," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Detroit, MI, May 1995, pp. II-1077–II-1080.

[6] J. E. Odegard, R. A. Gopinath, and C. S. Burrus, "Optimal wavelets for signal decomposition and the existence of scale limited signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, San Francisco, CA, Oct. 1992, pp. IV-597–IV-600.

[7] R. A. Gopinath, J. E. Odegard, and C. S. Burrus, "Optimal wavelet representation of signals and the wavelet sampling theorem," *IEEE Trans. Circiuts Syst. II*, vol. 41, pp. 262–277, Apr. 1994.

[8] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inform. Theory*, vol. 38, pp. 587–607, Mar. 1992.

[9] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992, Regional Conf. Series Appl. Math.

[10] F. Bao and N. Erdöl, "The optimal wavelet transform and translation invariance," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Adelaide, Australia, Apr. 1994, pp. III-13–III-16.

[11] H. Zou and A. H. Tewfik, "Parametrization of compactly supported orthonormal wavelets," *IEEE Trans. Signal Processing*, vol. 41, pp. 1428–1431, Mar. 1993.

[12] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Trans. Signal Processing*, vol. 40, pp. 2207–2232, Sept. 1992.

[13] S. A. Benno and J. M. F. Moura, "Shiftable representations and multipath processing," in *Proc. IEEE-SP Int. Symp. Time-Freq. Time-Scale Anal.*, Oct. 1994, pp. 397–400.

[14] ——, "Nearly shiftable scaling functions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Detroit, MI, May 1995, pp. II-1097–II-1100.

[15] ——, "On translation invariant subspaces and critically sampled wavelet transforms," *Multidimensional Syst. Signal Process., Special Issue on Wavelets Multiresolution Anal.*, vol. 8, pp. 89–110, Jan. 1997; invited paper.

[16] M. He, J. M. F. Moura, and S. Benno, "Gap detector for multipath," in *Proc. ICASSP, IEEE Int. Conf. Acoust., Speech, Signal Process.*, Atlanta, GA, May 1996, vol. V, pp. 2650–2653.

[17] C. He and J. M. F. Moura, "Robust detection with the gap metric," *IEEE Trans. Signal Processing*, vol. 45, pp. 1591–1604, June 1997.

[18] ——, "Focused detection via multiresolution analysis," *IEEE Trans. Signal Processing, Special Issue Theory and Applications of Filter Banks and Wavelet Transforms*, vol. 46, pp. 1094–1104, Apr. 1998.

[19] A. J. E. M. Janssen, "The Zak transform: A signal transform for sampled time-continuous signals," *Philips J. Res.*, vol. 43, pp. 23–69, 1988.

[20] L. Auslander and R. Tolimieri, "Radar ambiguity functions and group theory," *SIAM J. Math. Anal.*, vol. 16, pp. 577–601, May 1985.

[21] P. Heller, H. W. Resnikoff, and J. R. O. Wells, "Wavelet matrices and representation of discrete functions," in *Wavelets: A Tutorial in Theory and Applications*, C. K. Chui, Ed. New York: Academic, 1992, pp. 15–51.

[22] P. Steffen, P. Heller, R. A. Gopinath, and C. S. Burrus, "Theory of regular M-band wavelet bases," *IEEE Trans. Signal Processing,* vol. 41, pp. 3497–3511, Dec. 1993.

[23] A. H. Tewfik, D. Sinha, and P. Jorgensen, "On the optimal choice of a wavelet for signal representation," *IEEE Trans. Inform. Theory,* vol. 38, pp. 747–765, Mar. 1992.

[24] Y. Zhuang and J. S. Baras, "Optimal wavelet basis selection for signal representation," *Proc. SPIE,* pp. 200–210, 1994.

[25] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.,* vol. 41, pp. 909–996, Jan. 1988.

[26] S. G. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Pattern. Anal. Machine Intell.,* vol. 11, pp. 674–693, July 1989.

[27] A. Cohen, "Biorthogonal wavelets," in *Wavelet Analysis and its Applications*, Vol. 2 Wavelets: A Tutorial in Theory and Applications, C. K. Chui, Ed. New York: Academic, 1992, pp. 123–152.

[28] M. Vetterli, "Filter banks allowing perfect reconstruction," *Signal Process.,* vol. 10, no. 10, pp. 219–244, 1986.

[29] A. J. E. M. Janssen, "The Zak transform and sampling theorems for wavelet subspaces," *IEEE Trans. Signal Processing,* vol. 41, pp. 3360–3364, Dec. 1993.
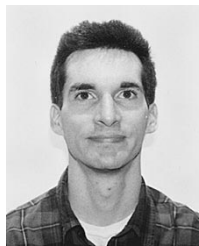
Steven A. Benno (S'86–M'96) received the B.Sc. degree in electrical engineering from Rutgers University, Piscataway, NJ, in 1987, the M.Sc. degree in electrical engineering from Columbia University, New York, NY, in 1989, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 1995.

From 1987 to 1991, he was with Bell Laboratories, Whippany, NJ, working in the area of underwater acoustics. After receiving the Ph.D. degree, he has been working for Lucent Technologies' Network Wireless Services Speech and Audio Group, Whippany, NJ, since 1996. His interests are in speech and audio compression, multimedia, and multiresolution processing.

José M. F. Moura (S'71–M'75–SM'90–F'94) received the Engenheiro Electrotécnico degree in 1969 from Instituto Superior Técnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and D.Sc. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 1973 and 1975, respectively.

He is presently a Professor of Electrical and Computer Engineering at Carnegie Mellon University (CMU), Pittsburgh, PA, which he joined in 1986. Prior to this, he was on the faculty of IST, where he was an Assistant Professor in 1975, Professor Agregado in 1978, and Professor Catedrático in 1979. He has had visiting appointments at several institutions, including MIT (Genrad Associate Professor of Electrical Engineering and Computer Science from 1984 to 1986) and the University of Southern California (Research Scholar, Department of Aerospace Engineering, Summers between 1978 and 1981). His research interests include statistical signal processing (one- and two-dimensional), digital communications, image and video processing, radar and sonar, and multiresolution techniques. He has organized and codirected two international scientific meetings on signal processing theory and applications. He has more than 190 published technical contributions and is coeditor of two books.

Dr. Moura is currently the Editor-in-Chief for the IEEE Transactions on Signal Processing, an elected member of the Board of Governors of the IEEE Signal Processing Society, and a member of the IEEE Signal Processing Society Publications Board. He is a member of the Underwater Acoustics Technical Committee and of the Multimedia Signal Processing Technical Committee of the IEEE Signal Processing Society. He was a member of the IEEE Press Board from 1991 to 1995, a Technical Associate Editor for the IEEE Signal Processing Letters from 1993 to 1995, and an Associate Editor for the IEEE Transactions on Signal Processing from 1988 to 1992. He was a program committee member for the IEEE International Conference on Image Processing (ICIP'95) and for the IEEE International Symposium on Information Theory (ISIT'93). He is a corresponding member of the Academy of Sciences of Portugal (Section of Sciences). He is affiliated with several IEEE societies, Sigma Xi, AMS, IMS, and SIAM.