Applied and Computational Harmonic Analysis **6,** 219–251 (1999)
Article ID acha.1998.0249, available online at http://www.idealibrary.com on IDEAL®

Approximation Error for Quasi-Interpolators and (Multi-)Wavelet Expansions

Thierry Blu

Swiss Federal Institute of Technology EPFL-DMT/JOR, CH-1015 Lausanne, Switzerland

E-mail: Thierry.Blu@epfl.ch

and

Michael Unser

Swiss Federal Institute of Technology, EPFL-DMT/IOA, CH-1015 Lausanne, Switzerland

E-mail: michael.unser@epfl.ch

Communicated by Charles K. Chui

Received April 7, 1997; revised January 19, 1998

We investigate the approximation properties of general polynomial preserving operators that approximate a function into some scaled subspace of $\mathbf{L}^2$ via an appropriate sequence of inner products. In particular, we consider integer shift-invariant approximations such as those provided by splines and wavelets, as well as finite elements and multi-wavelets which use multiple generators. We estimate the approximation error as a function of the scale parameter T when the function to approximate is sufficiently regular. We then present a generalized sampling theorem, a result that is rich enough to provide tight bounds as well as asymptotic expansions of the approximation error as a function of the sampling step T. Another more theoretical consequence is the proof of a conjecture by Strang and Fix, which states the equivalence between the order of a multi-wavelet space and the order of a particular subspace generated by a single function. Finally, we consider refinable generating functions and use the two-scale relation to obtain explicit formulae for the coefficients of the asymptotic development of the error. The leading constants are easily computable and can be the basis for the comparison of the approximation power of wavelet and multi-wavelet expansions of a given order L.

I. INTRODUCTION

Obtaining a discrete representation of a function is an unavoidable step if one wishes to develop numerical methods for solving problems that are formulated in the continuous

1063-5203/99 $30.00

domain. A general approach is to project the input function $f(x) \in \mathbf{L}^2$ onto an appropriate subspace V_T of $\mathbf{L}^2$, where T is a scale (or bandwidth) parameter that determines the quality of the approximation. One of the simplest forms of projection is interpolation where the function and its approximation are in perfect agreement at some specified grid points.

In digital signal processing (DSP), this discretization is obtained by sampling, using Shannon's sampling theorem [41], which allows one to represent a bandlimited function by a countable set of uniform samples. Usually, the considered functions do not satisfy the conditions of the sampling theorem exactly but this discrepancy is not detrimental: the important point is that the quality of the representation improves as the sampling frequency increases. Most of the research done in this field is thus made with data already discretized. However, the sampling theorem, which is at the basis of the theory, is still used regularly when continuous operations (such as a 2D rotation, or a noninteger time delay) have to be implemented in a discrete manner. In these cases, it is highly desirable to have a good handle on the approximation error, especially if one uses alternative signal representations such as splines and wavelets [48]. A precise characterization of the approximation power of wavelet bases may also be very valuable for coding applications [5, 38].

Interpolation methods have been studied extensively in approximation theory and areas of applied mathematics where the goal is to discretize continuous equations that cannot be solved easily without the help of a computer. These problems often involve several dimensions, which explains why most results in approximation theory are published for multivariate functions. The underlying process is otherwise very similar to the sampling scheme, with the important difference that here the approximation error is a parameter that has to be estimated, since in that case, the adequacy of fit between continuous and discrete data has to be controlled. The philosophy here is rather different from the one which prevails in DSP: emphasis is placed on the *regularity* of the function instead of on its bandlimited character.

In this paper, which we have divided into three parts, we present a detailed and quantitative characterization (Section III) of the $\mathbf{L}^2$ error introduced by such linear approximation schemes under very general conditions on the functions to be approximated and on the analysis and synthesis functions which define the approximation algorithm (Section II). An application (Section IV) to the special case of refinable generators shows the potential of our general results.

An aspect of our paper is the consideration of representation spaces generated from equidistant translates of several functions ("multi-wavelets"), instead of only one ("wavelets"). In order to preserve the sampling density, we use a shift increment that is a multiple of the number of generating functions q, so that our basic representation spaces are q-integer shift-invariant. We use a vector formalism well adapted to the study of multiwavelets, which have attracted much attention recently [1, 15, 27, 28, 51].

For our analysis, we have purposely chosen to consider a very broad class of linear approximation operators. An interesting subset of them includes the cases usually designated by quasi-interpolants [11, 20, 24, 42], the various types of projectors encountered in the context of the wavelet transform [14, 45], but also more general polynomial preserving operators that have been studied recently [9, 29, 33]. A general account of quasi-interpolation can be found in [21]. Here, we will see that the order constraint is translated into a simple moment condition for the analysis functions—a very weak form of biorthonormality, which we call "quasi-biorthonormality."

Our key result is a formula for the approximation error (Theorem 1) and its "phase average" (Theorem 2) which should be relevant for both DSP and approximation theory. Alternatively, Theorem 1 can be seen as a sampling theorem for which the sampling frequency, $\frac{1}{T}$, is the most important parameter or as an approximation theorem for which the regularity of the function is of greater concern. Using this result, we are also able to identify new conditions on the synthesis functions under which the Strang–Fix equivalence [42] is true. We finally show how to obtain asymptotic expansions and upper bounds for the expansion error: this constitutes a wide extension of what is already known in the literature [44–46, 48].

When there is an underlying multiresolution structure the computation of asymptotic constants is made much more tractable. Thus, our formulation applies directly to the various multiresolution spaces of the wavelet transform [19, 35, 36], which corresponds to the simpler case $q = 1$. The case is well known for wavelets, but is also of interest for multi-wavelets, which are a more recent construction [1, 15, 27, 51]. One advantage of multi-wavelets is that one can construct orthonormal, symmetric, and compactly supported basis functions, which is not possible otherwise. The multi-wavelet framework also allows for richer classes of piecewise polynomial functions including splines and Hermite polynomials [27]. Our present contribution is an exact computation of the asymptotic form of the approximation error as a function of the generating filters. This result is also an improvement for the wavelet case since it adds higher order terms to the asymptotic result in [45]. All these results, to our knowledge, are new.

Note that the unconstrained character of our hypotheses may also have some practical advantages: it may suggest alternative approximation procedures that are essentially as adequate as the least squares solution, but much easier to implement because the analysis functions can be much shorter (e.g., [46]).

To keep the presentation simple and understandable, we have chosen to concentrate on the case of univariate functions. There is no fundamental difficulty in extending our results to multiple dimensions: this can be checked theorem after theorem, at least for our theoretical results.

A. Notation

Our analysis relies heavily on Fourier techniques. We define the Fourier transform $\hat{f}$ of a function f by $\hat{f}(\nu) = \int f(x)e^{2i\pi\nu x}dx$. Also, we consider distributions [40] and, in particular, the Dirac mass $\delta(x)$ and its derivatives $\delta^{(s)}(x)$. With a similar notation, the Kronecker symbol δ_n is defined as the sequence which takes the value 1 for $n = 0$, and 0 for $n \neq 0$.

We recall the well-known Poisson summation formula which holds in the sense of distributions for every $\mathbf{L}^1$ function $f(x)$,

$$\sum_n \hat{f}(n)e^{2i\pi nx} - \sum_n f(x+n) = 0. \tag{1}$$

This means that, for any test (i.e., compactly supported and indefinitely differentiable) function $\psi(x)$, the scalar product between the lhs of (1) and $\psi(x)$ vanishes.

The usual operations acting on a complex number z, i.e., conjugation and real part, are denoted by, respectively, $\bar{z}$ and $\Re\{z\}$.

The conventional inner product $\int \overline{f(x)} g(x) dx$ between two complex $\mathbf{L}^2$ functions f, g is denoted $\langle f, g \rangle$, and the associated euclidean norm is $\| \cdot \|_{\mathbf{L}^2}$. We also denote by $\| \cdot \|_{\mathbf{L}^2(I)}$ the local $\mathbf{L}^2$ norm over an interval I.

The usual $\mathbf{L}^\infty$ norm is denoted by $\|f\|_\infty$ and equals $\sup_x |f(x)|$. For simplicity, we also denote $\|\mathrm{V}(\nu)\|_\infty = \sup_\nu \|\mathrm{V}(\nu)\|$ in the case where V is a vector function; no confusion should arise from the two similar notations, since we make a typographic distinction between scalar and vector functions (see below).

Let r be a positive real number. The Sobolev space $\mathbf{W}_2^r$ is defined as the collection of functions satisfying $\int (1 + \nu^2)^r |\hat{f}(\nu)|^2 d\nu < \infty$. In line with this definition of regularity, we extend $\|f^{(r)}\|_{\mathbf{L}^2}$ to noninteger values of r by equating it to the square root of $\int |2\pi\nu|^{2r} |\hat{f}(\nu)|^2 d\nu$.

For some estimates we need the Riemann zeta function defined as $\zeta(s) = \sum_{n \geq 1} n^{-s}$ for all real $s > 1$.

Filters are described either by polynomials (transfer function) or by their impulse response, i.e., the coefficients of these polynomials. To make the distinction clear, our convention is to use a lower-case letter for the coefficients, and upper-case for the corresponding polynomial. Knowing that scalar functions are italic, vector functions are roman, and matrix functions are boldface roman, we thus follow the notation described below:

$$\textit{scalar } H(z) = \sum_n h_n z^n$$

$$\text{vector } \mathrm{H}(z) = \sum_n \mathrm{h}_n z^n$$

$$\textbf{matrix } \mathbf{H}(z) = \sum_n \mathbf{h}_n z^n.$$

The square root of a positive Hermitian matrix $\mathbf{A}$ can be uniquely defined as the positive Hermitian matrix $\mathbf{B}$ such that $\mathbf{B}^2 = \mathbf{A}$: this square root is denoted by $\sqrt{\mathbf{A}}$, as if $\mathbf{A}$ were a scalar.

Most asymptotic expansions are presented with "$o(\,\cdot\,)$" and "$O(\,\cdot\,)$" terms, which allows us to give a more compact and understandable form to the results: writing $f(x) = o(x^n)$ is equivalent to writing $\lim_{x \to 0} \sup |f(x)/x^n| = 0$; in the same spirit, writing $f(x) = O(x^n)$ is equivalent to $\lim_{x \to 0} \sup |f(x)/x^n| < \infty$ (not necessarily 0).

II. APPROXIMATION BY MULTI-WAVELETS

The theories developed for digital signal processing would never have been possible without the tool of sampling and its dual, interpolation. The problem of the description of a continuous signal by a discrete sequence of numbers was first solved for the bandlimited case [41, 50]. In the classical result, a bandlimited function $f(t)$ can be written as a linear interpolation $\mathcal{I}_T f(t)$ of its equally spaced samples $f(nT)$, provided that the sampling frequency, T^{-1}, is at least twice the signal's bandwidth. This interpolation is written as

$$\mathcal{I}_T f(t) = \sum_n f(nT)\operatorname{sinc}\left(\frac{t}{T} - n\right), \tag{2}$$

where $\operatorname{sinc}(x) = \sin(\pi x)/(\pi x)$.

A. Invariances

The Shannon–Nyquist interpolation exhibits many "good" properties, in particular so far as scale changes and time shifts are concerned. For instance, if we denote by V_T the interpolation space generated by the sinc basis (i.e., $V_T = \operatorname{span}_{n\in\mathbb{Z}}\left\{\operatorname{sinc}\left(\frac{x}{T} - n\right)\right\}$ restricted to $\mathbf{L}^2$ functions) then for each f in V_T we have the following properties:

$$\forall \tau \in \mathbb{R} \qquad f_\tau = f(\cdot + \tau) \in V_T \qquad \text{(general shift invariance)} \tag{3}$$

$$\forall a \geq 1 \qquad f(a^{-1}\cdot) \in V_T \qquad \text{(general scale invariance).} \tag{4}$$

On the other hand, there are "bad" features, such as the infinite support of any nontrivial function of V_T, and more precisely the fact that the interpolation formula (2) is slowly convergent. This prevents us from giving the Nyquist function a local meaning.

These remarks led to a first generalization of the notion of interpolation through the loss of a certain amount of shift invariance [4, 42]. More specifically, it was recognized that defining an approximation space $V_T = \operatorname{span}_{n\in\mathbb{Z}}\left\{\varphi\left(\frac{t}{T} - n\right)\right\} \cap \mathbf{L}^2$, where φ is a function with acceptable frequency and time localization, can be more robust and useful in practical applications than the space of bandlimited functions. In that case, shift invariance is now satisfied only for integer increments; thus, the generalization operates by replacing (3) with the much weaker property

$$\forall \tau \in T\mathbb{Z} \qquad f_\tau = f(\cdot + \tau) \in V_T \qquad \text{(integer shift invariance)} \tag{5}$$

for all $f \in V_T$: this property is satisfied by wavelet-like approximation spaces. It can be further generalized by considering q generating functions instead of one: in order to keep the space density constant, we define $V_T = \operatorname{span}_{n\in\mathbb{Z}, i=0\ldots q-1}\left\{\varphi_i\left(\frac{t}{T} - nq\right)\right\} \cap \mathbf{L}^2$; now, (3) is replaced with

$$\forall \tau \in qT\mathbb{Z} \qquad f_\tau = f(\cdot + \tau) \in V_T \qquad (q\text{-integer shift invariance}) \tag{6}$$

for all $f \in V_T$: this property is satisfied by multi-wavelet-like approximation spaces.

Similarly, the weaker forms of scale invariance led to the concept of multiresolution analysis [35, 36]. In the usual case of interest where φ is compactly supported and where V_T is a wavelet-like space, it has been shown [13] that (4) must be replaced by

$$\forall n \in \mathbb{N} \qquad f(a_0^{-n}\,\cdot\,) \in V_T \qquad \text{(integer scale invariance)}, \tag{7}$$

where $a_0 \geq 2$ is a positive integer (scale factor). Yet, noninteger scale factors (of particular interest is the case $1 < a_0 < 2$) are allowed, if we give up the shift invariance property [7]. Among the integer scale-invariant functions such as Daubechies "father" wavelets [18] (where $a_0 = 2$), B-splines have the noteworthy feature of being integer scale-invariant for *any* positive integer scale factor a_0.

B. Multi-wavelets and Their Vector Formalism

In this paper, we concentrate on q-integer shift-invariant approximation spaces, which are generated by a finite number, q, of shifted functions φ_k. We shall refer to the $\varphi_k(x)$'s as the *generating functions* or *generators* of the representation, or also as *synthesis functions*. Note that some authors also use the term "finitely generated shift-invariant" (FSI) space [22]. In order to simplify some further expressions, we shall denote these generating functions (also qualified as multi-wavelets here, though they are of the "father" kind [36]) by φ_n where n spans $\mathbb{Z}$, instead of emphasizing the shift dependence of the φ_n, i.e., $\varphi_{n_0+n_1q}(x) = \varphi_{n_0}(x - n_1q)$, where $n_0 = 0, \ldots, q-1$ and $n_1 \in \mathbb{Z}$.

We do not distinguish between the cases of single and multiple generators, because the mathematics are exactly the same. As a matter of fact, it can be shown [51] that multi-wavelets behave like *vector* wavelets, and this is why it will also prove efficient to use the following equivalent vector notation: to the φ_n's, we associate a q-vector (i.e., a vector of length q) $\boldsymbol{\varphi}(x) = (\varphi_0, \varphi_1, \ldots, \varphi_{q-1})^{\mathrm{t}}$.

As in the wavelet case, it is possible to orthonormalize the q-shifted vector functions through matrix filtering. Indeed, if we let

$$\mathbf{A}(\nu) = \frac{1}{q}\sum_n \hat{\boldsymbol{\varphi}}\left(\frac{\nu+n}{q}\right)\overline{\hat{\boldsymbol{\varphi}}\left(\frac{\nu+n}{q}\right)}^{\mathrm{t}}, \tag{8}$$

then $\mathbf{A}(\nu)$ is Hermitian, positive and 1-periodic so that we can define $\mathbf{G}(e^{2i\pi\nu}) = \sqrt{\mathbf{A}(\nu)}$. Assuming that $\mathbf{A}(\nu)$ is invertible (which will ensue from our hypotheses), we can define the orthonormalized generating functions through their vector formulation [26]

$$\hat{\boldsymbol{\phi}}(\nu) = \mathbf{G}(e^{2i\pi q\nu})^{-1}\hat{\boldsymbol{\varphi}}(\nu) \tag{9}$$

$$\hat{\mathring{\boldsymbol{\phi}}}(\nu) = \mathbf{G}(e^{2i\pi q\nu})\hat{\mathring{\boldsymbol{\varphi}}}(\nu). \tag{10}$$

Expressed using this orthonormalized basis ϕ_n, our results take a particularly simple form.

C. Approximation Method

The error induced by the approximation process will be evaluated using the $\mathbf{L}^2$ norm: as will be seen later (Section III), this choice makes it possible to obtain an explicit computation of the approximation error.

We now need to define a linear functional operator $\mathcal{S}_T$, providing the coefficients of the linear decomposition (12) of the approximated function, on the φ_n's. The form of this operator is

$$\mathcal{S}_T\colon f \mapsto \left\{ \int f(\tau)\tilde{\varphi}_n\left(\frac{\tau}{T}\right) d\,\frac{\tau}{T} \right\}_{n\in\mathbb{Z}}, \tag{11}$$

where we assume that $\tilde{\varphi}_n$ are q-integer shift invariant distributions, i.e., once again $\tilde{\varphi}_{n+q}(x + q) = \tilde{\varphi}_n(x)$ for all $n \in \mathbb{Z}$. In analogy with Shannon's sampling theorem (2), we interpret $\mathcal{S}_T$ as a sampling operator, and we have the approximation operator

$$\begin{aligned} \mathcal{Q}_T f(t) &= \sum_n \mathcal{S}_T(f)_n \varphi_n\left(\frac{t}{T}\right) \\ &= \sum_n \int f(\tau)\tilde{\varphi}_n\left(\frac{\tau}{T}\right)\varphi_n\left(\frac{t}{T}\right) d\,\frac{\tau}{T}. \end{aligned} \tag{12}$$

Using the vector formalism for the sampling and synthesis sides, this fundamental approximation formula can be rewritten as

$$\mathcal{Q}_T f(t) = \sum_n \int f(\tau)\tilde{\varphi}\left(\frac{\tau}{T} - nq\right)^{\mathrm{t}} \varphi\left(\frac{t}{T} - nq\right) d\,\frac{\tau}{T}, \tag{13}$$

where $\tilde{\varphi}(x) = (\tilde{\varphi}_0, \tilde{\varphi}_1, \ldots, \tilde{\varphi}_{q-1})^{\mathrm{t}}$.

Of course, in order for these definitions to have a meaning (e.g., convergence of the summation in (12), existence of the scalar product in (11)) we must restrict somewhat the choice of f, φ_n, and $\tilde{\varphi}_n$. Our hypotheses will be given in subsection II-D.

Now, following the de Boor and others [20, 24, 45], we define the notion of "order."

DEFINITION 1. The set of synthesis functions is *of order* L iff there exist L real sequences $\{\lambda_n^{(s)}\}_{n\in\mathbb{Z}}$ such that, in the sense of distributions

$$x^s = \sum_n \lambda_n^{(s)} \varphi_n(x) \tag{14}$$

for $s = 0 \cdots L - 1$.

In general, (14) is assumed to hold pointwise, a property which is necessary when one considers $\mathbf{L}^\infty$ error measure; this is not our case, the $\mathbf{L}^2$ norm being *global*, so that we can use a weaker[1] condition: for additional hypotheses on the functions φ_n, so that the Poisson's summation formula hold pointwise, see [32].

In the approximation scheme considered here, unlike the minimal approximation

[1] It is not true that the pointwise version of (14) is stronger than its distribution version, but this becomes true once the pointwise convergence is uniform over any closed interval.

scheme, the sampling distributions and the synthesis functions are independent parameters: this is why we need to define another notion of order, based on both of them.

DEFINITION 2. We say that a set of sampling distributions $\{\tilde{\varphi}_n\}_{n\in\mathbb{Z}}$ and synthesis functions $\{\varphi_n\}_{n\in\mathbb{Z}}$ constitute a *quasi-biorthonormal* set of order L if and only if the following two conditions are met:

- the functions φ_n are of order L;
- the distributions $\tilde{\varphi}_n$ satisfy the moment conditions

$$\int x^s \tilde{\varphi}_n(x)\,dx = \lambda_n^{(s)} \qquad \text{for} \begin{cases} n \in \mathbb{Z} \\ s = 0 \cdots L-1, \end{cases} \tag{15}$$

where the $\lambda_n^{(s)}$ are assumed to satisfy (14).

It is easy to understand why this property is called "quasi-biorthonormality of order L": if the sampling and the interpolating functions were exactly biorthonormal then, under technical convergence hypotheses, it would be possible to use the approximation formula (12) to recover any function—including the monomials (14)—that lies in the span of the φ_n's. Here, we are primarily interested in the polynomials which can be reconstructed exactly under the much weaker condition (15).

Note that the Lth-order quasi-biorthonormality property puts a rather strong constraint on the synthesis functions, i.e., the reproduction of the polynomials of degree $L-1$ (cf. (14)). The moment condition (15) on the sampling distributions is much less constraining and leaves room for many design alternatives.

D. Hypotheses

We shall have to make substantial assumptions on the approximation scheme, which in turn will provide us with very sharp estimates together with the theoretical equivalence between quasi-biorthonormality of order L and approximation error of order L (see Theorem 3 below).

Hypotheses on the synthesis functions. We shall assume that the synthesis functions φ_n are in $\mathbf{L}^2$ and satisfy the Riesz condition $A\sum_n |c_n|^2 \le \|\sum_n c_n\varphi_n\|^2_{\mathbf{L}^2} \le B\sum_n |c_n|^2$ for any l^2 sequence $\{c_n\}_{n\in\mathbb{Z}}$, and where $0 < A \le B < \infty$. Following the proof of [2, 4], one finds that this requirement is equivalent to the condition

$$A\mathbf{I}_q \le \mathbf{A}(\nu) \le B\mathbf{I}_q \tag{16}$$

for almost every $\nu \in \mathbb{R}$. In this paper, we require slightly more, specifically that this inequality hold pointwise, i.e., for all $\nu \in \mathbb{R}$: actually, this subtlety is largely theoretical since in practical cases $\mathbf{A}(\nu)$ is always continuous. The important Riesz hypothesis means that the functions are linearly independent *whenever l^2 coefficients sequences are considered.* However, this does not rule out the possibility that non-l^2 sequences make the linear combination cancel for every value of x. Besides, a close examination of our proofs indicates that it might be sufficient to restrict the lower Riesz inequality to the invertibility of $\mathbf{A}(0)$, at least for our theoretical results (Section III).

In order to be able to derive some valuable equivalences in our theorems, we add another, more exotic assumption, which limits the amount of linear dependence of the basis functions. Let $\mathcal{M}$ be the vector space of all possible coefficients λ_n such that $\sum_n \lambda_n \varphi_n = 0$ in the sense of distributions. Then our hypothesis is that the dimension of $\mathcal{M}$ is finite. In practice, this encompasses most interesting situations, such as *strict* linear independence (i.e., $\dim \mathcal{M} = 0$) or compactly supported generating functions which satisfy the Riesz condition (proof in Appendix A).

Finally, we assume that when the φ_n are of order L then the integrals $\int |x - n|^k |\varphi_n(x)| dx$ are finite for every $n = 0, \ldots, q - 1$ and for $k = 0, \ldots, L$. This guarantees that the Fourier transforms of the synthesis functions are L times continuously differentiable with bounded derivatives.

Hypothesis on the sampling distributions. On the sampling side, we only assume that the $\tilde{\varphi}_n$'s have a *bounded* Fourier transform: this in particular allows us to consider distributions such as the Dirac mass.

Hypothesis on the functions to approximate. For (12) to generate a function that belongs to $\mathbf{L}^2$, we need to restrict the choice of f so that the sequence $\{\mathcal{S}_T(f)_n\}_{n\in\mathbb{Z}}$ is in l^2. We reduce the set of admissible functions to those which are in $\mathbf{W}_2^r$ with $r > \frac{1}{2}$, for it can be shown (see Appendix C) that this constraint ensures the convergence of $\sum_n |\mathcal{S}_T(f)_n|^2$. Note that this requirement is slightly stronger than continuity, since it implies that f is Hölder continuous with exponent $r - \frac{1}{2}$.

III. $\mathbf{L}^2$ APPROXIMATION ERROR

Independently of the properties of the sampling/synthesis functions, we are interested in evaluating the quantity

$$\epsilon_f = \| f - \mathcal{Q}_T f \|_{\mathbf{L}^2}. \tag{17}$$

As in the wavelet case, it is possible to express the Lth-order property of the synthesis functions equivalently in the Fourier domain of the vector functions.

LEMMA 1. *The three following properties are equivalent:*

(i) *The synthesis functions φ_n are of order L.*

(ii) *There exists a unique polynomial q-vector* P (*i.e., with vectors of size q as coefficients*) *of degree less than or equal to $L - 1$ such that*

$$\overline{\mathrm{P}(\nu)}^{\mathrm{t}} \hat{\varphi}\left(\nu + \frac{n}{q}\right) = q\delta_n + O(\nu^L) \tag{18}$$

for all integers n.

(iii) *There exists a unique coefficient sequence* $\{b_k\}$ *where* $k = 0, \ldots, qL - 1$ *such that the function defined by*

$$\psi = \sum_{k=0}^{qL-1} b_k \varphi_k \tag{19}$$

satisfies the Strang–Fix conditions $\hat{\psi}\left(\nu + \frac{n}{q}\right) = q\delta_n + O(\nu^L)$ *for all integers n.*

The proof is given in Appendix B.

This result can be compared to what is known in the approximation theory literature [23, 32, 42], where a form of this lemma is given; our form is slightly more general in that it does not assume that the coefficients $\lambda_n^{(s)}$ of (14) have a polynomial character (one speaks of "polynomial preservation" [32, 39], not of *polynomial reconstruction*): however, the finite dimension hypothesis on $\mathcal{M}$ implies (see Appendix B) that the $\lambda_n^{(s)}$ are actually polynomial; other hypotheses [37] also lead to the same conclusion. Our lemma goes even a little further: it provides the length of the minimum linear combination (see also [37]) of synthesis functions for building the function ψ and also shows the uniqueness of this *minimal* linear combination. Finally, note that the generating functions considered here are not necessarily compactly supported: their Fourier transforms must be L times differentiable.

Using the minimal polynomial $\mathrm{P}(\nu)$ of Lemma 1, we can construct many equivalent quasi-biorthonormal sets of order L such that $\hat{\tilde{\varphi}}(\nu) = \mathrm{P}(\nu) + O(\nu^L)$ (cf. proof of Lemma 1). As a consequence of the uniqueness of this minimal polynomial, a more precise result, namely the equivalence between Lth-order quasi-biorthonormality and this equation can be stated.

LEMMA 2. *The two following properties are equivalent:*

(i) *The multi-wavelets* φ_n *and* $\tilde{\varphi}_n$ *constitute a quasi-biorthonormal set of order L.*
(ii) *The Fourier transform of the vector functions satisfy*

$$\overline{\hat{\tilde{\varphi}}(\nu)}^{\mathrm{t}} \hat{\varphi}\left(\nu + \frac{n}{q}\right) = q\delta_n + O(\nu^L) \tag{20}$$

for all integers n.

A. Approximation Results

In the following, we state that the approximation error ϵ_f can be written as a main term plus a perturbation. The dominating component can be computed exactly by integration of $|\hat{f}(\nu)|^2$ against the kernel

$$E(\nu) = \frac{1}{q^2}\left|q - \overline{\hat{\tilde{\varphi}}(\nu)}^{\mathrm{t}}\hat{\varphi}(\nu)\right|^2 + \frac{1}{q^2}\sum_{n\neq 0}\left|\overline{\hat{\tilde{\varphi}}(\nu)}^{\mathrm{t}}\hat{\varphi}\left(\nu + \frac{n}{q}\right)\right|^2. \tag{21}$$

The additional correction, $e(f, T)$, can take positive or negative values and its magnitude depends on the (Sobolev-)regularity exponent of the function to be approximated. It becomes negligible when the sampling step T is small relative to the intrinsic smoothness scale of f.

Note that condition (ii) of Lemma 2 appears in a canonical way in (21), which suggests that the kernel $E(\nu)$ is thus intimately related to the quasi-biorthonormality of the sampling/synthesis scheme. In the orthonormalized basis (provided $\mathbf{A}(\nu)$ is invertible), this kernel takes the much simpler form

$$E(\nu) = 1 - \frac{2}{q}\,\Re\{\overline{\hat{\tilde{\phi}}(\nu)}^{\mathrm{t}}\hat{\phi}(\nu)\} + \frac{1}{q}\,\|\hat{\tilde{\phi}}(\nu)\|^2 \tag{22}$$

$$= 1 - \frac{1}{q}\,\|\hat{\phi}(\nu)\|^2 + \frac{1}{q}\,\|\hat{\phi}(\nu) - \hat{\tilde{\phi}}(\nu)\|^2. \tag{23}$$

The following theorem relates $E(\nu)$ to the approximation error ϵ_f and is the most important result of our paper.

THEOREM 1. *For all* $f \in \mathbf{W}_2^r$ *with* $r > \frac{1}{2}$, the approximation error is given by

$$\epsilon_f = \|f - \mathcal{Q}_T f\|_{\mathbf{L}^2} = \left[\int |\hat{f}(\nu)|^2 E(T\nu)\,d\nu\right]^{1/2} + e(f, T), \tag{24}$$

where the correction term $e(f, T) = o(T^r)$ *is bounded as*

$$|e(f, T)| \leq K T^r \|f^{(r)}\|_{\mathbf{L}^2} \tag{25}$$

with

$$K = 2\,\frac{q^r}{\pi^r}\,\sqrt{\zeta(2r)\|E\|_\infty}. \tag{26}$$

In addition, the term $e(f, T)$ *in* (24) *exhibits a double aliasing character* (*in* f *and in* φ) *and cancels whenever one of the conditions*

$$\hat{f}(\nu)\hat{f}\left(\nu + \frac{n}{qT}\right) = 0 \qquad \text{or} \qquad \hat{\tilde{\phi}}\left(\nu + \frac{n}{q}\right)^{\mathrm{t}}(\hat{\tilde{\phi}}(\nu) - 2\hat{\phi}(\nu)) = 0 \tag{27}$$

holds for all frequency ν *and all non-zero integers* n.

The proof is given in Appendix C.

There is one related result in the literature in [22, Eq. (4.2)] that applies to the orthogonal projection onto a principal (i.e., $q = 1$) shift-invariant space. In this special case, $\tilde{\phi}_n = \phi_n$, which implies that $E(\nu) = 1 - |\hat{\phi}(\nu)|^2$; this is precisely the form of the kernel appearing in [22, Thm. 2.20], which gives the minimum error for the approximation of a bandlimited function. However, these authors did not push their error analysis further because their motivation was different from ours. They were interested primarily in finding an accurate link between the order of the expansion error and the Fourier transform of the generating function. Our present error formula is sharper and also more complete. Not only does it apply to multiple generators, but it also covers a much wider class of approximation operators for which the sampling and synthesis functions need not be the dual of each other; such approximation schemes find numerous applications in finite element analysis and digital signal processing.

Our result can also be interpreted as a sampling theorem. For the particular case $\varphi(x) = \mathrm{sinc}(x)$ and $\tilde{\varphi}(x) = \delta(x)$, we recover Shannon's sampling theorem [41]. However, Theorem 1 is more elaborate for it provides an estimate of the error when the functions are not bandlimited. Another extension is to consider the approximation problem in the case of multiple analysis or generating functions: from the second part of the theorem, we can deduce that one possible condition for the minimal expansion error to cancel is that f be bandlimited with frequency support $\left[-\frac{1}{2T}, \frac{1}{2T}\right]$, and that the synthesis functions φ_k be bandlimited with frequency support (respectively) $\left[-\frac{k+1}{2q}, -\frac{k}{2q}\right] \cup \left[\frac{k}{2q}, \frac{k+1}{2q}\right]$ for $k = 0, \ldots, q-1$. Also note that the latter condition is sufficient for the second error term to vanish if $\tilde{\varphi}_k$ and φ_k have the same frequency support. These observations are especially relevant for digital signal processing, where the bandlimited hypothesis plays a central role in the theory.

From now on, however, we will leave these sampling issues aside and use Theorem 1 in connection with the Lth-order biorthonormality property. Instead, we will exploit the fact that for f sufficiently regular, the leading term of the error when the sampling step T is small, is given by the order of $E(\nu)$ in the neighborhood of $\nu = 0$, providing the asymptotic development of ε_f near $T = 0$ (see Theorem 4). We believe that such an asymptotic Fourier analysis is new; the first order of the development is known for some projection operators but it is typically obtained via a pointwise analysis in the signal domain [45].

In addition to asymptotics and other valuable upper bounds on the error, (24) also offers a deeper insight into the theoretical relation between Lth-order quasi-biorthonormality and Lth power of the approximation error (through the formulation (21) of $E(\nu)$). This double character should make Theorem 1 appealing to approximation theorists as well as to researchers working in more applied fields.

Averaged approximation error. Let us now examine the problem caused by the lack of shift invariance in our approximation space. For this purpose, we choose to approximate the shifted version $f_u(t) = f(t-u)$ of the function f. The resulting error is a qT-periodic function of the shift increment u, i.e., $\epsilon_{f_{u+qT}} = \epsilon_{f_u}$. Moreover, since the first term in (24) is shift-invariant, the influence of the phase factor u can only appear in the second error component, which is therefore a direct reflection of the shift variance of the approximation space.

As the initial phase of the sampler is somewhat arbitrary, we choose to evaluate an averaged form of the error

$$\eta_f^2 = \frac{1}{qT}\int_0^{qT} \epsilon_{f_u}^2 du = \frac{1}{qT}\int_0^{qT} \| f_u - \mathcal{Q}_T f_u\|_{\mathbf{L}^2}^2 du, \tag{28}$$

which takes the periodicity into account. A remarkable fact is that this quantity can be computed *exactly* and reduces to the first term in (24) (the second error term cancels on the average). Thus if we assume that the phase of the sampler is random and uniformly distributed, the term $\int |\hat{f}(\nu)|^2 E(T\nu)d\nu$ provides us with the expected value of the approximation error.

THEOREM 2. *The average approximation error of the sampling/synthesis scheme is given by*

$$\eta_f^2 = \int |\hat{f}(\nu)|^2 E(T\nu)d\nu, \tag{29}$$

provided the Sobolev exponent of the function f is strictly greater than $\frac{1}{2}$.

Proof. In the Fourier domain, the shift change $f \to f_u$ is equivalent to the multiplication of $\hat{f}$ by the phase term $e^{-2i\pi u\nu}$, which explains why the first term of (24) does not depend on u. If we consider the form of the term ϵ_2^2 given in Appendix C by (62), its integrand is multiplied by a phase term $e^{-2i\pi(nu/qT)}$ which does not depend on ν. Due to the absolute convergence of both the summations and the integrals, we can interchange the order of these operations. Finally, we note that the average of $e^{-2i\pi(nu/qT)}$ over $[0, qT]$ cancels for $n \neq 0$. ■

This is a rather powerful result given that we started with an error formula in Theorem 1 that had the flavor of an upper bound. Thanks to Theorem 2, we now have an interpretable measure η_f that is easy to compute exactly and gives a correct estimate of the error for many practical applications.

B. The Strang–Fix Equivalence for Multiple Generators

We will now use Theorem 1 to make the connection with the Strang–Fix theory of approximation [42]. In the case of a single generator, there is a very direct equivalence between the ability of φ to reproduce polynomials of degree $n = L - 1$ (or the so-called Strang–Fix conditions in the Fourier domain) and the rate of decay of the minimum expansion error [10, 42] (initially least squares solution, later extended to the other $\mathbf{L}^2$ measures [31]). In [42], Strang and Fix conjectured that such an equivalence would also hold for multiple generators but their initial claim was put into question by the construction of a (rather involved) counterexample by Jia [30], discussed below. Several authors worked around the problem for $q > 1$ by adding constraints and introducing sophisticated notions of controlled approximation [16, 25, 32].

In our case, the situation appears to be more favorable and there is no major difficulty in extending the Strang–Fix equivalence for $q \geq 1$ without adding to our initial assumptions.

Theorem 3. *We assume the boundedness of φ_n and of $\sum_n |x - n|^{2L}|\varphi_n(x)|$, and also require that $\int |\tilde{\varphi}_n| < \infty$ and $\int |x - n|^{2L}|\tilde{\varphi}_n| < \infty$. With these hypotheses we have the equivalence*

$$\forall f \in \mathbf{W}_2^L,\ \epsilon_f = O(T^L) \Leftrightarrow \{\varphi_n\}_{n\in\mathbb{Z}} \text{ and } \{\tilde{\varphi}_n\}_{n\in\mathbb{Z}} \text{ are order } L \text{ quasi-biorthonormal.} \tag{30}$$

The proof of this theorem is given in Appendix D. Note that the method is rather straightforward due to the particular form (21) of $E(\nu)$: the difficulty is essentially technical (one cannot bluntly differentiate an infinite sum of functions).

There are three points that distinguish this result from what has been reported before. First, Theorem 3 is a constructive result (for a related construction, see also [33]) that specifies a whole class of linear q-integer shift-invariant approximation procedures (not only the minimal one) in V_T that provide an $O(T^L)$ error decay rate. Second (minimal error case), though our hypotheses are weaker in other aspects, the use of the Riesz basis hypothesis is more restrictive than the condition of a controlled approximation initially proposed by Strang and Fix [42], and implies it. Finally, Theorem 3 clarifies the discussion around the Strang–Fix equivalence. Because of Lemma 1, Theorem 3 shows the existence of a special function $\psi(x)$—a *finite* linear combination of basis functions—such that the error between the function to be approximated and the optimal expansion using q-integer shifts of ψ is of the same order as the global minimal expansion error (using the full basis set). It is the existence of such a function that has been a major point of contention in the literature [30, 32]. Note that de Boor *et al.* have shown recently that this problem disappears if one drops the restriction of finite linear combinations [22].

In a counterexample to Strang–Fix conjecture, Jia [30] considers functions (smoothed box splines) of two variables and builds a set of $q = 4$ generating functions such that the expansion error is of order two, while no finite l^2 linear combination of the synthesis functions satisfies the Strang–Fix conditions. In fact, such a linear combination exists, but is infinite as is explicitly written in [30]. As can be checked readily, Jia's multi-wavelets satisfy neither the Riesz condition nor $\dim \mathcal{M} < \infty$, and thus they contradict the hypotheses of our equivalence theorem. This suggests that these conditions are essential for the equivalence (30) to hold.

In [23], it appears that a more general characterization (for compactly supported functions only) has been worked out: however, from the reading of the paper it is not clear what kind of hypotheses are implied by a "compactly supported generating set" [23], in particular in terms of linear independence of the basis functions.

C. Asymptotic Equivalents

We will now exploit the simple form of Theorem 1 to derive the asymptotic form of the error as the sampling step T becomes sufficiently small with reference to a "natural" time scale of f.

Theorem 4. *If $f \in \mathbf{W}_2^r$ and $E(\nu)$ is $2r - 1$ times differentiable with bounded $2r$th derivative, then we have the following development:*

$$\| f - \mathcal{Q}_T f \|_{\mathbf{L}^2}^2 = \sum_{k=0}^{r-1} \frac{E^{(2k)}(0)}{(2\pi)^{2k}(2k)!} \| f^{(k)} \|_{\mathbf{L}^2}^2 T^{2k} + O(T^{2r}). \tag{31}$$

In particular, the first nonvanishing order for multi-wavelets satisfying the Lth-order quasi-biorthonormality condition is given by

$$\| f - \mathcal{Q}_T f \|_{\mathbf{L}^2} = \frac{T^L \| f^{(L)} \|_{\mathbf{L}^2}}{q(2\pi)^L L!} \left[\sum_n \left| \frac{d^L}{d\nu^L} \left\{ \overline{\hat{\tilde{\varphi}}(\nu)}^{\mathrm{t}} \hat{\varphi}\left(\nu + \frac{n}{q} \right) \right\} \right|_{\nu=0}^2 \right]^{1/2} + O(T^{L+1}). \tag{32}$$

Proof. We develop E in Taylor series around 0 and bound its $2r$th term: since $E(\nu)$ is symmetric, all the odd powers cancel; the expansion (31) is therefore a straightforward consequence of (24). From Lemma 2, we know that Lth-order quasi-biorthonormal multi-wavelets must satisfy (20). Thus, all terms in (21) including the first one are $O(\nu^{2L})$, which implies all derivatives up to order $2L$ in (31) to vanish. The final form for $r = L + 1$ is then equivalent to (32). ■

Note that for the case of a single generator, we recover formula (17) of [45], which involves the infinite sum $\sum_{n\neq 0} |\hat{\varphi}^{(L)}(n)|^2$. Similarly, if we consider the definition of quasi-interpolation given in [48] (i.e., the sampling functions are Dirac masses $\tilde{\varphi}_n(x) = \delta(x - n)$), then again, we find Eq. (65) of [48] which, in this case, involves the unrestricted sum $\sum_n |\hat{\varphi}^{(L)}(n)|^2$. We can also go further and obtain the next terms of the development without great difficulty. The main point is that Theorem 1 allows for an asymptotic $\mathbf{L}^2$-error analysis simpler and more powerful than what has been obtained before by using Taylor series expansions in the time domain.

In the general (multi-wavelet) case, we can constrain $\hat{\tilde{\phi}}(\nu) = \hat{\phi}(\nu) + O(\nu^{L+1})$: this is equivalent to fixing the first $L + 1$ moments of $\tilde{\varphi}$ to appropriate values provided by φ; then the first order of the asymptotic error equals the first order of the *asymptotic minimum error* (least squares solution) that would be obtained with a dual set of basis functions. In fact, the form (23) of $E(\nu)$ implies that this condition is necessary and sufficient for the approximation to be asymptotically optimal. In particular, it is now possible to consider compactly supported sampling functions associated with any kind of synthesis function. This new freedom in the choice of $\tilde{\varphi}_n$ takes its full sense when we remember that, in general, the dual functions of a set of synthesis functions are not compactly supported, and often not so well behaved in the frequency domain. This fundamental remark has also been stated in [46] with slightly stronger hypotheses (namely $\tilde{\varphi} * \varphi$ has to be a quasi-interpolant of order at least $L + 1$, which implies additional constraints on $\hat{\tilde{\varphi}}(n)$ for $n \neq 0$). Thus, our present contribution is to weaken the requirements on the analysis functions and to extend the result to multiple generators.

Of course if we fix a greater number of moments of $\tilde{\varphi}_n$ through $\hat{\tilde{\phi}}(\nu) = \hat{\phi}(\nu) + \mathrm{O}(\nu^{\mathrm{L+N}})$ then we can force ϵ_f^2 to agree with the minimum expansion error up to order $2L + 2N - 1$.

D. Upper Bounds

Theorem 1 is also useful for deriving upper bounds for the approximation error.

THEOREM 5. *If the multi-wavelets are quasi-biorthonormal of order L, then the approximation error for $f \in \mathbf{W}_2^r$ with $r \geq L$ is bounded as*

$$\| f - \mathcal{Q}_T f \|_{\mathbf{L}^2} \leq CT^L \| f^{(L)} \|_{\mathbf{L}^2} + KT^r \| f^{(r)} \|_{\mathbf{L}^2}, \tag{33}$$

where K is given by (26), *and*

$$C = \frac{\sqrt{\|E^{(2L)}\|_\infty}}{(2\pi)^L \sqrt{(2L)!}}. \tag{34}$$

Proof. We use Theorem 1 and estimate $E(\nu)$ by its Taylor series expansion. Using the Lth-order quasi-biorthonormality property, we deduce that $E(\nu)$ is bounded by $(\nu^{2L}/(2L)!)\|E^{(2L)}\|_\infty$, which allows us to isolate $\| f^{(L)}\|_{\mathbf{L}^2}$. From the proof of Theorem 3, we also know that $E^{(2L)}(\nu)$ is bounded so that the constant C is finite. ∎

This type of result has to be compared with the more traditional error bounds of the simpler form: $\epsilon_f \leq C' T^L \| f^{(L)}\|_{\mathbf{L}^2}$. In the case of an extended formula, the bound can be sharper because the leading constant C tends to be smaller. However, the result in Theorem 5 is just an example of what can be obtained as upper bound. For example, using the same type of argument as in the proof of Theorem 5, we can find a bound similar to the one given in [46, Thm. 1]. In this particular setting with $q = 1$, the sampling and synthesis functions both satisfy the Riesz condition, plus an additional "quasi-projection" requirement: $\overline{\hat{\tilde{\varphi}}(\nu)}\hat{\varphi}(\nu) = 1 + O(\nu^{2L})$. Using (21), we can find a_0, a_1 such that $E(\nu) \leq a_0^2(2\pi\nu)^{2L} + a_1^2(2\pi\nu)^{4L}$; then the application of Theorem 1 yields

$$\| f - \mathcal{Q}_T f \|_{\mathbf{L}^2} \leq a_0 T^L \| f^{(L)} \|_{\mathbf{L}^2} + a_1 T^{2L} \| f^{(2L)} \|_{\mathbf{L}^2} + KT^r \| f^{(r)} \|_{\mathbf{L}^2},$$

where $r \geq 2L$ is the Sobolev exponent of f. Inspection of the leading constant a_0 reveals that it is slightly smaller than the constant C_2 in [46]. The second constant a_1 is exactly C_1 in [46]. The third term is not present in [46], but can be easily absorbed into the second by setting $r = 2L$.

To conclude the discussion on upper bounds, we consider the case of the least squares approximation with $q = 1$. We have $E(\nu) = \sum_{k\neq 0} |\hat{\varphi}(\nu + k)|^2/A(\nu)$, where $A(\nu) = \sum_k |\hat{\varphi}(\nu + k)|^2$ is bounded by the Riesz constants. Thus, we can obtain the following estimate of the constant C in Theorem 5:

$$C^2 = \frac{\sup_\nu \sum_{k\neq 0} (|\hat{\varphi}|^2)^{(2L)}(\nu + k)}{(2\pi)^{2L}(2L)!A}.$$

The corresponding bound (33) is then quite comparable to the inequality $\epsilon_f \leq CT^L \| f^{(L)}\|_{\mathbf{L}^2} + C^2 T^{2L} \| f^{(2L)}\|_{\mathbf{L}^2}$, which is given in [48] and uses the same definition of the constant C.

These examples are only meant to show that it is rather straightforward to find sharp upper bounds with the help of our theorem. Of course when the basis functions are known a priori, it is probably better to compare $E(\nu)$ exactly and obtain the most accurate computation of the approximation error.

IV. APPLICATION TO REFINABLE GENERATORS

The concept of a multiresolution analysis (MRA) [35, 36] has proven to be the key for the construction of a whole variety of wavelet bases. These include orthogonal [6, 18, 34, 35], semi-orthogonal [3, 12, 47], and biorthonormal wavelets [14, 49], as well as the more recent multi-wavelets [1, 15, 27]. MRA provides a simple geometrical interpretation of the decomposition process in terms of projections onto a sequence of nested subspaces and their complements.

In this part, we will consider the case of a dyadic multiresolution analysis with $T = 2^j$, where $j \in \mathbb{Z}$ represents the scale index. Consequently, we define a MRA as follows:

(i) A collection of embedded spaces $\{0\} \cdots \subset V_{2^2} \subset V_2 \subset V_1 \subset \cdots \mathbf{L}^2$.

(ii) $V_T = \operatorname*{span}_{n \in \mathbb{Z}} \left\{ \varphi_n\left(\frac{t}{T}\right) \right\}$ with $T = 2^j$.

(iii) There exists an integer q such that $\varphi_{n+q}(x + q) = \varphi_n(x)$ for all x real, and n integer.

(iv) $\lim_{T \to 0} \operatorname{Clos}(V_T) = \mathbf{L}^2$.

The only addition to our previous assumptions is the nestedness property (i). Other than that the functions spaces V_T are the same as in Section II-B, except that the generators are now called *scaling functions*. Also note that the completeness condition (iv) is automatically satisfied as long as the generating functions have at least a first order of approximation.

The signal approximations that we consider do not involve wavelets explicitly but rather scaling functions at a given scale T (this is an important methodological difference from [44], where the opposite is done). However, it is clear that a signal approximation at scale $T = 2^j$ can also be expressed as a truncated wavelet expansion in which all finer scale coefficients have been set to zero. Thus, a detailed characterization of the approximation error as a function of scale is important for the comparison of wavelet bases and the prediction of their coding performance.

The important point for this paper is that the multiresolution property (i) implies that the generating functions follow a two-scale linear relation

$$\varphi(x) = \sum_k \mathbf{h}_k \varphi(2x - kq), \tag{35}$$

where $\mathbf{H}$ is a $q \times q$ matrix filter. In fact, (35) provides the vector definition of the scaling multi-wavelet.

In order to compute the asymptotic approximation constants, we shall examine the case of wavelets and multi-wavelets separately: they do not exhibit the same degree of complexity. In the present section, we are interested only in the minimum $\mathbf{L}^2$ value of ϵ_f, which corresponds to the orthogonal projection of f into V_T, which we denote $\mathcal{P}_T f$: this least squares approximation can be obtained from (12) with $\tilde{\phi}_n = \phi_n$. If, instead, we were to use functions that are not exact duals of each other, we know from the results of the preceding section that we need to include the additional term $\frac{1}{q}\|\hat{\phi} - \hat{\tilde{\phi}}\|^2$ to the kernel $E(\nu)$. However, we also know that the order of this correction can be made as high as we want by choosing $\tilde{\varphi}_n$ appropriately. This type of consideration may be quite relevant for the design of initialization algorithms, especially in the multi-wavelet case.

A. Wavelet Constants

For $q = 1$, the refinement filter is $H(e^{-2i\pi\nu}) = \sum_n h_n e^{-2i\pi n\nu}$ with $H(1) = 2$, and it is not difficult to derive the following two-scale relation for $A(\nu) = \sum_n |\hat{\varphi}(\nu + n)|^2$:

$$A(\nu) = \frac{1}{4}\left[|H(e^{-i\pi\nu})|^2 A\left(\frac{\nu}{2}\right) + |H(-e^{-i\pi\nu})|^2 A\left(\frac{\nu+1}{2}\right)\right]. \tag{36}$$

This equation can be put in matrix form and solved *exactly* if H has a finite impulse response [43]. We shall thus assume that we have access to the values of $A(\nu)$. In addition, we know that the function φ is of order L iff $(z + 1)^L$ divides $H(z)$ (the analog of this property will be shown in the multi-wavelet subsection: see Lemma 3).

For refinable functions, we can obtain not only the first asymptotic error equivalent as a function of the generating filter coefficients, but also the following $2L - 1$ coefficients of the development in powers of T. This is stated in the next theorem:

THEOREM 6. *If φ is of order L then for all $f \in \mathbf{W}_2^{2L}$ the least squares approximation error in V_T has the asymptotic development*

$$\|f - \mathcal{P}_T f\|_{\mathbf{L}^2}^2 = \sum_{k=L}^{2L-1} \frac{a_k}{\pi^{2k}(4^k - 1)} \|f^{(k)}\|_{\mathbf{L}^2}^2 T^{2k} + O(T^{4L}), \tag{37}$$

where the constants a_k are the first $4L$ coefficients of the MacLaurin development

$$\sum_{k=L}^{2L-1} a_k \nu^{2k} = \frac{A\left(\frac{\nu+1}{2}\right)|H(-e^{-i\pi\nu})|^2}{4A(\nu)} + O(\nu^{4L}). \tag{38}$$

Proof. We apply Theorem 4. Since we consider the minimum $\mathbf{L}^2$ error, we have $E(\nu) = 1 - |\hat{\varphi}(\nu)|^2/A(\nu)$. Using (36) and the two-scale relation in the Fourier domain $\hat{\varphi}(\nu) = (1/2)H(e^{-i\pi\nu})\hat{\varphi}(\nu/2)$, we find, after some manipulations,

$$1 - \frac{|\hat{\varphi}(\nu)|^2}{A(\nu)} = 1 - \frac{\left|\hat{\varphi}\left(\frac{\nu}{2}\right)\right|^2}{A\left(\frac{\nu}{2}\right)} + \frac{A\left(\frac{\nu+1}{2}\right)|H(-e^{-i\pi\nu})|^2}{4A(\nu)} + O(\nu^{4L}).$$

Here, we have taken into account the fact that $H(-e^{-i\pi\nu}) = O(\nu^L)$. Substituting (38), we can then solve for the first $4L$ coefficients of the asymptotic development of $E(\nu)$ and, hence, of ϵ_f^2. ■

This result adds higher order terms to the asymptotic form of the error given in [45]. Moreover, if we let $H(z) = \left(\frac{z+1}{2}\right)^L Q(z)$ with $Q(1) = 2$, we obtain a closed form formula for the first-order constant C_φ^- of [45]:

$$C_\varphi^- = \frac{1}{(2\pi)^L L!}\,(\sum_{n\neq 0} |\hat{\varphi}^{(L)}(n)|^2)^{1/2} = \frac{|Q(-1)|\sqrt{A\left(\frac{1}{2}\right)}}{2^{L+1}\sqrt{4^L-1}}. \tag{39}$$

The right-hand side improves upon the infinite sum formula that has been reported previously. In [45], the calculation of the infinite sum formula was made possible through the use of an induction formula for the derivatives in the center of (39), which were shown to be proportional to $\hat{\varphi}(n/2)$.

B. Multi-wavelet Constants

In this case, the situation is more intricate, even if $E(\nu)$ remains rather simple. We start by writing the two-scale relation (35) in the Fourier domain: $\hat{\varphi}(\nu) = \frac{1}{2}\mathbf{H}(e^{-i\pi q\nu})\hat{\varphi}\left(\frac{\nu}{2}\right)$. It follows that 2 is an eigenvalue of $\mathbf{H}(1)$. As in the wavelet case, it is then not difficult to show that $\mathbf{A}(\nu)$ is the solution of the equation

$$\mathbf{A}(\nu) = \frac{1}{4}\left[\mathbf{H}(e^{-i\pi\nu})\mathbf{A}\left(\frac{\nu}{2}\right)\overline{\mathbf{H}(e^{-i\pi\nu})}^{\mathrm{t}} + \mathbf{H}(-e^{-i\pi\nu})\mathbf{A}\left(\frac{\nu+1}{2}\right)\overline{\mathbf{H}(-e^{-i\pi\nu})}^{\mathrm{t}}\right]. \tag{40}$$

Here again when $\mathbf{H}$ is FIR, this relation can be solved in the time domain to yield the expression of the polynomial matrix **A.** We shall from now on assume that this matrix is known.

As a first step, we must find out how the Lth-order property manifests itself on the generating filter **H.**

LEMMA 3. *The two following properties are equivalent:*

(i) *the multi-wavelets φ_n are of order L;*

(ii) *there exists a unique polynomial $\Lambda(\nu)$ of degree less than or equal to $L-1$ such that $\Lambda(0) \neq 0$ and satisfying together*

$$\overline{\mathbf{H}(-e^{-i\pi q\nu})}^{\mathrm{t}}\Lambda(\nu) = O(\nu^L) \tag{41}$$

and

$$\overline{\mathbf{H}(e^{-i\pi q\nu})}^{\mathrm{t}}\Lambda(\nu) = 2\Lambda\left(\frac{\nu}{2}\right) + O(\nu^L). \tag{42}$$

Proof. The main ingredients are the lower Riesz condition and the two-scale difference equation.

Let us first prove that (i) $\Rightarrow$ (ii). From Lemma 1 we know that there exists a unique polynomial $\Lambda(\nu)$ of degree lesser or equal to $L-1$ such that $\overline{\Lambda(\nu)}^{\mathrm{t}}\hat{\varphi}\left(\nu+\frac{n}{q}\right)=q\delta_n+O(\nu^L)$. Of course, $\Lambda(0)\neq 0$ since $\mathbf{A}(0)$ is invertible. Applying the vector two-scale relation with 2ν instead of ν, we find

- $\frac{1}{2}\overline{\Lambda(2\nu)}^{\mathrm{t}}\mathbf{H}(e^{-2i\pi q\nu})\hat{\varphi}\left(\nu+\frac{n}{2q}\right)=q\delta_n+O(\nu^L)$ if n is even;
- $\overline{\Lambda(2\nu)}^{\mathrm{t}}\mathbf{H}(-e^{-2i\pi q\nu})\hat{\varphi}\left(\nu+\frac{n}{2q}\right)=O(\nu^L)$ if n is odd.

Due to a uniqueness of Λ, (42) follows immediately from the first equation above. For the second equation, we use the same technique as Lemma 1 to prove the uniqueness of the polynomial P. We temporarily pose $\mathrm{P}(\nu)=\overline{\mathbf{H}(-e^{-2i\pi q\nu})}^{\mathrm{t}}\Lambda(2\nu)$ so that the equation above reads $\overline{\mathrm{P}(\nu)}^{\mathrm{t}}\hat{\varphi}\left(\nu+\frac{n}{2q}\right)=O(\nu^L)$ for n odd. If we let $\nu=0$ then right multiplying by $\overline{\hat{\varphi}\left(\frac{n}{2q}\right)}^{\mathrm{t}}$ and summing over odd n yields $\overline{\mathrm{P}(0)}^{\mathrm{t}}\mathbf{A}(\frac{1}{2})=0$. Due to the invertibility of $\mathbf{A}(\frac{1}{2})$, this implies that $\mathrm{P}(0)=0$. We then take the equation again for n odd with $\frac{\mathrm{P}(\nu)}{\nu}$ instead of $\mathrm{P}(\nu)$, and $L-1$ instead of L, which yields $\mathrm{P}'(0)=0$. Iteration of this argument L times leads to the result $\mathrm{P}(\nu)=O(\nu^L)$, which is (41).

For the converse implication (ii) $\Rightarrow$ (i), we simply observe that (ii) implies $\overline{\Lambda(\nu)}^{\mathrm{t}}\hat{\varphi}\left(\nu+\frac{n}{q}\right)=\overline{\Lambda\left(\frac{\nu}{2}\right)}^{\mathrm{t}}\hat{\varphi}\left(\frac{\nu}{2}+\frac{n}{2q}\right)+O(\nu^L)$ if n is even and $\overline{\Lambda(\nu)}^{\mathrm{t}}\hat{\varphi}\left(\nu+\frac{n}{2q}\right)=O(\nu^L)$ if n is odd. By induction, this proves that $\overline{\Lambda(\nu)}^{\mathrm{t}}\hat{\varphi}\left(\nu+\frac{n}{q}\right)=a\delta_n+O(\nu^L)$ for all n. Since $\Lambda(0)\neq 0$, the constant a cannot cancel. Applying Lemma 1, we thus find that the scaling functions are of order L. The uniqueness of such a Λ follows directly from Lemma 1. ■

The same result (but making the hypotheses of linear independence and compact support), expressed in a different way, can be found in [27, Thm. 1] and [15, Thm. 2.1]. See also [37] for relaxed hypotheses on the synthesis functions.

Condition (ii) of Lemma 3 is equivalent to the Lth divisibility of $H(z)$ at $z=-1$ in the usual case of wavelets where $H(1)=2$. In the vector (or multi-wavelet) case, however, the condition cannot be restricted to the values (or here the kernel) of $\mathbf{H}$ at half the sampling frequency (41) and must be completed by (42).

From now on, we also assume that the minimum expansion error is of order L, which implies that the scaling functions are of order L; hence the existence of such a Λ: by solving the linear system of Eqs. (41) and (42) we have access to its value. Consequently we can compute the two constants Δ_0 and Δ_1 which are implicit in (41) and (42):

$$\Delta_0=\lim_{\nu\to 0}\left[\overline{\mathbf{H}(e^{-i\pi q\nu})}^{\mathrm{t}}\Lambda(\nu)-2\Lambda\left(\frac{\nu}{2}\right)\right]\nu^{-L} \tag{43}$$

$$\Delta_1=\lim_{\nu\to 0}\overline{\mathbf{H}(-e^{-i\pi q\nu})}^{\mathrm{t}}\Lambda(\nu)\nu^{-L} \tag{44}$$

It will be of interest to give the exact equivalent of the Fourier transform of the dual vector function of φ up to the order L, in the neighborhood of 0. In order to simplify slightly the writing of the expression, let us pose $\mathbf{H}_0 = \mathbf{H}(1)$ and $\mathbf{H}_1 = \mathbf{H}(-1)$. Likewise, we shall denote by $\mathbf{A}_0$ and $\mathbf{A}_1$ the values of $\mathbf{A}(\nu)$ at $\nu = 0$ and $\nu = \frac{1}{2}$, respectively.

LEMMA 4. *We have the asymptotic development*

$$\mathbf{A}(q\nu)^{-1}\hat{\varphi}(\nu) = \Lambda(\nu) - \Omega\nu^L + O(\nu^{L+1}), \tag{45}$$

where

$$\Omega = \frac{1}{4}\mathbf{A}_0^{-1}\left[\mathbf{I}_q - \frac{1}{2^{L+1}}\mathbf{H}_0\right]^{-1}(\mathbf{H}_0\mathbf{A}_0\Delta_0 + \mathbf{H}_1\mathbf{A}_1\Delta_1) \tag{46}$$

Proof. In fact, since the expansion error is of order L, $E(\nu) = O(\nu^{2L})$ and this, in particular, means that if the $\tilde{\varphi}_n$ are dual to the φ_n, then by inspection of (21) we have $\overline{\hat{\tilde{\varphi}}(\nu)}^{\mathrm{t}}\hat{\varphi}\left(\nu + \frac{n}{q}\right) = q\delta_n + O(\nu^L)$. It follows that Λ is the order $L-1$ development of the dual function of φ, so that we need only to find the missing order L.

We thus define Ω by (45). This is, of course, equivalent to $\hat{\varphi}(\nu) = \mathbf{A}(q\nu)\Lambda(\nu) - \mathbf{A}_0\Omega\nu^L + O(\nu^{L+1})$. We can now use the Fourier two-scale relation on the left side of this equation, and with the use of (40), (43), and (44), and finally after some (tedious) rearrangements, we obtain the value of Ω; this validates (46).

Note that, as proven in [17], $\mathbf{H}_0$ has spectral radius 2 so that $\mathbf{I}_q - (1/2^{L+1})\mathbf{H}_0$ is never singular. ■

With these constants, we can now derive the asymptotic value of the approximation error in the multi-wavelet case.

THEOREM 7. *If the scaling multi-wavelets are of order L, then for all $f \in \mathbf{W}_2^L$ the asymptotic form of the least squares approximation error in V_T is*

$$\|f - \mathcal{P}_T f\|_{\mathbf{L}^2} = \frac{(\|\sqrt{\mathbf{A}_0}Q_0\|^2 + \|\sqrt{\mathbf{A}_1}Q_1\|^2)^{1/2}}{2\pi^L\sqrt{q(4^L-1)}}\|f^{(L)}\|_{\mathbf{L}^2}T^L + o(T^L), \tag{47}$$

where the vectors Q_0 and Q_1 have been defined by

$$\mathrm{Q}_0 = \Delta_0 - \overline{\mathbf{H}}_0^{\mathrm{t}}\Omega + 2^{-L+1}\Omega \tag{48}$$

$$\mathrm{Q}_1 = \Delta_1 - \overline{\mathbf{H}}_1^{\mathrm{t}}\Omega. \tag{49}$$

Proof. Once again, we use Theorem 4. The result will first be proven for orthonormal multi-wavelets. We start from $E(\nu) = 1 - \frac{1}{q}\|\hat{\phi}\|^2$, which can be rewritten using the two-scale equation $E(\nu) = 1 - \frac{1}{4q}\overline{\hat{\phi}\left(\frac{\nu}{2}\right)}^{\mathrm{t}}\overline{\mathbf{H}}^{\mathrm{t}}\mathbf{H}\hat{\phi}\left(\frac{\nu}{2}\right)$ if we denote by $\mathbf{H}$ the expression $\mathbf{H}(e^{-i\pi q\nu})$. Let us also denote by $\mathbf{U}$ the other expression $\mathbf{H}(-e^{-i\pi q\nu})$ (these intermediary notations are meant only to simplify the visual aspect of the expressions). We thus have (orthonormality) $\mathbf{H}\overline{\mathbf{H}}^{\mathrm{t}} + \mathbf{U}\overline{\mathbf{U}}^{\mathrm{t}} = 4\mathbf{I}_q$. One can verify that this implies the identity

$$\mathbf{I}_q - \frac{1}{4}\overline{\mathbf{H}}^{\mathrm{t}}\mathbf{H} = \frac{1}{16}\overline{\mathbf{H}}^{\mathrm{t}}\mathbf{U}\overline{\mathbf{U}}^{\mathrm{t}}\mathbf{H} + \left[\mathbf{I}_q - \frac{1}{4}\overline{\mathbf{H}}^{\mathrm{t}}\mathbf{H}\right]^2.$$

Multiplying this equation on both sides by (left) $\overline{\hat{\phi}\left(\frac{\nu}{2}\right)}^{\mathrm{t}}$ and (right) $\hat{\phi}\left(\frac{\nu}{2}\right)$, we obtain the following recursion equation between $E(\nu)$ and $E\left(\frac{\nu}{2}\right)$:

$$1 - \frac{1}{q}\|\hat{\phi}(\nu)\|^2 = 1 - \frac{1}{q}\left\|\hat{\phi}\left(\frac{\nu}{2}\right)\right\|^2 + \frac{1}{4q}\left[\|\overline{\mathbf{U}}^t\hat{\phi}(\nu)\|^2 + \left\|\overline{\mathbf{H}}^t\hat{\phi}(\nu) - 2\hat{\phi}\left(\frac{\nu}{2}\right)\right\|^2\right]. \tag{50}$$

We then compute the asymptotic equivalent of each element of this equation:

- $E(\nu) = 1 - \frac{1}{q}\|\hat{\phi}(\nu)\|^2 = a\nu^{2L} + o(\nu^{2L})$, where a is the positive constant providing the first asymptotic order: it is the goal of this theorem to formulate it explicitly.
- Using Lemma 4 and the definition (44) of Δ_1, we find $\overline{\mathbf{U}}^{\mathrm{t}}\hat{\phi}(\nu) = (\Delta_1 - \overline{\mathbf{H}}_1^{\mathrm{t}}\Omega)\nu^L + o(\nu^L)$.
- Using Lemma 4 and the definition (43) of Δ_0 we find $\overline{\mathbf{H}}^{\mathrm{t}}\hat{\phi}(\nu) - 2\hat{\phi}\left(\frac{\nu}{2}\right) = (\Delta_0 + [2^{-L+1}\mathbf{I}_q - \overline{\mathbf{H}}_0^{\mathrm{t}}]\Omega)\nu^L + o(\nu^L)$.

Replacing the values of these asymptotic equivalents in (50), we find the result for orthonormal multi-wavelets mentioned earlier.

For nonorthonormal scaling functions, where $\mathbf{A}$ may be different from $\mathbf{I}_q$, we remark that in (50), $\mathbf{H}(e^{-i\pi\nu})$ has to be replaced by $\mathbf{G}(e^{2i\pi\nu})^{-1}\mathbf{H}(e^{-i\pi\nu})\mathbf{G}(e^{-i\pi\nu})$, and $\hat{\phi}$ by $\mathbf{G}(e^{2i\pi q\nu})^{-1}\hat{\phi}$, where $\mathbf{G}(e^{2i\pi\nu}) = \sqrt{\mathbf{A}(\nu)}$. Equation (50) is thus replaced with

$$E(\nu) = E\left(\frac{\nu}{2}\right) + \frac{1}{4q}\left[\left\|\sqrt{\mathbf{A}\left(\frac{q\nu+1}{2}\right)}\overline{\mathbf{U}}^{\mathrm{t}}\hat{\tilde{\varphi}}(\nu)\right\|^2 + \left\|\sqrt{\mathbf{A}\left(\frac{q\nu}{2}\right)}\left(\overline{\mathbf{H}}^{\mathrm{t}}\hat{\tilde{\varphi}}(\nu) - 2\hat{\tilde{\varphi}}\left(\frac{\nu}{2}\right)\right)\right\|^2\right], \tag{51}$$

where $\hat{\tilde{\varphi}}(\nu) = \mathbf{A}(q\nu)^{-1}\hat{\varphi}(\nu)$ is the dual function of $\hat{\varphi}(\nu)$, whose asymptotic development, up to the order L near zero, is given by Lemma 4. The generalized result of Theorem 7 is thus a consequence of the introduction of this square root into the scalar product. ■

This result is in fact the exact extension of (39) to the multi-wavelet case, as can be readily checked when replacing matrices and vectors by the corresponding scalars. Specifically, we see that in the wavelet case, the vector/scalar Q_0 cancels, while Q_1 reduces to Δ_1, the equivalent of $H(-e^{i\pi\nu})$ in the neighborhood of $\nu = 0$.

In contrast with the wavelet case, we have no *easy* access to higher order asymptotic terms. This is because the multi-wavelet order property (41) and (42) is much weaker than the corresponding divisibility constraint by $(1 + z)^L$ in the wavelet case. However, if one wishes to find higher order terms for the asymptotic error, it is always possible to determine the development of the dual function of $\hat{\varphi}$ recursively, similar to what has been done in Lemma 4. One can verify that this leads to a triangular system of vector equations, whose diagonal matrix coefficients are all invertible, and which can thus be solved by induction; Ω is then the first input to the induction. Finally, this development can be substituted in Eq. (51) in order to obtain the desired asymptotic equivalents.

V. CONCLUSION

We have stated a powerful quantitative result concerning the approximation of $\mathbf{L}^2$ functions by the linear decomposition scheme given in Section II. We expect this formula to be of interest both for applied researchers (e.g., as a basic error estimation tool in geometric image processing), and for approximation theorists. In particular, our result has been used successfully to prove in a very natural way the Strang–Fix equivalence for multiple generators [42], while requiring very weak constraints. In addition, we could derive asymptotic formulae in the case where the approximation spaces belong to a multiresolution analysis: this application is of great interest to digital signal processing. We are also currently working on the explicit computation of asymptotic expansions for some representations of wide interest such as splines and Daubechies' wavelets [8].

APPENDIX A: ON THE FINITE DIMENSION OF $\mathcal{M}$

We prove here that, if the φ_n are compactly supported and satisfy the Riesz condition, then $\dim \mathcal{M} < \infty$. By definition, $\{\lambda_n\}_{n\in\mathbb{Z}} \in \mathcal{M}$ iff $\sum_n \lambda_n \varphi_n = 0$ in the sense of distributions. In vector notation, this is equivalent to $\sum_n \Lambda_n^{\mathrm{t}} \varphi(x - nq) = 0$, where $\Lambda_n = (\lambda_{nq}, \lambda_{nq+1}, \ldots, \lambda_{nq+q-1})^{\mathrm{t}}$. Right multiplying by $\varphi(x - kq)^{\mathrm{t}}$ and integrating over $x \in \mathbb{R}$ the resulting finite sum, we obtain

$$\sum_n \mathbf{b}_{k-n} \Lambda_n = 0, \tag{52}$$

where $\mathbf{b}_n = \int \varphi(x)\varphi(x + nq)^{\mathrm{t}} dx$ is a finite sequence $|n| \leq K < \infty$. Notice that $\mathbf{B}(\nu) = \sum_n \mathbf{b}_n e^{2i\pi n\nu} = \frac{1}{q}\mathbf{A}(\nu)$ in the sense of distributions, so that, due to the lower Riesz inequality, the finite degree polynomial $\det(\mathbf{B}(\nu))$ is not trivial. If we build $\Lambda(\nu) = \sum_{|n|\leq N+K-1} \Lambda_n e^{2i\pi n\nu}$, then (52) implies $\mathbf{B}(\nu)\Lambda(\nu) = z^{-N}\mathrm{Q}\left(\frac{1}{z}\right) + z^N \mathrm{R}(z)$, where Q and R are polynomials of degree $\leq 2K - 1$, and $z = e^{2i\pi\nu}$.

Using the matrix inversion formula $\tilde{\mathbf{B}}\mathbf{B} = \det(\mathbf{B})$, where $\tilde{\mathbf{B}}$ is a finite polynomial matrix, we thus have $\det(\mathbf{B}(\nu))\Lambda(\nu) = z^{-N}\tilde{\mathbf{B}}(\nu)\mathrm{Q}\left(\frac{1}{z}\right) + z^N\tilde{\mathbf{B}}(\nu)\mathrm{R}(z)$. Consequently, if we let $\det(\mathbf{B}(\nu)) = \sum_k d_k z^k$, then $\sum_n d_{k-n}\Lambda_n = 0$; the coefficients of the induction equation being now scalar, the finiteness of $\dim \mathcal{M}$ immediately follows.

APPENDIX B: PROOF OF LEMMA 1

The most difficult part is to prove the implication (i) $\Rightarrow$ (ii). For this purpose, we shall first study $\mathcal{M}$ and show that the coefficients of the polynomial expansions can be written in a simple form, involving a set of quasi-biorthonormal analysis functions. Finally, we will write the quasi-biorthonormality condition in the Fourier domain.

Characteristics of $\mathcal{M}$. We first prove that any coefficient sequence $\mu = \{\mu_n\}_{n\in\mathbb{Z}}$ belonging to $\mathcal{M}$, i.e., satisfying $\sum_n \mu_n\varphi_n = 0$, is such that there exists another coefficient sequence μ' belonging to $\mathcal{M}$ such that $\mu_n = \mu'_{n+q} - \mu'_n$ for all n. The finite dimension of $\mathcal{M}$ and the invertibility of $\mathbf{A}(0)$ both play a crucial role. Consider the linear operator Δ acting on elements of $\mathcal{M}$ through $\Delta\mu_n = \mu_{n+q} - \mu_n$. Clearly, due to the q-integer shift invariance, $\Delta\mu$ is in $\mathcal{M}$. The kernel of the endomorphism Δ is reduced to the coefficient sequences that are q-periodic. In vector notation, this means that, for the kernel to be nontrivial, it is necessary that there exist a constant vector Λ such that

$$\Lambda^{\mathrm{t}} \sum_n \varphi(x - nq) = 0. \tag{53}$$

This is equivalent to $\Lambda^{\mathrm{t}}\hat{\varphi}\left(\frac{n}{q}\right) = 0$ for all $n = \mathbb{Z}$ (using Poisson's formula). If we right multiply by $\overline{\hat{\varphi}\left(\frac{n}{q}\right)}^{\mathrm{t}}$ and sum over n, this yields $\Lambda^{\mathrm{t}}\mathbf{A}(0) = 0$, so that $\Lambda = 0$, due to the lower Riesz condition. Hence, the kernel of Δ reduces to 0. As a consequence, the endomorphism Δ is one-to-one, and $\mathcal{M} = \Delta\mathcal{M}$, thus proving our first claim.

Determination of $\Lambda^{(s)}$. Consider now the L equalities (14), which can also be expressed in vector notation: there exist L sets of vectors $\{\Lambda_n^{(s)}\}_{n\in\mathbb{Z}}$ for $s = 0, \ldots, L-1$ such that

$$\sum_n \Lambda_n^{(s)\mathrm{t}}\varphi(x - nq) = x^s \tag{54}$$

in the sense of distributions. If we define Δ' as the linear operator acting on polynomials through $\Delta' P(x) = P(x+q) - P(x)$ then we have $\sum_n \Delta\Lambda_n^{(s)\mathrm{t}}\varphi(x - nq) = \Delta' x^s$, which is a polynomial of degree strictly less than s. By iterating the operator, we find $\Delta^{s+1}\Lambda^{(s)} \in \mathcal{M}$. It is now important to notice that we can choose $\Lambda^{(s)}$ such that $\Delta^{s+1}\Lambda^{(s)} = 0$. To show this, let $\Delta^{s+1}\Lambda^{(s)} = \mathrm{M}$, where $\mathrm{M} \in \mathcal{M}$. We know that Δ is one-to-one on $\mathcal{M}$, as is Δ^{s+1}, and as a consequence, we can find $\mathrm{M}' \in \mathcal{M}$ such that $\Delta^{s+1}\mathrm{M}' = \mathrm{M}$; now, $\Lambda'^{(s)} = \Lambda^{(s)} - \mathrm{M}'$ satisfies the same monomial reconstruction property (54) as $\Lambda^{(s)}$, together with $\Delta^{s+1}\Lambda'^{(s)} = 0$. We can thus substitute $\Lambda'^{(s)}$ to $\Lambda^{(s)}$.

Assuming that the $\Lambda^{(s)}$ are chosen so that $\Delta^{s+1}\Lambda^{(s)} = 0$, we expand $(x \pm q)^s - x^s$. Replacing the monomials by their Λ-expansions (54), we obtain for each s an expression of the form $\sum_n \mathrm{K}_n^{\mathrm{t}}\varphi(x - nq) = 0$, so that K_n is in $\mathcal{M}$. On the other hand, we have $\Delta^{s+1}\mathrm{K}_n = 0$. Since Δ is invertible on $\mathcal{M}$, we find $\mathrm{K}_n = 0$. Explicitly writing the value of K_n, we thus find the following recursion for $\Lambda_n^{(s)}$:

$$\Lambda_{n\pm 1}^{(s)} = \sum_{k=0}^{s} \binom{s}{k} (\pm q)^{s-k} \Lambda_n^{(k)}.$$

Let us now choose a compactly supported vector function $\mathrm{F}(x)$ such that $\int x^s \mathrm{F}(x) dx = \Lambda_0^{(s)}$ for $s = 0, \ldots, L - 1$. This choice is always possible: it corresponds to imposing the value of the first $L - 1$ derivatives of $\hat{\mathrm{F}}(\nu)$ at $\nu = 0$. The induction equation for $\Lambda_n^{(s)}$ shows that in fact we have, as a general expression, $\Lambda_n^{(s)} = \int x^s \mathrm{F}(x - nq) dx$. This means that the analysis functions associated with F and the φ_n form a quasi-biorthonormal set of order L.

Equivalent Fourier condition. Taking the Fourier transform (in the sense of distributions) of the multi-wavelet equation that reproduces x^s, and using Poisson's formula (this is valid since, by hypothesis, $\mathrm{Q}(x)^{\mathrm{t}} \varphi(x) \in \mathbf{L}^1$ for any polynomial vector Q of degree less than or equal to $L - 1$), we find

$$\frac{1}{q} \sum_n \delta^{(s)}\left(\nu - \frac{n}{q}\right) \overline{\hat{\mathrm{F}}\left(\nu + \frac{n}{q}\right)}^{\mathrm{t}} \hat{\varphi}(\nu) = \delta^{(s)}(\nu)$$

for $s = 0, \ldots, L - 1$. Equivalently, we can write $\delta^{(s)}\left(\nu + \frac{n}{q}\right) \overline{\hat{\mathrm{F}}\left(\nu + \frac{n}{q}\right)}^{\mathrm{t}} \hat{\varphi}(\nu) = q \delta_n \delta^{(s)}(\nu)$ for all n integers, all $s = 0, \ldots, L - 1$, and all ν real. Finally, this is equivalent to saying that $(d^s/d\nu^s)\left[\overline{\hat{\mathrm{F}}(\nu)}^{\mathrm{t}} \hat{\varphi}\left(\nu + \frac{n}{q}\right)\right]_{\nu=0} = q \delta_n \delta_s$ for all integers n and $s = 0, \ldots, L - 1$. As we can see, this condition makes use only of $\hat{\mathrm{F}}(0)$ and of the $L - 1$ derivatives of $\hat{\mathrm{F}}$ at zero, which are otherwise directly related to $\Lambda_0^{(s)}$ by definition. A consequence is that if suffices to restrict $\hat{\mathrm{F}}$ (which is entire) to its Taylor development up to the order $L - 1$ and this provides the vector polynomial P of (18). This equation has been stated with "$o(\nu^{L-1})$" since we have characterized only the first $L - 1$ derivatives; however, since φ_n satisfies $\|\hat{\varphi}_n^{(L)}\|_\infty < \infty$ we automatically have $o(\nu^{L-1}) = O(\nu^L)$. This completes the proof of (18): (ii) implies (i) by reversing our set of equivalent assertions.

The uniqueness of the polynomial follows from the invertibility of $\mathbf{A}(0)$ once again: if the polynomial were not unique, it would be possible to find a polynmial $\mathrm{Q}(\nu) = \mathrm{P}(\nu) - \mathrm{P}'(\nu)$ such that $\overline{\mathrm{Q}(\nu)}^{\mathrm{t}} \hat{\varphi}\left(\nu + \frac{n}{q}\right) = O(\nu^L)$ for all n. In particular $\overline{\mathrm{Q}(0)}^{\mathrm{t}} \hat{\varphi}\left(\frac{n}{q}\right) = 0$ for all n: as shown earlier, this implies $\mathrm{Q}(0) = 0$. In that case, we can divide Q by ν and use the same argument, making the degree of Q decrease. Finally, the only possible Q is trivial, which proves $\mathrm{P} = \mathrm{P}'$ and thus the uniqueness.

The equivalence between (ii) and (iii) is straightforward and results from the fact that it is always possible to find a polynomial vector B such that $\mathrm{B}(e^{2i\pi q\nu}) = \mathrm{P}(\nu) + O(\nu^L)$: the wavelet defined through its Fourier transform $\hat{\psi}(\nu) = \overline{\mathrm{B}(e^{2i\pi q\nu})}^{\mathrm{t}} \hat{\varphi}(\nu)$ thus satisfies the Strang–Fix conditions.

APPENDIX C: PROOF OF THEOREM 1

We shall first show that under the hypotheses of the paper, $\mathcal{S}(f)_n$ is in l^2. Part of the argument will be used again to prove Theorem 1.

A. l^2 Convergence of the Samples

Let us define

$$\mathrm{U}(\nu) = \frac{1}{qT} \sum_n \overline{\hat{f}\left(\nu + \frac{n}{qT}\right)} \hat{\varphi}\left(\nu T + \frac{n}{q}\right) \tag{55}$$

$$\tilde{\mathrm{U}}(\nu) = \frac{1}{qT} \sum_n \overline{\hat{f}\left(\nu + \frac{n}{qT}\right)} \hat{\tilde{\varphi}}\left(\nu T + \frac{n}{q}\right). \tag{56}$$

We prove that these expressions are meaningful and that these $\frac{1}{qT}$ periodic vector functions belong to $\mathbf{L}^2(I)$, where $I = \left[-\frac{1}{2qT}, \frac{1}{2qT}\right]$. Of course, if this is the case, then using the well-known theorem about the development of periodic functions in Fourier series, we have the alternative definition of U and $\tilde{\mathrm{U}}$,

$$\mathrm{U}(\nu) = \sum_n \left[\int f(\tau)\varphi\left(\frac{\tau}{T} - nq\right) d\,\frac{\tau}{T}\right] e^{2i\pi nqT\nu} \qquad \text{a.e.} \tag{57}$$

$$\tilde{\mathrm{U}}(\nu) = \sum_n \left[\int f(\tau)\tilde{\varphi}\left(\frac{\tau}{T} - nq\right) d\,\frac{\tau}{T}\right] e^{2i\pi nqT\nu} \qquad \text{a.e.} \tag{58}$$

We start with $\tilde{\mathrm{U}}$ and define the functional sequence

$$\tilde{\mathrm{U}}_N(\nu) = \frac{1}{qT} \sum_{|n|\leq N} \overline{\hat{f}\left(\nu + \frac{n}{qT}\right)} \hat{\tilde{\varphi}}\left(\nu T + \frac{n}{q}\right), \tag{59}$$

where N is a positive integer; our goal is to prove that $\tilde{\mathrm{U}}_N$ is a Cauchy sequence, that is to say,

$$\lim_{N\to\infty} \sup_{N'>N} \|\tilde{\mathrm{U}}_{N'} - \tilde{\mathrm{U}}_N\|_{\mathbf{L}^2(I)} = 0.$$

By Fischer–Riesz theorem, this property will automatically imply the convergence of $\tilde{\mathrm{U}}_N$ toward an $\mathbf{L}^2(I)$ function $\tilde{\mathrm{U}}$.

We thus choose $N' > N$ and assume that $f \in \mathbf{W}_2^r$ with $r > 0.5$ and $\|\hat{\tilde{\varphi}}\|_\infty \leq K < \infty$. From f, we define a set of bandpass functions f_k through

$$\hat{f}_k(\nu) = \begin{cases} \hat{f}(\nu) & \text{if } 0 \le |\nu| - \dfrac{k}{2qT} < \dfrac{1}{2qT} \\ 0 & \text{elsewhere} \end{cases} \tag{60}$$

for $k \in \mathbb{N}$; of course, we have the partition equation $\sum_{k\ge 0} f_k = f$. Replacing f with this sum in (59), we can exchange the (finite–infinite) summations yielding

$$\tilde{\mathrm{U}}_{N'}(\nu) - \tilde{\mathrm{U}}_N(\nu) = \frac{1}{qT}\sum_{k\ge 0}\sum_{N<|n|\le N'} \overline{\hat{f}_k\left(\nu + \frac{n}{qT}\right)}\hat{\tilde{\varphi}}\left(\nu T + \frac{n}{q}\right).$$

Because of their support, not all the f_k contribute to the sum of the right-hand side. For our purpose, it is enough to say that only those k such that $k > 2N$ will be involved; using Minkowsky's inequality and the upper bound for $\|\hat{\tilde{\varphi}}\|$, we find

$$\|\tilde{\mathrm{U}}_{N'} - \tilde{\mathrm{U}}_N\|_{\mathbf{L}^2(I)} \le \frac{K}{qT}\sum_{k>2N}\|f_k\|_{\mathbf{L}^2}.$$

Due to the definition (60) of the f_k's, $|2\pi\nu|^{-r}$ is upper bounded by $\left(\frac{qT}{k\pi}\right)^r$ on the support of $\hat{f}_k$ so that $\|f_k\|_{\mathbf{L}^2} \le \left(\frac{qT}{k\pi}\right)^r \|f_k^{(r)}\|_{\mathbf{L}^2}$. Using Cauchy–Schwartz inequality for discrete sequences, we find that

$$\sum_{k>2N}\|f_k\|_{\mathbf{L}^2} \le \left(\frac{qT}{\pi}\right)^r \sqrt{\sum_{k>2N} k^{-2r}}\|f^{(r)}\|_{\mathbf{L}^2},$$

which tends to zero as N tends to infinity, whence $\tilde{\mathrm{U}}_N$ satisfies the Cauchy property. This ensures the convergence of the sum of the squared samples and proves that $\tilde{\mathrm{U}} \in \mathbf{L}^2(I)$. Of course, we have the same result for U: this will be used in the next subsection.

B. *Expression of* ϵ_f *in Fourier Variables*

Expanding the square norm ϵ_f^2, we obtain three terms: $\epsilon_f^2 = \|f\|_{\mathbf{L}^2}^2 - 2\langle f, \mathfrak{Q}_T f\rangle + \|\mathfrak{Q}_T f\|_{\mathbf{L}^2}^2$. Let us concentrate on the third. First, we manipulate (13) by substituting $\hat{\tilde{\varphi}}(\nu) = \int \tilde{\varphi}(x)e^{-2i\pi\nu x}dx$, which yields $\mathfrak{Q}_T f(t) = T\int \overline{\tilde{\mathrm{U}}(\nu)}^{\mathrm{t}}\hat{\varphi}(\nu T)e^{2i\pi\nu t}d\nu$. As can be seen from this formula, $\mathfrak{Q}_T f$ and $T\overline{\tilde{\mathrm{U}}(\nu)}^{\mathrm{t}}\hat{\varphi}(\nu T)$ are Fourier transforms of each other. A consequence is that their $\mathbf{L}^2$ norms are identical, $\|\mathfrak{Q}_T f\|_{\mathbf{L}^2} = \|T\overline{\tilde{\mathrm{U}}(\nu)}^{\mathrm{t}}\hat{\varphi}(\nu T)\|_{\mathbf{L}^2}$. Finally, because of the $\frac{1}{qT}$ periodicity of $\tilde{\mathrm{U}}$ we find

$$\begin{aligned}\|\mathfrak{Q}_T f\|_{\mathbf{L}^2}^2 &= qT^2\int_{-1/2qT}^{1/2qT}\overline{\tilde{\mathrm{U}}(\nu)}^{\mathrm{t}}\mathbf{A}(q\nu T)\tilde{\mathrm{U}}(\nu)d\nu \\ &= \frac{1}{q}\sum_n\int \overline{\hat{f}(\nu)}\hat{f}\left(\nu + \frac{n}{qT}\right)\overline{\hat{\tilde{\varphi}}\left(\nu T + \frac{n}{q}\right)}^{\mathrm{t}}\mathbf{A}(q\nu T)\hat{\tilde{\varphi}}\,(\nu T)d\nu,\end{aligned}$$

where **A** has been defined by (8). In the same spirit, we obtain

$$\begin{aligned}\langle f, \mathcal{Q}_T f\rangle &= qT^2 \int_{-1/2qT}^{1/2qT} \overline{\tilde{\mathrm{U}}(\nu)}^{\mathrm{t}} \mathrm{U}(\nu)\, d\nu \\ &= \frac{1}{q}\sum_n \int \overline{\hat{f}(\nu)}\hat{f}\left(\nu + \frac{n}{qT}\right) \overline{\hat{\tilde{\varphi}}\left(\nu T + \frac{n}{q}\right)}^{\mathrm{t}} \hat{\varphi}(\nu T)\, d\nu.\end{aligned}$$

From these formulæ, we identify two terms worthwhile to separate and write $\epsilon_f^2 = \epsilon_1^2 + \epsilon_2^2$, where

$$\epsilon_1^2 = \int |\hat{f}(\nu)|^2 E(\nu T)\, d\nu \tag{61}$$

$$\epsilon_2^2 = \frac{1}{q}\sum_{n\neq 0} \int \overline{\hat{f}(\nu)}\hat{f}\left(\nu + \frac{n}{qT}\right) \overline{\hat{\tilde{\varphi}}\left(\nu T + \frac{n}{q}\right)}^{\mathrm{t}} [\mathbf{A}(q\nu T)\hat{\tilde{\varphi}}(\nu T) - 2\hat{\varphi}(\nu T)]\, d\nu, \tag{62}$$

$E(\,\cdot\,)$ being defined by (21). Rewriting (62) using the orthonormalized basis functions further simplifies the formula. The second term exhibits a *double* form of aliasing; it vanishes whenever $\overline{\hat{f}(\nu)}\hat{f}\left(\nu + \frac{n}{qT}\right)\overline{\hat{\tilde{\phi}}\left(\nu T + \frac{n}{q}\right)}^{\mathrm{t}}(\hat{\tilde{\phi}}(\nu T) - 2\hat{\phi}(\nu T))$ cancels independently of ν and of $n \neq 0$.

C. Evaluation of ϵ_f

Due to the aliasing character of ϵ_2, we can benefit from decomposing f into its bandpass components f_k defined by (60): we necessarily have $\epsilon_{f_k}^2 = \int |\hat{f}_k|^2 E(\nu T)\, d\nu$. Using Minkowski's inequality we find

$$\epsilon_{f_0} - \sum_{k>0} \epsilon_{f_k} \leq \epsilon_f \leq \epsilon_{f_0} + \sum_{k>0} \epsilon_{f_k}. \tag{63}$$

Our first task is thus to estimate the sum of ϵ_{f_k}. As noticed earlier, due to (60), the frequency $|\nu|$ is lower bounded by $\dfrac{k}{2qT}$ when $k > 0$, on the support of $\hat{f}_k$. In particular, this implies that

$$\epsilon_{f_k} \leq M \frac{(qT)^r}{(\pi k)^r} \| f_k^{(r)} \|_{\mathbf{L}^2} \qquad \text{for } k > 0,$$

where the constant M is defined as $\sqrt{\|E\|_\infty}$. Thus, the Cauchy–Schwartz inequality on the discrete sums yields

$$\sum_{k>0} \epsilon_{f_k} \leq M\sqrt{\zeta(2r)}\left(\frac{qT}{\pi}\right)^r \| f^{(r)} - f_0^{(r)}\|_{\mathbf{L}^2}.$$

The second task is to bound ϵ_{f_0} in order to replace it with ϵ_1 in (63). Using the integral value for ϵ_{f_0} and the definition (61) of ϵ_1, then using Minkowski's inequality once again, we find

$$|\epsilon_1 - \epsilon_{f_0}| \leq M\left(\frac{qT}{\pi}\right)^r \| f^{(r)} - f_0^{(r)}\|_{\mathbf{L}^2}.$$

Finally, we combine these estimates,

$$\begin{aligned} |\epsilon_f - \epsilon_1| &\leq 2M\sqrt{\zeta(2r)}\,\frac{q^r}{\pi^r}\,T^r \| f^{(r)} - f_0^{(r)}\|_{\mathbf{L}^2} \\ &\leq K\alpha(f, T)T^r \| f^{(r)}\|_{\mathbf{L}^2} \\ &\leq KT^r \| f^{(r)}\|_{\mathbf{L}^2}, \end{aligned}$$

where we have bounded $1 + \zeta(2r)^{-1/2}$ by 2 to get the result of Theorem 1. Note that the middle inequality with

$$\alpha(f, T) = \frac{1}{\| f^{(r)}\|_{\mathbf{L}^2}}\left[\int_{|\nu|\geq 1/2qT} |2\pi\nu|^{2r} |\hat{f}(\nu)|^2 d\nu\right]^{1/2} \tag{64}$$

gives a sharper result because $\alpha(f, T) \leq 1$ vanishes as T goes to zero. Unfortunately, it is dependent on f; however, this proves that $\epsilon_f - \epsilon_1 = o(T^r)$. Also, it is possible to show that sharper bounds of ϵ_f can be obtained directly from (63) (i.e., using ϵ_{f_0} instead of ϵ_1); in that respect, Theorem 1 is essentially useful for studying the asymptotic behavior of ϵ_f.

APPENDIX D: PROOF OF THEOREM 3

Implication lhs $\Rightarrow$ rhs. $\epsilon_f = O(T^L)$ implies that the first term of (24) is $O(T^L)$. Due to the positivity of the integrand, and more precisely due to (21), this implies that $\int |\hat{f}(\nu)|^2 a(T\nu)d\nu$ is $O(T^{2L})$, where $a(\nu)$ is a template for $|q - \overline{\hat{\tilde{\varphi}}(T\nu)^{\mathrm{t}}}\hat{\varphi}(T\nu)|^2$ and $\left|\overline{\hat{\tilde{\varphi}}(T\nu)^{\mathrm{t}}}\hat{\varphi}\left(T\nu + \frac{n}{q}\right)\right|^2$ for $n \neq 0$. In addition, we note that our hypotheses imply that $\hat{\tilde{\varphi}}$ and $\hat{\varphi}$ can be differentiated $2L$ times and that all these derivatives are bounded: this property is thus also true for $a(\nu)$. We then show by induction that $a(\nu) = O(\nu^{2L})$: since a is bounded and continuous, we can apply Lebesgue's dominated convergence theorem to show that $\lim_{T\to 0} \int |\hat{f}(\nu)|^2 a(T\nu)d\nu = a(0)\int |\hat{f}|^2$, which proves that $a(0) = 0$; since $a(\nu)$ is symmetric in ν, and since a is at least twice differentiable, $a(\nu)\nu^{-2}$ is bounded, so that we can use the same reasoning as above for $\epsilon_f^2 T^{-2}$ this time yielding $\lim_{T\to 0} \int |\nu\hat{f}(\nu)|^2 a(T\nu)(T\nu)^{-2} d\nu = \frac{a''(0)}{2} \int |\nu \hat{f}|^2$. This induction can be repeated $2L$ times so that we find $a(\nu) = O(\nu^{2L})$. If we replace $a(\nu)$ with its templates, this is exactly condition (ii) of Lemma 2, which implies the quasi-biorthonormality of order L for the functions φ_n and $\tilde{\varphi}_n$.

Implication rhs ⇒ lhs. Let u be an arbitrary unit vector, and $u(x) = \bar{\mathrm{u}}^{\mathrm{t}}\varphi(x)$ a scalar function. A consequence of our hypotheses is that $\sup_x \sum_n |x - nq|^k |u(x - nq)| < \infty$ and $\int |x|^k |u(x)| dx < \infty$, for $k = 0, \ldots, 2L$, independent of the unitary vector u.

From $u(x)$, we define the functions $v_k(x) = \int \xi^k \bar{u}(\xi)(\xi + x)^k u(\xi + x) d\xi$ for $k = 0, \ldots, 2L$, which have the following properties: $v_k \in \mathbf{L}^1 \cap \mathbf{L}^\infty$, and consequently $v_k \in \mathbf{L}^2$, $\sum_n |v_k(nq)| < \infty$ and finally $\hat{v}_k = (1/(2\pi)^{2k})|\hat{u}^{(k)}|^2$.

Let us show that there exist constants $\{C_k\}_{k=0,\ldots,2L}$ such that for every $\mathbf{L}^2$ function g,

$$\sum_n \int |\hat{g}(\nu)|^2 \left| \hat{u}^{(k)}\left(\nu + \frac{n}{q}\right)\right|^2 d\nu \le C_k \|g\|_2^2. \tag{65}$$

For this, we first consider the case of indefinitely differentiable, compactly supported functions g, which are known to be dense in $\mathbf{L}^2$. In that case, it can easily be seen (using Poisson's summation formula for $g_2(x) = \int g(x + \xi) g(\xi) d\xi$) that

$$\int |\hat{g}(\nu)|^2 \sum_n \hat{v}_k\left(\nu + \frac{n}{q}\right) d\nu = q \sum_n g_2(nq) v_k(nq),$$

which proves that (65) is satisfied, if we let $q \sum_n |v_k(nq)| \le C_k < \infty$; note that C_k can be chosen independent of the unit vector u. In particular, (65) is still true if we consider only a finite summation on the lhs, since each term is positive. We can thus extend by density this result to the whole class of $\mathbf{L}^2$ functions g and, finally, prove (65).

Let us now consider the finite sum

$$a_N(\nu) = \sum_{0<|n|\le N} \left| \overline{\hat{\tilde{\varphi}}(\nu)}^{\mathrm{t}} \hat{\varphi}\left(\nu + \frac{n}{q}\right)\right|^2, \tag{66}$$

where N is an integer. a_N is $2L$ times differentiable and its $2L$th derivative is a finite (depending on L and not on N) sum (due to Leibnitz' rule) of terms of the form

$$\sum_{0<|n|\le N} \overline{\hat{\tilde{\varphi}}^{\alpha}(\nu)}^{\mathrm{t}} \hat{\varphi}^{(\beta)}\left(\nu + \frac{n}{q}\right) \hat{\tilde{\varphi}}^{(\gamma)}(\nu)^{\mathrm{t}} \overline{\hat{\varphi}^{(\delta)}\left(\nu + \frac{n}{q}\right)},$$

where $\alpha + \beta + \gamma + \delta = 2L$. Remembering that all the derivatives of $\hat{\tilde{\varphi}}$ are bounded up to the order $2L$ and using Cauchy–Schwartz inequality, we can thus claim that $\int |\hat{f}(\nu)|^2 \nu^{2L} a_N^{(2L)}(\nu T) d\nu$ is upper bounded by a constant $\times \| f^{(L)} \|_{\mathbf{L}^2}^2$; moreover this constant does not depend on N because of (65). Taking into account the quasi-biorthonormality hypothesis of the right-hand side of (30) (which provides $a_N^{(s)}(0) = 0$ for all $s = 0, \ldots, 2L - 1$), Taylor's theorem yields the inequality $\int |\hat{f}(\nu)|^2 a_N(\nu T) d\nu \leq C' T^{2L}$ where C' is a finite constant, independent of N. If we let N tend to infinity, this term remains $O(T^{2L})$ and on another side, its kernel tends to the second term of (21), the first term of (21) being $O(\nu^{2L})$ by hypothesis. This proves that $\epsilon_f = O(T^L)$.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers for their detailed reviews, resulting in a more rigorous redaction of this paper.

REFERENCES

1. A. Aldroubi, Oblique and hierarchical multiwavelet bases, *Appl. Comput. Harmon. Anal.* **4** (1997), 231–263.
2. A. Aldroubi, Oblique projections in atomic spaces, *Proc. Amer. Math. Soc.* **124** (1996), 2051–2060.
3. A. Aldroubi and M. Unser, Families of multiresolution and wavelet spaces with optimal properties, *Numer. Funct. Anal. Opt.* **14** (1993), 417–446.
4. A. Aldroubi and M. Unser, Sampling procedures in function spaces and asymptotic equivalence with Shannon's sampling theory, *Numer. Funct. Anal. Opt.* **15** (1994), 1–21.
5. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, Image coding using wavelet transform, *IEEE Trans. Image Process.* **1** (1992), 205–220.
6. G. Battle, A block spin construction of ondelettes. I. Lemarié functions, *Comm. Math. Phys.* **110** (1987), 601–615.
7. T. Blu, Iterated filter banks with rational rate changes—Connection with discrete wavelet transforms, *IEEE Trans. Signal Process.* **41** (1993), 3232–3244.
8. T. Blu and M. Unser, Quantitative Fourier analysis of approximation techniques, *IEEE Trans. Signal Process.*, submitted.
9. H. Burchard and J. Lei, Coordinate order of approximation by functional-based approximation operators, *J. Approx. Theory* **82** (1995), 240–256.
10. E. Cheney and W. Light, Quasi-interpolation with basis functions having non-compact support, *Constr. Approx.* **8** (1992), 35–48.
11. C. Chui and H. Diamond, A characterization of multivariate quasi-interpolation formulas and applications, *Numer. Math.* **57** (1990), 105–121.
12. C. Chui and J. Wang, On compactly supported spline wavelets and a duality principle, *Trans. Amer. Math. Soc.* **330** (1992), 903–915.
13. A. Cohen and I. Daubechies, Orthonormal bases of compactly supported wavelets. III. Better frequency resolution, *SIAM J. Math. Anal.* **24** (1993), 520–527.
14. A. Cohen, I. Daubechies, and J. Feauveau, Biorthogonal basis of compactly supported wavelets, *Comm. Pure Appl. Math.* **45** (1992), 485–560.
15. A. Cohen, I. Daubechies, and G. Plonka, Regularity of refinable function vectors, *J. Fourier Anal. Appl.* (1997), 295–324.
16. W. Dahmen and C. Micchelli, On the approximation order from certain multivariate spline spaces, *J. Austral. Math. Soc. Ser. B* **26** (1984), 233–246.

17. W. Dahmen and C. Micchelli, Biorthogonal wavelet expansions, *Constr. Approx.* **13** (1997), 293–328.
18. I. Daubechies, Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **41** (1988), 909–996.
19. I. Daubechies, "Ten Lectures on Wavelets," SIAM, Philadelphia, PA, 1992.
20. C. de Boor, The polynomials in the linear span of integer translates of a compactly supported function, *Constr. Approx.* **3** (1987), 199–208.
21. C. de Boor, Quasi-interpolation and approximation power of multivariate splines, *in* "Computation of Curves and Surfaces" (W. D. *et al.*, Eds.), pp. 313–345, Kluwer, Dordrecht, 1990.
22. C. de Boor, R. Devore, and A. Ron, Approximation from shift invariant subspaces of $\mathbf{L}^2(\mathbb{R}^d)$, *Trans. Amer. Math. Soc.* **341** (1994), 787–806.
23. C. de Boor, R. Devore, and A. Ron, The structure of finitely generated shift-invariant spaces in $\mathbf{L}^2(\mathbb{R}^d)$, *J. Funct. Anal.* **119** (1994).
24. C. de Boor and G. Fix, Spline approximation by quasi-interpolants, *J. Approx. Theory* **8** (1973), 19–45.
25. C. de Boor and R. Jia, Controlled approximation and a characterization of the local approximation order, *Proc. Amer. Math. Soc.* **95** (1985), 547–553.
26. T. Goodman and S. Lee, Wavelets of multiplicity r, *Trans. Amer. Math. Soc.* **342** (1993), 307–324.
27. C. Heil, G. Strang, and V. Strela, Approximation by translates of refinable functions, *Numer. Math* **73** (1996), 75–94.
28. L. Hervé, Multi-resolution analysis of multiplicity d. Application to dyadic interpolation, *Appl. Comput. Harmon. Anal.* **1** (1994), 299–315.
29. K. Jetter and D. Zhou, Order of linear approximation from shift invariant spaces, *Constr. Approx.* **11** (1995), 423–438.
30. R. Jia, A counterexample to a result concerning controlled approximation, *Trans. Amer. Math. Soc.* **97** (1986), 647–654.
31. R. Jia, Shift-invariant spaces on the real line, *Proc. Amer. Math. Soc.* **125** (1997), 785–793.
32. R. Jia and J. Lei, Approximation by multiinteger translates of functions having global support, *J. Approx. Theory* **72** (1993), 2–23.
33. J. Lei, $\mathbf{L}^p$-approximation by certain projection operators, *J. Math. Anal. Appl.* **185** (1994), 1–14.
34. P.-G. Lemarié, Ondelettes à localisation exponentielle, *J. Math. Pures Appl.* **67** (1988), 227–236. [In French]
35. S. Mallat, A theory for multiresolution signal decomposition: The wavelet decomposition, *IEEE Trans. Pattern Anal. Mach. Intell.* **11** (1989), 674–693.
36. Y. Meyer, "Ondelettes," Hermann, Paris, 1990. [In French]
37. G. Plonka, Approximation order provided by refinable function vectors, *Constr. Approx.* **13** (1997), 221–244.
38. O. Rioul, On the choice of wavelet filters for still image compression, *in* "Proc. IEEE Int. Conf. Acoust., Speech, Sig. Proc.," Vol. V, pp. 550–553, Minneapolis, MN, Apr. 1993.
39. I. Schoenberg, Contribution to the problem of approximation of equidistant data by analytic functions, *Quart. Appl. Math.* **4** (1946), 45–99, 112–141.
40. L. Schwartz, "Théorie des Distributions," Hermann, Paris, 1966. [In French]
41. C. Shannon, Communication in the presence of noise, *in* "Proc. IRE," Vol. 37, pp. 10–21, Jan. 1949.
42. G. Strang and G. Fix, A Fourier analysis of the finite element variational method, *in* "Constructive Aspect of Functional Analysis" (Cremonese, Ed.), pp. 796–830, Rome, 1971.
43. G. Strang and T. Nguyen, "Wavelets and Filter Banks," Wellesley–Cambridge Press, Cambridge, MA, 1996.
44. W. Sweldens and R. Piessens, Quadrature formulæ and asymptotic error expansions for wavelet approximations of smooth functions, *SIAM J. Math. Anal.* **31** (1994), 1240–1264.
45. M. Unser, Approximation power of biorthogonal wavelet expansions, *IEEE Trans. Signal Process.* **44** (1996), 519–527.
46. M. Unser, Quasi-orthogonality and quasi-projections, *Appl. Comput. Harmon. Anal.* **3** (1996), 201–214.

47. M. Unser, A. Aldroubi, and M. Eden, On the asymptotic convergence of B-spline wavelets to Gabor functions, *IEEE Trans. Inform. Theory* **38** (1992), 864–872.

48. M. Unser and I. Daubechies, On the approximation power of convolution-based least squares versus interpolation, *IEEE Trans. Signal Process.* **45,** (1997), 1697–1711.

49. M. Vetterli and C. Herley, Wavelets and filter banks: Theory and design, *IEEE Trans. Signal Process.* **40** (1992), 2207–2232.

50. E. Whittaker, The Fourier theory of the cardinal functions, *Proc. Math. Soc. Edinburgh* **1** (1929), 169–176.

51. X. Xia and B. Suter, Vector-valued wavelets and vector filter banks, *IEEE Trans. Signal Process.* **44** (1996), 508–518.