Vanishing moments and the approximation power of wavelet expansions

Michael Unser

Image Processing Group,
Biomedical Engineering and Instrumentation Program, NCRR, Bldg. 13, Room 3N17
National Institutes of Health, Bethesda, MD 20892-5766 USA.
E-mail: unser@helix.nih.gov

ABSTRACT

The order of a wavelet transform is typically given by the number of vanishing moments of the analysis wavelet. The Strang-Fix conditions imply that the error for an orthogonal wavelet approximation at scale $a = 2^{-i}$ globally decays as a^L, where L is the order of the transform. This is why, for a given number of scales, higher order wavelet transforms usually result in better signal approximations. We show that this result carries over for the general biorthogonal case and that the rate of decay of the error is determined by the order properties of the synthesis scaling function alone. We also derive asymptotic error formulas and show that biorthogonal wavelet transforms are equivalent to their corresponding orthogonal projector as the scale goes to zero. These results strengthen Sweldens' earlier analysis and confirm that the approximation power of biorthogonal and (semi-)orthogonal wavelet expansions is essentially the same. Finally, we compare the asymptotic performance of various wavelet transforms and briefly discuss the advantages of splines. We also indicate how the smoothness of the basis functions is beneficial in reducing the approximation error.

1. INTRODUCTION

For researchers working with multirate filterbanks, the mathematical theory of the wavelet transform brought about the new constraint of designing filterbanks with a certain number of zeros (multiplicity L) at $z = -1$ [1]. One of the initial justifications for selecting a zero of multiplicity L is that this condition is necessary for constructing regular wavelets with $L-1$ continuous derivatives [2]. Unfortunately, it is not sufficient and the regularity index of most wavelet bases is usually much smaller than L-1. Another motivation is that the order properties of the refinement filter get translated into a corresponding number of vanishing moments for the analysis wavelet. These vanishing moments can play a crucial role in the characterization of the local Hölder exponent of singularities [3]. These are all reasons why the order properties of the wavelet transform are generally believed to be useful in applications.

Beside the regularity of the basis functions themselves, there is also another compelling reason for using higher order wavelet decompositions, which takes its roots in approximation theory [4]. Specifically, if $P_a f$ denotes the least squares approximation (orthogonal projection) of a function in a multiresolution space at scale $a = 2^i$, then the theory indicates that the error must decay like $O(a^L)$ where L is the order of the representation. This fundamental result may explain why higher order wavelets are usually preferable for data compression. This particular aspect of the wavelet theory is not well known in signal/image processing, but it is probably very relevant to this particular area of application.

In this communication, we emphasize the relevance to wavelets of the Strang-Fix theory of approximation which was developed in the early 70s. We then present some new extensions for oblique projections that are directly applicable to biorthogonal wavelet expansions. We also provide asymptotic error formulas and give explicit bound constants which allow the comparison of various wavelet transforms. Finally, we consider the use of the two-scale relation for the calculation of the asymptotic bound constants and propose a new explanation of why the smoothness of the basis functions has a reducing effect on the approximation error. Some of these results are given without proof; for a complete mathematical treatment we refer to [5].

2. BASIC RESULTS

In the Strang-Fix theory of approximation, functions are represented in terms of the re-scaled translates of a generating function φ [4]. However, since there is no requirement for the two-scale relation, the setting is more general than that of the wavelet transform.

Definition 2.1 : An Lth order generating function is a function $\varphi \in L_2$ with the following properties.

(i) Riesz Basis Condition

$$0 < A \le \sum_{k\in Z} |\hat{\varphi}(\omega + 2\pi k)|^2 \le B < +\infty$$

(ii) Order Property

$$\hat{\varphi}(0) = 1 \text{ and } \hat{\varphi}^{(m)}(2\pi k) = 0,\ k \in Z,\ k \ne 0 \quad \text{for } (m = 0,\dots,L-1),$$

where $\hat{\varphi}(\omega)$ is the Fourier transform of φ, and $\hat{\varphi}^{(m)}(\omega)$ denotes its mth derivative with respect to ω.

Instead of powers of two, there is a continuous scale parameter a; the corresponding approximation space is

$$V_a(\varphi) = \left\{ f_a(x) = a^{-1/2} \sum_{k\in Z} c_a(k)\varphi(x/a - k) : c_a \in l_2 \right\}. \quad (1)$$

Since the spaces $V_a(\varphi)$ are rescaled versions of each other, we can discuss these conditions in terms of the basic space $V_1(\varphi) = \mathrm{span}\{\varphi(x-k)\}_{k\in Z}$. Condition (i) implies that $V_1(\varphi)$ is a closed subspace of L_2 with $\{\varphi(x-k)\}_{k\in Z}$ as its Riesz basis [6]. In other words, the basis functions are linearly independent and each function $f_1 \in V_1(\varphi)$ has a stable and unique representation in terms of its coefficients $c_1(k)$. Condition (ii) implies that the generating function φ reproduces all polynomials de degree n=L-1 [4]. This is also equivalent to say that there exists a function $\varphi_{QI}(x) \in V_1(\varphi)$ that interpolates all polynomials of degree n, including the monomials

$$x^n = \sum_{k\in Z} k^n \varphi_{QI}(x-k),\ n = 0,\dots,L-1. \quad (2)$$

Such a function, which is typically not unique, is called a quasi-interpolant of degree L-1 [7, 8]. The basic theoretical tool for establishing these polynomial reproducing properties is Poisson's summation formula. As an example, we derive the equivalence between Property (ii) with L=1 and the well known "partition of unity" condition (reproduction of a constant)

$$\sum_{k\in Z} \varphi(x-k) = \sum_{k\in Z} \hat{\varphi}(2\pi k) = 1. \quad (3)$$

The order condition (ii) has some important theoretical consequences on the rate of approximation of functions in L_2.

Theorem 2.2 (Strang-Fix) : If φ is an Lth order generating function with appropriate decay then the minimum approximation error at step size a for an arbitrary function f (sufficiently smooth in L_2) is bounded as follows

$$\inf_{f_a \in V_a(\varphi)} \|f - f_a\| \le C_\varphi \cdot a^L \cdot \|f^{(L)}\| \quad (4)$$

where C_φ is a constant that is independent of f and where $\|f^{(L)}\|$ denotes the norm of the Lth derivative of f.

In the particular case of the wavelet transform, the generator φ has the additional multiresolution property; the scale is also restricted to powers of two.

Definition 2.3 : An Lth order scaling function is an Lth order generating function if it satisfies the additional two-scale relation:

$$\varphi(x/2) = \sum_{k\in Z} h(k)\varphi(x-k). \quad (5)$$

Since Theorem 2.2 applies to the minimum error approximation (orthogonal projection), it is obviously also applicable for orthogonal and semi-orthogonal wavelet transforms [9]. In particular, this result implies that the error decays as $O(a^L)$ as the scale $a = 2^i$ becomes sufficiently small.

By applying the refinement equation (5) *ad infinitum*, we get the equivalent infinite product representation of the Fourier transform of the scaling function φ

$$\hat{\varphi}(\omega) = \prod_{i=1}^{+\infty} H(e^{j\omega/2^i}). \quad (6)$$

This relation is crucial for establishing the connection between the order condition (ii) and the better known wavelet properties mentioned in the introduction.

Proposition 2.4: If $H(e^{j\omega})$ has zeros of multiplicity L at $\omega = \pi$ (or if $H(z) = 2^{-L}(1+z)^L \cdot Q(z)$ where $Q(z)$ is a stable transfer function) then φ is an Lth order scaling function.

Proof: Let us write $k = 2^n q$ with q odd, a representation that always exists for any non-zero integer k. Using (6), we factorize and differentiate $\hat{\varphi}(\omega)$ as follows

$$\hat{\varphi}^{(m)}(\omega) = \sum_{k=0}^{m} H^{(k)}(e^{j\omega/2^n}) \left[\prod_{\substack{i=1 \\ i\ne n}}^{+\infty} H(e^{j\omega/2^i}) \right]^{(m-k)}.$$

For $\omega = 2\pi k = 2\pi \cdot (2^n q)$ with q odd, the leading factors in the sum all vanish because $H^{(k)}(e^{j\pi q}) = H^{(k)}(e^{j\pi}) = 0$ (recall that $H^{(k)}(e^{j\xi})$ is 2π-periodic). Thus, $\hat{\varphi}^{(m)}(2\pi k) = 0$ for $k \in Z,\ k \ne 0$. ❑

Proposition 2.5: For an *Lth* order wavelet transform (i.e., the synthesis scaling function has an Lth order of approximation), the biorthogonal analysis wavelet $\tilde{\psi}$ has L vanishing moments.

Proof: Let $V(\varphi_2)$ denote the corresponding synthesis space. Since φ_2 is an Lth order function, there also exists a quasi-interpolant $\varphi_{QI} \in V(\varphi_2)$ such that (2) is verified. For m=0,..., $L-1$, we then write

$$\int_{-\infty}^{+\infty} x^m \tilde{\psi}(x)dx = \int_{-\infty}^{+\infty} \sum_{k\in Z} k^m \varphi_{QI}(x-k)\tilde{\psi}(x)dx$$
$$= \sum_{k\in Z} k^m \langle \varphi_{QI}(x-k), \tilde{\psi}(x)\rangle dx = 0,$$

where the inner products on the right are zero because $\tilde{\psi}$ is perpendicular $V(\varphi_2)$ by construction. For the proof to be complete, we also require φ_2 and φ_{QI} to have sufficient decay (e.g. $|\varphi_2(x)| \le C\cdot|1+x|^{-L-\varepsilon}$ with ε>0) so that we can safely permute the infinite sum and the integral. ❑

3. NEW THEORETICAL RESULTS FOR BIORTHOGONAL EXPANSIONS

The Strang-Fix results only applies to the (semi-) orthogonal case (orthogonal projection). To obtain more general bounds and formulas, we consider the oblique projection operator

$$\left(\tilde{P}_a f\right)(x) = a^{-1}\sum_{k\in Z}\langle f, \tilde{\varphi}(x/a-k)\rangle\cdot\varphi(x/a-k) \qquad (7)$$

that computes the projection of $f\in L_2$ into the synthesis space $V_a(\varphi)$ perpendicular to the analysis space $V_a(\tilde{\varphi})$, where φ and $\tilde{\varphi}$ are two biorthogonal (or dual) generating functions [10].

We can then prove the following, which is the generalization of Theorem 2.2 for the biorthogonal case.

Theorem 3.1 : If φ is an Lth order generating function then the oblique projection error at scale a for an arbitrary function f (sufficiently smooth in L_2) is bounded as follows

$$\inf_{f_a\in V_a(\varphi)}\|f-f_a\| \le \|f-\tilde{P}_a f\| \le C_{\varphi,\tilde{\varphi}}\cdot a^L\cdot\|f^{(L)}\|, \qquad (8)$$

where $C_{\varphi,\tilde{\varphi}} = C/\cos\theta$ is a constant that is independent of f; θ is also the angle between the spaces $V_a(\varphi)$ and $V_a(\tilde{\varphi})$.

Proof: This result can be derived as a corollary Strang's L_2-bound (4). For this purpose, we make use of a rescaled version of Theorem 3 in [10] which provides a direct bound between the orthogonal and oblique projection errors

$$\|f-P_a f\| \le \|f-\tilde{P}_a f\| \le \frac{1}{\cos\theta}\|f-P_a f\|, \qquad (9)$$

where $\cos\theta$ is given by

$$\cos\theta = \operatorname*{ess\,inf}_{\omega\in(0,1/2)} \frac{\sum_{k\in Z}\hat{\varphi}(\omega+2\pi k)\cdot\overline{\hat{\tilde{\varphi}}(\omega+2\pi k)}}{\sqrt{\sum_{k\in Z}|\hat{\varphi}(\omega+2\pi k)|^2}\cdot\sqrt{\sum_{k\in Z}|\hat{\tilde{\varphi}}(\omega+2\pi k)|^2}}. \qquad (10)$$

Since the generating functions are biorthogonal, we can easily show that $\cos\theta \ge (B\tilde{B})^{-1} > 0$, where B and $\tilde{B}$ are the upper frame bounds for φ and $\tilde{\varphi}$ in Definition 2.1, respectively. Hence, combining (4) with (9), we get

$$\|f-P_a f\| \le \|f-\tilde{P}_a f\| \le \frac{1}{\cos\theta}\|f-P_a f\| \le C_{\varphi,\tilde{\varphi}}\cdot a^N\cdot\|f^{(N)}\|,$$

with a finite constant $C_{\varphi,\tilde{\varphi}} = C_\varphi/\cos\theta$. ❑

This error bound shows that oblique and orthogonal projection operators are qualitatively equivalent.

3.1 Asymptotic error analysis

In practice, the agreement between orthogonal and biorthogonal projections is usually much better than what is suggested by the bound (8) which reflects the worst case scenario (e.g., the safety factor 1/cosθ). In fact, we can use a Taylor series argument to derive the following asymptotic equivalence result, which also provides an explicit formula for the smallest possible bound constant C.

Theorem 3.2 : If the analysis function $\tilde{\varphi}$ satisfies the partition of unity condition (i.e., if it is a first order function), then the Lth order oblique projection operator (1) is asymptotically optimal (i.e., equivalent to the orthogonal projection or minimum error solution). The asymptotic error is given by

$$\|f-\tilde{P}_a f\| = C_\varphi^- \cdot a^L\cdot\|f^{(L)}\| + O(a^{L+1}) \quad \text{as } a\to 0, \qquad (11)$$

where

$$C_\varphi^- = \frac{1}{L!}\left(\sum_{\substack{k\in Z\\ k\neq 0}}|\hat{\varphi}^{(L)}(2\pi k)|^2\right)^{1/2}. \qquad (12)$$

This result is consistent with an earlier asymptotic error analysis by Sweldens [11]. However, it provides some improvement in two important respects. First, the computation is more direct in the sense that it avoids using wavelet expansions. The main benefit is a much simpler formula for the bound constant. Second, the present error estimate is sharper (smaller constant) and asymptotically exact. The complete proof can be found in [5].

Theorem 3.2 is a quantitative result that is applicable to all wavelet transforms. It essentially shows that their performance primarily depends on the synthesis space — in other words, orthogonal and biorthogonal wavelet transforms are asymptotically equivalent.

3.2 Bound constant computation

The computation of the bound constant C_φ^- can serve as a basis for the comparison of different wavelet transforms. While the explicit relation (12) is directly applicable in certain special cases such as splines, we also need a mechanism to compute the required derivatives when the scaling function is specified indirectly through the refinement filter h. To derive such a relation, we rewrite the two-scale relation (5) in the Fourier domain and differentiate L times, applying the chain rule:

$$\hat{\varphi}^{(L)}(\omega)=\sum_{m=1}^{L}\binom{L}{m}\frac{1}{2^m}H^{(m)}(e^{j\omega/2})\cdot\frac{1}{2^{L-m}}\hat{\varphi}^{(L-m)}(\omega/2) \quad (13)$$

Using the property $H^{(m)}(e^{j(2l+1)\pi})=0$, m=0,...,L-1, we obtain a direct expression for the odd indexed derivatives in (12):

$$\hat{\varphi}^{(L)}(2\pi k)=\frac{1}{2^L}H^{(L)}(e^{j\pi})\cdot\hat{\varphi}(\pi k) \qquad \text{(for } k \text{ odd).}$$

Next, we apply the basic factorization $H(z)=2^{-L}(1+z)^L\cdot Q(z)$ (cf. Proposition 2.4), and evaluate $H^{(L)}(e^{j\omega})$ explicitly

$$H^{(L)}(e^{j\omega})\Big|_{\omega=\pi}=\frac{L!}{2^L}\cdot Q(e^{j\pi}),$$

which finally yields for k=2l+1:

$$\hat{\varphi}^{(L)}(2\pi k)=\left(\frac{L!}{2^{2L}}\right)\cdot Q(e^{j\pi})\cdot\hat{\varphi}(\pi k) \qquad (k \text{ odd}). \qquad (14)$$

Likewise, using the order property (ii) and the fact that $H(e^{j2\pi l})=1$ in (13), we determine the remaining even indexed coefficients using the recursive rule

$$\hat{\varphi}^{(L)}(2\pi k)=\hat{\varphi}^{(L)}(4\pi l)=\frac{1}{2^L}\cdot\hat{\varphi}^{(L)}(2\pi l), \quad (k \text{ even}). \qquad (15)$$

These last two equations are then used to compute the bound constant in (12) numerically, summing up over a sufficient number of terms.

Table I provides a comparison for various wavelet transforms. Note that the results for splines are applicable for any type of spline wavelet transform (orthogonal Battle-Lemarié, semi-orthogonal Chui-Wang / Unser-Aldroubi, and biorthogonal Cohen-Daubechies-Feauveau). For a given order L, we see that the Daubechies wavelets have the worst performance; splines are by far the best. These results suggest that splines at half the resolution can provide as good an approximation as Daubechies wavelets at twice the rate. In general, the performance is better for the scaling functions that are the most regular.

3.3 Discussion

The results in Table 1 indicate that some representations — splines, in particular— are more favorable than others for approximating smooth functions. If we look at (14), we can easily identify the ingredients that are important for good asymptotic performance. Clearly, it is preferable to have $Q(e^{j\pi})$ small and $\hat{\varphi}$ decaying fast. This last property is primarily dependent on the regularity of the scaling function. Specifically, if $\varphi \in C^m$ (m times continuously differentiable) then its Fourier transform decays at least as $O(\omega^{-m})$. Thus, we can conclude that the smoothest scaling functions should have the better performance. Once again, this strongly points toward splines which are among the better behaved functions of a given order L. While it is possible to construct examples of scaling functions with even more smoothness, B-splines turn out to be the shortest ones for a given order L. They therefore appear to be optimal if we include the filter length constraint in the design. This is perhaps one of the primary reasons why biorthogonal spline wavelets perform so well in coding applications.

TABLE 1 : RESCALED BOUND CONSTANT $A_L^-=C_\varphi^-\cdot L!$ FOR DIFFERENT WAVELET FAMILIES.

L	Daubechies	closest-to-linear phase	coiflets	spline	Deslauriers-Dubuc
1	0.2887	0.2887		0.2887	
2	0.2236	0.2236	0.2124	0.07454	0.07454
3	0.2988	0.2988		0.03450	
4	0.5557	0.5557	0.4953	0.02182	0.1871
5	1.316	1.316		0.01734	
6	3.779	3.779	3.231	0.01655	1.212
7	12.74	12.74		0.01844	
8	49.35	49.35	40.92	0.02347	15.06
9	215.8	215.8		0.03362	

References

[1] M. Vetterli and J. Kovacevic, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice Hall, 1995.

[2] I. Daubechies, *Ten lectures on wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.

[3] S. Mallat and W.L. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 617-643, March 1992.

[4] G. Strang and G. Fix, "A Fourier analysis of the finite element variational method," in *Constructive Aspect of Functional Analysis*, Rome: Edizioni Cremonese, pp. 796-830, 1971.

[5] M. Unser, "Approximation power of biorthogonal wavelet expansions," *IEEE Trans. Signal Processing*, vol. 44, no. 3, pp. 519-527, March 1995.

[6] A. Aldroubi and M. Unser, "Sampling procedures in function spaces and asymptotic equivalence with Shannon's sampling theory," *Numer. Funct. Anal. and Optimiz.*, vol. 15, no. 1&2, pp. 1-21, February 1994.

[7] C. de Boor and G. Fix, "Spline approximation by quasi-interpolants," *J. Approximation Theory*, vol. 8, pp. 19-45, 1973.

[8] M. Unser, "Quasi-orthogonality and quasi-projections," *Applied and Computational Harmonic Analysis*, in press.

[9] G. Strang, "Wavelets and dilation equations: a brief introduction," *SIAM Review*, vol. 31, pp. 614-627, 1989.

[10] M. Unser and A. Aldroubi, "A general sampling theory for non-ideal acquisition devices," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 2915-2925, November 1994.

[11] W. Sweldens and R. Piessens, "Asymptotic error expansions for wavelet approximations of smooth functions II," *Numer. Math.*, vol. 68, no. 3, pp. 377-401, 1994.