DISCRETE FINITE VARIATION: A NEW MEASURE OF SMOOTHNESS FOR THE DESIGN OF WAVELET BASIS

Jan E. Odegard   *C. Sidney Burrus*

Department of Electrical and Computer Engineering
Rice University, Houston, TX 77251-1892, USA
odegard@rice.edu   *csb@rice.edu*

ABSTRACT

A new method for measuring and designing smooth wavelet basis which dispenses with the need for having a large number of zero moments of the wavelet is given. The method is based on minimizing the "discrete finite variation", and is a measure of the local "roughness" of a *sampled* version of the scaling function giving rise to "visually smooth" wavelet basis. Smooth wavelet basis are deemed to be important for several applications and in particularly for image compression where the goal is to limit spurious artifacts due to non-smooth basis functions in the presence of quantization of the individual subbands. The definition of smoothness introduced here gives rise to new algorithms for designing smooth wavelet basis with only one vanishing moment leaving free parameters, otherwise used for setting moments to zero, for optimization.

1. INTRODUCTION

Hölder and Sobolev exponents are fundamental mathematical measures of smoothness giving precise definition of differentiability. Several recent publications (in particular [6, 2]) has provided algorithms for estimating both the Hölder and the Sobolev exponent for a given wavelet basis given the associated scaling filter. However, to obtain good estimates it is necessary to provide the regularity or more specifically the number of zero moments of the wavelet function (or filter). Although this is a simple task if an estimate of the exponent for a given filter, designed to have a fixed number of zero moments, is desired, the task is significantly harder if one first have to estimate the number of zero moments. In fact, estimating the number of zero moments is exactly what has to be done if the Hölder or Sobolev exponents were to be used as cost functions in an optimization algorithm for finding the optimally smooth wavelet of a given support. It should be noticed that the problem of estimating the number of zero moments is equivalent to finding roots of a polynomial which is known to be a numerically ill conditioned problem. Due to the numerical problems neither the Hölder nor the Sobolev exponent are suited cost functions for designing smooth wavelets using nonlinear optimization. Recently a method for designing "more" differentiable wavelets (compared to the K-regular Daubechies solution) has been developed by Lang and Heller [4]. However, it does necessarily require that a large number of moments be set to zero.

First recall that the continuous (time) wavelet basis functions are obtained by iterating the associated filter bank an infinite number of times. Now realizing that any practical application will only involve a *finite* number of iterations, the associated functions are not continuous but discrete for which the strict definition of differentiability given by the Hölder and Sobolev exponents are not (necessarily) meaningful. As a result alternative descriptions of smoothness incorporating "visual" smoothness behavior (e.g., *local variation*) should be considered.

This work was supported by ARPA and Texas ATP.
Appears in *Proc. ICASSP-96*, Atlanta, GA.

Furthermore, one can show [5] that a good approximation to the discrete wavelet transform (DWT) can be implemented efficiently by approximate orthogonal rotations. However, by applying approximate orthogonal rotations one can only make the zeroth moment vanish exactly and only hope that higher order moments are small. Hence, with the goal of implementing the DWT efficiently, the fact that exact zero moments of higher order can not be achieved should be incorporate in to the design specifications. If in addition to only requiring that the first moment be set to zero it is imposed that the lattice angles be of the form

$$\beta_i = \sum_{j=0}^{W-1} b_{i,j} \arctan(1/2^j) \tag{1}$$

where W is the specified word-length and $b_{i,j} \in \{0,1\}$ one can in theory design (e.g., by integer programming) wavelet basis which can be implemented *exactly* by a small number of μ-rotations.

Based on these observations and the fact that prescribing only one zero moment (the zeroth moment of $\psi_1(t)$) the wavelet basis can at best be C^1 and hence measures of smoothness such as the Hölder and Sobolev exponents are insufficient for our purpose. The results presented in this paper is building on ideas first seen in [1] where total variation was introduced as a way to optimize the smoothness of the prototype filter for M-band cosine modulated orthonormal wavelets basis. The new measure, discrete finite variation, can be optimized independent of the number of vanishing moments giving rise to smooth wavelet basis. Furthermore, since the new measure does not depend on having a large number of vanishing moments for obtaining smooth basis it can in principle be used to solve the lattice angle quantization problem prescribed by (1).

1.1. Wavelets

In the remainder of the paper the discussion will be restricted to 2-band orthonormal wavelet basis. However, most of the theory easily applies to the M-band solution as well as the biorthogonal solution. A 2-band compactly supported orthonormal wavelet bases are characterized by a scaling filter, $h_0(k)$, and wavelet filter, $h_1(k)$, both of finite length $N = 2K$, satisfying the linear condition

$$\sum_{n=0}^{N-1} h_0(n) = \sqrt{2} \tag{2}$$

and for all $k \in \mathbb{Z}$ the quadratic condition

$$\sum_{n=0}^{N-1} h_i(n) h_j(n+2k) = \delta(i-j)\delta(k) \quad i,j \in \{0,1\}\,. \tag{3}$$

The scaling function, $\psi_0(t)$, and the wavelet function, $\psi_1(t)$, is then defined for $i = 0, 1$ by the dyadic difference equation

$$\psi_i(t) = \sqrt{2}\sum_k h_i(k)\psi_0(2t-k). \tag{4}$$

Exact hardware implementations of a discrete wavelet transform (DWT) must satisfy both the orthogonality as well as the wavelet zero moment condition. In fact, for the maximally vanishing moment solution (Daubechies) there are $K = \frac{N}{2}$ vanishing moments. It is well known that implementing the DWT by lattice rotations orthogonality is structurally imposed through the basic building blocks (2×2 orthogonal rotations) [7]. It is also well know that by imposing that the sum of the lattice angles be $-\frac{\pi}{4}$ the zeroth moment is guaranteed to be zero (e.g., one can easily guarantee the implementation of an orthogonal wavelet basis). In fact, choosing any set of lattice angles as long as the sum is constrained to be $-\frac{\pi}{4}$ an orthogonal wavelet transformation is guaranteed independent of the properties (e.g., smoothness, regularity, stop band characteristics etc.) of the corresponding scaling and wavelet filters/functions.

If more than one vanishing moment (to high numerical accuracy) is required this can be achieved by implementing the lattice filter with floating point multipliers. However, if efficient implementations are desired [5] one has to give up the moment approximation property in favor of efficient arithmetic (replace floating point multiplies with a few shifts and adds).

The goal is to obtain an algorithm for designing "smooth" wavelet basis that can be implement efficiently without sacrificing properties obtained by the design. The first step in this direction is to obtain a new measure of smoothness coupling the constraints given by the lattice architecture and the design algorithm.

2. SMOOTHNESS

2.1. The Hölder and Sobolev measure

Definition 1. ***[Hölder continuity]***
Let $\varphi : \mathbb{R} \to \mathbb{C}$ and let $0 < \alpha \le 1$. Then the function φ is Hölder continuous of order α if there exist a constant c such that

$$|\varphi(x) - \varphi(y)| \le c\,|x-y|^{\alpha} \quad \text{for all } x, y \in \mathbb{R} \tag{5}$$

Based on the above definition, φ has to be a constant if $\alpha > 1$. This is not very useful for determining continuity of order $\alpha > 1$. However, using the above definition Hölder continuity of any order $r > 0$ is defined as follows:

Definition 2. ***[Hölder exponent]***
A function $\varphi : \mathbb{R} \to \mathbb{C}$ is continuous of order $r = P + \alpha$ ($0 < \alpha \le 1$) if $\varphi \in C^P$ and its Pth derivative is Hölder continuous of order α. Then r is said to be the Hölder exponent of φ

Similarly the Sobolev exponent is a measure of the decay of the Fourier transform and is given by the following definition.

Definition 3. ***[Sobolev exponent]***
Let $\varphi : \mathbb{R} \to \mathbb{C}$, then φ is said to belong to the Sobolev space of order s ($\varphi \in H^s$) if

$$\int_{\mathbb{R}} (1+|\omega|^2)^s\,|\hat{\varphi}(\omega)|^2\,d\omega < \infty \tag{6}$$

Then s is said to be the Sobolev exponent of φ

Notice, that although Sobolev exponent does not give explicit order of differentiability it does yield an lower and upper bound on r, the Hölder exponent, and hence the differentiability of φ. This can be seen from the following inclusions:

$$H^{s+1/2} \subset C^r \subset H^s \tag{7}$$

2.2. Bounded variation and total variation

In the following discussion we will consider vector-valued functions, $f : [a,b] \to \mathbb{R}^n$, and the Euclidean norm on $\mathbb{R}^n$.

Definition 4 (Total variation). *Let $f(x)$ be a vector-valued function such that $f : [a,b] \to \mathbb{R}^m$ and let $\mathcal{X}$ be the set of m points partitioning $[a,b]$ such that*

$$\mathcal{X} = \{x_i : a = x_0 < x_1 < \cdots < x_{m-1} = b\}.$$

Form the sum

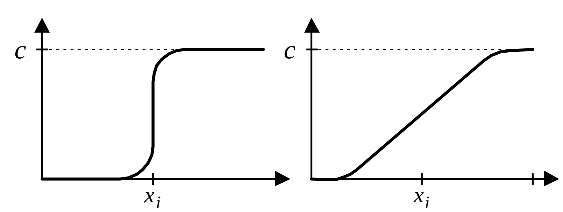
$$V_f(a,b) = \sum_{i=0}^{m-1} |f(x_i) - f(x_{i-1})| \tag{8}$$

then

$$\mathcal{V} = \sup_{\mathcal{X}} V_f(a,b) \tag{9}$$

is defined to be the total variation of f on $[a,b]$. Furthermore, in case $\mathcal{V} < \infty$ then we say the function f is of bounded variation.

Functions of bounded variation have a number of interesting properties [3, pp. 530-543], however, for the purpose of applying ideas from the theory of bounded variation to the design of smooth wavelets these properties are not relevant and in fact "total variation" is not the exact quantity that should be minimized. Also, noticing that (9) involves finding the least upper bound (e.g., sup) over all possible partitions on $[a,b]$ (requiring that the scale tend to infinity resulting in a computational intractable problem). However, using intuition and practical experience we notice that it is not the "global behavior" of the function we care about but rather the local variation. This is illustrated by the example in Fig. 1. Consider $f(x)$ in Fig. 1a and $g(x)$ in Fig. 1b and notice that both $f(x)$ and $g(x)$ trivially have the same total variation, namely $\mathcal{V} = c$, and yet the functions are different with respect to smoothness "locally." In fact, while the derivative of $g(x)$ in Fig. 1b is "small" on the entire support, $f(x)$ in Fig. 1a has a derivative tending to infinity at x^*. A measure of smoothness should take these practical issues in to account and objectively quantify the differences observed in Figs. 1a and 1b.



(a) (b)

Figure 1: Bounded variation and locally smooth functions.

2.3. Discrete finite variation

Based on the above discussion the desired property of a new smoothness measure is as follows:

locally samples of the wavelet should not change too fast

This is exactly why Fig. 1b is prefered over Fig. 1a where at x^* the rate of change is approaching infinity. Also, notice the use of "samples" in the above description. This comes from the fact that typically only a few stages (scales) of the wavelet transform is computed in any real application and hence we only care about "finite scale" smoothness.

If $f : \mathbb{R} \to \mathbb{R}$ and we let $f(x_i)$ for $i = 0, \dots, m-1$ be m equally spaced samples of f then the first order discrete difference operator $D_m : \mathbb{R}^m \to \mathbb{R}^{m-1}$ is defined as

$$D_m f(x_i) = f(x_i) - f(x_{i-1}) \quad 1 < i \leq m. \tag{10}$$

Using (10) the nth order discrete difference operator $D_m^n : \mathbb{R}^m \to \mathbb{R}^{m-n}$ is defined by

$$D_m^n f(x_i) = \left(\prod_{k=0}^{n-1} D_{m-k} \right) f(x_i) \tag{11}$$

and *discrete finite variation* (DFV) is defined as follows:

Definition 5 (Discrete finite variation). *Let J be the number of stages of the iterated filter bank and let $\psi_0^J(x_i)$, supported on $[0, N-1]$, be the length L sequence of samples of the scaling function such that $0 = x_0 < x_1 < \cdots < x_{L-1} = N-1$ with $x_i - x_{i-1} = \Delta x$ for all $i = 1 \dots, L-1$. Then the discrete finite variation of order n is defined as*

$$\mathcal{V}^n = \left| D_L^n \psi_0^J(x_i) \right| \tag{12}$$

The term DFV is used to emphasize that the measure is based on the (local) variation of discrete samples of the finite scale wavelet. The class of finite scale smooth wavelets are obtained numerically by solving the following constraint minimization problem

$$\min_{h_0} \mathbf{w}^T \nu \tag{13}$$

subject to

$$\text{a)} \quad \sum_{n=0}^{N-1} h_0(n) = \sqrt{2}$$

$$\text{b)} \quad \sum_{n=0}^{N-1} h_0(n) h_0(n+2l) = \delta(l)$$

where for $i_j \in \mathfrak{I}$ and $\mathfrak{I}$ the set of desired ith order differences to be optimized then

$$\nu = \left[\left\| \mathcal{V}^{i_1} \right\|_p, \left\| \mathcal{V}^{i_2} \right\|_p, \dots, \left\| \mathcal{V}^{i_{|\mathfrak{I}|}} \right\|_p \right]^T$$

and $w(n)$ is a set of weights

$$\mathbf{w} = [w(1), \dots, w(|\mathfrak{I}|)]^T .$$

For shorthand notation we will denote the optimal solution of (13) by $DFV_N^{n,p}$ where N is the filter length, n is the order of discrete difference operator and p is the desired order of the Euclidean norm to apply to $\mathcal{V}^n$ in (12).

3. EXAMPLES

This section gives results of the new design and compare the results with the K-regular solutions due to Daubechies. Table 1 compares DFV and Hölder exponent for the K-regular solution and the $\mathcal{DFV}_1^2$ solution. The solution of the optimal DFV was obtained by performing a constrained minimization using the MATLAB constrained optimization algorithm: `constr`. In Fig. 2 the corresponding scaling function for $N = 12$ is plotted. Notice, that although the Hölder regularity is significantly different (2.1553 versus 0.8964) the new design is "visually" smoother. In Fig. 3 the corresponding first and second order numerical derivatives are plotted. Notice that higher Hölder regularity does indeed correspond to smoother first and second order derivatives and yet one could not have guessed that by observing the functions in Fig. 2 only. In Table 2 scaling filter coefficients for the optimal $N = 12$ solution is given and Table 3 show the first 6 moments of both filters.

Table 1: Comparison of DFV and Hölder regularity for a length $N = 2K$ maximally regular (K-regular) wavelet basis and optimal DFV solution.

N	DFV		Hölder	
	K-regular	$\text{DFV}_N^{1,2}$	K-regular	$\text{DFV}_N^{1,2}$
6	0.2024	0.1962	1.0816	0.8498
8	0.1804	0.1735	1.6008	0.9132
10	0.1731	0.1680	1.9336	0.8976
12	0.1697	0.1655	2.1553	0.9124
14	0.1678	0.1649	2.4261	0.8955

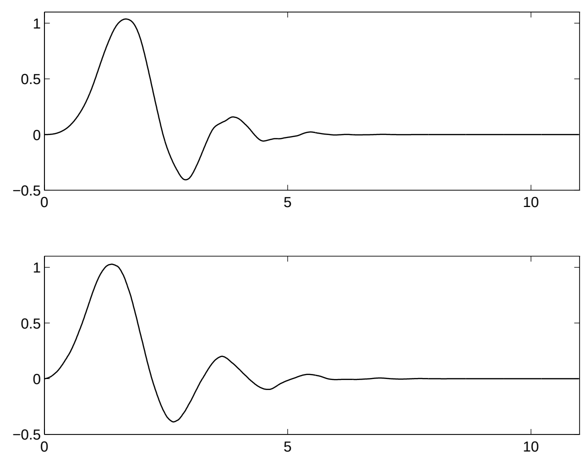


Figure 2: Scaling function $\psi_0(t)$ for length $N = 12$. Top: 6-regular (Daubechies). Bottom: optimal DFV.

4. SUMMARY

This paper introduces a new definition of smoothness which is more meaningful and less restrictive than optimization of Hölder and Sobolev exponents. The method dispenses with the traditional measures of smoothness which requires that a large number of moments be set to zero in favor of a measure that describes the "visual smoothness" of the finitely iterated filters. Results show that minimization of discrete finite variation (a measure of local roughness) results in wavelet basis which are visually smoother than the K-regular solutions due to Daubechies without having to impose a large number of zero moments of $\psi_1(x)$.

REFERENCES

[1] R. A. Gopinath and C. S. Burrus. On cosine-modulated wavelet orthonormal bases. *IEEE Trans. on Image Processing*, 4(2):162–176, February 1995. Also Tech. Report No. CML TR92-06, 1992.

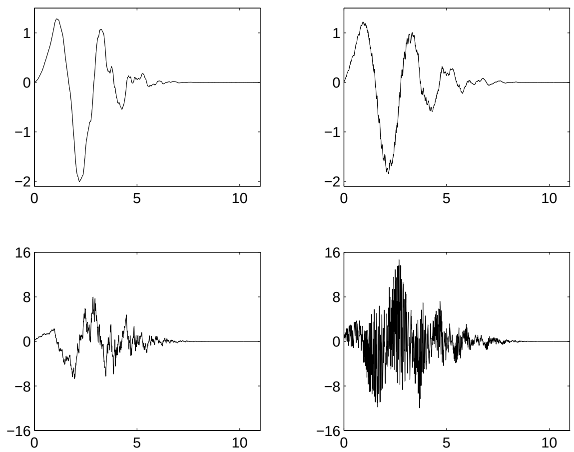


Figure 3: Numerical derivatives for $N = 12$. Top: first derivative for 6-regular solution (left) and optimal DFV (right). Bottom: corresponding numerical second derivative.

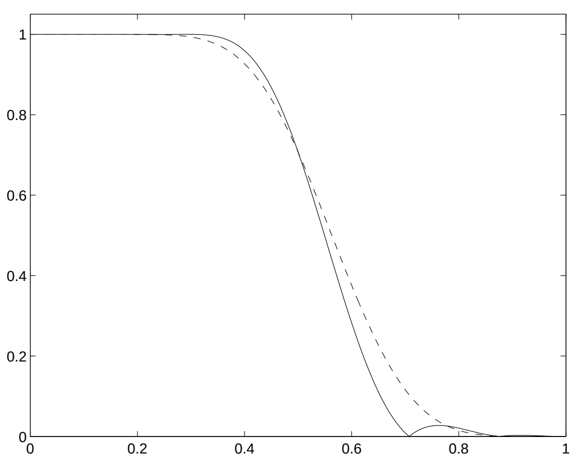


Figure 4: Frequency response of the length $N = 12$ scaling filters. Solid line: optimal DFV solution. Dashed line: 6-regular solution.

Table 2: Coefficients for the 6-regular Daubechies and optimal DFV of length $N = 12$ scaling filter, $h_0(i)$

i	6-regular	$\min \mathcal{DFV}_1^2$
0	0.11154074335011	0.19245354387245
1	0.49462389039847	0.61548666606234
2	0.75113390802111	0.68371407583713
3	0.31525035170920	0.15741065475483
4	−0.22626469396546	−0.25034258010171
5	−0.12976686756727	−0.09471248100885
6	0.09750160558732	0.12424589905097
7	0.02752286553031	0.03213128171802
8	−0.03158203931749	−0.05910916266348
9	0.00055384220116	0.00183896311853
10	0.00477725751095	0.01614500519119
11	−0.00107730108531	−0.00504830345833

Table 3: Discrete wavelet moments m_i of the 6-regular Daubechies and optimal DFV of length $N = 12$ wavelets

i	6-regular	$\mathrm{DFV}_{12}^{1,2}$
0	8.8818e-16	1.6653e-16
1	6.6613e-15	−1.5449e-05
2	5.3291e-14	2.5649e-01
3	3.4106e-13	4.9725e+00
4	2.7285e-12	9.0241e+01
5	2.1828e-11	1.5307e+03

[2] P. Heller and R. O. Wells, Jr. Spectral theory of multiresolution operators and applications. In *Proceedings of the International Conference on Wavelets*, Taormina, Italy, 1993. Also Rice University Tech. Report CML TR93-09.

[3] F. Jones. *Lebesgue Integration on Euclidean Space*. Jones and Bartlett, Boston, MA, 1993.

[4] M. Lang and P. N. Heller. The design of maximally smooth wavelets. In *IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing*, volume 3, pages 1463–1466, Atlanta, GA, May 1996.

[5] P. Rieder, K. Gerganoff, J. Götze, and J. A. Nossek. Parameterization and implementation of orthonormal wavelet transforms. In *IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing*, Atlanta, GA, May 1996.

[6] O. Rioul. Simple regularity criteria for subdivision schemes. *SIAM J. Math Anal.*, 23(6):1544–1576, November 1992.

[7] P. P. Vaidyanathan and Phuong-Quan Hoang. Lattice structures for optimal design and robust implementation of two-channel perfect reconstruction QMF banks. *IEEE Trans. ASSP*, 36(1):81–93, January 1988.